

**BIOINFORMATIC ANALYSIS OF SOME PLANT
GLUTAMATE RECEPTOR LIKE GENES FROM
FAMILY BRASSICACEAE**

A thesis submitted
in Partial Fulfillment of Requirements
for the Degree of
Doctor of Philosophy
by

Sri Bidhan Chandra Roy

INDEX NO.: 30/18/Life Sc./25



To the
DEPARTMENT OF LIFE SCIENCE AND BIOTECHNOLOGY
JADAVPUR UNIVERSITY, KOLKATA-700032

JUNE, 2024

CERTIFICATE FROM THE SUPERVISOR(S)

This is to certify that the thesis entitled “**BIOINFORMATIC ANALYSIS OF SOME PLANT GLUTAMATE RECEPTOR LIKE GENES FROM FAMILY BRASSICACEAE**”, submitted by Sri Bidhan Chandra Roy who got his name registered on 08/02/2018 with registration number (Index No: 30/18/Life Sc./25) for the award of Ph. D (Science) degree of Jadavpur University, is absolutely based upon his own work under the supervision of **Professor Ratan Gachhui** and co-supervision of **Dr. Ashutosh Mukherjee** and that neither this thesis nor any part of it has been submitted for either any degree/diploma or any other academic award anywhere before.

Ratan Gachhui.

(Signature of the Supervisor/

05.06.2024

Date with official seal)

Ratan Gachhui, Ph. D.
Professor
Life Sc. & Biotechnology Dept.
Jadavpur University
Kolkata - 700 032

Ashutosh Mukherjee

(Signature of the Co-Supervisor/

05.06.2024

Date with official seal)

DR. ASHUTOSH MUKHERJEE
Associate Professor
Department of Botany
Vivekananda College, Kol-63

Declaration Certificate by the Candidate

I declare that the thesis entitled “**BIOINFORMATIC ANALYSIS OF SOME PLANT GLUTAMATE RECEPTOR LIKE GENES FROM FAMILY BRASSICACEAE**” Submitted by me for the degree of Doctor of Philosophy is the record of work carried out by me under the guidance of **Professor Ratan Gachhui**, Department of Life Science & Biotechnology, Jadavpur University, Kolkata-700032 and **Dr. Ashutosh Mukherjee**, Department of Botany, Vivekananda College, 269, Diamond Harbour Road, Thakurpukur, Kolkata- 700063, has not formed the basis for the award of any degree, diploma, fellowship, titles in this or any other university or other Institution of Higher learning. I further declare that the material obtained from other sources has been duly acknowledged in the thesis.

Bidhan Chandra Roy.

Date: 05.06.2024

Place: Kolkata

Bidhan Chandra Roy
Department of Life Science & Biotechnology
Jadavpur University, Kolkata-700032

**This Thesis is dedicated to my cute
twin daughters Ronita & Nikita**

ACKNOWLEDGEMENT

I am very indebted to “**GOD ALMIGHTY**” for HIS grace and blessings without which the study would not have been fruitful. Undertaking doctoral research has been really a lifetime experience for me and it would never have been possible to do without continuous and spontaneous support, supervision and guidance that I received from numerous people. Hence, it gives me immense pleasure to acknowledge all those who have guided, contributed and supported me in carrying out my doctoral research entitled “**BIOINFORMATIC ANALYSIS OF SOME PLANT GLUTAMATE RECEPTOR LIKE GENES FROM FAMILY BRASSICACEAE**” at the Jadavpur University, Kolkata, India. There are lots of people whom I would like to convey my humble gratitude without whom this thesis might not have been written and completed. First and foremost, I would like to express my heartfelt gratefulness to my supervisor **Professor Ratan Gachhui**, Department of Life Science & Biotechnology, Jadavpur University, Kolkata, West Bengal-700032, without whom I would not have come so far. His sincere guidance, supervision, constructive criticism and words of encouragement always enlightened my way ever since the inception of my research career. He taught me the meaning of research and developed in me the passion for the same. I am rather indebted to my Co-Supervisor **Dr. Ashutosh Mukherjee**, Department of Botany, Vivekananda College, 269, Diamond Harbour Road, Thakurpukur, Kolkata, West Bengal-700063, India, Under University of Calcutta for providing me with all the necessary facilities whenever I asked for. He is my pillar of strength and biggest motivator throughout my Ph.D life, without his constant support my journey would not be successful. He is also my friend, philosopher, guide. Besides my Supervisors, I would like to thank my Research Advisory Committee Member **Dr. Joydeep Mukherjee**, School of Environmental Studies, Jadavpur University for his timely guidance, suggestions and encouragement throughout my work. My special thanks to the faculty members of our department **Prof. Parimal Karmakar**, **Prof. Biswadip Das**, **Dr. Paltu Kumar Dhal**, **Dr. Arunima Sengupta**, **Dr. Arghya Adhikary**, and **Dr. Sougata Roy Chowdhury** for their whole hearted and unfailing assistance throughout my research work. I am always grateful to all non-teaching staff, librarian of this department for their time and support at the various phases of my Ph.D. I am grateful to the **University Grant Commission (UGC)** and **Jadavpur University**, Kolkata for providing the working space.

I like to extend my sincere thanks to all my seniors, **Semantee Bhattacharya, Soumyadev Sarkar, Avishek Mukherjee**. I express my sincere thanks to all my lab mates, **Jhilam Majumder, Rubia Parvin, Rinky Bhadra, Sayanta Sarkar, Sontu Halder**.

I am thankful to **Dr. Biswajit Ghosh**, Principal, Dinabandhu Mahavidyalaya, Bongaon, North-24-Parganas, West Bengal-743235, India, who have supported me along the way.

I would like to acknowledge my gratitude to my teacher **Dr. Bablu Roy**, HoD of Department of Botany, Dinabandhu Mahavidyalaya, Bongaon, West Bengal. who helped me with his valuable suggestions whenever I asked for. I thank all my teachers who have provided me with the gift of knowledge throughout my student life.

I am also grateful to my colleague cum friends **Santanu Halder, Amit Gangapadhyay, Debraj Mondal, Rakesh Mondal** for their constant help, support and encouragement in every thick and thin during the entire tenure of my Ph.D. work.

I would be failing my duty if I do not acknowledge the kind cooperation of **my parents**, my wife **Pallabi**, who have always been a source of inspiration for me to work hard. I would also like to thank my friends, all my relatives, and well-wishers for their blessings, support, wishes and everything they did for me so that I could carry out this research work successfully.

Date: 05.06.2024

Place: Kolkālā

Bidhan Chandra Roy.

Bidhan Chandra Roy

List of Abbreviations

Abbreviation	Expanded from
• Glu	➤ Glutamate
• GABA	➤ Gamma-aminobutyric acid
• mGluR	➤ Metabotropic glutamate receptor
• iGluR	➤ Ionotropic glutamate receptors
• GLRs	➤ Glutamate receptors
• NSCC	➤ Non-selective cation channels
• ABA	➤ Absciscic acid
• SA	➤ Salicylic acid
• LUCA	➤ Last universal common ancestor
• PBP	➤ Periplasmic binding protein
• ATD	➤ Amino-terminal domain
• LBD	➤ Ligand-binding domain
• TMD	➤ Transmembrane domain
• CTD	➤ Carboxy-terminal domain
• NTD	➤ N-terminal domain
• LIVBP	➤ Leucine-isoleucine-valine binding protein
• PBPs	➤ Periplasmic binding proteins
• NMDA	➤ N-methyl d-aspartate
• GA	➤ Gibberellic acid
• ABA	➤ Absciscic acid
• CNQX	➤ 6-cyano-7-nitroquinoxaline-2,3-dione
• DNQX	➤ 6,7-dinitroquinoxaline-2,3-dione
• MAPK	➤ Mitogen activated protein kinase
• CBF	➤ C-repeat binding factor
• DREB	➤ Dehydration responsive component binding
• SAR	➤ Systemic acquired resistance

Abbreviation	Expanded from
• SAA	➤ Systemic acquired acclimation
• CHI	➤ Chitinase
• PAL	➤ Phenylalanine ammonialyase
• PR	➤ Pathogenesis-related protein
• GABA	➤ Gamma-aminobutyric acid
• JA	➤ Jasmonic acid
• GSLs	➤ Glucosinolates
• WGD	➤ Whole-genome duplication
• WGT	➤ Whole-genome triplication
• CDD	➤ Conserved Domain Database
• ML	➤ Maximum Likelihood
• GSDS	➤ Gene Structure Display Server
• TSS	➤ Transcription start site
• PlantCARE	➤ Plant Cis-Acting Regulatory Element
• PAP	➤ Purple acid phosphatase
• GRAVY	➤ Grand average hydropathy
• PDB	➤ Protein data bank
• 3D	➤ Three-dimensional
• RMSD	➤ Root mean square deviation Similarity Analysis
• APBS	➤ Adaptive Poisson-Boltzmann Solver
• ATD	➤ Amino-terminal domain
• LBD	➤ Ligand-binding domain
• TMD	➤ Transmembrane domain
• CTD	➤ Carboxy-terminal domain
• PIPSA	➤ Protein Interaction Property Similarity Analysis
• mbSUS	➤ Mating-based split ubiquitin system

ABSTRACT

Index No: 30/18/Life Sc./25

Title: BIOINFORMATIC ANALYSIS OF SOME PLANT GLUTAMATE RECEPTOR LIKE GENES FROM FAMILY BRASSICACEAE

Submitted by: Sri Bidhan Chandra Roy

Plant Glutamate receptor like channels (GLRs) are homologous to Ionotropic glutamate receptors (iGluRs) that mediate neurotransmission in mammals. GLRs are nonselective cation channels involved in various physiological processes. Thus, knowledge about *GLR* genes is crucial which can lead to effective way of improving agronomic traits in crop plants. The plant family Brassicaceae possess the model plant *Arabidopsis thaliana* as well as some major crops of the genus *Brassica*. *A. thaliana* possesses 20 *GLR* genes (*AtGLRs*) in its genome which are involved in many functions including light signal transduction and calcium homeostasis. This study reports the physico-chemical, structural and functional aspects of GLRs of *A. thaliana* as well as three important crops i.e. *B. napus*, *B. rapa* and *B. oleracea*. Gene ontology analysis showed the participation of *AtGLRs* in various biological processes including different stress responses. This study have identified 11, 27 and 65 *GLR* genes in *B. oleracea*, *B. rapa* and *B. napus*, respectively showing an expansion of this gene family in *B. napus*. Chromosomal locations revealed several tandemly duplicated *GLR* genes in all the three species. The phylogenetic analysis showed that the GLRs are classified into three main groups. The exon-intron structures of these genes are not very conserved and showed wide variation in intron numbers. However, protein sequences are much conserved as shown by the presence of ten short amino acid sequence motifs. Predicted *cis*-acting elements in promoters of *GLR* genes are mainly involved in light, stress and hormone responses. RNA-seq analysis showed that in *B. oleracea* and *B. rapa*, some *GLRs* are more tissue specific than others. In *B. napus*, some *GLRs* are differentially expressed under cold stress. The three-dimensional structure of GLRs has been developed *in silico* by homology modeling. The modeled proteins were predicted to be membrane bound homo tetramers with a three-layer domain architecture. This bioinformatic study will act as a platform to gain knowledge about the structure-function relationship of plant GLRs.

Ratan Gachhui.
05.06.2024
Ratan Gachhui, Ph. D.
Professor
Life Sc. & Biotechnology Dept.
Jadavpur University
Kolkata - 700 032

Dr. Ananta Kumar
05.06.2024
DR. ANANTA KUMAR
Associate Professor
Department of Botany
Vivekananda College, Kol-63

VIII

Bidhan Chandra Roy.
05.06.2024

CONTENTS

SL No.	Topic	Page No.
A.	<i>Certificate from the Supervisors</i>	I
B.	<i>Declaration Certificate by the Candidate</i>	II
C.	<i>Dedication</i>	III
D.	<i>Acknowledgement</i>	IV-V
E.	<i>List of Abbreviations</i>	VI-VII
F.	<i>Abstract</i>	VIII
G. Chapters		
Chapter 1	<i>Recent Advances in the field of Plant Glutamate Receptors</i>	1-24
Chapter 2	<i>Computational analysis of the Glutamate receptor gene family of Arabidopsis thaliana</i>	25-104
Chapter 3	<i>Genome-wide analysis of glutamate receptor gene family in allopolyploid Brassica napus and its diploid progenitors</i>	105-129
Chapter 4	<i>Homology modeling and in silico structural characterization of glutamate receptors of allopolyploid Brassica napus and its diploid progenitors</i>	130-152
H.	<i>Summary & Conclusions</i>	153-155
I.	<i>References</i>	156-182
K.	<i>List of Publications</i>	183-184

Chapter I.

Recent Advances in the field of Plant Glutamate Receptors

Chapter I.

Recent Advances in the field of Plant Glutamate Receptors

- 1.1 Introduction
- 1.2 Origin and Diversity of Plant Glutamate receptors
- 1.3 Subcellular localization of GLRs
- 1.4 Role of GLRs in Plant Growth and Development
- 1.5 Role of GLRs in Stress response
- 1.6 Role of GLRs In Plant Defense and Immune Response
- 1.7 Structural Characteristics of the GLRs
- 1.8 The plant family *Brassicaceae*
- 1.9 Genome Evolution and Polyploidy in *Brassicaceae*
- 1.10 The specific objective of the present research work are as follows

1.1 Introduction

Glutamate (Glu), a neurotransmitter in animals, has been found to function as signaling molecule in plants, controlling cellular metabolism, seed germination, plant growth, development, and long-distance information transmission (Forde 2014; Kong et al. 2015; Toyota et al. 2018). It is well known that Glu is a protein amino acid, which is a precursor of the mixture of proteins and polypeptides. Besides, Glu also is a common precursor of many organic compounds like glutamine, proline, arginine and histidine; non-protein amino acid like gamma-aminobutyric acid (GABA), antioxidant tripeptide (glutathione, GSH), heme, chlorophyll, and so forth (Brosnan and Brosnan 2013; Reiner and Levitz 2018). Therefore, Glutamate is a multifunctional (at least metabolite and signaling molecule) amino acid (Brosnan and Brosnan 2013; Reiner and Levitz 2018).

Functioning as a major excitatory neurotransmitter in animals, glutamate activates two receptor families: metabotropic glutamate receptor proteins (mGluR), which activate biochemical cascades, and ionotropic glutamate receptors (iGluR), which form cation selective ion channels that are members of the pore-loop subfamily (Mayer 2006). Glutamate Receptors are synaptic and non-synaptic receptors situated primarily on the membranes of neuronal and glial cells (Brassai et al. 2015). In plants, Glutamate generally plays its physiological functions by its receptors called plant glutamate receptor-like proteins or plant glutamate receptors (*GLRs*) which are multimeric, ionotropic, ligand-gated receptors (Lam et al. 1998), located in the plasma membrane (Forde and Lea 2007). First discovered in the model plant *Arabidopsis thaliana*, they are homologs of mammalian ionotropic glutamate receptors (*iGluRs*) which are ligand-gated ion channels (Singh and Chang 2018). They are non-specific ion channels, with a absorbency for calcium (Michard et al. 2011).

Although many works have focused on highly selective ion-channels in plants, a variety of cations are transported via non-selective cation channels (NSCC) (Davenport 2002), NSCCs are ion channels that are selectively permeable for cations over anions, but do not strongly distinguish between monovalent cations (Demidchik et al. 2002; Davenport 2002). Several studies suggest that plant *GLRs* function in Ca^{2+} and monovalent cation transport and may form constitutively active ion channels (Davenport 2002). Plant

GLRs have been found to function as non-selective cationic channels (NSCC), allowing the flux of calcium as well as sodium and potassium ions (Bortoli et al. 2016). Dedicated attempts are currently being invested into elucidating plant *iGLRs* physiological function and ion channel properties (Weiland et al. 2015). Although the ion-selectivity of plant *GLRs* is not very clear (Wudick et al. 2018a), *AtGLR1.4*, *AtGLR 3.4* and the moss *PpGLR1* were shown to have a poor ionic selectivity when heterologously expressed in mammalian cells or *Xenopus oocytes* (Vincill et al. 2012; Tapken et al. 2013; Ortiz-Ramírez et al. 2017). In *Arabidopsis* seedlings, *AtGLR1.4* has been shown to function as a nonselective Ca^{2+} -permeable cation channel and mediate methionine-induced depolarization (Tapken et al. 2013). *GLRs* have been shown to signal through calcium fluxes connected with male gamete function during fertilization (Michard et al. 2011; Ortiz-Ramírez et al. 2017; Wudick et al. 2018a).

Plant *GLRs* take part in the generation and transmission of electrical signals (Feng et al. 2021). These are a rapid-response signals generated by the instantaneous change in the electric potential gradient on the plasma membrane or alveolar leaf membrane (Sukhova et al. 2017). All the plants can produce long-distance electrical signals which play a key role in the communication of diverse tissues and organs leading to the integration of environmental responses (Garcia-Servin et al. 2021). *GLRs* are also associated to a wide array of processes including carbon and nitrogen metabolism, root development, defense response, stomatal aperture regulation, pollen tube growth and stress response (Simon et al. 2023); abscisic acid (ABA) biosynthesis and signaling, water loss, seed germination (Kong et al. 2015); root gravitropism (Miller et al. 2010), lateral root initiation (Vincill et al. 2013), root development (Singh et al. 2016), innate immune responses (Forde and Roberts 2014), Stomatal closure (Cho et al. 2009), Pollen germination and Pollen tube growth (Wudick et al. 2018b), Self-incompatibility (Iwano et al. 2015), wound-induced leaf to leaf electric signaling (Mousavi et al. 2013), sperm chemotaxis in *Physcomitrella patens* (Ortiz-Ramírez et al. 2017), adaptation to abiotic stress (Cheng et al. 2018; Li et al. 2019a) and so on. *GLR* mutants also show defects in defense against biotrophic pathogens, suggesting that they may also have a role in the salicylic acid (SA) pathway (Forde and Roberts 2014). Therefore, some of the plant's rapid responses to environmental and developmental inputs are mediated by *GLRs*.

1.2 Origin and Diversity of Plant Glutamate receptors

Plant *GLRs* share common ancestry with iGluRs (Stroebe and Paoletti 2021) but diverged very early from animal iGluRs, before the divergence of different classes of iGluRs (Chiu et al. 1999; Chiu et al. 2002). Primitive GluR signaling mechanisms existed even before the separation of plants and animals, as is indicated by the presence of GluRs in the prokaryotes, cyanobacteria *Synechocystis* (*GluR0*) and *Anabaena* (Chiu et al. 2002). Thus, this receptor family is not unique to animals and evolved in both plants and animals. Price et al. (2012) suggested that separation of plant, eukaryotic, and bacterial *GLRs* probably dates back to perhaps as far as the separation of the last universal common ancestor (LUCA). High similarity of functional domains in glutamate receptors with other proteins suggests that glutamate receptors originated by the fusion of a periplasmic binding protein (PBP) with a potassium ion channel (Wo and Oswald 1995).

The iGluRs have four subunits each containing four main domains: the amino-terminal domain (ATD) implicated in receptor assembly, trafficking, and regulation; the ligand-binding domain (LBD) that harbours binding sites for agonists, antagonists, and allosteric modulators; the transmembrane domain (TMD) that forms an ion channel; and the cytosolic carboxy-terminal domain (CTD), which is involved in receptor localization and regulation (Sobolevsky 2015; Twomey and Sobolevsky 2018). The N-terminal domain (NTD) *i.e.* ATD and LTD are both related to periplasmic binding proteins (PBPs). While ATD is related to leucine-isoleucine-valine binding protein (LIVBP or PBP type I); LBD is related to glutamine-binding protein (QBP or PBP type II). Fusion of a PBP (PBP type II) with an inverted potassium channel to form proteins like GluR0 of *Synechocystis* (Chen et al. 1999). Eukaryotic iGluRs evolved from the proteins like GluR0 by addition of NTD (corresponding to the fusion of a PBP type I) and a supplementary transmembrane segment right after the LBD (Stroebe and Paoletti 2021). So, the GluR0 has been expanded at both N and C terminal end (Figure 1.1).

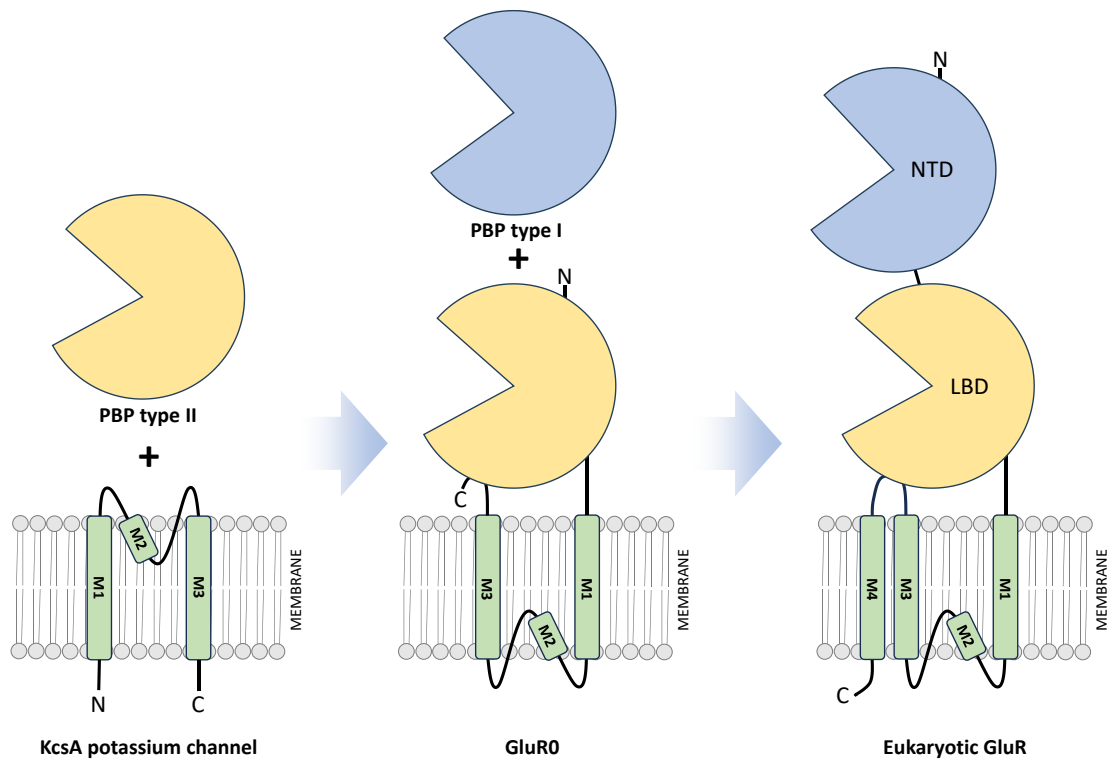


Figure 1.1: The proposed origin of eukaryotic glutamate receptor. The proposed gene fusion events are represented by arrows. Bacterial periplasmic binding protein (PBP) of type II and an inverted potassium channel fused to form prokaryotic iGluR homolog GluR0. Expansion of GluR0 at both N and C terminal end by fusion of another PBP (type I) and a supplementary transmembrane segment formed the eukaryotic Glutamate receptor (GluR). LBD: ligand binding domain; NTD: N-terminal domain; M1, M3, M4: transmembrane helices; M2: partial re-entry pore loop; N and C represent the amino- and carboxy-terminal ends, respectively. See text for details.

GLRs originated during evolution of the plant kingdom, in the genomes of gymnosperms and angiosperms and *Chlamydomonas*, chlorophytes, mosses, ferns, and flowering plants (Chiu et al. 1999; Bortoli et al. 2016; Wudick et al. 2018a). However, GLRs are thus far not found in yeast, eubacteria, fungi, and archaeobacteria (Chiu et al. 1999; Wudick et al. 2018a). Remarkably, GLRs have also been recognized in the nematode *Caenorhabditis elegans* as well as in marine invertebrates, such as the sea anemone *Nematostella vectensis*, and the ctenophora *Pleurobrachia bachei* (Wudick et al. 2018a). There is a wide-ranging diversity in the GLR superfamily, with the highest arrangement variability between the algal sequences, includes diatoms (large contributors to the world's oxygen), brown algae, cryptophytes, and green algae

(Bortoli et al. 2016). 16 glutamate receptor-like channels (GLR) fitted to two groups were identified in the *Brassica rapa* genome by bioinformatic investigation (Yang et al. 2022). 43 GLR genes, including 34 in *Saccharum spontaneum* and 9 in the *Saccharum* hybrid cultivar R570, were identified and characterized, which could be divided into three clades (clade I, II, and III). They had diverse evolutionary mechanisms, the former was mainly on the segmental duplication, while the latter mainly on the proximal duplication. Those sugarcane *GLR* proteins in the same clade had a similar gene structure and motif circulation (Zhang et al. 2022). It is striking that *GLRs* have not been found from fungi (including yeast), eubacteria, and archaeobacteria (Chiu et al. 1999) suggesting kingdom-specific roles for *iGLuRs* and *GLRs* in animals and plants, respectively (Wudick et al. 2018a). *GLRs* have been identified in other plant species, including mosses, rice and tomato (Li et al. 2006; Aouini et al. 2012; Lu et al. 2014; Ortiz-Ramírez et al. 2017). *Arabidopsis thaliana* possesses 20 glutamate receptor-like genes (*AtGLRs*), which are grouped into three clades based on sequence similarity, with clade I and II being sister clades (Chiu et al. 2002). Tomato contains 13 *GLRs*, which are also grouped into three clades. Comparison of tomato and *Arabidopsis GLRs* shows homology between two of the three clades (clade II and III), whereas clade I of tomato is distinct from *Arabidopsis* clade I (Aouini et al. 2012). The rice *GLRs* also divide into three clades, but one of the rice *GLR* clades does not group with any of the three clades in *Arabidopsis* (Ni et al. 2016). These studies indicated the presence of a distinct clade in plants apart from the three clades in *Arabidopsis*, which was later identified by (Bortoli et al. 2016). The fourth clade contains sequences from tomato and rice that did not group with the three *Arabidopsis* clades. Besides these, clade IV also contains sequences from *Phoenix dactylifera*, *E. guineensis*, *Musa acuminata*, and *Cucumis melo* (Bortoli et al. 2016). The *GLRs* in mosses (*Physcomitrella patens*) and gymnosperms (*Ginkgo biloba*) group with the clade III in *Arabidopsis* indicating that clade III is the most ancient among all the clades (Bortoli et al. 2016). The clade classification of mammalian glutamate receptors into NMDA, AMPA/KA and delta receptors conforms to their division based on ligand binding and electrophysiological properties. However, no such distinctions have been established for plant *GLRs*. The differentiation of clades in plants is independent of specific function or expression patterns (Chiu et al. 2002). While clade I and III genes in *Arabidopsis* are ubiquitously expressed in leaves, siliques, roots, and flowers, clade II genes are mostly root specific (Chiu et al. 2002). In plants,

GLRs are at least classified into three clades; clade-1 (*GLRs* 1.1-1.4), clade-2 (*GLRs* 2.1-2.9) and clade-3 (*GLRs* 3.1-3.7) (Lam et al. 1998; López-Bucio et al. 2018; Wudick et al. 2018a).

All members of the *AtGLR* family were most likely gated by amino acid residues, as per the homology modeling of their ligand-binding domains, and the broad ligand specificity observed in experiments is consistent with the high degree of sequence diversity found within the region of the ligand-binding domain that is predicted to bind the agonist side chain (Tapken et al. 2013). Even though only a few of the 20 *AtGLRs* have been well characterized, it appears logical to assume that as a whole, the plant *GLR* family is probably gated by many of the 20 proteinogenic amino acids in addition to some of their related molecules, such as small peptides like GSH (Forde and Roberts 2014). In the NMDA (N-methyl D-aspartate) subgroup of mammalian *iGluRs* (the most closely related subgroup to plant *GLRs*), the ATD possesses sequences related to the bacterial periplasmic-binding domain that binds various small molecules and ions which are crucial for both negative and positive allosteric regulation of the receptor (Kumar and Mayer 2013). Plant *GLRs* possess a similar conserved sequence in their ATD (Turano et al. 2001; Acher and Bertrand 2005), suggesting the potential for additional sensory complexity arising from allosteric interactions between this domain and as-yet-unknown regulatory molecules (Forde and Roberts 2014). Thus, it now appears that if the plant *GLR* gene family had been identified prior to their mammalian *iGluR* homologues, they would have been referred to be general or non-specific amino acid receptors rather than glutamate receptors (Forde and Roberts 2014).

1.3 Subcellular localization of *GLRs*

GLRs are articulated throughout the entire *Arabidopsis* plant, and have been reported in petiole tissue, leaves, stem, and roots, with subcellular localization in plastids, vacuolar system, endoplasmic reticulum, and like *iGluRs*, in the plasma membrane (Teardo et al. 2011; Vincill et al. 2013; Teardo et al. 2015; Nguyen et al. 2018; Wudick et al. 2018a). Various localization prediction tools suggested that some of the plant *iGLRs* might have chloroplast or mitochondrial targeting sequences (Teardo et al. 2010). The subcellular localization of *GLRs* can be a step towards better understanding their specific role in a cell. The analysis of the N terminal domain (NTD) sequence indicates their involvement in the secretory pathway. Thus, they are most likely located in the plasma membrane (Meyerhoff et al. 2005; Vincill et al. 2012). However, they are also predicted to be

targeted to other cellular compartments (<http://aramemnon.botanik.uni-koeln.de/>). *AtGLR3.4* is localized in both the plastid and the plasma membranes (Teardo et al. 2011). Dual targeting to mitochondria and chloroplast was observed for *AtGLR3.5* guided by alternative splicing (Teardo et al. 2015). (Kong et al. 2016) reported *AtGLR3.5* localization in the plasma membrane as well as chloroplast. *AtGLR2.1* and *AtGLR3.3* are the two GLRs highly expressed in the pollen. Subcellular localization studies with these two GLRs demonstrated *AtGLR2.1*-GFP localized to the tonoplast, whereas *AtGLR3.3*-GFP is found in the sperm plasma membrane as well as the endomembranes (Wudick et al. 2018b). GLRs sense amino acids at the exterior of the membrane in which the GLRs are localized (Price et al. 2012). In contrast to iGLRs have not only been located in the plasma membrane but also located in the plant mitochondria, chloroplast, tonoplast, endoplasmic reticulum and sperm cell (endo) membranes (Vincill et al. 2013; Teardo et al. 2015; Wudick et al. 2018a).

1.4 Role of GLRs in Plant Growth and Development

a) Seed Growth: In seed plants, seed development is the first and a main step in plant life cycle, which is an important basis of fruitful plant cloning and proliferation. In general, seed germination is strongly controlled by the synergistic action of endogenous (mainly phytohormones such as abscisic acid, ABA; gibberellin, GA; ethylene etc.) and exogenous indications (such as light, temperature, water, and oxygen etc.). Because seed germination is sensitive & breakable to environmental pressure, which has become a significant research mode of plant signaling transduction. This signaling method is involved in many signaling molecules (i.e. Calcium; nitricoxide: NO; hydrogen sulfide; plant hormones etc.) and their interactions (Duermeyer et al. 2018; Prakash et al. 2019). In *A. thaliana*, the gene expression of GLR homolog3.5 (*AtGLR3.5*, constituting Ca^{2+} channels) was significantly up-regulated in developing seeds, which in turn increased cytosolic Ca^{2+} concentration, ultimately stimulating seed germination via responding the inhibiting effect of ABA. Unfavorably, inhibiting expression of *AtGLR3.5* blocked Ca^{2+} increase in cytosol, followed by postponing germination and increasing ABA sensitivity. In addition, the over expression of *AtGLR3.5* lowered the compassion of seeds to ABA and reduced the time of early germination. Also, Ca^{2+} could inhibit the gene expression of Absciscic Acid INSENSITIVE4 (*ABI4*, a transcription factor), which is involving in ABA response in seeds (Kong et al. 2015). These outcomes provide molecular evidence that the Ca^{2+}

influx mediated by AtGLR3.5 indorses seed germination via releasing the inhibitory effects of ABA by inhibiting ABI4 action.

b) Root Architecture & Growth: The plant root organization accounts for 80% of the entire plant, which not only anchors the plant into the soil, but also finds, uptakes, and carriages nutrient. In general, the root architecture is controlled by a various of signaling molecules, such as Ca^{2+} , ROS, NO, H_2S , Glutamate, and plant hormones (ABA, auxin, cytokinin etc.), as well as their crosstalk (Chen et al. 2012; López-Bucio et al. 2018). Recently, (Singh and Chang 2018) reported that, in model plant *Arabidopsis*, iGluR competitive rivals [6-cyano-7- nitroquinoxaline-2,3-dione (CNQX) and 6,7-dinitroquinoxali-ne-2,3-dione acted (DNQX)] dramatically inhibited the development of primary root and the thickness of lateral root by damaging root meristem size. On the contrary, the treatment with iGluR agonist Glutamate recovered the growth of *Arabidopsis* roots. In addition, a low volume of Ca^{2+} chelator [ethylene glycol-bis (b-aminoethylether)-N,N,N',N'-tetraacetic acid, EGTA] radically inhibited the root elongation of *Arabidopsis*, while this inhibition was partially improved by Glutamate (Singh and Chang 2018). The further experiments showed that 1-N-naphthylphthalamic acid (NPA, a polar auxin transport inhibitor) inhibited the development of primary root and reduced the density of lateral roots, whereas this adverse effect of NPA was rescued by exogenous Ca^{2+} and Glu, respectively (Singh and Chang 2018), indicating that synergistic action of Ca^{2+} , auxin, and Glutamate signaling in the structure of root architecture in *Arabidopsis*. Equally, Glutamate inhibited the growth of primary root and stimulated the outgrowth of lateral roots, forming a shorter and more branched root system in *Arabidopsis* (Walch-Liu and Forde 2007). *GLR3.6* of *A. thaliana* controls primary and lateral root development through adaptable the cell cycle control gene *KRP4*, a Kip-related protein (Singh et al. 2016), and *GLR3.2* and *GLR3.4* may form a protein complex that regulates adjacent root initiation (Vincill et al. 2013). These findings show that *GLRs* take part in the development of root architecture by regulating meristem upkeep in *Arabidopsis*. The additional pharmacological, electro-physiological, and molecular-genetic methods also confirm that *GLRs* participate in the establishment of root architecture in *Arabidopsis* by communication with Ca^{2+} , mitogen activated protein kinase (MAPK), ROS, and auxin (Forde 2014; López-Bucio et al. 2018). In *Arabidopsis*, the knockout mutant plants of *AtGLR3.2* or *AtGLR3.4* genes chief to the robust increase in the number of lateral root

primordia, but no important effect on the number of visible lateral roots, suggesting the *AtGLR3.2* and *AtGLR3.4* control the initiation of lateral roots. In addition, the *AtGLR3.2/AtGLR3.4* receptors may locate on the surface of the phloem cells to control the root architecture by arbitrating Ca^{2+} signaling in response to extra-cellular amino acids including Glutamate (Vincill et al. 2013). Similarly, in rice (*Oryza sativa*), the OsGLR3.1 knockout mutant plants also show a short root, and the additional experiments indicate that OsGLR3.1 absolutely controls cell propagation and survival in the root apical meristem (Li et al. 2006). In *Arabidopsis*, GLR3.4 and GLR3.2 are involved in lateral root initiation (Vincill et al. 2013). Mutations in *GLR3.4* and *GLR3.2* produced an increased number and abnormal placement of lateral root primordia as compared to wild type. Both GLR3.2 and 3.4 are localized in the phloem, particularly in the vicinity of the sieve plates and physically interact with each other. It is proposed that apoplastic amino acids act through heteromeric GLR3.2/GLR3.4 channels which affect lateral root development via Ca^{2+} signaling in the phloem (Vincill et al. 2013). Li et al. (2006) demonstrated the role of GLR3.1 in regulating cell division in root apical meristem of rice. Loss of function mutation in GLR3.1 due to T-DNA insertion inhibits root elongation giving rise to a short root phenotype (Li et al. 2006).

c) Stomatal opening and closure:

AtGLR3.1 and *AtGLR3.5* are two GLRs highly expressed in the guard cells suggesting their role in guard cell signaling (Cho et al. 2009; Yoshida et al. 2016). Transgenic plants overexpressing *AtGLR3.1* show an impaired stomatal closure in response to external Ca^{2+} (Cho et al. 2009). (Yoshida et al. 2016) showed the involvement of GLR3.5 in Glu-induced stomatal closure in both *Arabidopsis* and fava bean.

d) Pollen Tube Growth & Pollen Germination: The higher plants, pollen tube Growth & Pollen Germination are basis for attaining dual fertilization, which is a key stage in the entire life cycle of plants. The procedure needs to observe and integrate all types of signaling molecules, such as Ca^{2+} , ROS, NO, H_2S , & plant hormones, eventually forming a sophisticated signaling system regulating pollen tube growth and pollen germination. In overall view Calcium playing a vital role in Pollen Tube Growth & Pollen Germination an increase in cytosolic free calcium concentration is an essential signaling transduction. But the formation mechanism of Calcium signaling in pollen tube growth & pollen germination is not clear (Michard et al. 2011). The *AtGLR1.2* and *AtGLR3.7* mutants can weaken Ca^{2+} signaling in pollen tube, the *AtGLR1.2* mutants

show unusually distorted tips and tubes, while the *AtGLR3.7* mutant tubes grow slower than wild-type (Michard et al. 2011). These outcomes show a novel signaling mechanism in pollen germination and tube growth between male gametophyte and pistil tissue, which is comparable to animal nervous systems based on amino acid-mediated communication (Qiu et al. 2020). Michard et al. (2011) reported the role of GLRs in pollen tube growth. In this study, GLR specific antagonists significantly inhibited the pollen tube growth rate in tobacco. Moreover, in *Arabidopsis*, the knockout mutants of *Atglr2.1* displayed abnormal pollen tube phenotype with deformed tips. Another mutant of *Atglr3.7* showed a decrease in the pollen tube growth rate. This study also showed that the GLRs present at the apical region of the pollen tubes are activated by D-Serine which allow Ca^{2+} ions to permeate into the cytoplasm.

e) Light responsiveness:

Lam et al. (1998) demonstrated the role of *AtGLRs* in transmitting light signals in *Arabidopsis*. Hypocotyl length is an important indicator of light-driven development in plants. *Arabidopsis* seedlings grown in normal light conditions exhibit short hypocotyls, whereas limiting light leads to hypocotyl elongation. *Arabidopsis* seedlings treated with DNQX, an antagonist of animal *iGluRs*, show an increase in hypocotyl length and reduction in light-induced chlorophyll accumulation when grown in light. Dark grown plants do not show any effect on either of these traits. The treatment of light-grown *Arabidopsis* plants with BMAA, an agonist of animal *iGluRs*, also promotes hypocotyl elongation and inhibits cotyledon opening (Brenner et al. 2000). Hebda et al. (2022) showed that both phytochromes and cryptochromes may be involved in the upregulation of *AtGLR* gene expression. This work indicates that mRNA levels of most *AtGLRs* in leaves increase upon irradiation and that this upsurge is likely to be controlled by photoreceptors, chiefly by phytochromes, rather than by reversing signaling from the photosynthetic apparatus. One of the highest strengths of this study stems from the use of whole mature plants and their irradiation with defined monochromatic light commands. This enabled us to look into the mechanism of light parameter of *GLR* transcript levels at the major stage of plant ontogenesis (Hebda et al. 2022).

1.5 Role of GLRs in Stress response

In accumulation to seed germination, root architecture, and Pollen Tube Growth & Pollen Germination mentioned overhead that Glutamate as a signaling molecule is also

involved in the response and adaptation to cold, heat, salt, drought, pathogen, and wound stress, which are describe as follows in details:

a) Cold Stress: In plants, many studies have established that cell membrane is the primary site perceiving cold stress. The observation of cold stress by cell membrane can activate Ca^{2+} channels and triggers Ca^{2+} -centered signaling transduction pathways, eventually responding and acclimating to cold stress (Miura and Furumoto 2013; Raju et al. 2018; Ding et al. 2019; Friedrich et al. 2019). GLRs are also involved in response to cold stress. For example, in *Arabidopsis* seedlings, *AtGLR1.2* and *AtGLR1.3* mutants were sensitive to cold stress, whereas treatment with jasmonate (JA, a plant hormone) assuaged the sensitivity of *AtGLR1.2* and *AtGLR1.3* mutants to cold stress. The overexpression of *AtGLR1.2* or *AtGLR1.3* could recover the tolerance of mutants to cold stress by creating endogenous JA (Zheng et al. 2018). In addition, under cold stress, both mutants displayed a lower expression of C-repeat binding factor/dehydration responsive component binding1 (CBF/DREB1) compared with the wild-type plants, while the expression of *AtGLR1.2*-OE [*AtGLR1.2*-GFP (green fluorescence protein)] and *AtGLR1.3*-OE [*AtGLR1.2*-GFP] was upregulated in transgenic plants (Zheng et al. 2018). The additional experiments confirmed that *AtGLR1.2* and *AtGLR1.3* autonomously exerted similar role in JA-induced cold tolerance (Zheng et al. 2018). Recently, in tomato (*Solanum lycopersicum*), (Li et al. 2019a) reported that cold acclimatization at 12°C upregulated the expression of GLR3.3 and GLR3.5, which in turn amplified the resistance of tomato plants to a subsequent chilling stress at 4°C. These results indicate that plant GLRs (mainly GLR1.2, GLR1.3, GLR3.3, and GLR3.5) positively control the cold tolerance of plants. Two-week-old *Arabidopsis* plants exposed to touch or cold show a three to six-fold increase in the expression of *AtGLR3.4* (Meyerhoff et al. 2005).

b) Heat stress: In high temperature, above the ideal growth temperature, is a main stress factor that impacts on the entire life cycle of plants; Ca^{2+} , crossing membrane into the cells via Ca^{2+} channels, is occupied in the development of heat tolerance in plants (Gong et al. 1997; Wahid et al. 2007; Li et al. 2012; Asthir 2015; Li et al. 2019a; Serrano et al. 2019; Li et al. 2019b). Recently, the positive effect of Glutamate in heat tolerance had been found by (Li et al. 2019b), who first reported that in maize (*Zea mays* L.) seedlings treatment with Glutamate raised the survival percentage of seedlings under heat stress. Also, heat tolerance induced by Glutamate was impaired by exogenous

CaCl₂ (may produce Ca²⁺ overloading), LaCl₃ (plasma membrane Ca²⁺ channel blocker), EGTA (Ca²⁺ chelator), as well as trifluoperazine and chlorpromazine (calmodulin antagonists) (Li et al. 2019b). Similarly, the Glutamate induced heat tolerance was debilitated by GLR antagonist (DNQX) and its blocker (MgCl₂) (Li et al. 2019b). This work has for the first time reported that Glutamate is able to increase the heat tolerance of maize seedlings by activating GLRs-mediated calcium signaling, but the detailed physiological and molecular mechanisms need to be further explored in the future aspect (Qiu et al. 2020). Glutamate can also enhance the heat tolerance in maize seedlings by GLR mediated calcium signaling (Li et al. 2019b).

c) Salt stress: The Salt stress normally triggers early (osmotic phase) and late (ionic phase) stress phase. The former refers to the hasty production of osmotic stress within several minutes due to the quick decline in extra-cellular water potential, while the latter is ion and oxidative stress as well as nutrient imbalance as a result of accumulation of Na⁺ and Cl⁻ with time passing by. The symptoms of late stress chapter can be detected within several days (Deinlein et al. 2014; Muchate et al. 2016; Saddhe et al. 2019). Generally, plant response and adaptation to salt stress is implicated in a sophisticated signaling interaction network based on Ca²⁺ signaling. Therefore, Ca²⁺/calmodulin-centered Ca²⁺ messenger system plays a vital role in the response and version to salt stress in plants (Deinlein et al. 2014; Muchate et al. 2016; Saddhe et al. 2019). As describe overhead, seed germination is sensitive and fragile to environmental stress including NaCl stress, and Ca²⁺ and ABA control seed germination under NaCl stress. In *A. thaliana*, the results of (Cheng et al. 2018) showed that, compared with the wild-type plants, the AtGLR3.4-1 and AtGLR3.4-2 mutant plants were more sensitive to NaCl in both seed germination and post-germination seedling growth and development. Recently, (Wang et al. 2019) also observed that AtGLR3.7 communication with 14-3-3 proteins partaken in salt stress response in *Arabidopsis* by affecting calcium signaling pathways. The data suggest that Ca²⁺ influx mediated by AtGLRs (mainly AtGLR3.4 and AtGLR3.7) regulates seed growth and germination under salt stress (Qiu et al. 2020).

d) Drought Stress: Drought stress caused by the deterioration in soil water potential hinders with many physiological methods at different levels, such as water uptake, transpiration, photosynthesis, seed germination, plant growth, and development (Deinlein et al. 2014; Tardieu et al. 2018; Hasanuzzaman et al. 2019; Laxa et al. 2019).

Under unadorned conditions, drought stress lead concurrently or consecutively to osmotic and oxidative stress. Therefore, the regulation of water balance by regulating stomatal movement is one of the major strategies that resist and adapt to drought stress in plants (Deinlein et al. 2014; Tardieu et al. 2018; Hasanuzzaman et al. 2019; Laxa et al. 2019). In *Brassica napus*, exogenous glutamate treatment triggered both cytosolic Ca^{2+} levels and salicylic acid levels with the highest expression of calcium dependent protein kinase (*CPK5*) and salicylic acid synthesis-related gene isochorismate synthase 1 (*ICS1*), which in turn enhanced proline synthesis by enhancing the proline synthesis-related genes, thus improving drought tolerance by regulating cellular redox potential (La et al. 2020). The results of Yang et al. (2022) suggest that *GLRs* of *B. rapa* (*BraGLRs*) play significant role in abiotic stress tolerance and sexual reproduction. The expression of two *GLRs* namely *BraGLR5* and *BraGLR12* was significantly enlarged in drought-sensitive plants after being subjected to drought stress. The investigation of promoter rudiments showed that *BraGLR5* contained drought inducible and MeJA-responsive elements, and *BraGLR12* contained drought-inducible and ABA-responsive elements. The authors suggested that *BraGLRs* have a role in regulating the growth and development of stamens, and *BraGLR3* may be involved in the growth and development of the pistil (Yang et al. 2022).

e) Wound and Pathogen Stress: Though the plants, have no nervous system, they can swiftly reply to environmental stimulation including wound stress by a erudite signaling transduction pathway. In alternative words, plants can intellect local stimulation (such as herbivore attack, wound, drought, and high light) and convey this information to the whole plant body, thus triggering the defense responses, systemic acquired resistance (SAR), or systemic acquired acclimation (SAA) (Choi et al. 2017; Sukhov et al. 2019; Takahashi and Shinozaki 2019; Wasternack 2019). (Toyota et al. 2018) reported that Glutamate, as a wound signaling, could be quickly released from phloem cells when *Arabidopsis* was wound by herbivores or external force. The released Glutamate bound to *GLRs* locating on the surface of the phloem cells, and then triggered an increase in $[\text{Ca}^{2+}]_{\text{cyt}}$, thus propagating long distance, Ca^{2+} -based defense response, and even dispersal over the whole plant body.

Compared to the abiotic stress, biotic stress reply and resistance also is involved in numerous signaling molecules (Ca^{2+} , ROS, NO, H_2S , and Glutamate) and their cross talks (Seifi et al. 2013). Glutamate pre-treatment enlarged the resistant to *Penicillium*

expansum in pear (*Pyrus bretschneideri*) fruit under normal or cold environments. In the addition of Glutamate also reduced the spore germination of *P. expansum* equally in fruit wounds and in vitro culture (Jin et al. 2019). Also, Glutamate boosted the activities of β -1,3-glucanase (GLU), chitinase (CHI), phenylalanine ammonia-lyase (PAL), peroxidase, and polyphenol oxidase, and also the gene expression of pathogenesis-related proteins1 (PR1), GLU, CHI3, CHI4, and PAL (Jin et al. 2019). In addition to these, Glutamate pre-treatment encourage the metabolism of amino acids, specially the accumulation of Glutamate, GABA, and arginine (Jin et al. 2019). Also, in tomato fruits, Glutamate pre-treatment could increase the gene expression of nine GLRs (GLR1.1, GLR1.2, GLR2.1, GLR2.4, GLR2.5, GLR2.6, GLR3.1, GLR3.2, and GLR3.3) and four PRs (PR1a, PR2a, PR3a, and PR3b), and also the accumulation of eight amino acids (Glutamate; GABA; methionine: Met; asparagine: Asn; phenylalanine: Phe; His; Lys; and Arg). Sequentially, in which induced the resistance of tomato fruits to *Botrytis cinerea* (Sun et al. 2019). These results supported that the GLRs (specially AtGLR3.3) are involved in the response and adaptation to pathogen stress in whole plants of the World (Qiu et al. 2020). Glutamate receptors play both direct and indirect roles in plant defense response. Over-expression of a GLR from small radish, *Raphanus sativus* (RsGLR) in *Arabidopsis* improves resistance to necrotic fungal pathogen *Botrytis cinerea* (Kang et al. 2006). *AtGLR3.3* is involved in mediating resistance to downy mildew, caused by the oomycete, *Hyaloperonospora arabidopsidis* (Manzoor et al. 2013). In addition, *AtGLR3.3* plays a role in glutathione (GSH) mediated defense response against the bacterial pathogen, *Pseudomonas syringae* (Li et al. 2013) as well as in defense response to mechanical wounding (Mousavi et al. 2013).

1.6 Role of GLRs In Plant Defense and Immune Response

In recent years, the function of electrical signals in GLR-mediated plant resistance to biotic stress has attracted much attention (Li et al. 2020; Feng et al. 2021). It has been shown that one way to inhibit plant pathogen expansion and minimize Mold rot on fruit, such as *Penicillium* growth on pears, is with a topical glutamate treatment (Jin et al. 2019). Plant GLRs also have a well-documented role in defense downstream of jasmonic acid (JA) (Mousavi et al. 2013; Nguyen et al. 2018). In defense, GLRs have been concerned in calcium signaling that activates JA-dependent systemic signaling to

prepare distant organs for a hypothetically forthcoming attack (Mousavi et al. 2013; Nguyen et al. 2018).

Plants usage calcium signaling for defense, rebirth, and systemic immune replies upon wounding (Shanmukhan et al. 2020). The antioxidant glutathione (GSH) has been shown to persuade defense response and increase cytosolic calcium concentration, $[Ca^{2+}]_{cyt}$. In *Arabidopsis*, GSH-modulated defense genes in a GLR-dependent method (exactly *AtGLR3.3*) (Li et al. 2013). An example of one of the many defense genes upregulated is Lipoxxygenase 4 for jasmonic acid biosynthesis (Li et al. 2013).

In rice (*Oryza sativa* L.), the GLR gene *OsGLR3.4*, an ortholog to *AtGLR3.4*, has been reported to be essential for plant growth and systemic wound signaling (Yu et al. 2022). Like most GLRs, *OsGLR3.4* has been designated to be a Ca^{2+} -permeable channel. In an agonist outline assay, multiple amino acids have been shown to trigger temporary Ca^{2+} influx in an *OsGLR3.4*-dependent manner (Yu et al. 2022). Total 20 amino acids tested, the amino acids that persuaded a Ca^{2+} flux via *OsGLR3.4* were Ala, Arg, Asn, Cys, Glu, Gly, Leu, Lys, Pro, Ser, and Thr . (Yu et al. 2022), reported *OsGLR3.4* mediates long-distance hurtful response and is required for root-to-shoot systemic wound signaling in rice. When *OsGLR3.4* was lacking, they detected a missing downstream response to leaves upon root damage, an breakdown of actin filament organization and growth imperfections (Yu et al. 2022).

1.7 Structural Characteristics of the GLRs

The vertebrate iGluRs are large tetrameric (homo- or heteromeric) complexes in which the four subunits are assembled around a single membrane-embedded ion channel with a fourfold or pseudo-fourfold symmetrical arrangement with dimer of dimers organization at the extracellular region (Price et al. 2012; Stroebel and Paoletti 2021). For some iGluRs such as AMPA and kainate receptors, initial dimerization is facilitated by compatibility and interaction between the NTDs of two subunits. A subsequent dimerization takes place through interactions of the ligand-binding domains and transmembrane domains (Ayalon and Stern-Bach, 2001; Tichelaar et al. 2004). Functional channels are formed by assembling the subunits from within the same iGluR class (Hansen et al. 2021). The N-terminal domain and the ligand-binding domains are on the extracellular side of the membrane, and the C-terminal domain is intracellular. Between these lie the transmembrane domains that form the channel for the cations to pass through (Twomey and Sobolevsky 2018). Studies suggest that glutamate receptors form

multimeric complexes in plants but the subunits that make up plant GLRs are not well known. Roy et al. (2008) proposed the possibility of heteromer formation of GLRs when they observed co-expression of an average of five to six *AtGLRs* in epidermal and mesophyll cells obtained from leaves. *AtGLRs* 3.2 and 3.4 physically interact with each other when expressed in mammalian HEK293 cells as well as *Nicotiana benthamiana* (Vincill et al. 2012; Vincill et al. 2013). Homomer formation for *AtGLRs* 1.1 and 3.4 has also been reported (Price et al. 2013). In this study, a modified yeast-2-hybrid system approach known as “yeast mating-based split ubiquitin system (mbSUS)” identified various subunits (1.1, 2.1, 2.9, 3.2, and 3.4) which interact with some other subunits. This study also highlighted that in contrast to the mammalian iGluRs, which only interact with subunits within their own clade; the interactions between plant GLRs are not clade-specific (Price et al. 2013). Although the plant and animal iGluRs diverged very early during evolution (Chiu et al. 1999), they do share a similar structure consisting of an N terminal domain (NTD), a ligand-binding domains (LBD), a transmembrane domain with three membrane spanning helices (M1, M3 and M4), a partial re-entry pore loop (M2) and a Carboxy terminal domain (Simon et al. 2023; Wudick et al. 2018a).

In 2021, the structure of *Arabidopsis thaliana* GLR 3.4 has been revealed by X-ray crystallography and single-particle cryo-electron microscopy (Green et al. 2021). The structure is Y-shaped and resembles the shape of non-NMDA subtype of iGluR structures (Green et al. 2021). The structure, like all iGluRs, has a three-layer architecture, including the ATD at the top, LBD at the middle and TMD at the base of the ‘Y’-shape. The ATDs and LBDs of the four subunits form dimers showing two-fold rotational symmetry. Unlike major iGluR subtypes, GLR3.4 did not show domain swapping between the ATD and LBD dimers. ATD and LBD layers maintain a dimer-of-dimers organization but the TMD layer shows a pseudo-four-fold symmetrical organization. However, unlike all major iGluR subtypes (except orphan delta receptors), the *AtGLR3.4* showed lack of domain swapping between the ATD and LBD dimers (Green et al. 2021).

The ATD of *AtGLR3.4* possess a clamshell architecture as found in other similar proteins including iGluRs, bacterial periplasmic amino acid binding proteins, GABA_B receptors, the extracellular domain of the natriuretic peptide receptor and LBDs of

family C G protein-coupled receptors (Green et al. 2021). The ATDs of AtGLR3.4 have been found to be bound to glutathione (GSH) through S-glutathionylation of cysteine C205. This explains the observation of Qi et al. (2006) who reported GSH as an activator of GLRs. It is suggested that GSH cause conformational changes in the GLR resulting in gating of its ion channel (Green et al. 2021).

The LBDs of different *A. thaliana* GLRs have been studied in detail. The LBDs of *A. thaliana* GLR3.2 and GLR3.3 provided the first glimpse of the fragment of GLR structure responsible for ligand binding (Alfieri et al. 2020; Gangwar et al. 2021). They displayed the predicted bilobed structure resembling prokaryotic and eukaryotic LBDs (Mayer 2006; Alfieri et al. 2020; Gangwar et al. 2021). GLR3.2 and GLR3.3 LBD structures identified similar overall folds to iGluRs LBD structures, with the distinct features in the ligand-binding pocket that are responsible for the promiscuous GLR activation by various amino acids. Later, the GLR3.4 of *A. thaliana* has been studied and found to be structurally similar to the GLR3.2 and GLR3.3 (Green et al. 2021). All of them have a clamshell architecture like iGluRs. All these structures showed that the ligand-binding pocket has evolved to bind different-sized amino acids and also explains the ligand-binding promiscuity of GLRs.

The structure of the ion channel has been studied in detail in GLR3.4 of *A. thaliana* (Green et al. 2021). Here, the ion-channel is assembled of the TMDs of all the subunits. Each TMD contributes the transmembrane helices M1, M3 and M4 as well as a re-entrant intracellular loop M2. The intracellular part is responsible for the ion selectivity and is lined up by the non-helical part of M2. The extracellular parts of the M3 helices are responsible for the formation of the narrow constriction of the pore. This M3 is the most highly conserved domain of GLRs and contains the consensus sequence SYTAxLxxx (Chiu et al. 1999; Simon et al. 2023). This motif is represented by SYTANLAAF in *Homo sapiens* and essential for iGluR gating (Traynelis et al. 2010).

The Activation of the channel requires the binding of a specific ligand(s) to the ligand-binding domains, which leads to conformational changes causing the channel to open. L-glutamate is the primary ligand that binds to the GluR channels in mammals. Some of the other amino acids, like glycine, aspartate, and D-serine, have also been reported to bind to these channels. The similarity of the GLR structure in animals and plants led to the belief that plants GLRs might be gated by the same amino acids as their

mammalian counterparts. However, studies indicate that plant GLRs have a broader ligand specificity than animal iGluRs (Vincill et al. 2012; Tapken et al. 2013; Forde and Roberts 2014). Glutamate is still the most common ligand that activates the channel. However, membrane depolarizations and Ca^{2+} influxes have also been observed with glycine, asparagine, serine, alanine, cysteine, methionine and even glutathione (Michard et al. 2011; Vincill et al. 2012; Tapken et al. 2013; Forde and Roberts 2014).

From the above discussions, it is clear that *GLRs* play some crucial roles in plants, and majority of work regarding plant *GLRs* has been reported from the model plant *Arabidopsis thaliana* as this is the first plant species from which *GLR* homologs have been identified (Lam et al. 1998). Although gene specific data is available for some *GLRs* of this species, a genome-wide analysis of *GLRs* from this model plant is required to get a wholistic picture of gene structure, expression and structural features of this gene family in *A. thaliana*. Thus, in this study, *GLRs* of *A. thaliana* has been studied in detail with the help of several computational tools. We have also studied *GLRs* from different crop species of *Brassica* as this genus belongs to the plant family Brassicaceae which also contains *A. thaliana*. As this family has undergone a series of whole genome duplications, study of *GLR* gene family from the model plant *A. thaliana* and species of the genus *Brassica* should highlight the evolution of this gene family in the light of gene and genome evolution. For a better understanding of the above facts, taxonomic and economic importance of the family Brassicaceae along with the role of whole genome duplication in its genome evolution is briefly discussed below.

1.8 The plant family *Brassicaceae*

The family *Brassicaceae* (Cruciferae) is among the fifth largest angiospermic monophyletic families belonging to the order *Brassicales*. It contains 12–15 tribes with 338–360 genera and about 3709 species distributed all around the earth in all landmasses excluding Antarctica (Al-Shehbaz 1973; Appel and Al-Shehbaz 2003; Warwick et al. 2006). Besides diverse geographic distribution of the family from the northern to the southern hemisphere, its distribution has been informed to be sparse on mountainous and alpine regions in the tropics. Most of the morphological characters in the family are efficacious from an agricultural point of view (Davis et al. 1998). This family is highly blessed with plants that have an extensive range of economic

importance from food to fodder, medicines to use, as research model plants, to high yielding crops to ornamental plants. Thus, it infers and concealments all possible sources that a plant can be of a benefit to humankind. The *Brassicaceae* family comprises numerous edible species such as fodder, oilseed, vegetables, and condiments. This family is crucial for vitamins, A, B1–2, B6, C, E, K, and minerals such as magnesium, iron, and calcium (Wink and Van Wyk 2008). The most common vegetable crops are turnip, broccoli, cauliflower, kohlrabi, and cabbage supplemented with nutritional fibers (Shankar et al. 2019). *Brassicaceae* members like cabbage, kale, green-mustard, cauliflower and broccoli decrease the risk of several types of cancer (Wang et al. 2004). A significant constituent of *Brassicaceae* vegetables known as indole- 3-carbinol acts as an anticarcinogenic compound through dissimilar hormonal and metabolic changes (Hanf and Gonder 2005), and also decreases the tumour specifically in reproductive organs (Staub et al. 2002). The family consists of numerous important genera have diverse economic as well as agronomic use in discovering the world of knowledge using them as model plants. The plants, which revolutionized modern biology i.e. *Arabidopsis* and *Brassica*. Therefore, it is discussed that there would be no life for model plants, or we would deficiency many more crop plants without this particular, *Brassicaceae* family (Franzke et al. 2011). *Brassicaceae* is considered to be a model plant family as this family consists of many economically important crop plants including many species and cultivars of the genus *Brassica*, model plant *Arabidopsis thaliana* and also a perfect family to study plant evolution and domestication (Hasanuzzaman 2020; Huang et al. 2020).

1. *Arabidopsis thaliana* (L.) Heynh

Arabidopsis is one of the most important genera in the *Brassicaceae* family, and various aspects of its taxonomic history have been explored and reported on the basis of its morphology and molecular phylogeny (Al-Shehbaz et al. 1999). There are approximately 9 species and 8 subspecies of *Arabidopsis* (Koch et al. 2008). *Arabidopsis* has no well-documented economic importance, but its small genome and short life cycle covered the way to better study, understand, and discover the development, physiology, and gene function of plants through genetic approaches, thus making it a highly valued and reared model plant. *Arabidopsis thaliana* was the first

plant genome to be sequenced, selected partly for its phylogenetic relation to agriculturally significant *Brassica* species, thus importantly inspiring proportionate genomics and evolutionary research in this family. It has a small genome that has been completely sequenced (The Arabidopsis Genome Initiative 2000). Thus, *Arabidopsis thaliana* contributes to the exploration of various aspects of plant biology and provides insights into growing crop production through various molecular and biotechnological approaches (Buell and Last 2010).

2. *Brassicas*

Brassica is the important genera of this family, consisting of 37 species (Gomez-Campo 1980). Most of the crop plants are of this genus which has edible buds, flowers, leaves, roots, seeds, and stems. Most of the wild species are applied for hybrid seed development. These hybrid seeds and modified species are therefore of major importance in developing resistant varieties against various pests (Rakow 2004). *Brassica* plants included many amino acids such as L-glutamine, L-histidine, L-alanine, L-aspartic acid, L-valine, L-tryptophan, L-threonine, L-phenylalanine, and L-methionine; moreover, different phytochemicals such as indole phytoalexins (N-Methoxyspirobrassinol methyl ether, 1-methoxyspirobrassinol, 1-methoxyspirobrassinin, camalexin, brassilexin, spirobrassinin, brassinin, glucosinolates (neoglucobrassicin, gluconapin, glucobrassicin, gluconasturtiin, gluco*Brassic*anapin, glucoalyssin, glucoraphanin, and glucoiberin), and phenolics (kameferol, anthocyanins, p-coumaric, quercetin, sinapic acid, ferulic, caffeic, chlorogenic, neochlorogenic, hydroxybenzoic, isoferuloylcholine, and feruloyl) have antioxidant, cardiovascular protective, and anticarcinogenic properties (Jahangir et al. 2009). *Brassica* is widely used as a biocontrol agent (Ahuja et al. 2011).

a. *Brassica rapa* L.

Brassica rapa contains turnip, turnip rape, fast plants, field mustard, and turnip mustard; largely cultivated as root and leaf vegetable, as well as an oilseed. *Brassica* vegetables afford soluble fibers, vitamin C, and many compounds having anticancer properties, which controls innate immune system response by anticancer, antibacterial, and antiviral activity. *Brassica rapa* has 10 chromosomes. *B. rapa* has been reported to be of Mediterranean source excluding the coastal areas. These areas are mountainous and

have a very cold temperature; therefore, the species show preferred fast growth to low-temperature conditions. It is spread all over Germany, Eastern Europe, while announced in the Western Asian countries. In south Asian countries, such as India, it is also cultivated (OECD 2016). It is cultivated as an oilseed crop in China, India, Canada, and Finland (Rakow 2004).

b. *Brassica oleracea* L.

Brassica oleracea is the most significant agronomically and diploid crop species covering different crop; i.e., cauliflower, Brussels sprout, kale, cabbage, kohlrabi, and cabbage. *Brassica* crops are grown because of their inflorescence, stems, leaves, and axillary buds, also present unbelievable morphological diversity (Golicz et al. 2016). Besides their morphological inconsistency, these cultivars along with their interrelated wild species have a collection of secondary metabolites, which have been involved in supporting human health like different compounds functions as antioxidant and anti-carcinogens (Parkin et al. 2014). *Brassica oleracea* shows a very differentiated phenotype with 9 chromosomes. It is broadly distributed in Asia, Turkey, and Southern Greece. It has corpulent leaves. The plant has a woody outlook and is perennial (Rakow 2004). It is mainly used as vegetables, oil, and as a fodder crop (Rakow 2004).

c. *Brassica napus* L.

B. napus (oilseed rape) growing all over the earth, has the diverse variety of consumptions comprising animal feed, plant oil as well as fuel substitute (Abdallah et al. 2010). It is an amphidiploid species and consists of 19 chromosomes, which has been found as a result of an interspecific cross between *B. rapa* ($n = 9$) and *B. oleracea* ($n = 9$) (Parkin et al. 1995; Rakow 2004). Wild species other than the cultivated forms of *B. napus* could be obtained on beaches, coastal areas of Britain, Gotland, Netherlands, and Sweden (Rakow 2004). In some countries, *B. napus* is the most productive plant grown as an oilseed. The probable of yield is much high as the photosynthetic activity is high concerning per unit of the leaf which is further positive and proportional to numerous chloroplasts present per unit area of a leaf (Liu et al. 2009).

1.9 Genome Evolution and Polyploidy in *Brassicaceae*

Polyploidy or whole-genome duplication plays an important role in innovation, and it is broadly distributed across the plant's life, although its long-term evolutionary

importance is still arguable. While polyploidy expanded much slower than diploidy, polyploidy has very fast extinction and biological evolution rate. It is also demonstrated, that polyploidy has played a crucial part in the shaping of the current-day differences among species' richness within the *Brassicaceae* family, along with the forthcoming intimations in explaining the pattern of genetic diversity beyond the plant's life (Van de Peer et al. 2017; Román-Palacios et al. 2020). Although the potential role of whole genome duplication in plant evolution remains debatable, there are some evidences that the long-term establishment of polyploidy might be associated with environmental changes, stress conditions, or periods of extinction (Yao et al. 2019). The two diploid sub-genomes, i.e., *B. oleracea* and *B. rapa*, known for historical polyploids, still have ancestry of triplicated genome structure, connected from *B. napus* (oilseed rape). However, no evidence of wild canola varieties has been found; its genome evolution has only been artificially manipulated by human needs and agricultural systems (Chalhoub et al. 2014). *Arabidopsis* as a model plant, the genomic evolution of the genus *Brassica* (particularly *B. napus*) has been genome-wide explored via polyploidy. A comparative study of *Arabidopsis* and *Brassicaceae* was highly appreciable for some significant traits, i.e., flowering stage and susceptibility against stress (Lagercrantz et al. 1996; Osborn et al. 1997; Robert et al. 1998; Sillito et al. 2000; Tadege et al. 2001; Mun et al. 2009) to detect candidate genes by sequencing data considering *Arabidopsis* as a reference genome (The Arabidopsis Genome Initiative 2000). The positive and close genomic relationships between *Brassica* and *Arabidopsis* anticipates comparing similarities or differences by their sequences (Parkin et al. 2005), and genetic explanation (Wang et al. 2011).

Considering the importance of GLRs as well as the family *Brassicaceae*, this study aims to characterise the glutamate receptors of *A. thaliana* (a model plant) and some species of *Brassicaceae* (*B. napus*, *B. oleracea* and *B. rapa*) keeping in mind that whole genome sequences are available for these plants.

1.10 The specific objective of the present research work are as follows:

1. To Study the physiochemical characterization, motif and domain analysis of the *AtGLR* genes.
2. The Chromosomal distribution, duplication, gene structure and phylogeny of *AtGLR* genes.
3. To Expression analysis and functional annotation of *AtGLR* genes.
4. The Three-dimensional structure of *AtGLRs*.
5. To Identify *GLR* genes in *B. napus*, *B. oleracea* and *B. rapa* and their phylogenetic analysis.
6. The Chromosomal localization of *Brassica GLR* genes.
7. To Analysis of *GLR* gene expression in *Brassica*.
8. To study the microRNAs which controls the expression of *GLR* genes in *Brassica*.
9. Homology modeling and structural features of Glutamate receptors of the three species of *Brassica*.

Chapter **II.**

Computational analysis of the
Glutamate receptor gene family
of *Arabidopsis thaliana*

Chapter II.

Computational analysis of the Glutamate receptor gene family of *Arabidopsis thaliana*

- 2.1 Introduction
- 2.2 Materials and methods
- 2.3 Results and Discussion
- 2.4 Conclusion

2.1 Introduction

Membranes of plant cells possess a variety of ion channels which play vital role in transport of ions used as signaling molecules, nutrients, metabolites and osmotica (Davenport 2002). Membrane proteins at the surface of the cell serve as the communication interface between the cell's internal and external environments (Venkatakrishnan et al. 2013). Although many works have focused on highly selective ion channels in plants, a variety of cations is transported via non-selective cation channels (NSCC) (Davenport 2002). NSCCs are ion channels that are selectively permeable for cations over anions, but do not strongly distinguish between monovalent cations (Demidchik et al. 2002). Some of these putative NSCC genes of plant origin are related to animal glutamate receptors (Demidchik et al. 2002). Using the genome sequence and expressed sequence tag (EST) database of the model plant *Arabidopsis thaliana* (L.) Heynh., (Lam et al. 1998) identified two complementary DNA clones known as *GLR1* and *GLR2* as well as two genomic sequences known as *GLR3* and *GLR4*. In addition to primary sequence similarity, the *Arabidopsis* GLR (AtGLR) genes are also predicted to be similar to animal iGluRs (Lam et al. 1998). To date, 20 GLRs have been identified in the *Arabidopsis* genome (Davenport 2002). Study of glutamate-like receptor homologs from sequenced genomes of *A. thaliana*, lycophytes and bryophytes revealed that these plant genes share homology with mammalian ionotropic glutamate receptors (iGluRs), which are involved in neurotransmission (Price et al. 2012). iGluRs are non-selective cation channels that function predominantly as glutamate-gated Na^+ and Ca^{2+} influx pathways at neuronal (in vertebrates) or neuromuscular (in invertebrates) junctions (Dingledine et al. 1999). GLR-like genes have also been identified in other dicotyledonous and monocotyledonous species as well as in gymnosperms indicating that these genes are of general importance in plant physiology (Davenport 2002).

Following the discovery of the *Arabidopsis* AtGLR family, (Dennison and Spalding 2000) demonstrated that addition of glutamate (1mM) to the medium around *Arabidopsis* seedlings stimulated a rapid increase in cytosolic Ca^{2+} and caused membrane depolarization of root tip cells. These effects were attributed to activation of plasma membrane Ca^{2+} permeable channels, even though the possibility of secondary activation of Ca^{2+} influx by glutamate uptake was not excluded (Davenport 2002). To date, GLRs are found to be involved in many biological processes in plants (Price et al. 2012). These include Carbon/Nitrogen balance (J. Kang and Turano

2003), photosynthesis (Teardo et al. 2010, 2011), abiotic stress response (J. Kang et al. 2004; Meyerhoff et al. 2005), growth of pollen tube (Michard et al. 2011), morphogenesis of root (J. Li et al. 2006; Miller et al. 2010), plant-pathogen interaction (S. Kang et al. 2006; Kwaaitaal et al. 2011) and regulation of cellular Ca^{2+} kinetics (Cho et al. 2009; Dubos et al. 2003; S. Kang et al. 2006; Kim et al. 2001; Qi et al. 2006; Vincill et al. 2012).

Similar to the animal GluRs, plant GLRs are predicted to have a long N-terminal domain of unidentified function, with homology to the ligand-binding domain of mGluRs (metabotropic glutamate receptors), GABA_B receptors (receptors of gamma-aminobutyric acid) and extracellular calcium sensors (Davenport 2002). This domain can be deleted from vertebrate iGluRs without loss of function, although it has been implicated in desensitization membrane targeting and allosteric modulation by Zn^{2+} (F. Zheng et al. 2001). Moreover, plant GLRs preserve the predicted channel structure of animal iGluRs and show high amino acid sequence identity in the ligand binding domains and in the M1 and M2 transmembrane domains (J. Chiu et al. 1999).

In spite of several studies concerning AtGLRs, there are several areas to explore. For example, analysis of *cis*-acting elements can throw light on the expression and possible role of these proteins in response to different environmental conditions. Moreover, evolutionary history and functional divergence can be inferred from the study of gene structure, chromosomal location and evolutionary rates along with expression analysis. Additionally, studies of three-dimensional (3D) structures are crucial for deciphering the actual process of their functioning.

In the post genomic era, several areas of *Arabidopsis* research are easy to perform because of the enormous information and databases associated with *Arabidopsis* (D. Li et al. 2014). Because of the lack of experimental three-dimensional structural data for AtGLR proteins, it is necessary to use comparative modeling and structural prediction to study their molecular properties. The utility of such approaches in designing experiments and interpreting experimental results is now widely recognized (Petrey and Honig 2005; Zhang 2009). Therefore, in the present study, we have carried out detailed *in silico* analysis of the AtGLR proteins to gain insights into their structural and functional attributes. We have analyzed physico-chemical properties, gene structure and gene ontology of AtGLRs. Three-dimensional structures were predicted with homology modeling, which would facilitate deciphering the molecular function of these proteins. These results may be helpful in designing of experiments

concerning GLR proteins of *A. thaliana* for their further characterization. Subsequently, we analysed their expression profiles in different tissues and developmental stages, which will further help future studies for elucidating the precise roles of AtGLRs in growth, development and environmental interactions.

2.2 Materials and methods

2.2.1 Identification, sequence retrieval and phylogenetic analysis

The previously reported (J. C. Chiu et al. 2002) AtGLR genes have been retrieved using the BioMart tool (Kinsella et al. 2011) of EnsemblPlants database (Kersey et al. 2014). Additionally, the *Arabidopsis thaliana* genes dataset of EnsemblPlants database was searched for the Pfam (Finn et al. 2016) domain PF00060 (Ligand-gated ion channel) which is characteristic of ionotropic glutamate receptors to find any new member. However, no new member could be found and thus, the final dataset was comprised of the previously reported 20 AtGLR genes. The gene and protein sequences were retrieved using the BioMart tool of EnsemblPlants database. Some genes showed more than one protein products and they were also studied. Multiple sequence alignment was carried out with MUSCLE (Edgar 2004). The resulting multiple alignment was converted into nexus format with ALTER (Glez-Peña et al. 2010). Phylogeny reconstruction using the Bayesian inference was conducted with MrBayes 3.2.6 (Ronquist et al. 2012).

2.2.2 Physicochemical characterization, motif and domain analysis of the AtGLR proteins

Different physico-chemical parameters viz. amino acid composition, theoretical pI (isoelectric point) and grand average of hydropathicity (GRAVY) were measured using ProtParam (Gasteiger et al. 2005) server (<http://web.expasy.org/protparam/>). Different phosphorylation, Asn-glycosylation and myristoylation sites have been predicted using motifscan tool of MyHits (Pagni et al. 2007). Protein domains were identified with Pfam (Finn et al. 2016). Protein disorder content have been studied using IUPred (Dosztányi et al. 2005a, 2005b). The different *cis*-elements of promoter regions of the AtGLR genes and corresponding transcription factors (TFs) of these binding sites have been predicted from AGRIS, the *Arabidopsis* gene regulatory information server (Yilmaz et al. 2011).

2.2.3 Chromosomal distribution and gene structure prediction

Using the BioMart tool (Kinsella et al. 2011) of EnsemblPlants database (Kersey et al. 2014), the chromosomal locations of each gene were retrieved. Subsequently, the genes were displayed onto the five chromosomes of *Arabidopsis thaliana* according to their ascending order of physical position (Mbp) using MapChart (Voorrips 2002). Tandem duplications were characterized on the basis of the criteria that tandem duplicated genes are placed within 10 predicted open reading frames or within 30 kb of each other as described previously (Shiu and Bleecker 2003). The exon-intron structures of the genes were determined using Gene structure display server (<http://gsds.cbi.pku.edu.cn/>) (Guo et al. 2007) by comparing the full-length CDSs (predicted coding sequence) of AtGLRs with their corresponding genomic sequences.

2.2.4 Estimation of Synonymous and Non-synonymous Substitution Rates and Divergence Time

For the calculation of the synonymous (d_N) and non-synonymous (d_S) substitution rates between the tandemly duplicated genes, ClustalW-based (Larkin et al. 2007) multiple sequence alignment was performed using their amino acid sequences. The aligned amino acid sequences and their corresponding original cDNA sequences were analysed using PAL2NAL (Suyama et al. 2006), which used the CODEML program of PAML (Yang 2007), to estimate d_N and d_S . The orthologous gene pairs (with 1:1 orthology) among *A. thaliana*, *A. lyrata* (L.) O'Kane & Al-Shehbaz and *Brassica rapa* L. as well as their corresponding d_N and d_S values have been retrieved using the BioMart tool (Kinsella et al. 2011) of EnsemblPlants database (Kersey et al. 2014). Time (million years ago, Mya) of duplication and divergence of each gene were calculated using a synonymous mutation rate of λ substitutions per synonymous site per year as $T = K_S/2\lambda$ (Puranik et al. 2013) where $\lambda = 6.1 \times 10^{-9}$ (Lynch and Conery 2000).

2.2.5 Expression analysis and functional annotation

The normalized expression values of the 20 AtGLR genes has been obtained from PLEXdb (Dash et al. 2012). For obtaining the expression profiles of these genes in different plant parts and developmental stages, normalized expression data of “Expression Atlas of *Arabidopsis* Development (AtGenExpress)” (accession No. AT40) (Schmid et al. 2005) has been downloaded from PLEXdb. To examine the expression pattern of AtGLRs under different biotic stresses, we have downloaded the data regarding “Transcriptome changes of *Arabidopsis* during pathogen and insect

attack” (accession No. AT49) (De Vos et al. 2005) and “Response to *Botrytis cinerea* infection” (accession No. AT51). In AT49, effect of gene expression of *A. thaliana* has been studied in response to different pathogens and insects namely *Pseudomonas syringae* pv. *tomato* (a pathogenic leaf bacterium), *Alternaria brassicicola* (a pathogenic leaf fungus), *Pieris rapae* (a tissue-chewing caterpillar), *Frankliniella occidentalis* (a cell-content-feeding thrip) and *Myzus persicae* (a phloem-feeding aphid). To examine the light dependent expression pattern, we have downloaded expression data of “Gene expression from *Arabidopsis* under high light conditions” (accession No. AT123) (González-Pérez et al. 2011). In this experiment, *Arabidopsis* cell suspension cultures were treated under control light (50microE/m²/s), high light (1800microE/m²/s) and dark conditions.

The tissue specificity index τ of the AtGLR genes was calculated following Yanai et al. (2005) as follows:

$$\tau = \frac{\sum_{i=1}^N (1-x_i)}{N-1},$$

where N is the number of tissues and x_i is the expression profile component normalized by the maximal component value. The index τ ranges from 0 to 1, with a higher value suggesting higher tissue specificity (Mukherjee et al. 2016). EN_c (effective number of codons) (Wright 1990) and GC_3 (Guanine and Cytosine content of the third codon position) were calculated from the coding sequences of the corresponding transcripts of the 20 AtGLR genes using CodonW (Peden 1999) obtained from <http://codonw.sourceforge.net/>.

The functional annotation of AtGLR gene sequences and subsequent analyses were performed using Blast2GO (Conesa et al. 2005; Conesa and Götz 2008). The gene sequences of AtGLRs were submitted into Blast2GO program to perform three subsequent steps namely (i) BLASTx search (Altschul et al. 1990) against the nucleotide database of NCBI, (ii) mapping of GO (Gene Ontology Consortium, 2004) terms associated with the BLAST results and (iii) annotation of GO terms associated with each query to relate the sequences to known function. Three categories of GO classification namely biological processes, molecular functions and cellular components were then analysed and graphically represented.

2.2.6 Homology modeling and structure validation

Secondary structure prediction was carried out with SOPMA (Geourjon and Deleage 1995). Three dimensional structures of AtGLR proteins were predicted by homology

modeling. To identify templates for homology modeling, the amino acid sequence of each protein was submitted to HHpred (Söding et al. 2005) with default parameters against the pdb70 database which is based on the protein data bank (PDB) (Berman et al. 2000) (<http://www.rcsb.org/>) which contains all publicly available 3D structures of proteins. On the basis of the results obtained from HHpred, the crystal structure of the AMPA subtype ionotropic glutamate receptor from *Rattus norvegicus* (PDB ID: 3KG2 chain A) at a resolution of 3.60Å (Sobolevsky et al. 2009) was used as the template for generating the 3D models of AtGLR proteins. The models were built using MODELLER version 9.10 auto-model program (Šali and Blundell 1993) within Chimera (Pettersen et al. 2004). Loop refinement of the modeled proteins was done with loop modeling protocol of MODELLER within Chimera with default parameters. Energy minimization was carried out with Gromos force field (van Gunsteren et al. 1996) in Swiss-PdbViewer (Guex and Peitsch 1997).

Superposition and calculation of the root mean square deviation (RMSD) of the modeled protein and template was performed using with Chimera (Pettersen et al. 2004) by fitting carbon backbones of both of the proteins. Model validation was carried out with PROCHECK (Laskowski et al. 1993). Using PROCHECK, computationally calculated phi/psi values and allowed and disallowed regions of the model were made using Ramachandran plot, which is mainly performed for predicting range of helices, strands and coils of furnished model. To estimate the evolutionary conservation of amino acid positions in the AtGLR proteins, we have generated a multiple sequence alignment of the 25 AtGLR proteins using ClustalW (Larkin et al. 2007). Then, we have provided this multiple sequence alignment file and the three-dimensional structure of AT1G05200.1 for the ConSurf analysis using the ConSurf server (Ashkenazy et al. 2010; Celniker et al. 2013; Glaser et al. 2003; Landau et al. 2005) (<http://consurf.tau.ac.il/>).

2.2.7 Surface-property analysis: hydrophobicity and electrostatics

Surface hydrophobicity was obtained using PLATINUM (Pyrkov et al. 2009) (<http://model.nmr.ru/platinum/>), which has been designed to calculate match or mismatch in receptor-ligand complexes (Efremov et al. 2007; Pyrkov et al. 2009). The program allows the calculation and visualization of the molecular hydrophobic/hydrophilic surface properties using the concept of molecular hydrophobicity potential (MHP) (Efremov et al. 2007; Legrand et al. 2011; Pyrkov et al. 2009). However, the total hydrophobic and hydrophilic surfaces for a molecule can

also be obtained with this tool (Legrand et al. 2011). Three dimensional structures of AtGLRs were submitted, and the PLATINUM website provided the calculated surface hydrophobicity.

Electrostatic potentials of the proteins were compared with the web-based version of PIPSA (Protein Interaction Property Similarity Analysis) (Blomberg, Gabdoulline et al. 1999; Gabdoulline et al. 2007; Richter et al. 2008; R. C. Wade et al. 2001). The models were superposed with Chimera and submitted to the PIPSA web server and Adaptive Poisson-Boltzmann Solver (APBS) software (Baker et al. 2001) was chosen to calculate the electrostatic potentials. For pair-wise comparisons of electrostatic potentials of two proteins, Hodgkin (Good et al. 1992; Hodgkin and Richards 1987) similarity index (SI) for two proteins (labeled 1 and 2) was used as follows:

$$SI_{12} = \frac{2(\mathbf{p}_1, \mathbf{p}_2)}{(\mathbf{p}_1, \mathbf{p}_1) + (\mathbf{p}_2, \mathbf{p}_2)},$$

where $(\mathbf{p}_1, \mathbf{p}_1)$, $(\mathbf{p}_1, \mathbf{p}_2)$ and $(\mathbf{p}_2, \mathbf{p}_2)$ are the scalar products of the electrostatic potentials over the region where the potentials are compared (Blomberg et al. 1999). If the electrostatic potentials (ϕ) are computed on a three-dimensional grid over the two superimposed molecules, the scalar product is

$$(\mathbf{p}_1, \mathbf{p}_2) = \sum_{i,j,k} \phi_1(i, j, k) \phi_2(i, j, k),$$

Where summation is carried out over the grid points (i,j,k) that are within the region of interest (Blomberg et al. 1999). The resulting similarity indices have been converted by PIPSA into distances as follows:

$$D_{a,b} = \sqrt{2 - 2SI_{a,b}}$$

where $D_{a,b}$ and $SI_{a,b}$ are the respective distance and similarity index of two proteins (labeled a and b) (Richter et al. 2008). Finally, a cluster analysis has been performed with PIPSA. Electrostatic surface of the proteins were calculated using APBS using PDB2PQR (Dolinsky et al. 2004) with default parameters. Visualization of the electrostatic potentials on the molecular surfaces of the proteins was produced using PyMOL (Schrödinger, LLC, 2010).

Statistical analysis

Wilcoxon Rank Sum test (Wilcoxon 1945) which is equivalent to Mann-Whitney *U* test (Mann and Whitney 1947) was used for pair-wise comparisons of different variables since the values were not normally distributed in our dataset. For correlation analysis, we performed the Spearman's rank correlation coefficient ρ (Spearman

1904), where the significant correlations were denoted by $P < 0.05$. All these statistical analyses were performed with R package (R Core Team 2013).

2.3 Results and discussion

2.3.1 Physico-chemical properties and domain organization of AtGLR family

The gene length of the AtGLRs ranged from 3002 bp (AT5G48400) to 4603 bp (AT2G29120) with an average ($\pm$ SE) of 3686.75 ($\pm$ 105.48) bp. The length of the coding sequence ranged from 2253 bp (AT5G48400.1) to 2880 bp (both variants of AT1G05200) with an average ($\pm$ SE) of 2694.36 ($\pm$ 30.2) bp. The length of the proteins ranged from 750 amino acids (AT5G48400.1) to 959 amino acids (both variants of AT1G05200) with an average ($\pm$ SE) of 897.12 ($\pm$ 10.07) amino acids (Table 2.1).

Table 2.1: Details of the AtGLR genes and proteins used in this study

Gene name	Gene stable ID	Gene start (bp)	Gene end (bp)	Gene length (bp)	Transcript stable ID	Coding sequence length (bp)	Protein stable ID	Amino acid length
ATGLR3.4	AT1G05200	1,504,960	1,509,152	4,192	AT1G05200.1	2,880	AT1G05200.1	959
					AT1G05200.2	2,880	AT1G05200.2	959
ATGLR3.3	AT1G42540	15,973,054	15,976,703	3,649	AT1G42540.1	2,802	AT1G42540.1	933
ATGLR3.1	AT2G17260	7,504,756	7,508,394	3,638	AT2G17260.1	2,856	AT2G17260.1	951
ATGLR2.3	AT2G24710	10,516,880	10,520,549	3,669	AT2G24710.1	2,688	AT2G24710.1	895
ATGLR2.2	AT2G24720	10,522,783	10,525,840	3,057	AT2G24720.1	2,763	AT2G24720.1	920
ATGLR2.9	AT2G29100	12,501,092	12,504,912	3,820	AT2G29100.1	2,823	AT2G29100.1	940
ATGLR2.8	AT2G29110	12,506,880	12,510,552	3,672	AT2G29110.1	2,844	AT2G29110.1	947
ATGLR2.7	AT2G29120	12,511,292	12,515,895	4,603	AT2G29120.1	2,859	AT2G29120.1	952
ATGLR3.5	AT2G32390	13,748,441	13,751,933	3,492	AT2G32390.1	2,688	AT2G32390.1	895
					AT2G32390.2	2,556	AT2G32390.2	851
					AT2G32390.3	2,697	AT2G32390.3	898
ATGLR3.7	AT2G32400	13,752,584	13,756,853	4,269	AT2G32400.1	2,766	AT2G32400.1	921
ATGLR1.1	AT3G04110	1,077,328	1,080,405	3,077	AT3G04110.1	2,427	AT3G04110.1	808
ATGLR1.4	AT3G07520	2,394,891	2,398,291	3,400	AT3G07520.1	2,586	AT3G07520.1	861
ATGLR3.6	AT3G51480	19,100,676	19,104,825	4,149	AT3G51480.1	2,712	AT3G51480.1	903
ATGLR2.4	AT4G31710	15,349,121	15,352,962	3,841	AT4G31710.1	2,697	AT4G31710.1	898
ATGLR3.2	AT4G35290	16,789,549	16,793,685	4,136	AT4G35290.1	2,739	AT4G35290.1	912
					AT4G35290.2	2,739	AT4G35290.2	912
ATGLR2.6	AT5G11180	3,557,261	3,561,575	4,314	AT5G11180.1	2,721	AT5G11180.1	906
ATGLR2.5	AT5G11210	3,571,214	3,574,537	3,323	AT5G11210.1	2,490	AT5G11210.1	829
ATGLR2.1	AT5G27100	9,535,059	9,538,333	3,274	AT5G27100.1	2,706	AT5G27100.1	901
ATGLR1.2	AT5G48400	19,616,244	19,619,246	3,002	AT5G48400.1	2,253	AT5G48400.1	750
					AT5G48400.2	2,604	AT5G48400.2	867
ATGLR1.3	AT5G48410	19,620,267	19,623,425	3,158	AT5G48410.1	2,583	AT5G48410.1	860

The molecular weight of the proteins ranged from 84.29 kDa (AT5G48400.1) to 107.41 kDa (AT2G29120.1) with an average ($\pm$ SE) of 100.71 (1.11) kDa. To further

understand the nature of ATGLRs proteins, their physicochemical properties were analyzed (Table 2.2).

Table 2.2: Physico-chemical properties of AtGLR proteins

Protein name	Molecular weight (kD)	Theoretical pI	GRAVY	Percent disorder
AT1G05200.1	107.21	8.45	-0.07	7.090719499
AT1G05200.2	107.21	8.45	-0.07	7.090719499
AT1G42540.1	104.69	8.92	-0.037	4.608788853
AT2G17260.1	106.18	7.97	0.066	4.100946372
AT2G24710.1	100.81	7.81	0.077	2.122905028
AT2G24720.1	102.85	7.51	0.033	0.652173913
AT2G29100.1	106.27	6.63	-0.107	5.319148936
AT2G29110.1	107.19	6.45	-0.094	1.583949314
AT2G29120.1	107.41	6.98	-0.051	4.831932773
AT2G32390.1	100.93	7.99	-0.119	4.581005587
AT2G32390.2	95.9	7.17	-0.141	4.81786134
AT2G32390.3	99.95	6.44	-0.143	4.565701559
AT2G32400.1	103.51	7.66	0.02	2.823018458
AT3G04110.1	90.51	8.69	0.014	1.608910891
AT3G07520.1	97.76	8.9	-0.075	2.322880372
AT3G51480.1	100.53	7.64	0.112	0.996677741
AT4G31710.1	99.93	8.36	0.058	1.113585746
AT4G35290.1	101.66	8.03	0.028	2.302631579
AT4G35290.2	101.66	8.03	0.028	2.302631579
AT5G11180.1	101.71	6.64	-0.137	9.602649007
AT5G11210.1	93.47	5.67	-0.052	3.860072376
AT5G27100.1	100.9	5.87	0.066	3.218645949
AT5G48400.1	84.29	6.17	0.051	0.533333333
AT5G48400.2	98.15	6.78	-0.018	0.92272203
AT5G48410.1	97.02	6.89	0.022	1.279069767
Average	100.71	7.44	-0.022	3.37010726
SE	1.11	0.19	0.015	0.461

The GRAVY value for a peptide is the sum of hydropathy values (Kyte and Doolittle 1982). Maximum and minimum GRAVY score were recorded as 0.112 (AT3G51480.1) and -0.143 (AT2G32390.3) with an average ($\pm$ SE) value of -0.022 ($\pm$ 0.015). Twelve of these proteins showed positive GRAVY scores while the rest showed negative values. Kyte and Doolittle (1982) proposed that GRAVY score of membrane bound proteins are higher than the mean value of sequenced soluble proteins (-0.4). Marmagne et al. (2004) reported that The GRAVY score of the *Arabidopsis* plasma membrane proteins analyzed ranged from -1.6 to +1. Our results, thus, confirmed the membrane-protein nature of AtGLRs. The value of pI ranged from 5.67 (AT5G11210.1) and 8.92 (AT1G42540.1) with an average ($\pm$ SE) of 7.44 ($\pm$ 0.19). The amino acid composition of the AtGLR proteins showed the abundance of leucine, valine and serine while scarcity of cysteine (Figure 2.1).

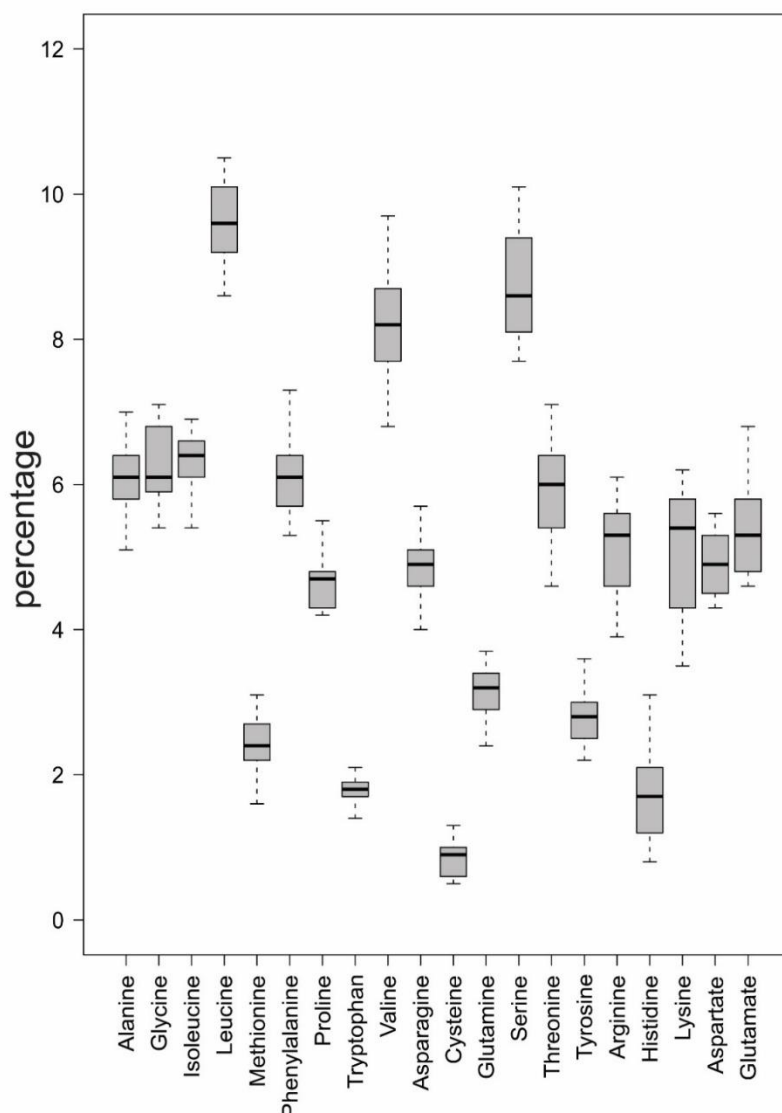


Figure 2.1: Boxplot showing percentages of different amino acids in AtGLR proteins.

The disorder content ranged from 0.53% (AT5G48400.1) to 9.6% (AT5G11180.1) with an average ($\pm$ SE) of 3.37% ($\pm$ 0.461). The disorder content showed significant negative correlation with GRAVY (Spearman's $\rho = -0.626$, $P = 8.1 \times 10^{-4}$). This is probably due to the tight packing of transmembrane part of the AtGLRs with the membrane which makes them more structured. Moreover, this also indicates that the ligand binding domains are also structured. Naturally, the thermostability is also negatively associated with disorder.

Several phosphorylation, myristoylation and Asn-glycosylation sites have also been found (Table 2.3).

Table 2.3: List and position of different phosphorylation, myristoylation and Asn-glycosylation sites found in 25 AtGLR proteins.

Protein name	Type of site	Start	Stop
AT1G05200.1	Asn-glycosylation	38	41
	Asn-glycosylation	42	45
	Asn-glycosylation	108	111
	Asn-glycosylation	365	368
	Asn-glycosylation	378	381
	Asn-glycosylation	404	407
	Asn-glycosylation	443	446
	Asn-glycosylation	461	464
	Asn-glycosylation	576	579
	Asn-glycosylation	689	692
	cAMP- and cGMP-dependent protein kinase phosphorylation	219	222
	CK2 (Caesin kinase II) phosphorylation	174	177
	CK2 (Caesin kinase II) phosphorylation	238	241
	CK2 (Caesin kinase II) phosphorylation	290	293
	CK2 (Caesin kinase II) phosphorylation	503	506
	CK2 (Caesin kinase II) phosphorylation	664	667
	CK2 (Caesin kinase II) phosphorylation	707	710
	CK2 (Caesin kinase II) phosphorylation	717	720
	CK2 (Caesin kinase II) phosphorylation	846	849
	CK2 (Caesin kinase II) phosphorylation	891	894
	CK2 (Caesin kinase II) phosphorylation	914	917
	Myristoylation	64	69
	Myristoylation	97	102
	Myristoylation	114	119
	Myristoylation	233	238
	Myristoylation	274	279
	Myristoylation	402	407
	Myristoylation	723	728
	Myristoylation	765	770
	Myristoylation	792	797
	Myristoylation	949	954
	Protein kinase C phosphorylation	50	52
	Protein kinase C phosphorylation	225	227
	Protein kinase C phosphorylation	238	240
	Protein kinase C phosphorylation	320	322
	Protein kinase C phosphorylation	416	418
	Protein kinase C phosphorylation	575	577
	Protein kinase C phosphorylation	664	666
	Protein kinase C phosphorylation	701	703
	Protein kinase C phosphorylation	706	708
	Protein kinase C phosphorylation	854	856
	Protein kinase C phosphorylation	906	908
	Protein kinase C phosphorylation	914	916
	Protein kinase C phosphorylation	937	939

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT1G05200.2	Asn-glycosylation	38	41
	Asn-glycosylation	42	45
	Asn-glycosylation	108	111
	Asn-glycosylation	365	368
	Asn-glycosylation	378	381
	Asn-glycosylation	404	407
	Asn-glycosylation	443	446
	Asn-glycosylation	461	464
	Asn-glycosylation	576	579
	Asn-glycosylation	689	692
	cAMP- and cGMP-dependent protein kinase phosphorylation	219	222
	CK2 (Caesin kinase II) phosphorylation	174	177
	CK2 (Caesin kinase II) phosphorylation	238	241
	CK2 (Caesin kinase II) phosphorylation	290	293
	CK2 (Caesin kinase II) phosphorylation	503	506
	CK2 (Caesin kinase II) phosphorylation	664	667
	CK2 (Caesin kinase II) phosphorylation	707	710
	CK2 (Caesin kinase II) phosphorylation	717	720
	CK2 (Caesin kinase II) phosphorylation	846	849
	CK2 (Caesin kinase II) phosphorylation	891	894
	CK2 (Caesin kinase II) phosphorylation	914	917
	Myristoylation	64	69
	Myristoylation	97	102
	Myristoylation	114	119
	Myristoylation	233	238
	Myristoylation	274	279
	Myristoylation	402	407
	Myristoylation	723	728
	Myristoylation	765	770
	Myristoylation	792	797
	Myristoylation	949	954
	Protein kinase C phosphorylation	50	52
	Protein kinase C phosphorylation	225	227
	Protein kinase C phosphorylation	238	240
	Protein kinase C phosphorylation	320	322
	Protein kinase C phosphorylation	416	418
	Protein kinase C phosphorylation	575	577
	Protein kinase C phosphorylation	664	666
	Protein kinase C phosphorylation	701	703
	Protein kinase C phosphorylation	706	708
	Protein kinase C phosphorylation	854	856
	Protein kinase C phosphorylation	906	908
	Protein kinase C phosphorylation	914	916
	Protein kinase C phosphorylation	937	939

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT1G42540.1	Asn-glycosylation	76	79
	Asn-glycosylation	336	339
	Asn-glycosylation	341	344
	Asn-glycosylation	415	418
	Asn-glycosylation	434	437
	Asn-glycosylation	549	552
	Asn-glycosylation	662	665
	Asn-glycosylation	712	715
	Asn-glycosylation	909	912
	cAMP- and cGMP-dependent protein kinase phosphorylation	300	303
	cAMP- and cGMP-dependent protein kinase phosphorylation	857	860
	CK2 (Caesin kinase II) phosphorylation	21	24
	CK2 (Caesin kinase II) phosphorylation	58	61
	CK2 (Caesin kinase II) phosphorylation	142	145
	CK2 (Caesin kinase II) phosphorylation	154	157
	CK2 (Caesin kinase II) phosphorylation	260	263
	CK2 (Caesin kinase II) phosphorylation	360	363
	CK2 (Caesin kinase II) phosphorylation	475	478
	CK2 (Caesin kinase II) phosphorylation	611	614
	CK2 (Caesin kinase II) phosphorylation	687	690
	CK2 (Caesin kinase II) phosphorylation	701	704
	CK2 (Caesin kinase II) phosphorylation	887	890
	CK2 (Caesin kinase II) phosphorylation	895	898
	CK2 (Caesin kinase II) phosphorylation	907	910
	Myristoylation	32	37
	Myristoylation	65	70
	Myristoylation	82	87
	Myristoylation	244	249
	Myristoylation	304	309
	Myristoylation	334	339
	Myristoylation	535	540
	Myristoylation	722	727
	Myristoylation	738	743
	Myristoylation	764	769
	Myristoylation	913	918
	Protein kinase C phosphorylation	23	25
	Protein kinase C phosphorylation	187	189
	Protein kinase C phosphorylation	193	195
	Protein kinase C phosphorylation	438	440
	Protein kinase C phosphorylation	475	477
	Protein kinase C phosphorylation	548	550
	Protein kinase C phosphorylation	687	689
	Protein kinase C phosphorylation	821	823
	Protein kinase C phosphorylation	876	878
	Protein kinase C phosphorylation	879	881
	Protein kinase C phosphorylation	899	901
	Protein kinase C phosphorylation	914	916
	Tyrosine kinase phosphorylation	689	697

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT2G17260.1	Asn-glycosylation	335	338
	Asn-glycosylation	367	370
	Asn-glycosylation	385	388
	Asn-glycosylation	445	448
	Asn-glycosylation	463	466
	Asn-glycosylation	514	517
	Asn-glycosylation	764	767
	Asn-glycosylation	838	841
	CK2 (Caesin kinase II) phosphorylation	155	158
	CK2 (Caesin kinase II) phosphorylation	237	240
	CK2 (Caesin kinase II) phosphorylation	292	295
	CK2 (Caesin kinase II) phosphorylation	360	363
	CK2 (Caesin kinase II) phosphorylation	390	393
	CK2 (Caesin kinase II) phosphorylation	505	508
	CK2 (Caesin kinase II) phosphorylation	564	567
	CK2 (Caesin kinase II) phosphorylation	668	671
	CK2 (Caesin kinase II) phosphorylation	732	735
	CK2 (Caesin kinase II) phosphorylation	754	757
	Myristoylation	45	50
	Myristoylation	51	56
	Myristoylation	63	68
	Myristoylation	365	370
	Myristoylation	377	382
	Myristoylation	765	770
	Myristoylation	791	796
	Myristoylation	819	824
	Myristoylation	844	849
	Protein kinase C phosphorylation	54	56
	Protein kinase C phosphorylation	224	226
	Protein kinase C phosphorylation	264	266
	Protein kinase C phosphorylation	310	312
	Protein kinase C phosphorylation	318	320
	Protein kinase C phosphorylation	378	380
	Protein kinase C phosphorylation	483	485
	Protein kinase C phosphorylation	505	507
	Protein kinase C phosphorylation	517	519
	Protein kinase C phosphorylation	579	581
	Protein kinase C phosphorylation	668	670
	Protein kinase C phosphorylation	723	725
	Protein kinase C phosphorylation	746	748
	Protein kinase C phosphorylation	899	901
	Protein kinase C phosphorylation	902	904
	Protein kinase C phosphorylation	922	924
	Protein kinase C phosphorylation	947	949
	Tyrosine kinase phosphorylation	181	189

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT2G24710.1	Asn-glycosylation	52	55
	Asn-glycosylation	203	206
	Asn-glycosylation	266	269
	Asn-glycosylation	330	333
	Asn-glycosylation	342	345
	Asn-glycosylation	477	480
	Asn-glycosylation	542	545
	Asn-glycosylation	848	851
	Asn-glycosylation	885	888
	cAMP- and cGMP-dependent protein kinase phosphorylation	879	882
	CK2 (Caesin kinase II) phosphorylation	29	32
	CK2 (Caesin kinase II) phosphorylation	37	40
	CK2 (Caesin kinase II) phosphorylation	54	57
	CK2 (Caesin kinase II) phosphorylation	76	79
	CK2 (Caesin kinase II) phosphorylation	144	147
	CK2 (Caesin kinase II) phosphorylation	174	177
	CK2 (Caesin kinase II) phosphorylation	205	208
	CK2 (Caesin kinase II) phosphorylation	268	271
	CK2 (Caesin kinase II) phosphorylation	344	347
	CK2 (Caesin kinase II) phosphorylation	384	387
	CK2 (Caesin kinase II) phosphorylation	424	427
	CK2 (Caesin kinase II) phosphorylation	545	548
	CK2 (Caesin kinase II) phosphorylation	681	684
	CK2 (Caesin kinase II) phosphorylation	716	719
	CK2 (Caesin kinase II) phosphorylation	807	810
	CK2 (Caesin kinase II) phosphorylation	813	816
	CK2 (Caesin kinase II) phosphorylation	860	863
	Myristoylation	23	28
	Myristoylation	82	87
	Myristoylation	731	736
	Myristoylation	748	753
	Protein kinase C phosphorylation	3	5
	Protein kinase C phosphorylation	134	136
	Protein kinase C phosphorylation	339	341
	Protein kinase C phosphorylation	424	426
	Protein kinase C phosphorylation	580	582
	Protein kinase C phosphorylation	669	671
	Protein kinase C phosphorylation	753	755
	Protein kinase C phosphorylation	790	792
	Protein kinase C phosphorylation	822	824

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT2G24720.1	Asn-glycosylation	53	56
	Asn-glycosylation	204	207
	Asn-glycosylation	267	270
	Asn-glycosylation	331	334
	Asn-glycosylation	342	345
	Asn-glycosylation	477	480
	Asn-glycosylation	542	545
	Asn-glycosylation	702	705
	Asn-glycosylation	896	899
	Asn-glycosylation	913	916
	cAMP- and cGMP-dependent protein kinase phosphorylation	868	871
	CK2 (Caesin kinase II) phosphorylation	43	46
	CK2 (Caesin kinase II) phosphorylation	55	58
	CK2 (Caesin kinase II) phosphorylation	77	80
	CK2 (Caesin kinase II) phosphorylation	145	148
	CK2 (Caesin kinase II) phosphorylation	175	178
	CK2 (Caesin kinase II) phosphorylation	187	190
	CK2 (Caesin kinase II) phosphorylation	206	209
	CK2 (Caesin kinase II) phosphorylation	265	268
	CK2 (Caesin kinase II) phosphorylation	269	272
	CK2 (Caesin kinase II) phosphorylation	335	338
	CK2 (Caesin kinase II) phosphorylation	344	347
	CK2 (Caesin kinase II) phosphorylation	384	387
	CK2 (Caesin kinase II) phosphorylation	424	427
	CK2 (Caesin kinase II) phosphorylation	545	548
	CK2 (Caesin kinase II) phosphorylation	716	719
	CK2 (Caesin kinase II) phosphorylation	807	810
	CK2 (Caesin kinase II) phosphorylation	813	816
	CK2 (Caesin kinase II) phosphorylation	915	918
	Myristoylation	24	29
	Myristoylation	83	88
	Myristoylation	258	263
	Myristoylation	348	353
	Myristoylation	420	425
	Myristoylation	528	533
	Myristoylation	731	736
	Myristoylation	748	753
	Protein kinase C phosphorylation	21	23
	Protein kinase C phosphorylation	61	63
	Protein kinase C phosphorylation	135	137
	Protein kinase C phosphorylation	292	294
	Protein kinase C phosphorylation	339	341
	Protein kinase C phosphorylation	753	755
	Protein kinase C phosphorylation	790	792
	Protein kinase C phosphorylation	891	893

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT2G29100.1	Asn-glycosylation	4	7
	Asn-glycosylation	25	28
	Asn-glycosylation	39	42
	Asn-glycosylation	115	118
	Asn-glycosylation	338	341
	Asn-glycosylation	345	348
	Asn-glycosylation	528	531
	Asn-glycosylation	771	774
	Asn-glycosylation	776	779
	Asn-glycosylation	805	808
	Asn-glycosylation	875	878
	cAMP- and cGMP-dependent protein kinase phosphorylation	908	911
	CK2 (Caesin kinase II) phosphorylation	75	78
	CK2 (Caesin kinase II) phosphorylation	143	146
	CK2 (Caesin kinase II) phosphorylation	303	306
	CK2 (Caesin kinase II) phosphorylation	406	409
	CK2 (Caesin kinase II) phosphorylation	702	705
	CK2 (Caesin kinase II) phosphorylation	838	841
	CK2 (Caesin kinase II) phosphorylation	913	916
	CK2 (Caesin kinase II) phosphorylation	932	935
	Myristoylation	19	24
	Myristoylation	339	344
	Myristoylation	410	415
	Myristoylation	717	722
	Myristoylation	817	822
	Protein kinase C phosphorylation	49	51
	Protein kinase C phosphorylation	133	135
	Protein kinase C phosphorylation	303	305
	Protein kinase C phosphorylation	386	388
	Protein kinase C phosphorylation	406	408
	Protein kinase C phosphorylation	417	419
	Protein kinase C phosphorylation	467	469
	Protein kinase C phosphorylation	702	704
	Protein kinase C phosphorylation	745	747
	Protein kinase C phosphorylation	801	803
	Protein kinase C phosphorylation	867	869
	Protein kinase C phosphorylation	881	883
	Tyrosine kinase phosphorylation	670	677

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT2G29110.1	Asn-glycosylation	7	10
	Asn-glycosylation	42	45
	Asn-glycosylation	118	121
	Asn-glycosylation	333	336
	Asn-glycosylation	341	344
	Asn-glycosylation	348	351
	Asn-glycosylation	420	423
	Asn-glycosylation	478	481
	Asn-glycosylation	524	527
	Asn-glycosylation	659	662
	Asn-glycosylation	704	707
	Asn-glycosylation	723	726
	Asn-glycosylation	779	782
	cAMP- and cGMP-dependent protein kinase phosphorylation	857	860
	cAMP- and cGMP-dependent protein kinase phosphorylation	916	919
	CK2 (Caesin kinase II) phosphorylation	78	81
	CK2 (Caesin kinase II) phosphorylation	146	149
	CK2 (Caesin kinase II) phosphorylation	363	366
	CK2 (Caesin kinase II) phosphorylation	510	513
	CK2 (Caesin kinase II) phosphorylation	713	716
	CK2 (Caesin kinase II) phosphorylation	781	784
	CK2 (Caesin kinase II) phosphorylation	846	849
	CK2 (Caesin kinase II) phosphorylation	921	924
	CK2 (Caesin kinase II) phosphorylation	939	942
	Myristoylation	24	29
	Myristoylation	344	349
	Myristoylation	525	530
	Myristoylation	724	729
	Protein kinase C phosphorylation	78	80
	Protein kinase C phosphorylation	136	138
	Protein kinase C phosphorylation	389	391
	Protein kinase C phosphorylation	419	421
	Protein kinase C phosphorylation	481	483
	Protein kinase C phosphorylation	753	755
	Protein kinase C phosphorylation	809	811
	Protein kinase C phosphorylation	813	815
	Protein kinase C phosphorylation	875	877
	Tyrosine kinase phosphorylation	144	151
	Tyrosine kinase phosphorylation	681	688

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT2G29120.1	Asn-glycosylation	10	13
	Asn-glycosylation	34	37
	Asn-glycosylation	60	63
	Asn-glycosylation	355	358
	Asn-glycosylation	428	431
	Asn-glycosylation	546	549
	Asn-glycosylation	664	667
	Asn-glycosylation	728	731
	Asn-glycosylation	748	751
	Asn-glycosylation	784	787
	Asn-glycosylation	809	812
	Asn-glycosylation	888	891
	cAMP- and cGMP-dependent protein kinase phosphorylation	921	924
	CK2 (Caesin kinase II) phosphorylation	62	65
	CK2 (Caesin kinase II) phosphorylation	84	87
	CK2 (Caesin kinase II) phosphorylation	152	155
	CK2 (Caesin kinase II) phosphorylation	277	280
	CK2 (Caesin kinase II) phosphorylation	397	400
	CK2 (Caesin kinase II) phosphorylation	517	520
	CK2 (Caesin kinase II) phosphorylation	718	721
	CK2 (Caesin kinase II) phosphorylation	734	737
	CK2 (Caesin kinase II) phosphorylation	753	756
	CK2 (Caesin kinase II) phosphorylation	786	789
	CK2 (Caesin kinase II) phosphorylation	851	854
	CK2 (Caesin kinase II) phosphorylation	927	930
	CK2 (Caesin kinase II) phosphorylation	934	937
	Myristoylation	32	37
	Myristoylation	353	358
	Myristoylation	392	397
	Myristoylation	420	425
	Myristoylation	532	537
	Myristoylation	729	734
	Protein kinase C phosphorylation	74	76
	Protein kinase C phosphorylation	295	297
	Protein kinase C phosphorylation	343	345
	Protein kinase C phosphorylation	670	672
	Protein kinase C phosphorylation	749	751
	Protein kinase C phosphorylation	758	760
	Protein kinase C phosphorylation	859	861

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT2G32390.1	Asn-glycosylation	165	168
	Asn-glycosylation	313	316
	Asn-glycosylation	339	342
	Asn-glycosylation	378	381
	Asn-glycosylation	396	399
	Asn-glycosylation	511	514
	Asn-glycosylation	624	627
	CK2 (Caesin kinase II) phosphorylation	4	7
	CK2 (Caesin kinase II) phosphorylation	103	106
	CK2 (Caesin kinase II) phosphorylation	167	170
	CK2 (Caesin kinase II) phosphorylation	219	222
	CK2 (Caesin kinase II) phosphorylation	398	401
	CK2 (Caesin kinase II) phosphorylation	599	602
	CK2 (Caesin kinase II) phosphorylation	642	645
	CK2 (Caesin kinase II) phosphorylation	781	784
	CK2 (Caesin kinase II) phosphorylation	849	852
	CK2 (Caesin kinase II) phosphorylation	881	884
	Myristoylation	162	167
	Myristoylation	203	208
	Myristoylation	298	303
	Myristoylation	316	321
	Myristoylation	658	663
	Myristoylation	699	704
	Myristoylation	727	732
	Myristoylation	838	843
	Myristoylation	879	884
	Protein kinase C phosphorylation	35	37
	Protein kinase C phosphorylation	154	156
	Protein kinase C phosphorylation	266	268
	Protein kinase C phosphorylation	308	310
	Protein kinase C phosphorylation	351	353
	Protein kinase C phosphorylation	398	400
	Protein kinase C phosphorylation	438	440
	Protein kinase C phosphorylation	510	512
	Protein kinase C phosphorylation	599	601
	Protein kinase C phosphorylation	641	643
	Protein kinase C phosphorylation	839	841
	Protein kinase C phosphorylation	844	846
	Protein kinase C phosphorylation	849	851
	Protein kinase C phosphorylation	872	874

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT2G32390.2	Asn-glycosylation	121	124
	Asn-glycosylation	269	272
	Asn-glycosylation	295	298
	Asn-glycosylation	334	337
	Asn-glycosylation	352	355
	Asn-glycosylation	467	470
	Asn-glycosylation	580	583
	CK2 (Caesin kinase II) phosphorylation	59	62
	CK2 (Caesin kinase II) phosphorylation	123	126
	CK2 (Caesin kinase II) phosphorylation	175	178
	CK2 (Caesin kinase II) phosphorylation	354	357
	CK2 (Caesin kinase II) phosphorylation	555	558
	CK2 (Caesin kinase II) phosphorylation	598	601
	CK2 (Caesin kinase II) phosphorylation	737	740
	CK2 (Caesin kinase II) phosphorylation	805	808
	CK2 (Caesin kinase II) phosphorylation	837	840
	Myristoylation	118	123
	Myristoylation	159	164
	Myristoylation	254	259
	Myristoylation	272	277
	Myristoylation	614	619
	Myristoylation	655	660
	Myristoylation	683	688
	Myristoylation	794	799
	Myristoylation	835	840
	Protein kinase C phosphorylation	110	112
	Protein kinase C phosphorylation	222	224
	Protein kinase C phosphorylation	264	266
	Protein kinase C phosphorylation	307	309
	Protein kinase C phosphorylation	354	356
	Protein kinase C phosphorylation	394	396
	Protein kinase C phosphorylation	466	468
	Protein kinase C phosphorylation	555	557
	Protein kinase C phosphorylation	597	599
	Protein kinase C phosphorylation	795	797
	Protein kinase C phosphorylation	800	802
	Protein kinase C phosphorylation	805	807
	Protein kinase C phosphorylation	828	830

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT2G32390.3	Asn-glycosylation	33	36
	Asn-glycosylation	90	93
	Asn-glycosylation	218	221
	Asn-glycosylation	366	369
	Asn-glycosylation	392	395
	Asn-glycosylation	431	434
	Asn-glycosylation	449	452
	Asn-glycosylation	564	567
	CK2 (Caesin kinase II) phosphorylation	156	159
	CK2 (Caesin kinase II) phosphorylation	220	223
	CK2 (Caesin kinase II) phosphorylation	272	275
	CK2 (Caesin kinase II) phosphorylation	451	454
	CK2 (Caesin kinase II) phosphorylation	645	648
	CK2 (Caesin kinase II) phosphorylation	784	787
	CK2 (Caesin kinase II) phosphorylation	852	855
	CK2 (Caesin kinase II) phosphorylation	884	887
	Myristoylation	46	51
	Myristoylation	79	84
	Myristoylation	96	101
	Myristoylation	215	220
	Myristoylation	256	261
	Myristoylation	351	356
	Myristoylation	369	374
	Myristoylation	661	666
	Myristoylation	702	707
	Myristoylation	730	735
	Myristoylation	841	846
	Myristoylation	882	887
	Protein kinase C phosphorylation	207	209
	Protein kinase C phosphorylation	319	321
	Protein kinase C phosphorylation	361	363
	Protein kinase C phosphorylation	404	406
	Protein kinase C phosphorylation	451	453
	Protein kinase C phosphorylation	491	493
	Protein kinase C phosphorylation	563	565
	Protein kinase C phosphorylation	644	646
	Protein kinase C phosphorylation	842	844
	Protein kinase C phosphorylation	847	849
	Protein kinase C phosphorylation	852	854
	Protein kinase C phosphorylation	875	877

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT2G32400.1	Asn-glycosylation	214	217
	Asn-glycosylation	300	303
	Asn-glycosylation	330	333
	Asn-glycosylation	369	372
	Asn-glycosylation	396	399
	Asn-glycosylation	478	481
	Asn-glycosylation	568	571
	Asn-glycosylation	660	663
	cAMP- and cGMP-dependent protein kinase phosphorylation	189	192
	cAMP- and cGMP-dependent protein kinase phosphorylation	424	427
	cAMP- and cGMP-dependent protein kinase phosphorylation	465	468
	cAMP- and cGMP-dependent protein kinase phosphorylation	770	773
	CK2 (Caesin kinase II) phosphorylation	110	113
	CK2 (Caesin kinase II) phosphorylation	144	147
	CK2 (Caesin kinase II) phosphorylation	171	174
	CK2 (Caesin kinase II) phosphorylation	181	184
	CK2 (Caesin kinase II) phosphorylation	264	267
	CK2 (Caesin kinase II) phosphorylation	334	337
	CK2 (Caesin kinase II) phosphorylation	468	471
	CK2 (Caesin kinase II) phosphorylation	675	678
	CK2 (Caesin kinase II) phosphorylation	715	718
	Myristoylation	4	9
	Myristoylation	34	39
	Myristoylation	84	89
	Myristoylation	328	333
	Myristoylation	380	385
	Myristoylation	510	515
	Myristoylation	528	533
	Myristoylation	704	709
	Myristoylation	731	736
	Protein kinase C phosphorylation	195	197
	Protein kinase C phosphorylation	266	268
	Protein kinase C phosphorylation	270	272
	Protein kinase C phosphorylation	287	289
	Protein kinase C phosphorylation	294	296
	Protein kinase C phosphorylation	299	301
	Protein kinase C phosphorylation	336	338
	Protein kinase C phosphorylation	381	383
	Protein kinase C phosphorylation	422	424
	Protein kinase C phosphorylation	480	482
	Protein kinase C phosphorylation	578	580
	Protein kinase C phosphorylation	679	681
	Protein kinase C phosphorylation	789	791
	Protein kinase C phosphorylation	822	824
	Protein kinase C phosphorylation	871	873

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT3G04110.1	Asn-glycosylation	288	291
	Asn-glycosylation	339	342
	Asn-glycosylation	504	507
	Asn-glycosylation	622	625
	cAMP- and cGMP-dependent protein kinase phosphorylation	526	529
	cAMP- and cGMP-dependent protein kinase phosphorylation	763	766
	CK2 (Caesin kinase II) phosphorylation	22	25
	CK2 (Caesin kinase II) phosphorylation	75	78
	CK2 (Caesin kinase II) phosphorylation	103	106
	CK2 (Caesin kinase II) phosphorylation	112	115
	CK2 (Caesin kinase II) phosphorylation	145	148
	CK2 (Caesin kinase II) phosphorylation	202	205
	CK2 (Caesin kinase II) phosphorylation	301	304
	CK2 (Caesin kinase II) phosphorylation	487	490
	CK2 (Caesin kinase II) phosphorylation	534	537
	CK2 (Caesin kinase II) phosphorylation	747	750
	CK2 (Caesin kinase II) phosphorylation	754	757
	CK2 (Caesin kinase II) phosphorylation	766	769
	CK2 (Caesin kinase II) phosphorylation	792	795
	Myristoylation	59	64
	Myristoylation	189	194
	Myristoylation	386	391
	Myristoylation	476	481
	Myristoylation	574	579
	Myristoylation	641	646
	Myristoylation	652	657
	Myristoylation	684	689
	Protein kinase C phosphorylation	114	116
	Protein kinase C phosphorylation	131	133
	Protein kinase C phosphorylation	160	162
	Protein kinase C phosphorylation	254	256
	Protein kinase C phosphorylation	290	292
	Protein kinase C phosphorylation	331	333
	Protein kinase C phosphorylation	347	349
	Protein kinase C phosphorylation	394	396
	Protein kinase C phosphorylation	405	407
	Protein kinase C phosphorylation	438	440
	Protein kinase C phosphorylation	487	489
	Protein kinase C phosphorylation	503	505
	Protein kinase C phosphorylation	524	526
	Protein kinase C phosphorylation	600	602
	Protein kinase C phosphorylation	625	627
	Protein kinase C phosphorylation	648	650
	Protein kinase C phosphorylation	693	695
	Protein kinase C phosphorylation	762	764
	Protein kinase C phosphorylation	766	768
	Tyrosine kinase phosphorylation	475	483

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT3G07520.1	Asn-glycosylation	306	309
	Asn-glycosylation	534	537
	Asn-glycosylation	627	630
	Asn-glycosylation	646	649
	Asn-glycosylation	707	710
	cAMP- and cGMP-dependent protein kinase phosphorylation	411	414
	CK2 (Caesin kinase II) phosphorylation	91	94
	CK2 (Caesin kinase II) phosphorylation	119	122
	CK2 (Caesin kinase II) phosphorylation	162	165
	CK2 (Caesin kinase II) phosphorylation	174	177
	CK2 (Caesin kinase II) phosphorylation	218	221
	CK2 (Caesin kinase II) phosphorylation	351	354
	CK2 (Caesin kinase II) phosphorylation	436	439
	CK2 (Caesin kinase II) phosphorylation	517	520
	CK2 (Caesin kinase II) phosphorylation	530	533
	CK2 (Caesin kinase II) phosphorylation	554	557
	CK2 (Caesin kinase II) phosphorylation	697	700
	CK2 (Caesin kinase II) phosphorylation	786	789
	Myristoylation	55	60
	Myristoylation	398	403
	Myristoylation	471	476
	Myristoylation	606	611
	Myristoylation	662	667
	Myristoylation	694	699
	Myristoylation	807	812
	Protein kinase C phosphorylation	244	246
	Protein kinase C phosphorylation	270	272
	Protein kinase C phosphorylation	308	310
	Protein kinase C phosphorylation	314	316
	Protein kinase C phosphorylation	362	364
	Protein kinase C phosphorylation	381	383
	Protein kinase C phosphorylation	407	409
	Protein kinase C phosphorylation	415	417
	Protein kinase C phosphorylation	428	430
	Protein kinase C phosphorylation	436	438
	Protein kinase C phosphorylation	517	519
	Protein kinase C phosphorylation	554	556
	Protein kinase C phosphorylation	767	769
	Protein kinase C phosphorylation	798	800
	Protein kinase C phosphorylation	838	840
	Tyrosine kinase phosphorylation	506	513

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT3G51480.1	Asn-glycosylation	65	68
	Asn-glycosylation	207	210
	Asn-glycosylation	300	303
	Asn-glycosylation	333	336
	Asn-glycosylation	411	414
	Asn-glycosylation	425	428
	Asn-glycosylation	652	655
	Asn-glycosylation	897	900
	cAMP- and cGMP-dependent protein kinase phosphorylation	858	861
	CK2 (Caesin kinase II) phosphorylation	83	86
	CK2 (Caesin kinase II) phosphorylation	125	128
	CK2 (Caesin kinase II) phosphorylation	143	146
	CK2 (Caesin kinase II) phosphorylation	259	262
	CK2 (Caesin kinase II) phosphorylation	514	517
	CK2 (Caesin kinase II) phosphorylation	627	630
	CK2 (Caesin kinase II) phosphorylation	680	683
	CK2 (Caesin kinase II) phosphorylation	713	716
	CK2 (Caesin kinase II) phosphorylation	885	888
	Myristoylation	33	38
	Myristoylation	243	248
	Myristoylation	331	336
	Myristoylation	400	405
	Myristoylation	478	483
	Myristoylation	525	530
	Myristoylation	728	733
	Myristoylation	754	759
	Myristoylation	898	903
	Protein kinase C phosphorylation	24	26
	Protein kinase C phosphorylation	68	70
	Protein kinase C phosphorylation	106	108
	Protein kinase C phosphorylation	187	189
	Protein kinase C phosphorylation	194	196
	Protein kinase C phosphorylation	280	282
	Protein kinase C phosphorylation	377	379
	Protein kinase C phosphorylation	384	386
	Protein kinase C phosphorylation	442	444
	Protein kinase C phosphorylation	473	475
	Protein kinase C phosphorylation	538	540
	Protein kinase C phosphorylation	627	629
	Protein kinase C phosphorylation	745	747
	Protein kinase C phosphorylation	856	858
	Protein kinase C phosphorylation	864	866
	Protein kinase C phosphorylation	885	887
	Protein kinase C phosphorylation	899	901

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT4G31710.1	Asn-glycosylation	49	52
	Asn-glycosylation	69	72
	Asn-glycosylation	76	79
	Asn-glycosylation	227	230
	Asn-glycosylation	290	293
	Asn-glycosylation	354	357
	Asn-glycosylation	364	367
	Asn-glycosylation	536	539
	Asn-glycosylation	543	546
	cAMP- and cGMP-dependent protein kinase phosphorylation	811	814
	CK2 (Caesin kinase II) phosphorylation	78	81
	CK2 (Caesin kinase II) phosphorylation	85	88
	CK2 (Caesin kinase II) phosphorylation	100	103
	CK2 (Caesin kinase II) phosphorylation	167	170
	CK2 (Caesin kinase II) phosphorylation	229	232
	CK2 (Caesin kinase II) phosphorylation	292	295
	CK2 (Caesin kinase II) phosphorylation	358	361
	CK2 (Caesin kinase II) phosphorylation	366	369
	CK2 (Caesin kinase II) phosphorylation	372	375
	CK2 (Caesin kinase II) phosphorylation	406	409
	CK2 (Caesin kinase II) phosphorylation	546	549
	CK2 (Caesin kinase II) phosphorylation	702	705
	CK2 (Caesin kinase II) phosphorylation	781	784
	CK2 (Caesin kinase II) phosphorylation	794	797
	CK2 (Caesin kinase II) phosphorylation	868	871
	Myristoylation	41	46
	Myristoylation	47	52
	Myristoylation	64	69
	Myristoylation	106	111
	Myristoylation	202	207
	Myristoylation	401	406
	Myristoylation	418	423
	Myristoylation	439	444
	Myristoylation	458	463
	Myristoylation	479	484
	Myristoylation	586	591
	Myristoylation	673	678
	Myristoylation	718	723
	Protein kinase C phosphorylation	2	4
	Protein kinase C phosphorylation	84	86
	Protein kinase C phosphorylation	158	160
	Protein kinase C phosphorylation	315	317
	Protein kinase C phosphorylation	440	442
	Protein kinase C phosphorylation	446	448
	Protein kinase C phosphorylation	522	524
	Protein kinase C phosphorylation	641	643
	Protein kinase C phosphorylation	665	667
	Protein kinase C phosphorylation	777	779
	Protein kinase C phosphorylation	809	811

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT4G35290.1	Asn-glycosylation	306	309
	Asn-glycosylation	338	341
	Asn-glycosylation	378	381
	Asn-glycosylation	417	420
	Asn-glycosylation	435	438
	Asn-glycosylation	445	448
	Asn-glycosylation	532	535
	Asn-glycosylation	734	737
	Asn-glycosylation	808	811
	Asn-glycosylation	813	816
	cAMP- and cGMP-dependent protein kinase phosphorylation	431	434
	CK2 (Caesin kinase II) phosphorylation	72	75
	CK2 (Caesin kinase II) phosphorylation	125	128
	CK2 (Caesin kinase II) phosphorylation	207	210
	CK2 (Caesin kinase II) phosphorylation	262	265
	CK2 (Caesin kinase II) phosphorylation	362	365
	CK2 (Caesin kinase II) phosphorylation	477	480
	CK2 (Caesin kinase II) phosphorylation	483	486
	CK2 (Caesin kinase II) phosphorylation	638	641
	CK2 (Caesin kinase II) phosphorylation	702	705
	CK2 (Caesin kinase II) phosphorylation	724	727
	CK2 (Caesin kinase II) phosphorylation	815	818
	Myristoylation	33	38
	Myristoylation	39	44
	Myristoylation	353	358
	Myristoylation	536	541
	Myristoylation	735	740
	Myristoylation	761	766
	Myristoylation	789	794
	Protein kinase C phosphorylation	194	196
	Protein kinase C phosphorylation	207	209
	Protein kinase C phosphorylation	234	236
	Protein kinase C phosphorylation	277	279
	Protein kinase C phosphorylation	283	285
	Protein kinase C phosphorylation	291	293
	Protein kinase C phosphorylation	349	351
	Protein kinase C phosphorylation	434	436
	Protein kinase C phosphorylation	455	457
	Protein kinase C phosphorylation	477	479
	Protein kinase C phosphorylation	488	490
	Protein kinase C phosphorylation	549	551
	Protein kinase C phosphorylation	638	640
	Protein kinase C phosphorylation	693	695
	Protein kinase C phosphorylation	724	726
	Protein kinase C phosphorylation	872	874
	Protein kinase C phosphorylation	892	894
	Protein kinase C phosphorylation	905	907
	Tyrosine kinase phosphorylation	151	159

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT4G35290.2	Asn-glycosylation	306	309
	Asn-glycosylation	338	341
	Asn-glycosylation	378	381
	Asn-glycosylation	417	420
	Asn-glycosylation	435	438
	Asn-glycosylation	445	448
	Asn-glycosylation	532	535
	Asn-glycosylation	734	737
	Asn-glycosylation	808	811
	Asn-glycosylation	813	816
	cAMP- and cGMP-dependent protein kinase phosphorylation	431	434
	CK2 (Caesin kinase II) phosphorylation	72	75
	CK2 (Caesin kinase II) phosphorylation	125	128
	CK2 (Caesin kinase II) phosphorylation	207	210
	CK2 (Caesin kinase II) phosphorylation	262	265
	CK2 (Caesin kinase II) phosphorylation	362	365
	CK2 (Caesin kinase II) phosphorylation	477	480
	CK2 (Caesin kinase II) phosphorylation	483	486
	CK2 (Caesin kinase II) phosphorylation	638	641
	CK2 (Caesin kinase II) phosphorylation	702	705
	CK2 (Caesin kinase II) phosphorylation	724	727
	CK2 (Caesin kinase II) phosphorylation	815	818
	Myristoylation	33	38
	Myristoylation	39	44
	Myristoylation	353	358
	Myristoylation	536	541
	Myristoylation	735	740
	Myristoylation	761	766
	Myristoylation	789	794
	Protein kinase C phosphorylation	194	196
	Protein kinase C phosphorylation	207	209
	Protein kinase C phosphorylation	234	236
	Protein kinase C phosphorylation	277	279
	Protein kinase C phosphorylation	283	285
	Protein kinase C phosphorylation	291	293
	Protein kinase C phosphorylation	349	351
	Protein kinase C phosphorylation	434	436
	Protein kinase C phosphorylation	455	457
	Protein kinase C phosphorylation	477	479
	Protein kinase C phosphorylation	488	490
	Protein kinase C phosphorylation	549	551
	Protein kinase C phosphorylation	638	640
	Protein kinase C phosphorylation	693	695
	Protein kinase C phosphorylation	724	726
	Protein kinase C phosphorylation	872	874
	Protein kinase C phosphorylation	892	894
	Protein kinase C phosphorylation	905	907
	Tyrosine kinase phosphorylation	151	159

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT5G11180.1	Asn-glycosylation	45	48
	Asn-glycosylation	57	60
	Asn-glycosylation	121	124
	Asn-glycosylation	336	339
	Asn-glycosylation	345	348
	Asn-glycosylation	424	427
	Asn-glycosylation	550	553
	Asn-glycosylation	734	737
	cAMP- and cGMP-dependent protein kinase phosphorylation	343	346
	cAMP- and cGMP-dependent protein kinase phosphorylation	674	677
	CK2 (Caesin kinase II) phosphorylation	29	32
	CK2 (Caesin kinase II) phosphorylation	59	62
	CK2 (Caesin kinase II) phosphorylation	149	152
	CK2 (Caesin kinase II) phosphorylation	209	212
	CK2 (Caesin kinase II) phosphorylation	273	276
	CK2 (Caesin kinase II) phosphorylation	327	330
	CK2 (Caesin kinase II) phosphorylation	346	349
	CK2 (Caesin kinase II) phosphorylation	351	354
	CK2 (Caesin kinase II) phosphorylation	525	528
	CK2 (Caesin kinase II) phosphorylation	553	556
	CK2 (Caesin kinase II) phosphorylation	616	619
	CK2 (Caesin kinase II) phosphorylation	664	667
	CK2 (Caesin kinase II) phosphorylation	703	706
	CK2 (Caesin kinase II) phosphorylation	736	739
	CK2 (Caesin kinase II) phosphorylation	759	762
	CK2 (Caesin kinase II) phosphorylation	810	813
	Myristoylation	87	92
	Myristoylation	416	421
	Myristoylation	427	432
	Myristoylation	634	639
	Myristoylation	679	684
	Protein kinase C phosphorylation	52	54
	Protein kinase C phosphorylation	81	83
	Protein kinase C phosphorylation	139	141
	Protein kinase C phosphorylation	165	167
	Protein kinase C phosphorylation	342	344
	Protein kinase C phosphorylation	346	348
	Protein kinase C phosphorylation	375	377
	Protein kinase C phosphorylation	394	396
	Protein kinase C phosphorylation	420	422
	Protein kinase C phosphorylation	434	436
	Protein kinase C phosphorylation	623	625
	Protein kinase C phosphorylation	708	710
	Protein kinase C phosphorylation	740	742
	Protein kinase C phosphorylation	888	890
	Protein kinase C phosphorylation	894	896
	Tyrosine kinase phosphorylation	290	296
	Tyrosine kinase phosphorylation	523	531

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT5G11210.1	Asn-glycosylation	48	51
	Asn-glycosylation	262	265
	Asn-glycosylation	266	269
	Asn-glycosylation	472	475
	Asn-glycosylation	717	720
	cAMP- and cGMP-dependent protein kinase phosphorylation	564	567
	CK2 (Caesin kinase II) phosphorylation	9	12
	CK2 (Caesin kinase II) phosphorylation	76	79
	CK2 (Caesin kinase II) phosphorylation	254	257
	CK2 (Caesin kinase II) phosphorylation	267	270
	CK2 (Caesin kinase II) phosphorylation	272	275
	CK2 (Caesin kinase II) phosphorylation	279	282
	CK2 (Caesin kinase II) phosphorylation	401	404
	CK2 (Caesin kinase II) phosphorylation	447	450
	CK2 (Caesin kinase II) phosphorylation	599	602
	CK2 (Caesin kinase II) phosphorylation	647	650
	CK2 (Caesin kinase II) phosphorylation	686	689
	CK2 (Caesin kinase II) phosphorylation	719	722
	CK2 (Caesin kinase II) phosphorylation	742	745
	Myristoylation	6	11
	Myristoylation	110	115
	Myristoylation	285	290
	Myristoylation	342	347
	Myristoylation	617	622
	Myristoylation	662	667
	Protein kinase C phosphorylation	66	68
	Protein kinase C phosphorylation	92	94
	Protein kinase C phosphorylation	267	269
	Protein kinase C phosphorylation	272	274
	Protein kinase C phosphorylation	301	303
	Protein kinase C phosphorylation	320	322
	Protein kinase C phosphorylation	346	348
	Protein kinase C phosphorylation	355	357
	Protein kinase C phosphorylation	606	608
	Protein kinase C phosphorylation	691	693
	Tyrosine kinase phosphorylation	445	453

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT5G27100.1	Asn-glycosylation	46	49
	Asn-glycosylation	53	56
	Asn-glycosylation	204	207
	Asn-glycosylation	267	270
	Asn-glycosylation	331	334
	Asn-glycosylation	342	345
	Asn-glycosylation	461	464
	Asn-glycosylation	477	480
	Asn-glycosylation	536	539
	cAMP- and cGMP-dependent protein kinase phosphorylation	563	566
	CK2 (Caesin kinase II) phosphorylation	55	58
	CK2 (Caesin kinase II) phosphorylation	62	65
	CK2 (Caesin kinase II) phosphorylation	73	76
	CK2 (Caesin kinase II) phosphorylation	77	80
	CK2 (Caesin kinase II) phosphorylation	93	96
	CK2 (Caesin kinase II) phosphorylation	145	148
	CK2 (Caesin kinase II) phosphorylation	175	178
	CK2 (Caesin kinase II) phosphorylation	206	209
	CK2 (Caesin kinase II) phosphorylation	258	261
	CK2 (Caesin kinase II) phosphorylation	269	272
	CK2 (Caesin kinase II) phosphorylation	384	387
	CK2 (Caesin kinase II) phosphorylation	424	427
	CK2 (Caesin kinase II) phosphorylation	800	803
	CK2 (Caesin kinase II) phosphorylation	806	809
	Myristoylation	41	46
	Myristoylation	348	353
	Myristoylation	407	412
	Myristoylation	457	462
	Myristoylation	510	515
	Myristoylation	720	725
	Myristoylation	741	746
	Myristoylation	885	890
	Protein kinase C phosphorylation	93	95
	Protein kinase C phosphorylation	135	137
	Protein kinase C phosphorylation	298	300
	Protein kinase C phosphorylation	535	537
	Protein kinase C phosphorylation	561	563
	Protein kinase C phosphorylation	783	785
	Protein kinase C phosphorylation	815	817

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT5G48400.1	Asn-glycosylation	301	304
	Asn-glycosylation	400	403
	Asn-glycosylation	496	499
	Asn-glycosylation	619	622
	Asn-glycosylation	676	679
	Asn-glycosylation	688	691
	Asn-glycosylation	699	702
	CK2 (Caesin kinase II) phosphorylation	85	88
	CK2 (Caesin kinase II) phosphorylation	95	98
	CK2 (Caesin kinase II) phosphorylation	113	116
	CK2 (Caesin kinase II) phosphorylation	155	158
	CK2 (Caesin kinase II) phosphorylation	211	214
	CK2 (Caesin kinase II) phosphorylation	216	219
	CK2 (Caesin kinase II) phosphorylation	269	272
	CK2 (Caesin kinase II) phosphorylation	316	319
	CK2 (Caesin kinase II) phosphorylation	339	342
	CK2 (Caesin kinase II) phosphorylation	352	355
	CK2 (Caesin kinase II) phosphorylation	411	414
	CK2 (Caesin kinase II) phosphorylation	433	436
	CK2 (Caesin kinase II) phosphorylation	468	471
	CK2 (Caesin kinase II) phosphorylation	509	512
	CK2 (Caesin kinase II) phosphorylation	689	692
	Myristoylation	49	54
	Myristoylation	317	322
	Myristoylation	364	369
	Myristoylation	401	406
	Myristoylation	497	502
	Myristoylation	543	548
	Myristoylation	666	671
	Myristoylation	686	691
	Protein kinase C phosphorylation	263	265
	Protein kinase C phosphorylation	303	305
	Protein kinase C phosphorylation	378	380
	Protein kinase C phosphorylation	404	406
	Protein kinase C phosphorylation	428	430
	Protein kinase C phosphorylation	433	435
	Protein kinase C phosphorylation	437	439
	Protein kinase C phosphorylation	449	451
	Protein kinase C phosphorylation	509	511
	Protein kinase C phosphorylation	525	527
	Protein kinase C phosphorylation	678	680

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT5G48400.2	Asn-glycosylation	301	304
	Asn-glycosylation	400	403
	Asn-glycosylation	496	499
	Asn-glycosylation	619	622
	Asn-glycosylation	676	679
	Asn-glycosylation	688	691
	Asn-glycosylation	699	702
	Asn-glycosylation	748	751
	CK2 (Caesin kinase II) phosphorylation	85	88
	CK2 (Caesin kinase II) phosphorylation	95	98
	CK2 (Caesin kinase II) phosphorylation	113	116
	CK2 (Caesin kinase II) phosphorylation	155	158
	CK2 (Caesin kinase II) phosphorylation	211	214
	CK2 (Caesin kinase II) phosphorylation	216	219
	CK2 (Caesin kinase II) phosphorylation	269	272
	CK2 (Caesin kinase II) phosphorylation	316	319
	CK2 (Caesin kinase II) phosphorylation	339	342
	CK2 (Caesin kinase II) phosphorylation	352	355
	CK2 (Caesin kinase II) phosphorylation	411	414
	CK2 (Caesin kinase II) phosphorylation	433	436
	CK2 (Caesin kinase II) phosphorylation	468	471
	CK2 (Caesin kinase II) phosphorylation	509	512
	CK2 (Caesin kinase II) phosphorylation	689	692
	CK2 (Caesin kinase II) phosphorylation	777	780
	Myristoylation	49	54
	Myristoylation	317	322
	Myristoylation	364	369
	Myristoylation	401	406
	Myristoylation	497	502
	Myristoylation	543	548
	Myristoylation	666	671
	Myristoylation	686	691
	Myristoylation	798	803
	Protein kinase C phosphorylation	263	265
	Protein kinase C phosphorylation	303	305
	Protein kinase C phosphorylation	378	380
	Protein kinase C phosphorylation	404	406
	Protein kinase C phosphorylation	428	430
	Protein kinase C phosphorylation	433	435
	Protein kinase C phosphorylation	437	439
	Protein kinase C phosphorylation	449	451
	Protein kinase C phosphorylation	509	511
	Protein kinase C phosphorylation	525	527
	Protein kinase C phosphorylation	678	680
	Protein kinase C phosphorylation	759	761
	Protein kinase C phosphorylation	829	831

Table continued -

Table continued -

Protein name	Type of site	Start	Stop
AT5G48410.1	Asn-glycosylation	302	305
	Asn-glycosylation	346	349
	Asn-glycosylation	398	401
	Asn-glycosylation	493	496
	Asn-glycosylation	523	526
	Asn-glycosylation	616	619
	Asn-glycosylation	696	699
	Asn-glycosylation	745	748
	CK2 (Caesin kinase II) phosphorylation	29	32
	CK2 (Caesin kinase II) phosphorylation	87	90
	CK2 (Caesin kinase II) phosphorylation	97	100
	CK2 (Caesin kinase II) phosphorylation	115	118
	CK2 (Caesin kinase II) phosphorylation	157	160
	CK2 (Caesin kinase II) phosphorylation	213	216
	CK2 (Caesin kinase II) phosphorylation	271	274
	CK2 (Caesin kinase II) phosphorylation	350	353
	CK2 (Caesin kinase II) phosphorylation	431	434
	CK2 (Caesin kinase II) phosphorylation	495	498
	CK2 (Caesin kinase II) phosphorylation	506	509
	CK2 (Caesin kinase II) phosphorylation	526	529
	CK2 (Caesin kinase II) phosphorylation	653	656
	CK2 (Caesin kinase II) phosphorylation	686	689
	CK2 (Caesin kinase II) phosphorylation	774	777
	CK2 (Caesin kinase II) phosphorylation	841	844
	Myristoylation	51	56
	Myristoylation	399	404
	Myristoylation	663	668
	Myristoylation	683	688
	Myristoylation	795	800
	Protein kinase C phosphorylation	28	30
	Protein kinase C phosphorylation	52	54
	Protein kinase C phosphorylation	83	85
	Protein kinase C phosphorylation	265	267
	Protein kinase C phosphorylation	304	306
	Protein kinase C phosphorylation	376	378
	Protein kinase C phosphorylation	388	390
	Protein kinase C phosphorylation	402	404
	Protein kinase C phosphorylation	426	428
	Protein kinase C phosphorylation	435	437
	Protein kinase C phosphorylation	447	449
	Protein kinase C phosphorylation	506	508
	Protein kinase C phosphorylation	522	524
	Protein kinase C phosphorylation	675	677
	Protein kinase C phosphorylation	756	758
	Protein kinase C phosphorylation	824	826

The number of phosphorylation sites ranged from 20-35 (average $\pm$ SE of 26.64 ± 0.83). These phosphorylation sites mainly include CK2 (Caesin kinase II) phosphorylation sites (ranging from 8-18 with an average $\pm$ SE of 12.20 ± 0.61) and Protein kinase C phosphorylation sites (ranging from 7-19 with an average $\pm$ SE of 12.96 ± 0.69). However, 18 proteins (72%) showed the presence of cAMP- and cGMP-dependent protein kinase phosphorylation sites and 10 proteins (40%) showed the presence of tyrosine kinase phosphorylation sites. Phosphorylation plays important roles in plants in a variety of crucial processes (Cheng et al. 2014) including ABA signaling (Kusakina and Dodd 2012) and stress response (Rampitsch and Bykova 2012). High number of phosphorylation sites in AtGLRs indicates their role in signal transduction especially in stress. The number of myristoylation sites ranged from 4-13 (average $\pm$ SE of 7.92 ± 0.47). N-terminal myristoylation has a vital role in signal transduction and membrane targeting in plants in responses to environmental stress conditions (Podell and Gribskov 2004). Myristoylation influence the conformational stability of a protein as well as their interacting capability with membranes or the hydrophobic domains of other proteins (Olsen and Kaarsholm 2000; Resh, 1999; J. Zheng et al. 1993). Thus, the presence of myristoylation sites in AtGLRs indicates their role in interaction with membranes. The number of Asn-glycosylation sites ranges from 4-13 (average $\pm$ SE of 8.52 ± 0.42). Asn-glycosylation is a co-translational protein modification reaction that influence either the protein folding process or the stability of the native glycoprotein conjugate (Imperiali and O'Connor 1999). Hence, presence of several of these sites indicates that the stability of these membrane proteins is maintained by glycosylation. Domain organization with Pfam showed that the N-terminal end of all these proteins is composed of ANF_receptor (extracellular ligand binding domain; PF01094). The C-terminal region, on the other hand, consists of two overlapping domains namely SBP_bac_3 (bacterial extracellular solute-binding proteins, family 3; PF00497) and Lig_chan (transmembrane region of ligand-gated ion channel; PF00060). The SBP_bac_3 domain is often associated with ABC-type transporters (Tam and Saier 1993). ABC transporters are ubiquitous membrane proteins that couple ATP hydrolysis to the translocation of various substrates across membranes (Locher et al. 2002).

2.3.2 Chromosomal distribution, duplication, gene structure and phylogeny of AtGLR genes

The AtGLR genes are distributed unevenly in the 5 chromosomes of *A. thaliana* (Figure 2.2).

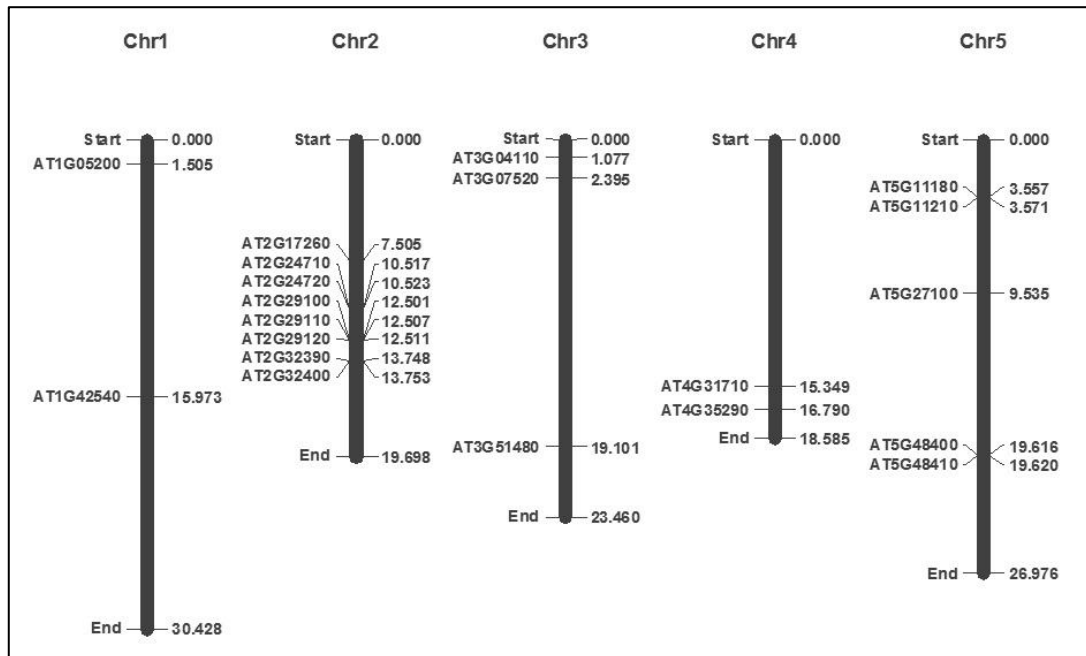


Figure 2.2 : Graphical representation of physical locations for each AtGLR gene on the five chromosomes (numbered Chr1-Chr5) of *Arabidopsis thaliana*.

Note: Chromosomal distances are given in Mb.

Chromosome 2 contains maximum number of AtGLR genes (8 genes i.e. 40%) while chromosome 1 and 4 contains minimum number (2 i.e. 10% each). Distribution of AtGLRs genes on chromosomal loci revealed that 11 of 20 genes are in tandem repeats suggesting local duplication of which 7 and 4 genes were located on chromosome 2 and 5, respectively. Duplicated genes often evolved for partitioning of existing functions (subfunctionalization) or for acquiring new functions (neofunctionalization), which enhance plants' adaptability (Airoidi and Davies 2012; Lee and Irish 2011). Chromosome 2 contains three groups of tandem repeats while chromosome 5 contains two pairs. Three direct tandem repeats were found in chromosome 2 (AT2G29100, AT2G29110 and AT2G29120). Using a divergence rate of 6.1×10^{-9} mutations per synonymous site per year (Lynch and Conery 2000), the divergence time between the

tandemly repeated genes were calculated. All the locally duplicated genes showed divergence time ranging from 22.65-51.95 Mya (Table 2.4).

Table 2.4 : -Estimated divergence time of tandemly arranged AtGLR genes in *Arabidopsis thaliana*.

Gene pairs	d_N	d_S	d_N/d_S	Divergence time (Mya)
AT5G11180 vs AT5G11210	0.1415	0.3820	0.37042	31.31
AT5G48400 vs AT5G48410	0.1326	0.3312	0.40036	27.15
AT2G24710 vs AT2G24720	0.1518	0.3351	0.453	27.47
AT2G29100 vs AT2G29110	0.1459	0.2763	0.52805	22.65
AT2G29100 vs AT2G29120	0.2301	0.5361	0.42921	43.94
AT2G29110 vs AT2G29120	0.1944	0.6339	0.30667	51.95

However, d_S value of the gene pair AT2G32390/AT2G39400 was very large (more than 3) which can give erroneous results. Also, the sequence similarity of these two genes was only 38%. Thus, we have excluded this pair from our study. Maybe, these two genes diverged long ago thus showed such low sequence similarity. It was reported that the divergence time between *A. thaliana* and *A. lyrata* was around 13 Mya (Beilstein et al. 2010). Thus, the tandemly arranged AtGLR gene pairs diverged before the separation of these two species. Among the orthologous gene pairs of AtGLRs with those of other members of the family Brassicaceae, the average d_N/d_S value was not significantly different (Wilcoxon Rank Sum test, $P = 0.62$) between *A. thaliana* and *A. lyrata* (average $\pm$ SE = 0.26 ± 0.02) as well as between *A. thaliana* and *Brassica rapa* (average $\pm$ SE = 0.27 ± 0.02). This showed that AtGLR genes

encountered intense purifying selection. However, the significantly higher (Wilcoxon Rank Sum test, $P = 4.77 \times 10^{-7}$) average rate of synonymous substitution between *A. thaliana* and *B. rapa* (average $\pm$ SE = 0.58 ± 0.08) than between *A. thaliana* and *A. lyrata* (average $\pm$ SE = 0.17 ± 0.01) indicated the early divergence of AtGLRs from the common ancestors of *Arabidopsis* and *B. rapa*. The divergence time (Mya) calculated from the synonymous substitution rate indicated *B. rapa* and *Arabidopsis* diverged around 14.5-20.4 Mya ago (Yang et al. 1999). This suggests that the AtGLR genes in tandem repeats duplicated before the divergence of *Brassica* and *Arabidopsis*. The phylogenetic tree obtained with MrBayes showed that AtGLR genes are primarily grouped into two clusters (Figure 2.3). The first cluster was further divided into two subclusters. Notably, the tandemly arranged AtGLR genes also grouped except AT2G32390 and AT2G32400 (although they belong to the same subcluster). Majority (10, 50%) of the AtGLR genes contains 4 introns followed by 5 introns (8, 40%) (Figure 2.3).

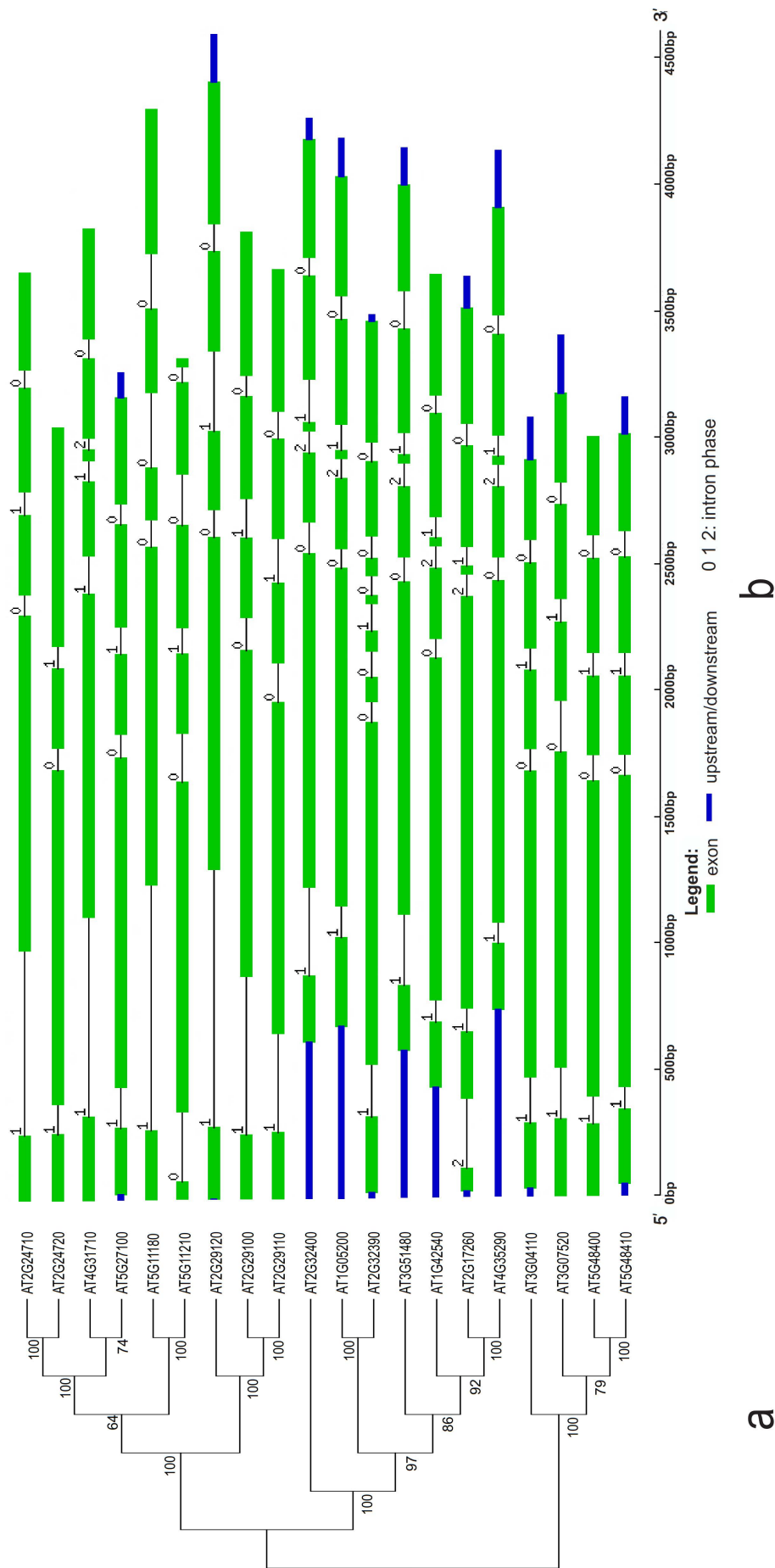


Figure 2.3 : Phylogenetic tree of *Arabidopsis thaliana* glutamate receptor (AtGLR) genes along with exon/intron and upstream/downstream sequence organization.
Notes : Exons and introns are represented by green bars and black lines, respectively. Blue regions represent upstream/downstream sequences. 0, 1 and 2 on each intron indicates intron phase 0, 1 and 2, respectively.

One each of the genes contains 3 and 7 introns. There are three possible locations for an intron within a gene: between two consecutive codons (phase 0), between the first and second nucleotide of a codon (phase 1) and between the second and third nucleotides of a codon (phase 2) (Sharp, 1981). The percentage of phase 0 introns ranged from 20-80% with 12 genes with at least 50% phase 0 introns. The percentage of phase 1 introns ranged from 20-66.66% with 11 genes with at least 50% phase 1 introns. Thirteen genes did not show the presence of any phase 2 introns. It was previously reported that majority of the new introns in *A. thaliana* are phase 0 (Sverdlov et al. 2003). Thus, it is apparent that the phase 0 introns are newer in AtGLRs. The first intron of majority (18, 90%) of the AtGLR genes showed phase 1 intron. Moreover, the 5'-end of all these 18 genes have a similar exon-intron structure *i.e.* a short exon followed by a phase 1 intron and a long exon (Figure 2. 3). This indicates that these AtGLR genes have a common origin. The two tandemly arranged genes on chromosome 5 namely AT5G48400 and AT5G48410 showed similar arrangement of exons and introns. Another two tandemly arranged genes on chromosome 2 namely AT2G24710 and AT2G24720 differed only in their first intron length and absence of the last intron in AT2G24720. Three tandemly arranged genes on chromosome 2 namely AT2G29100, AT2G29110 and AT2G29120 also differed mainly in their first intron length. However, two tandemly repeated genes on chromosome 5 namely AT5G11180 and AT5G11210 showed many differences in their exon/intron arrangements.

2.3.3 Expression of AtGLR genes

We have analyzed the expression pattern of twenty AtGLR genes in fifteen plant parts and seventeen developmental stages to get a clear picture of their overall expression profiles. In general, expression of three genes namely AT2G29100, AT2G24710 and AT2G24720 were much lower than the other genes in different plant parts and in different developmental stages (Figure 2.4a, Table 2.5). Regarding expression in different plant parts, the highest expression was shown by AT1G05200 (in cauline leaf) followed by AT3G51480 (in hypocotyls). It has been previously reported that AT1G05200 (AtGLR 3.4) plays an important role in the Ca^{2+} -based rapid transmission of environmental stress (Meyerhoff et al. 2005). It was also reported that this gene is located in plasma membrane and plastids (Teardo et al. 2011). The present study showed that this gene is expressed more in dark than control light and high light

conditions (Figure 2.4b). Moreover, under different biotic stresses, it showed lower expression than control (Figure 2.4c). All these indicate that this gene helps the plant to receive environmental signals through leaf and helps to combat that stress by reducing its expression. Notably, this gene is also highly expressed in senescence, the reason of which is currently not clear. AT3G51480 (GLR 3.6) has been shown to be necessary for the production of both slow and fast (action potential like) remote depolarization waves in caterpillar induced wound response (Salvador-Recatalà et al. 2014). High expression in hypocotyls indicates the role of this gene in protecting the seedling from insects. Further laboratory-based research based on this *in silico* studies can open new avenues in the role of glutamate receptors in plant stress response. Regarding expression in different developmental stages, four genes (AT1G05200, AT2G29110, AT2G29120 and AT5G48410) showed higher expression than others in senescence (Figure 2.4a, Table 2.5).

Except AT1G05200, all the other three genes showed much less expression in other developmental stages showing their involvement specifically in senescence. Glutamate has been shown to increase just at the onset of senescence and then decrease considerably (Diaz et al. 2005; Forde and Lea 2007). The expression profiles indicate the roles of these four proteins in regulating the level of glutamate in senescence.

It was previously reported that redundant duplicated genes are lost from the genome due to random mutation and loss of function, except when neo- or sub-functionalization occur (Blanc and Wolfe 2004; Thomas et al. 2006). Among the tandem repeat pairs consisted of AT5G48400 and AT5G48410, AT5G48410 showed more expression than AT5G48400 in all the plant parts as well as throughout the plant development (Figure 2.4a).

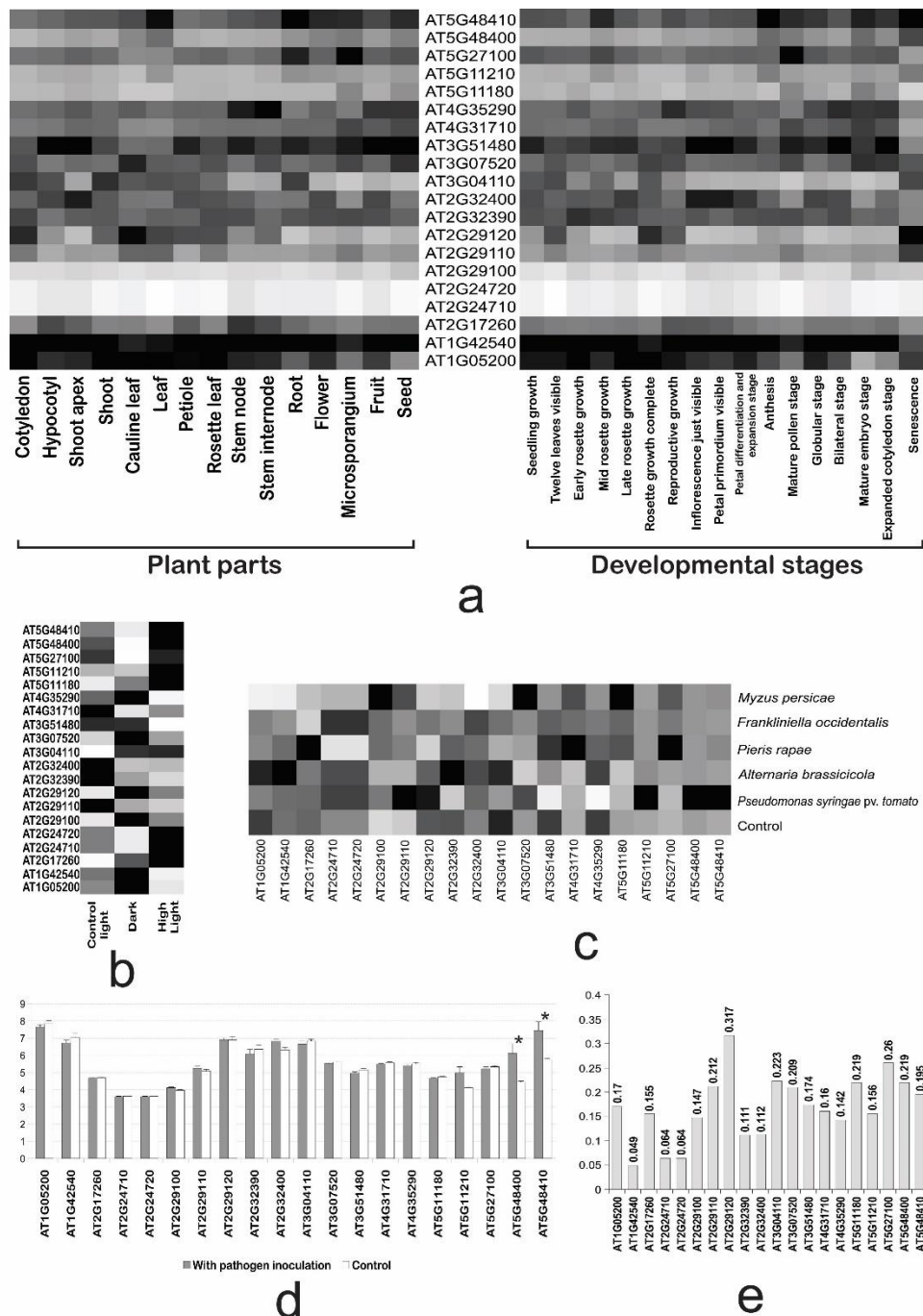


Figure 2.4 : Expression pattern of AtGLR genes. (a) Heatmap representation of the expression profiles of the 20 AtGLR genes in different plant parts and developmental stages. (b) Heatmap representation of the expression profiles of the 20 AtGLR genes under different light intensities. (c) Heatmap representation of the expression profiles of the 20 AtGLR genes in response to infection of different pathogens and insects. In all heatmap representations, the dark and light colour shadings represent relatively high or low expression levels, respectively. (d) Expression patterns of the 20 AtGLR genes in response to *Botrytis cinerea* infection. *Represents significant difference at $p < 0.01$ level. (e) Tissue specificity index (τ) of the 20 AtGLR genes.

Table – 2.5: Normalized expression values (as obtained from PlexDB) of twenty AtGLR genes in different plant parts and different developmental phases of *Arabidopsis thaliana*

Gene name	AT1G05200	AT1G42540	AT2G17260	AT2G47100	AT2G24720	AT2G29100	AT2G29110	AT2G29120	AT2G32390	AT2G32400	AT3G04110	AT3G07050	AT3G34500	AT4G31710	AT4G33290	AT5G11180	AT5G11210	AT5G27100	AT5G48400	AT5G48410
Plant parts	Cotyledon	7.84566	7.793143333	5.575983333	3.594486667	3.394486667	4.065036667	5.99961	7.308733333	6.934143333	6.37277	7.087156667	6.61227	5.879233333	6.16342	4.8923	4.778063333	6.063806667	4.736666667	6.032026667
	Hypocotyl	6.870266667	7.596006667	6.48915	3.474233333	3.474233333	3.714973333	4.631563333	5.11	5.700073333	6.590086667	6.359083333	5.592156667	8.158813333	5.622286667	5.986733333	4.355066667	4.638736667	5.791403333	4.60038
	Shoot apex	6.689471667	7.510233333	5.895458333	3.471215	3.471215	3.762403333	4.789155	4.183352	5.902658333	6.817041667	4.590380833	5.676719167	7.04776667	5.714613333	5.980283333	4.49141667	4.528405	5.808198333	4.651016667
	Shoot	7.485444	7.676394	5.516024667	3.42510667	3.934408667	5.013248667	6.136273333	6.27642867	6.243036667	6.889762	5.845178667	6.534818667	5.940548667	5.940014667	4.82218667	4.642606667	5.889794	4.836648	6.264018667
	Cauline leaf	8.202903333	7.686066667	5.994613333	3.59933	3.59933	3.96116	6.5297	7.9845	6.364983333	6.208866667	6.79303	6.799113333	6.329893333	5.774633333	6.13111	5.733496667	5.853496667	4.850466667	7.035716667
	Leaf	8.09087	7.853235	5.671511667	3.462536667	3.462536667	4.219065	6.58851667	7.140666667	6.634166667	6.172653333	6.942451667	6.710188333	6.416786667	5.72516667	6.338073333	5.621688333	5.652466667	6.376266667	7.69925
	Petiole	7.15362	7.340136667	6.118543333	3.459733333	3.459733333	3.908313333	4.82773	6.706386667	6.004973333	6.7384	6.235266667	6.050166667	6.945133333	5.66106	5.97099	4.595743333	5.833653333	4.762383333	5.770523333
	Rosette leaf	7.558955897	7.491505897	5.542044615	3.485841282	3.485841282	3.929155641	5.707831795	6.473272564	5.62670377	6.41255	5.992964872	6.43125359	5.887611538	6.07142436	4.72829231	4.690645385	6.084414872	4.732367949	6.04862805
	Stem node	7.019456667	7.821753333	6.827766667	3.508586667	3.508586667	3.626733333	4.74761	5.19037	6.380576667	6.05443	4.548433333	5.66705	7.110063333	5.722083333	7.108713333	4.273526667	4.270043333	5.630446667	6.217806667
	Stem internode	6.776666667	7.529633333	6.312733333	3.500066667	3.500066667	3.840063333	4.87608	5.833446667	5.821616667	5.553033333	6.268966667	6.252506667	5.680423333	7.165626667	4.890303333	4.629383333	5.82447	4.71624	6.25336
	Root	5.929415714	7.02253	5.782682381	3.58494	3.58494	3.795798095	4.84031429	4.21073	6.12028571	5.56242086	6.48704086	6.79524524	5.816592857	5.468916667	4.78804762	5.144026667	6.793574762	5.093525238	7.16094
Developmental stages	Flower	6.555006667	7.70371905	5.409929524	3.464740476	3.464740476	3.767961429	4.957890714	4.543483065	6.16437857	6.20591948	4.467035952	5.804318095	5.717316429	5.80483381	4.37036592	4.940803571	5.826376667	5.261979286	6.65091429
	Microsporangium	6.116246667	7.356273333	5.778643333	3.720483333	3.720483333	4.5547	6.039803333	5.25145	6.05589	5.634566667	4.565023333	6.8854	7.461166667	6.894106667	5.90867	5.362673333	7.989536667	6.0046	7.281443333
	Fruit	6.2297	7.22057778	5.33072778	3.425895556	3.425895556	3.670204444	4.659666667	4.097445556	5.955656667	6.057563333	4.350611111	5.58842222	6.843103333	7.78534	6.32466778	4.392395556	4.99242222	5.687613333	4.692434444
	Seed	5.004788	6.859652667	5.409216	3.345596667	3.345596667	4.174349333	4.880730667	4.459634	5.632049333	5.303126	4.246252333	6.297451333	6.688673333	6.243730667	4.943641333	4.781556667	5.678477333	5.204433333	6.361722
	Seedling growth	7.021718989	7.570642121	5.74823442	3.530355152	3.530355152	3.881007879	4.943041818	5.125714545	6.29654818	6.344039485	6.219897516	5.827226861	6.3305781818	5.848014242	5.834586667	4.80796152	4.673990661	4.721729697	6.27581818
	Twelve leaves visible	7.13364222	7.25321111	5.649178444	3.480258889	3.480258889	3.901783111	5.630449556	6.501863333	6.25794889	6.110218	6.47393	6.34827778	6.522781778	5.874793333	5.944851111	4.676580222	4.853589111	6.159298444	6.348621111
	Early rosette growth	7.966358333	8.033045	5.546666667	3.471103333	3.471103333	3.954633333	4.959733333	4.996533333	6.99623	6.327961667	5.480106667	6.016588333	6.77613	5.845766667	6.376291667	4.776576667	4.580795	4.726383333	6.65801
	Mid rosette growth	6.739865	7.449183333	5.437533333	3.459801667	3.459801667	3.786533333	4.773181667	4.330213333	6.489205	5.98882	5.79941667	5.80397	6.707573333	5.859733333	5.663395	4.355881667	5.00186	6.344631667	4.92175
	Late rosette growth	7.28424	7.741613333	5.51375	3.49649	3.49649	3.844023333	4.917233333	4.708733333	6.267033333	6.530433333	4.816846667	5.829713333	6.766353333	5.95024	5.7581	4.920333333	4.598973333	5.903726667	4.637016667
	Rosette growth complete	7.727611111	7.344094444	5.713692222	3.464342222	3.464342222	3.885735556	5.753492222	7.148021111	6.347886667	6.479824444	6.493835556	6.853886667	6.618191111	5.825796667	6.289755556	4.655106667	4.632367778	6.015445556	4.825451111
Developmental stages	Reproductive growth	7.333008889	7.679261111	6.378384444	3.535944444	3.535944444	3.811298889	5.384463333	6.33605556	6.18902222	5.98716667	5.37128889	6.544386667	6.655807778	5.736566667	6.993816667	4.385004444	4.536883333	5.729471111	5.031015556
	Inflorescence just visible	6.56206667	7.446636667	5.925363333	3.338966667	3.338966667	3.731683333	4.71929	3.957883333	6.007916667	6.812076667	4.38915	5.498203333	5.68737	6.067313333	4.325933333	4.355823333	4.550966667	5.678483333	5.83066
	Petal primordium visible	6.377636667	7.48249	5.83259	3.51455	3.51455	3.718066667	4.801836667	4.001836667	5.726743333	6.770276667	4.289883333	5.63895	7.079673333	5.488466667	6.131793333	4.285023333	5.701433333	4.693966667	6.00463
	Petal differentiation and expansion stage	6.688805	7.70893625	5.399287083	3.420282917	3.420282917	3.69521667	4.83388125	4.18094167	6.042233333	6.384217917	4.36244075	5.81807	6.6455	5.71928075	5.679000417	4.529348075	5.753625417	4.812284583	6.00593875
	Anthesis	6.376608889	7.705286111	5.424119444	3.52407222	3.52407222	3.88494778	5.124303333	5.026935	6.327100556	5.98818722	4.606485556	6.05248889	6.210679889	5.714686667	5.98341667	5.51362	5.923733333	5.861572222	7.58961667
	Mature pollen stage	6.116246667	7.356273333	5.778643333	3.720483333	3.720483333	4.5547	6.039803333	5.25145	6.05589	5.634566667	4.565023333	6.8854	7.461166667	6.894106667	5.90867	5.362673333	7.989536667	6.0046	7.281443333
	Global stage	6.544833333	7.29377	5.325476667	3.355803333	3.355803333	3.637386667	4.671376667	4.011893333	5.75349	6.112706667	4.467183333	5.312953333	6.4756	5.64066	6.00756667	4.495026667	5.833883333	4.844586667	6.275563333
	Bilateral stage	6.072283333	7.183651667	5.333533333	3.462686667	3.462686667	3.686613333	4.653811667	4.148621667	6.06554	6.026961667	7.028955	6.476233333	4.310251667	4.90135	5.624473333	4.916398333	4.916398333	5.935748333	6.350366667
	Mature embryo stage	4.642	7.295146667	4.405833333	3.49363	3.49363	4.46481	5.021286667	4.526873333	5.688346667	4.960923333	4.18568	6.46766667	6.35163	6.26219	5.351253333	5.328966667	4.924833333	5.909716667	5.650163333
	Expanded cotyledon stage	5.005792222	6.656586667	5.579481111	3.309688889	3.309688889	4.040483333	4.794255556	4.341294444	5.605173333	5.488918889	4.257304444	6.165401111	5.848341111	6.225712222	4.744707778	4.659084444	5.8003	4.961147778	6.216733333
	Senescence	8.128936667	7.73887	5.901986667	3.41132	3.41132	4.48756	8.1765	9.18596667	6.294733333	6.213893333	7.844416667	7.525	6.20766667	5.45696667	6.23	4.234023333	6.603183333	5.669786667	7.891033333

However, in response to light stimulus or different biotic stresses, they showed similar expression patterns. They both showed increased expression under high light intensity and less expressed under dark conditions (Figure 2.4b). They also showed significantly more expression in response to *Botrytis cinerea* infection than control (Wilcoxon Rank Sum test, $P = 0.004$ for both AT5G48400 and AT5G48410) (Figure 2.4d) than control. This finding will also help creating improved crop plants against this pathogen. Similar diversification patterns were also observed in another duplicate pair AT2G32390 and AT2G32400. Here, expression of AT2G32390 was found to be more than AT2G32400 in response to *Alternaria brassicicola* while the opposite was observed in response to *Pseudomonas syringae* pv. *tomato* infection (Figure 2.4c). These differences among different tandemly repeated AtGLR genes indicate that after duplication, some functional diversification took place. However, this phenomenon has not been observed in another tandem pair *i.e.* AT2G24710 and AT2G24720. They showed similar expression pattern in different plant parts, developmental phases as well as in response to light stimulus and different biotic stresses.

Six genes (AT2G17260, AT2G29100, AT2G29120, AT3G04110, AT3G07520 and AT5G11180) showed increased expression both under high light (Figure 2.4b) and dark conditions than control. However, four genes (AT2G29110, AT2G32390, AT2G32400 and AT4G31710) showed lower expression levels both in high light and dark conditions than control. All these results showed that the relationship of light intensity and gene expression of these genes are non-linear. Notably, different genes showed varied level of expression under the attack of different pathogens (Figure 2.4d). For example, both AT5G48400 and AT5G48410 showed higher level of expression in response to the infection of *Pseudomonas syringae* pv. *tomato* while, in response to *Pieris rapae* infection, three other genes (AT5G27100, AT4G31710 and AT2G17260) showed higher expression. However, more researches are needed to determine the functions of the AtGLR genes in these different conditions by additional laboratory experiments.

The tissue specificity index (τ) ranged from 0.049 (AT1G42540) to 0.317 (AT2G29120) (Figure 2.4e). The three tandemly arranged genes namely AT2G29100, AT2G29110 and AT2G29120 showed difference in tissue specificity. While AT2G29120 showed the highest tissue specificity, AT2G29100 showed the lowest. Table 2.4, showed that AT2G29120 and AT2G29110 diverged much earlier than the

divergence of AT2G29110 and AT2G29100. These results indicate that among these genes, tissue specificity decreased following divergence. Moreover, the expression levels of AT2G29100 in different tissues throughout the developmental phases are also less than AT2G29110 and AT2G29120 (Figure 2.4a). According to Li et al. (2005), the most likely chance for a redundant duplicate is nonfunctionalization. The lower expression and tissue specificity of AT2G29100 may indicate gradual nonfunctionalization of this gene after duplication. However, further study is needed to confirm this.

GC₃ of the transcripts ranged from 0.355 (AT5G11180.1) to 0.437 (AT2G29120.1) (Figure 2.5a). According to Straussman et al. (2009), methylation and GC₃ are correlated. GC₃-rich genes provide more targets for *de novo* methylation which can serve as an additional mechanism of transcriptional regulation that ultimately increases adaptability to a species under external stresses (Elhaik and Tatarinova 2012). Our analysis showed that one of the three transcripts (AT2G32390.3) of AT2G32390 showed lower GC₃ than the other two. This indicates that AT2G32390.3 may have lower targets for methylation than the other two transcripts.

In several genomes, some codons are preferred over others by selection as they are, probably translated more accurately and efficiently (Bulmer 1991; Duret 2002; Hershberg and Petrov 2008; Hershberg and Petrov 2009; Shields and Sharp 1987; Shields et al. 1988). Favoured codons should be strongly selected for genes that are more highly expressed (Hershberg and Petrov 2009). Thus, highly expressed genes show high levels of codon bias (Duret and Mouchiroud 1999; Goetz and Fuglsang 2005; Ikemura 1985). We have measured the effective number of codons (EN_c) to get a clear picture of codon usage bias and thus, gene expression. Moreover, some of the AtGLR genes have more than one transcript. By using EN_c, we have got some clues about their expression patterns. EN_c values ranged from 48.4 (AT5G48400.1) to 57.3 (AT2G32390.1) (Figure 2.5 b).

The gene AT2G32390 has three transcripts of which one transcript (AT2G32390.3) showed higher codon bias than the other two. This indicates that AT2G32390.3 shows somewhat higher expression than the other two transcripts. AT2G29100.1 (the only transcript of the gene AT2G29100) has lower EN_c value than the transcripts of AT2G29110 and AT2G29120. One notable observation is that AT2G24710 showed slightly lower EN_c than AT2G24720 although they showed similar expression patterns. The reason is not clear yet. Overall, analysis of EN_c not only provided us the

codon usage bias of all the AtGLR genes, but also some indirect evidence of differential expression of the transcripts of the same gene which was not clear from the expression analysis.

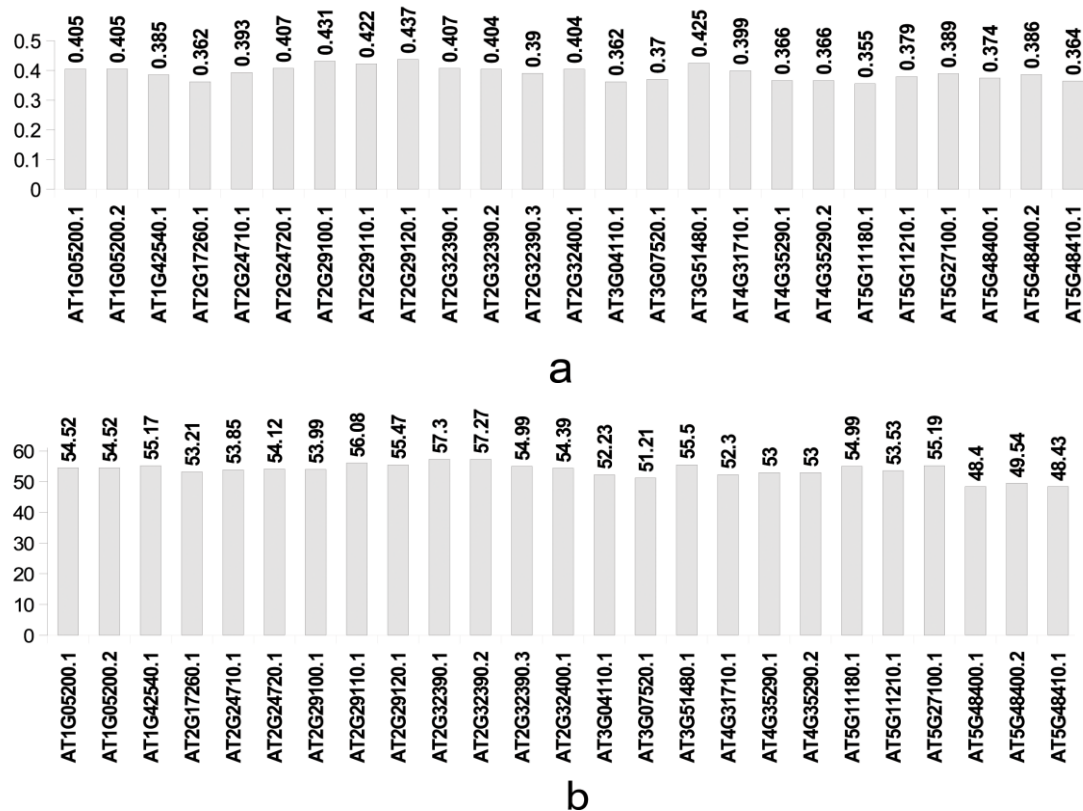


Figure 2.5 : Graphs showing (b) EN_c of different transcripts of AtGLR genes.

2.3.4 Gene ontology annotation

The GO analysis performed with Blast2GO showed the participation of AtGLRs in diverse biological processes (Figure 2.6 a). It was found that all these proteins play their role in three major processes viz. cellular calcium ion homeostasis, ionotropic glutamate receptor signaling pathway and response to light stimulus. Seven of these proteins showed involvement in G-protein coupled receptor signaling pathway. G protein-coupled receptor mediated ligand signaling is a well-known mechanism of signal perception in eukaryotes and in plants, heterotrimeric G protein signaling has been implicated in the control of cell division as well as in transduction of a number of different phytohormone signals (Pandey and Assmann 2004). Our results, thus, showed the importance of some AtGLR genes in this important mode of signal transduction. Other than these, some members participate in cellular response to water deprivation, phosphate starvation and cold as well as response to wound and nematode

attack. This shows that AtGLRs have role in plant defense and stress response. Molecular function analysis identified several functions (Figure 2.6 b).

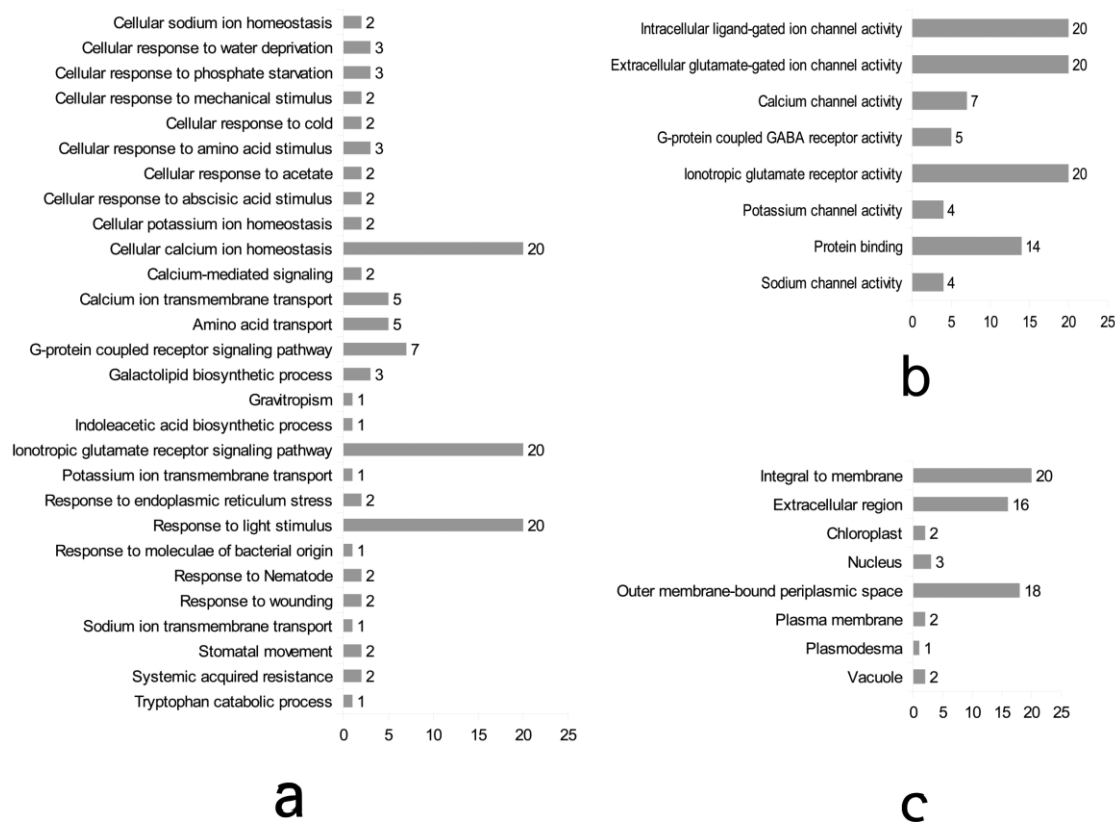


Figure 2.6 : Gene Ontology (GO) distributions for the AtGLR proteins as revealed by the Blast2GO program. The gene ontology is represented by three categories, (a) biological processes, (b) molecular functions and (c) cellular components.

It was found that all these proteins serve three major molecular functions viz. intracellular ligand-gated ion channel activity, extracellular glutamate-gated ion channel activity and ionotropic glutamate receptor activity. Other functions include sodium channel activity (4, 20%), potassium channel activity (4, 20%), protein binding (14, 70%) and calcium channel activity (7, 35%). Existence of plant GLRs has been reported to be correlated with increases in $[Ca^{2+}]_{cyt}$ (Demidchik et al. 2004; Dennison and Spalding 2000; Dubos et al. 2003; Meyerhoff et al. 2005; Michard et al. 2011). It has also been shown that GLRs are involved in the generation of Ca^{2+} influx oscillations in *Arabidopsis* pollen tubes (Michard et al. 2011). According to Qi et al. (2006), sensing of six amino acids (Glutamate, Glycine, alanine, serine, asparagine, and cysteine) in the rhizosphere as well as possibly extracellular peptides is coupled

with calcium ion signaling through a glutamate receptor-dependent mechanism which is homologous to a fundamental component of animal neuronal signaling. Notably, 5 (40%) of AtGLRs (AT1G05200, AT1G42540, AT2G32390, AT2G32400 and AT3G51480) have shown to comprise G-protein coupled GABA (γ -aminobutyric acid) receptor activity. Although GABA is an inhibitory neurotransmitter in mammalian systems, it has been found to be involved in pollen tube growth (Bouché and Fromm 2004). It also inhibits root growth in *Allium sativum* (Mukherjee 2014). All these data suggest a possible interaction between AtGLRs and GABA. Further studies are needed to better understand this aspect. AtGLRs are also shown to occupy different cellular compartments (Figure 2.6c). Naturally, they were all integral to membrane. Notably, one gene (AT2G32400) showed its presence in plasmodesma.

2.3.5 *Cis*-acting elements in the AtGLR gene promoters and corresponding transcription factors

Investigation of *cis*-acting elements in the promoter which regulate gene expression at the transcriptional level are essential for better understanding of gene regulation (Hernandez-Garcia and Finer 2014). The *cis*-acting regulatory elements provide binding site for transcription factors, which synergistically regulates transcription of genes (Manuka et al. 2015). Thus, knowledge of the regulatory *cis*-elements in the promoter regions as well as the transcription factors that bind with those *cis*-elements is necessary to understand the spatial and temporal expression pattern of AtGLR genes. So, we have investigated the *cis*-acting elements as well as their corresponding transcription factors from AGRIS database. It was found that RAV1-A binding site motif was the most frequent motif (6.1 motifs per gene) followed by GATA promoter motif [LRE] (5.75 motifs per gene) (Figure 2.7, Table 2.6). The results obtained with AGRIS also suggested that the transcription factor associated with RAV1-A binding site is ABI3/VP1 (Table 2.6).

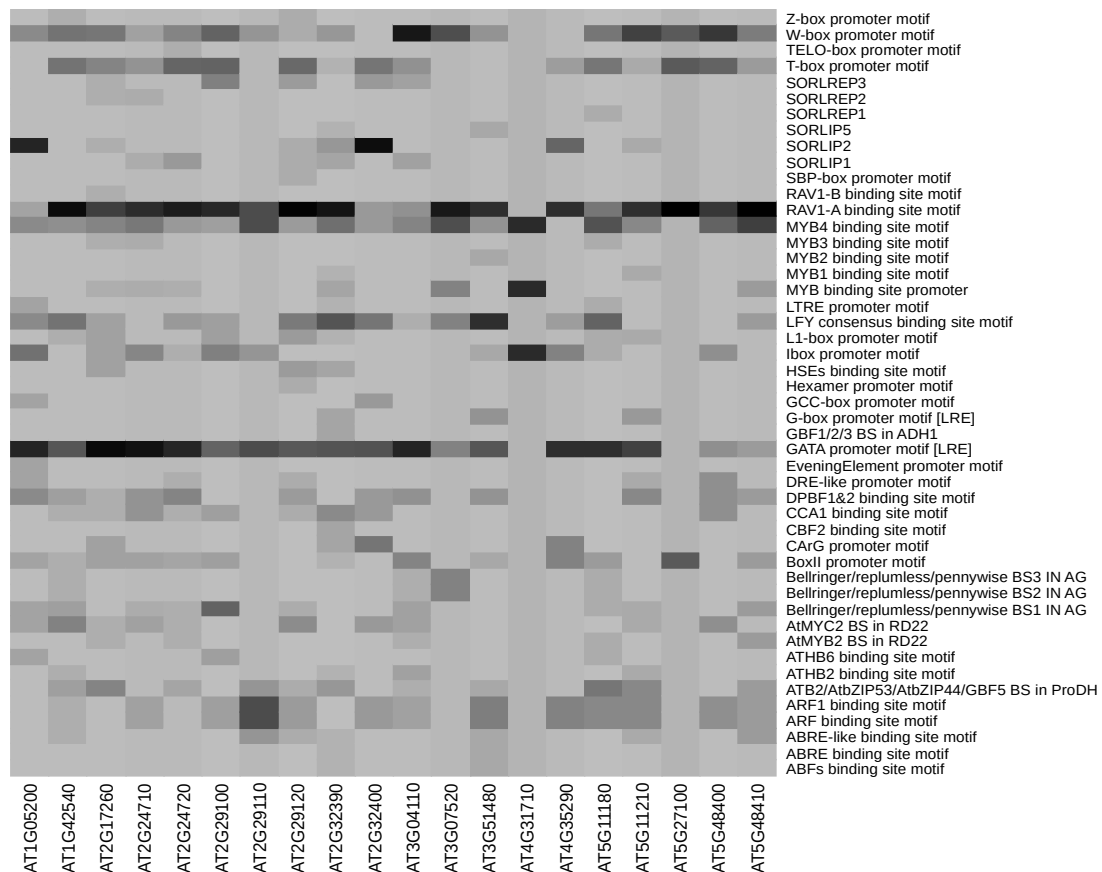


Figure 2.7: Heatmap representation of the number of *cis*-acting elements in different AtGLR genes.

ABI3/VP1 proteins act as intermediaries in regulating abscisic acid (ABA)-responsive genes during seed development (Lazarova et al. 2002). Moreover, Feng et al. (2014) showed that *Arabidopsis* RAV1 transcription factor plays an important role in ABA signaling. ABA is well recognized for its role in stomatal closure (Chen et al. 2008) and our Blast2GO results show that two genes (AT2G17260 and AT4G35290) participate in stomatal closing. These two genes possess RAV1-A binding sites in their promoters. This suggests their role in stomatal movement through ABA signaling. It is also noteworthy that RAV1 is a cold-responsive transcription factor (Fowler and Thomashow 2002). Our Blast2GO results also showed cellular response to cold by two AtGLR genes. Thus, the role of RAV1 in ABA and cold responsiveness can open a new avenue of research. GATA is a type of LRE (Light Responsive Element) (Chattopadhyay et al. 1998). GATA LRE motifs have been shown to be associated with light and nitrate-dependent control of transcription in plants (Reyes et al. 2004; Terzaghi and Cashmore 1995). Moreover,

GATA-box binding TFs are thought to play role in light signaling (Gangappa and Chattopadhyay 2013). However, in AT4G31710, no GATA promoter motif has been found. In this gene, one Ibox promoter motif has been found which is a conserved motif found upstream of light regulated genes (Giuliano et al. 1988). Another gene, AT5G27100 also lack GATA promoter motif although contains a Box II promoter motif. Box II is a promoter *cis*-element found in light-regulated genes (Green et al. 1987). All these indicated that all the AtGLR genes are light responsive. This is also supported by Gene Ontology results which showed that all the 20 AtGLR genes are responsive to light stimulus (Figure 2.6a). Chen et al. (2008) showed integration of light and ABA signaling that help young seedlings to better adapt to environmental stresses. Presence of both RAV1-A binding site motifs and GATA promoter motifs in majority of the AtGLR genes indicates their light and ABA dependent functioning. Based on these results, further laboratory experiments are necessary for better clarification of this aspect.

Several of the AtGLR genes showed the presence of WRKY, MYB and LFY binding sites (Figure 2.7). WRKY proteins are associated with defense against pathogens (Chen et al. 2002; Dong et al. 2003), responses to the abiotic stresses of wounding (Cheong et al. 2002; Hara et al. 2000), drought and heat (Rizhsky et al. 2002) as well as cold (Huang and Duman 2002; Pnueli et al. 2002). MYB transcription factors take part in development, hormone signal transduction, secondary metabolism, disease resistance and abiotic stress tolerance in plants (Allan et al. 2008; Cominelli and Tonelli 2009). LFY is a transcription factor with role in floral development (Liu et al. 2009; Moyroud et al. 2010). Indeed, our Blast2GO results as well as expression profiles showed involvement of AtGLRs in many of the above processes. This is further supported by the presence of binding sites for transcription factors related to these processes. While role of AtGLRs in floral development is not fully clear at this moment, based on our bioinformatic studies which showed the presence of LFY binding sites, future studies should emphasize on their role in floral development.

Table 2.6 : Details of the cis-acting regulatory sequences and corresponding transcription factors of the 20 AtGLR genes (... represents unknown TF).

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT1G05200	AtMYC2 BS in RD22	1504488	1504493	cacatg	BHLH
	Bellringer/replumless/pennywise BS1 IN AG	1505181	1505188	aaattaaa	Homeobox
	W-box promoter motif	1504287	1504292	ttgacc	WRKY
	W-box promoter motif	1504057	1504062	ttgact	WRKY
	ATHB6 binding site motif	1504677	1504685	caattatta	HB
	DPBF1&2 binding site motif	1504832	1504838	acacaag	bZIP
	DPBF1&2 binding site motif	1504488	1504494	acacatg	bZIP
	MYB4 binding site motif	1504762	1504768	aactacc	MYB
	MYB4 binding site motif	1505127	1505133	aacaaac	MYB
	RAV1-A binding site motif	1504217	1504221	caaca	ABI3VP1
	LFY consensus binding site motif	1504025	1504030	ccaatg	LFY
	LFY consensus binding site motif	1504349	1504354	ccactg	LFY
	BoxII promoter motif	1503905	1503910	gggttaa	...
	DRE-like promoter motif	1504327	1504335	taccgacaa	...
	EveningElement promoter motif	1504554	1504562	aaaatatct	...
	GATA promoter motif [LRE]	1503738	1503743	tgataa	...
	GATA promoter motif [LRE]	1504777	1504782	tgatag	...
	GATA promoter motif [LRE]	1504954	1504959	tgataa	...
	GATA promoter motif [LRE]	1504507	1504512	tgataa	...
	GATA promoter motif [LRE]	1503972	1503977	agataa	...
	GATA promoter motif [LRE]	1503949	1503954	agataa	...
	GCC-box promoter motif	1504630	1504635	gccgcc	...
	Ibox promoter motif	1503739	1503744	gataag	...
	Ibox promoter motif	1504955	1504960	gataag	...
	Ibox promoter motif	1503820	1503825	gataag	...
	LTRE promoter motif	1504328	1504334	accgaca	...
	SORLIP2	1504190	1504194	gggcc	...
	SORLIP2	1505290	1505294	gggcc	...
	SORLIP2	1504190	1504194	gggcc	...
	SORLIP2	1503989	1503993	gggcc	...
	SORLIP2	1503893	1503897	gggcc	...
	SORLIP2	1503871	1503875	gggcc	...

Table continued -

Table continued -

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT1G42540	AtMYC2 BS in RD22	15975380	15975385	cacatg	BHLH
	AtMYC2 BS in RD22	15975538	15975543	cacatg	BHLH
	AtMYC2 BS in RD22	15974061	15974066	cacatg	BHLH
	AtMYC2 BS in RD22	15973326	15973331	cacatg	BHLH
	Bellringer/replumless/pennywise BS1 IN AG	15974489	15974496	aaattaa	Homeobo x
	Bellringer/replumless/pennywise BS1 IN AG	15974972	15974979	aaattaa	Homeobo x
	Bellringer/replumless/pennywise BS2 IN AG	15975899	15975906	aaattagt	Homeobo x
	Bellringer/replumless/pennywise BS3 IN AG	15975900	15975907	actaattt	Homeobo x
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	15974688	15974693	actcat	bZIP
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	15974321	15974326	actcat	bZIP
	W-box promoter motif	15973445	15973450	ttgact	WRKY
	W-box promoter motif	15974545	15974550	ttgact	WRKY
	W-box promoter motif	15974901	15974906	ttgacc	WRKY
	W-box promoter motif	15974931	15974936	ttgacc	WRKY
	W-box promoter motif	15974607	15974612	ttgact	WRKY
	ARF1 binding site motif	15974783	15974788	tgtctc	ARF
	ATHB2 binding site motif	15975341	15975349	taataatta	HB
	CCA1 binding site motif	15975556	15975563	aaaaatct	MYB-related
	DPBF1&2 binding site motif	15974763	15974769	acacttg	bZIP
	DPBF1&2 binding site motif	15975379	15975385	acacatg	bZIP
	MYB4 binding site motif	15973681	15973687	aacaaac	MYB
	MYB4 binding site motif	15974831	15974837	aacaaac	MYB
	MYB4 binding site motif	15974590	15974596	aacaaac	MYB
	RAV1-A binding site motif	15974119	15974123	caaca	ABI3VP1
	RAV1-A binding site motif	15974751	15974755	caaca	ABI3VP1
	RAV1-A binding site motif	15974761	15974765	caaca	ABI3VP1
	RAV1-A binding site motif	15974852	15974856	caaca	ABI3VP1
	RAV1-A binding site motif	15975137	15975141	caaca	ABI3VP1
	RAV1-A binding site motif	15975549	15975553	caaca	ABI3VP1
	RAV1-A binding site motif	15974928	15974932	caaca	ABI3VP1
	RAV1-A binding site motif	15974908	15974912	caaca	ABI3VP1
	RAV1-A binding site motif	15974845	15974849	caaca	ABI3VP1
	RAV1-A binding site motif	15974593	15974597	caaca	ABI3VP1
	RAV1-A binding site motif	15974366	15974370	caaca	ABI3VP1
	RAV1-A binding site motif	15974357	15974361	caaca	ABI3VP1
	LFY consensus binding site motif	15973012	15973017	ccaatg	LFY
	LFY consensus binding site motif	15973417	15973422	ccattg	LFY
	LFY consensus binding site motif	15973881	15973886	ccaatg	LFY
	LFY consensus binding site motif	15974905	15974910	ccaatg	LFY

Table continued -

Table continued -

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT1G42540	LFY consensus binding site motif	15973417	15973422	ccaatg	LFY
	ABRE-like binding site motif	15973937	15973944	tacgtgta	...
	ARF binding site motif	15974783	15974788	tgtctc	...
	BoxII promoter motif	15974308	15974313	gggttaa	...
	GATA promoter motif [LRE]	15973931	15973936	tgataa	...
	GATA promoter motif [LRE]	15975919	15975924	tgataa	...
	GATA promoter motif [LRE]	15975397	15975402	agatag	...
	GATA promoter motif [LRE]	15975167	15975172	tgataa	...
	GATA promoter motif [LRE]	15974726	15974731	tgataa	...
	GATA promoter motif [LRE]	15974683	15974688	tgataa	...
	GATA promoter motif [LRE]	15974037	15974042	tgatag	...
	L1-box promoter motif	15974499	15974506	taaagtga	...
	T-box promoter motif	15974870	15974875	actttg	...
	T-box promoter motif	15975631	15975636	actttg	...
	T-box promoter motif	15975854	15975859	actttg	...
	T-box promoter motif	15974738	15974743	actttg	...
	T-box promoter motif	15973256	15973261	actttg	...
	Z-box promoter motif	15973936	15973943	atacgtgt	...

Table continued -

Table continued -

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT2G17260	AtMYB2 BS in RD22	7509705	7509711	ctaacca	MYB
	AtMYC2 BS in RD22	7510030	7510035	cacatg	BHLH
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	7510056	7510061	actcat	bZIP
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	7511413	7511418	actcat	bZIP
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	7511757	7511762	actcat	bZIP
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	7510441	7510446	actcat	bZIP
	W-box promoter motif	7509117	7509122	ttgact	WRKY
	W-box promoter motif	7509838	7509843	ttgact	WRKY
	W-box promoter motif	7511192	7511197	ttgact	WRKY
	W-box promoter motif	7511742	7511747	ttgacc	WRKY
	W-box promoter motif	7511703	7511708	ttgact	WRKY
	CARg promoter motif	7509769	7509778	ccatatttgg	MADS
	CARg promoter motif	7509768	7509777	cctaatatgg	MADS
	CCA1 binding site motif	7509742	7509749	aaaaatct	MYB-related
	DPBF1&2 binding site motif	7510248	7510254	acacctg	bZIP
	HSEs binding site motif	7509532	7509541	agaaagttct	HSF
	HSEs binding site motif	7509531	7509540	agaactttct	HSF
	MYB binding site promoter	7509603	7509610	caccaaac	MYB
	MYB3 binding site motif	7510796	7510803	taactaac	MYB
	MYB4 binding site motif	7509128	7509134	aacaacc	MYB
	MYB4 binding site motif	7509604	7509610	accaaac	MYB
	MYB4 binding site motif	7509886	7509892	aactaac	MYB
	MYB4 binding site motif	7510796	7510802	aactaac	MYB
	RAV1-A binding site motif	7509052	7509056	caaca	ABI3VP1
	RAV1-A binding site motif	7509127	7509131	caaca	ABI3VP1
	RAV1-A binding site motif	7509262	7509266	caaca	ABI3VP1
	RAV1-A binding site motif	7509966	7509970	caaca	ABI3VP1
	RAV1-A binding site motif	7510678	7510682	caaca	ABI3VP1
	RAV1-A binding site motif	7510905	7510909	caaca	ABI3VP1
	RAV1-A binding site motif	7509820	7509824	caaca	ABI3VP1
	RAV1-A binding site motif	7509789	7509793	caaca	ABI3VP1
	RAV1-A binding site motif	7509441	7509445	caaca	ABI3VP1
	RAV1-B binding site motif	7510249	7510254	cacctg	ABI3VP1

Table continued -

Table continued -

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT2G17260	LFY consensus binding site motif	7511099	7511104	ccaatg	LFY
	LFY consensus binding site motif	7510530	7510535	ccactg	LFY
	BoxII promoter motif	7511299	7511304	gggttaa	...
	BoxII promoter motif	7511541	7511546	gggttaa	...
	GATA promoter motif [LRE]	7509351	7509356	tgataa	...
	GATA promoter motif [LRE]	7510163	7510168	tgatag	...
	GATA promoter motif [LRE]	7510472	7510477	tgataa	...
	GATA promoter motif [LRE]	7510479	7510484	tgataa	...
	GATA promoter motif [LRE]	7511246	7511251	agataa	...
	GATA promoter motif [LRE]	7511270	7511275	tgataa	...
	GATA promoter motif [LRE]	7511307	7511312	tgataa	...
	GATA promoter motif [LRE]	7511568	7511573	agataa	...
	GATA promoter motif [LRE]	7511493	7511498	tgataa	...
	GATA promoter motif [LRE]	7511255	7511260	tgataa	...
	GATA promoter motif [LRE]	7510666	7510671	agatag	...
	GATA promoter motif [LRE]	7509681	7509686	agataa	...
	GATA promoter motif [LRE]	7508902	7508907	agatag	...
	Ibox promoter motif	7509352	7509357	gataag	...
	Ibox promoter motif	7511370	7511375	gataag	...
	L1-box promoter motif	7511500	7511507	taaagtga	...
	L1-box promoter motif	7510918	7510925	taaagtga	...
	T-box promoter motif	7509481	7509486	actttg	...
	T-box promoter motif	7509557	7509562	actttg	...
	T-box promoter motif	7509841	7509846	actttg	...
	T-box promoter motif	7510576	7510581	actttg	...
	SORLREP2	7510870	7510878	ataaacgt	...
	SORLREP3	7510839	7510847	tgtatatat	...
	SORLIP2	7510176	7510180	gggcc	...

Table continued -

Table continued -

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT2G24710	AtMYC2 BS in RD22	10528496	10528501	cacatg	BHLH
	AtMYC2 BS in RD22	10527807	10527812	cacatg	BHLH
	Bellringer/replumless/pennywise BS1 IN AG	10528057	10528064	aaattaaa	Homeobox
	W-box promoter motif	10528320	10528325	ttgact	WRKY
	W-box promoter motif	10529243	10529248	ttgact	WRKY
	ARF1 binding site motif	10528045	10528050	tgtctc	ARF
	ARF1 binding site motif	10528531	10528536	tgtctc	ARF
	CCA1 binding site motif	10528902	10528909	aacaatct	MYB-related
	CCA1 binding site motif	10529848	10529855	aaaaatct	MYB-related
	CCA1 binding site motif	10529212	10529219	aaaaatct	MYB-related
	DPBF1&2 binding site motif	10529536	10529542	acacttg	bZIP
	DPBF1&2 binding site motif	10528496	10528502	acacatg	bZIP
	DPBF1&2 binding site motif	10528079	10528085	acacacg	bZIP
	MYB binding site promoter	10527662	10527669	aacaaac	MYB
	MYB3 binding site motif	10528065	10528072	taactaac	MYB
	MYB4 binding site motif	10527663	10527669	acaaac	MYB
	MYB4 binding site motif	10528066	10528072	aactaac	MYB
	MYB4 binding site motif	10528784	10528790	aacaaac	MYB
	MYB4 binding site motif	10528472	10528478	aactaac	MYB
	MYB4 binding site motif	10527950	10527956	aacaaac	MYB
	RAV1-A binding site motif	10528191	10528195	caaca	ABI3VP1
	RAV1-A binding site motif	10528390	10528394	caaca	ABI3VP1
	RAV1-A binding site motif	10528440	10528444	caaca	ABI3VP1
	RAV1-A binding site motif	10528465	10528469	caaca	ABI3VP1
	RAV1-A binding site motif	10528783	10528787	caaca	ABI3VP1
	RAV1-A binding site motif	10529150	10529154	caaca	ABI3VP1
	RAV1-A binding site motif	10528481	10528485	caaca	ABI3VP1
	RAV1-A binding site motif	10528140	10528144	caaca	ABI3VP1
	RAV1-A binding site motif	10527981	10527985	caaca	ABI3VP1
	RAV1-A binding site motif	10527972	10527976	caaca	ABI3VP1
	RAV1-A binding site motif	10527953	10527957	caaca	ABI3VP1
	ARF binding site motif	10528045	10528050	tgtctc	...
	ARF binding site motif	10528531	10528536	tgtctc	...
	BoxII promoter motif	10528645	10528650	gggttaa	...
	BoxII promoter motif	10528352	10528357	gggttaa	...
	GATA promoter motif [LRE]	10529089	10529094	agatag	...
	GATA promoter motif [LRE]	10529110	10529115	tgataa	...
	GATA promoter motif [LRE]	10529467	10529472	agataa	...
	GATA promoter motif [LRE]	10529826	10529831	tgataa	...
	GATA promoter motif [LRE]	10529819	10529824	tgatag	...
	GATA promoter motif [LRE]	10529766	10529771	agatag	...

Table continued -

Table continued -

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT2G24710	GATA promoter motif [LRE]	10529711	10529716	agataa	...
	GATA promoter motif [LRE]	10528715	10528720	agataa	...
	GATA promoter motif [LRE]	10528630	10528635	agataa	...
	GATA promoter motif [LRE]	10528559	10528564	tgatag	...
	GATA promoter motif [LRE]	10528324	10528329	agataa	...
	GATA promoter motif [LRE]	10528088	10528093	tgataa	...
	GATA promoter motif [LRE]	10527733	10527738	tgatag	...
	Ibox promoter motif	10529825	10529830	gataag	...
	Ibox promoter motif	10529710	10529715	gataag	...
	Ibox promoter motif	10528804	10528809	gataag	...
	Ibox promoter motif	10528323	10528328	gataag	...
	T-box promoter motif	10529690	10529695	actttg	...
	T-box promoter motif	10529202	10529207	actttg	...
	T-box promoter motif	10529120	10529125	actttg	...
	SORLREP2	10529248	10529256	ataaaacgt	...
	SORLIP1	10529124	10529129	agccac	...
AT2G24720	AtMYB2 BS in RD22	10534782	10534788	ctaacca	MYB
	AtMYC2 BS in RD22	10533169	10533174	cacatg	BHLH
	Bellringer/replumless/pennywise BS1 IN AG	10534242	10534249	aaattaaa	Homeobox
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	10534569	10534574	actcat	bZIP
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	10535078	10535083	actcat	bZIP
	W-box promoter motif	10533970	10533975	ttgact	WRKY
	W-box promoter motif	10534000	10534005	ttgact	WRKY
	W-box promoter motif	10534298	10534303	ttgact	WRKY
	W-box promoter motif	10534502	10534507	ttgacc	WRKY
	W-box promoter motif	10534079	10534084	ttgacc	WRKY
	CCA1 binding site motif	10534237	10534244	aaaaatct	MYB-related
	DPBF1&2 binding site motif	10532951	10532957	acactag	bZIP
	DPBF1&2 binding site motif	10533168	10533174	acacatg	bZIP
	DPBF1&2 binding site motif	10533240	10533246	acacacg	bZIP
	DPBF1&2 binding site motif	10535493	10535499	acacaag	bZIP
	DPBF1&2 binding site motif	10533020	10533026	acacacg	bZIP
	MYB binding site promoter	10533962	10533969	aaccaaac	MYB
	MYB4 binding site motif	10533225	10533231	aacaacc	MYB
	MYB4 binding site motif	10533397	10533403	aacaaac	MYB
	MYB4 binding site motif	10533962	10533968	accaaac	MYB
	RAV1-A binding site motif	10532922	10532926	caaca	ABI3VP1
	RAV1-A binding site motif	10533433	10533437	caaca	ABI3VP1
	RAV1-A binding site motif	10534579	10534583	caaca	ABI3VP1
	RAV1-A binding site motif	10534614	10534618	caaca	ABI3VP1

Table continued -

Table continued -

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT2G24720	RAV1-A binding site motif	10535021	10535025	caaca	ABI3VP1
	RAV1-A binding site motif	10535421	10535425	caaca	ABI3VP1
	RAV1-A binding site motif	10535491	10535495	caaca	ABI3VP1
	RAV1-A binding site motif	10535800	10535804	caaca	ABI3VP1
	RAV1-A binding site motif	10535759	10535763	caaca	ABI3VP1
	RAV1-A binding site motif	10535041	10535045	caaca	ABI3VP1
	RAV1-A binding site motif	10534899	10534903	caaca	ABI3VP1
	RAV1-A binding site motif	10534854	10534858	caaca	ABI3VP1
	RAV1-A binding site motif	10534818	10534822	caaca	ABI3VP1
	RAV1-A binding site motif	10534282	10534286	caaca	ABI3VP1
	RAV1-A binding site motif	10533997	10534001	caaca	ABI3VP1
	LFY consensus binding site motif	10533122	10533127	ccaatg	LFY
	LFY consensus binding site motif	10533133	10533138	ccattg	LFY
	LFY consensus binding site motif	10535434	10535439	ccactg	LFY
	BoxII promoter motif	10533988	10533993	gggtta	...
	BoxII promoter motif	10534863	10534868	gggtta	...
	DRE-like promoter motif	10533495	10533503	gaccgacta	...
	GATA promoter motif [LRE]	10533059	10533064	agataa	...
	GATA promoter motif [LRE]	10533695	10533700	tgatag	...
	GATA promoter motif [LRE]	10534056	10534061	agataa	...
	GATA promoter motif [LRE]	10534591	10534596	tgataa	...
	GATA promoter motif [LRE]	10535045	10535050	tgataa	...
	GATA promoter motif [LRE]	10535212	10535217	agatag	...
	GATA promoter motif [LRE]	10535235	10535240	tgataa	...
	GATA promoter motif [LRE]	10535884	10535889	tgataa	...
	GATA promoter motif [LRE]	10535877	10535882	tgatag	...
	GATA promoter motif [LRE]	10535748	10535753	agatag	...
	GATA promoter motif [LRE]	10535011	10535016	agatag	...
	GATA promoter motif [LRE]	10534650	10534655	agataa	...
	GATA promoter motif [LRE]	10533182	10533187	tgatag	...
	GATA promoter motif [LRE]	10533074	10533079	agatag	...
	Ibox promoter motif	10533201	10533206	gataag	...
	T-box promoter motif	10533034	10533039	actttg	...
	T-box promoter motif	10533602	10533607	actttg	...
	T-box promoter motif	10535231	10535236	actttg	...
	T-box promoter motif	10535729	10535734	actttg	...
	T-box promoter motif	10535815	10535820	actttg	...
	T-box promoter motif	10535327	10535332	actttg	...
	T-box promoter motif	10535245	10535250	actttg	...
	T-box promoter motif	10533958	10533963	actttg	...
	TELO-box promoter motif	10533376	10533384	aaaccctaa	...
	SORLIP1	10535249	10535254	agccac	...
	SORLIP1	10533105	10533110	agccac	...
	SORLIP1	10533025	10533030	agccac	...

Table continued -

Table continued -

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT2G29100	Bellringer/replumless/pennywise BS1 IN AG	12512916	12512923	aaattaaa	Homeobox
	Bellringer/replumless/pennywise BS1 IN AG	12513195	12513202	aaattaaa	Homeobox
	Bellringer/replumless/pennywise BS1 IN AG	12512961	12512968	aaattaaa	Homeobox
	W-box promoter motif	12512775	12512780	ttgacc	WRKY
	W-box promoter motif	12513728	12513733	ttgact	WRKY
	W-box promoter motif	12513759	12513764	ttgact	WRKY
	ARF1 binding site motif	12513794	12513799	tgtctc	ARF
	ATHB6 binding site motif	12513349	12513357	caattatta	HB
	CCA1 binding site motif	12512921	12512928	aaaaatct	MYB-related
	MYB4 binding site motif	12512848	12512854	aacaaac	MYB
	RAV1-A binding site motif	12512572	12512576	caaca	ABI3VP1
	RAV1-A binding site motif	12512577	12512581	caaca	ABI3VP1
	RAV1-A binding site motif	12513763	12513767	caaca	ABI3VP1
	RAV1-A binding site motif	12512804	12512808	caaca	ABI3VP1
	RAV1-A binding site motif	12512729	12512733	caaca	ABI3VP1
	LFY consensus binding site motif	12513877	12513882	ccactg	LFY
	ARF binding site motif	12513794	12513799	tgtctc	...
	BoxII promoter motif	12513528	12513533	ggttaa	...
	GATA promoter motif [LRE]	12512020	12512025	agatag	...
	GATA promoter motif [LRE]	12512383	12512388	agataa	...
	GATA promoter motif [LRE]	12513440	12513445	agataa	...
	Ibox promoter motif	12511989	12511994	gataag	...
	Ibox promoter motif	12512026	12512031	gataag	...
	L1-box promoter motif	12512648	12512655	taaatgta	...
	T-box promoter motif	12513833	12513838	actttg	...
	T-box promoter motif	12512986	12512991	actttg	...
	T-box promoter motif	12512082	12512087	actttg	...
	SORLREP3	12513400	12513408	tgtatatat	...
	SORLREP3	12513881	12513889	tgtatatat	...
AT2G29110	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	12518227	12518232	actcat	bZIP
	W-box promoter motif	12518187	12518192	ttgact	WRKY
	ARF1 binding site motif	12517779	12517784	tgtctc	ARF
	ARF1 binding site motif	12518149	12518154	tgtctc	ARF
	ARF1 binding site motif	12518344	12518349	tgtctc	ARF
	MYB4 binding site motif	12517737	12517743	aactaac	MYB
	MYB4 binding site motif	12517867	12517873	aacaaac	MYB
	MYB4 binding site motif	12518082	12518088	aactacc	MYB
	RAV1-A binding site motif	12517812	12517816	caaca	ABI3VP1
	RAV1-A binding site motif	12518191	12518195	caaca	ABI3VP1
	RAV1-A binding site motif	12517727	12517731	caaca	ABI3VP1
	ABRE-like binding site motif	12517788	12517795	gacgtgta	...
	ARF binding site motif	12517779	12517784	tgtctc	...
	ARF binding site motif	12518149	12518154	tgtctc	...
	ARF binding site motif	12518344	12518349	tgtctc	...
	GATA promoter motif [LRE]	12517913	12517918	tgataa	...
	GATA promoter motif [LRE]	12518350	12518355	agataa	...
	GATA promoter motif [LRE]	12518210	12518215	tgataa	...
	Ibox promoter motif	12518349	12518354	gataag	...

Table continued -

Table continued -

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT2G29120	AtMYC2 BS in RD22	12523396	12523401	cacatg	BHLH
	AtMYC2 BS in RD22	12524361	12524366	cacatg	BHLH
	AtMYC2 BS in RD22	12523405	12523410	cacatg	BHLH
	Bellringer/replumless/pennywise BS1 IN AG	12523259	12523266	aaattaa	Homeobox
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	12525079	12525084	actcat	bZIP
	W-box promoter motif	12524405	12524410	ttgact	WRKY
	ARF1 binding site motif	12524062	12524067	tgtctc	ARF
	ARF1 binding site motif	12524807	12524812	tgtctc	ARF
	CCA1 binding site motif	12525539	12525546	aaaaatct	MYB-related
	DPBF1&2 binding site motif	12523178	12523184	acacgag	bZIP
	DPBF1&2 binding site motif	12523405	12523411	acacatg	bZIP
	HSEs binding site motif	12525110	12525119	agaagcttct	HSF
	HSEs binding site motif	12525109	12525118	agaagcttct	HSF
	MYB4 binding site motif	12525665	12525671	aacaaac	MYB
	MYB4 binding site motif	12524735	12524741	aacaaac	MYB
	RAV1-A binding site motif	12523176	12523180	caaca	ABI3VP1
	RAV1-A binding site motif	12524155	12524159	caaca	ABI3VP1
	RAV1-A binding site motif	12524428	12524432	caaca	ABI3VP1
	RAV1-A binding site motif	12525140	12525144	caaca	ABI3VP1
	RAV1-A binding site motif	12525886	12525890	caaca	ABI3VP1
	RAV1-A binding site motif	12524738	12524742	caaca	ABI3VP1
	RAV1-A binding site motif	12524568	12524572	caaca	ABI3VP1
	RAV1-A binding site motif	12523725	12523729	caaca	ABI3VP1
	RAV1-A binding site motif	12523245	12523249	caaca	ABI3VP1
	RAV1-A binding site motif	12523228	12523232	caaca	ABI3VP1
	RAV1-A binding site motif	12523083	12523087	caaca	ABI3VP1
	LFY consensus binding site motif	12523541	12523546	ccaatg	LFY
	LFY consensus binding site motif	12524986	12524991	ccaatg	LFY
	LFY consensus binding site motif	12525583	12525588	ccattg	LFY
	LFY consensus binding site motif	12524986	12524991	ccattg	LFY
	ABRE-like binding site motif	12523153	12523160	tacgtgtc	...
	ARF binding site motif	12524062	12524067	tgtctc	...
	ARF binding site motif	12524807	12524812	tgtctc	...
	DRE-like promoter motif	12523432	12523440	agccgacct	...
	GATA promoter motif [LRE]	12523683	12523688	agatag	...
	GATA promoter motif [LRE]	12524547	12524552	tgataa	...
	GATA promoter motif [LRE]	12525703	12525708	agataa	...
	GATA promoter motif [LRE]	12525243	12525248	agataa	...
	GATA promoter motif [LRE]	12524886	12524891	agatag	...
	GATA promoter motif [LRE]	12524681	12524686	tgataa	...
	Hexamer promoter motif	12524460	12524465	ccgtcg	...
	L1-box promoter motif	12524854	12524861	ttaatgca	...
	L1-box promoter motif	12524741	12524748	ttaatgca	...
	SBP-box promoter motif	12524258	12524266	ttcgtacaa	...
	T-box promoter motif	12523036	12523041	actttg	...
	T-box promoter motif	12524335	12524340	actttg	...
	T-box promoter motif	12525860	12525865	actttg	...
	T-box promoter motif	12525547	12525552	actttg	...
	T-box promoter motif	12523046	12523051	actttg	...
	Z-box promoter motif	12523152	12523159	atacgtgt	...
	SORLREP3	12525717	12525725	tgatatatat	...
	SORLREP3	12525745	12525753	tgatatatat	...
	SORLIP1	12523687	12523692	agccac	...
	SORLIP2	12525134	12525138	gggcc	...

Table continued -

Table continued -

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT2G32390	GBF1/2/3 BS in ADH1	13759376	13759383	ccacgtgg	...
	GBF1/2/3 BS in ADH1	13759375	13759382	ccacgtgg	...
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	13761548	13761553	actcat	bZIP
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	13761596	13761601	actcat	bZIP
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	13761256	13761261	actcat	bZIP
	W-box promoter motif	13760553	13760558	ttgact	WRKY
	W-box promoter motif	13761322	13761327	ttgact	WRKY
	W-box promoter motif	13759159	13759164	ttgacc	WRKY
	CARg promoter motif	13761473	13761482	ccaaattgg	MADS
	CARg promoter motif	13761472	13761481	ccaaattgg	MADS
	ABFs binding site motif	13759374	13759381	cacgtggc	bZIP
	ATHB2 binding site motif	13761564	13761572	taatcatta	HB
	CBF2 binding site motif	13759376	13759383	ccacgtgg	AP2-EREBP
	CBF2 binding site motif	13759375	13759382	ccacgtgg	AP2-EREBP
	CCA1 binding site motif	13761799	13761806	aaaaatct	MYB-related
	CCA1 binding site motif	13761923	13761930	aaaaatct	MYB-related
	CCA1 binding site motif	13760675	13760682	aacaatct	MYB-related
	CCA1 binding site motif	13759415	13759422	aacaatct	MYB-related
	HSEs binding site motif	13759125	13759134	agaaacttct	HSF
	HSEs binding site motif	13759124	13759133	agaagtttct	HSF
	MYB binding site promoter	13759069	13759076	aaccaaac	MYB
	MYB binding site promoter	13759170	13759177	aaccaaac	MYB
	MYB1 binding site motif	13761511	13761518	atccaacc	MYB
	MYB4 binding site motif	13759070	13759076	accaaac	MYB
	MYB4 binding site motif	13761201	13761207	aacaacc	MYB
	MYB4 binding site motif	13761557	13761563	accaaac	MYB
	MYB4 binding site motif	13761433	13761439	accaaac	MYB
	MYB4 binding site motif	13761010	13761016	aacaaac	MYB
	MYB4 binding site motif	13759170	13759176	accaaac	MYB
	RAV1-A binding site motif	13759024	13759028	caaca	ABI3VP1
	RAV1-A binding site motif	13759102	13759106	caaca	ABI3VP1
	RAV1-A binding site motif	13760301	13760305	caaca	ABI3VP1
	RAV1-A binding site motif	13760417	13760421	caaca	ABI3VP1
	RAV1-A binding site motif	13760830	13760834	caaca	ABI3VP1
	RAV1-A binding site motif	13761348	13761352	caaca	ABI3VP1
	RAV1-A binding site motif	13761455	13761459	caaca	ABI3VP1
	RAV1-A binding site motif	13761882	13761886	caaca	ABI3VP1

Table continued -

Table continued -

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT2G32390	RAV1-A binding site motif	13760679	13760683	caaca	ABI3VP1
	RAV1-A binding site motif	13760610	13760614	caaca	ABI3VP1
	RAV1-A binding site motif	13759907	13759911	caaca	ABI3VP1
	RAV1-A binding site motif	13759757	13759761	caaca	ABI3VP1
	RAV1-A binding site motif	13759419	13759423	caaca	ABI3VP1
	LFY consensus binding site motif	13759409	13759414	ccattg	LFY
	LFY consensus binding site motif	13760619	13760624	ccattg	LFY
	LFY consensus binding site motif	13760811	13760816	ccattg	LFY
	LFY consensus binding site motif	13761371	13761376	ccagtg	LFY
	LFY consensus binding site motif	13761897	13761902	ccattg	LFY
	LFY consensus binding site motif	13761411	13761416	ccaatg	LFY
	LFY consensus binding site motif	13760811	13760816	ccaatg	LFY
	LFY consensus binding site motif	13759104	13759109	ccactg	LFY
	ABRE binding site motif	13759374	13759381	cacgtggc	...
	ABRE-like binding site motif	13759374	13759381	cacgtggc	...
	BoxII promoter motif	13760509	13760514	ggftaa	...
	GATA promoter motif [LRE]	13759702	13759707	tgataa	...
	GATA promoter motif [LRE]	13760049	13760054	tgatag	...
	GATA promoter motif [LRE]	13760446	13760451	tgatag	...
	GATA promoter motif [LRE]	13761427	13761432	tgatag	...
	GATA promoter motif [LRE]	13761725	13761730	agatag	...
	GATA promoter motif [LRE]	13761765	13761770	agataa	...
	GATA promoter motif [LRE]	13760835	13760840	agatag	...
	GATA promoter motif [LRE]	13759211	13759216	tgataa	...
	G-box promoter motif [LRE]	13759377	13759382	cacgtg	...
	G-box promoter motif [LRE]	13759376	13759381	cacgtg	...
	L1-box promoter motif	13760209	13760216	taaagtga	...
	LTRE promoter motif	13759242	13759248	accgaca	...
	T-box promoter motif	13760084	13760089	actttg	...
	SORLIP1	13759374	13759379	agccac	...
	SORLIP1	13759710	13759715	agccac	...
	SORLIP2	13759984	13759988	gggcc	...
	SORLIP2	13760591	13760595	gggcc	...
	SORLIP2	13761656	13761660	gggcc	...
	SORLIP5	13759434	13759440	gagtgag	...

Table continued -

Table continued -

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT2G32400	AtMYC2 BS in RD22	13764594	13764599	cacatg	BHLH
	CArG promoter motif	13764121	13764130	cctaatttgg	MADS
	CArG promoter motif	13764120	13764129	ccaaattagg	MADS
	ARF1 binding site motif	13764099	13764104	tgtctc	ARF
	CCA1 binding site motif	13764444	13764451	aaaaatct	MYB-related
	DPBF1&2 binding site motif	13764593	13764599	acacatg	bZIP
	MYB4 binding site motif	13764706	13764712	aacaaac	MYB
	RAV1-A binding site motif	13763928	13763932	caaca	ABI3VP1
	LFY consensus binding site motif	13764686	13764691	ccactg	LFY
	LFY consensus binding site motif	13763962	13763967	ccactg	LFY
	ARF binding site motif	13764099	13764104	tgtctc	...
	GATA promoter motif [LRE]	13764213	13764218	agataa	...
	GATA promoter motif [LRE]	13764407	13764412	tgataa	...
	GATA promoter motif [LRE]	13764742	13764747	agatag	...
	GCC-box promoter motif	13764437	13764442	gccgcc	...
	T-box promoter motif	13764749	13764754	actttg	...
	T-box promoter motif	13764138	13764143	actttg	...
	SORLREP3	13764299	13764307	tgtatatat	...
	SORLIP2	13764651	13764655	gggcc	...
	SORLIP2	13764666	13764670	gggcc	...
	SORLIP2	13764681	13764685	gggcc	...
	SORLIP2	13764681	13764685	gggcc	...
	SORLIP2	13764666	13764670	gggcc	...
AT3G04110	AtMYB2 BS in RD22	1076862	1076868	ctaacca	MYB
	AtMYC2 BS in RD22	1076724	1076729	cacatg	BHLH
	AtMYC2 BS in RD22	1074439	1074444	cacatg	BHLH
	Bellringer/replumless/pennywise BS1 IN AG	1074542	1074549	aaattaaa	Homeobox
	Bellringer/replumless/pennywise BS1 IN AG	1077081	1077088	aaattaaa	Homeobox
	Bellringer/replumless/pennywise BS2 IN AG	1075131	1075138	aaattagt	Homeobox
	Bellringer/replumless/pennywise BS3 IN AG	1075130	1075137	actaattt	Homeobox
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	1076564	1076569	actcat	bZIP
	W-box promoter motif	1074579	1074584	ttgact	WRKY
	W-box promoter motif	1075061	1075066	ttgact	WRKY
	W-box promoter motif	1076408	1076413	ttgacc	WRKY
	W-box promoter motif	1076965	1076970	ttgact	WRKY
	W-box promoter motif	1077172	1077177	ttgact	WRKY
	W-box promoter motif	1077253	1077258	ttgacc	WRKY
	W-box promoter motif	1077142	1077147	ttgact	WRKY

Table continued -

Table continued -

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT3G04110	W-box promoter motif	1075135	1075140	ttgact	WRKY
	W-box promoter motif	1075104	1075109	ttgact	WRKY
	W-box promoter motif	1075005	1075010	ttgacc	WRKY
	W-box promoter motif	1074693	1074698	ttgact	WRKY
	W-box promoter motif	1074629	1074634	ttgact	WRKY
	ARF1 binding site motif	1075175	1075180	tgtctc	ARF
	ARF1 binding site motif	1075559	1075564	tgtctc	ARF
	ATHB2 binding site motif	1074353	1074361	taataatta	HB
	ATHB2 binding site motif	1077280	1077288	taataatta	HB
	DPBF1&2 binding site motif	1075021	1075027	acaccag	bZIP
	DPBF1&2 binding site motif	1075282	1075288	acactag	bZIP
	DPBF1&2 binding site motif	1074439	1074445	acacatg	bZIP
	MYB4 binding site motif	1075261	1075267	aacaaac	MYB
	MYB4 binding site motif	1076860	1076866	aactaac	MYB
	MYB4 binding site motif	1077197	1077203	aactaac	MYB
	MYB4 binding site motif	1076156	1076162	aactaac	MYB
	RAV1-A binding site motif	1075108	1075112	caaca	ABI3VP1
	RAV1-A binding site motif	1077068	1077072	caaca	ABI3VP1
	RAV1-A binding site motif	1075693	1075697	caaca	ABI3VP1
	LFY consensus binding site motif	1077318	1077323	ccattg	LFY
	ARF binding site motif	1075175	1075180	tgtctc	...
	ARF binding site motif	1075559	1075564	tgtctc	...
	BoxII promoter motif	1075544	1075549	ggtaa	...
	BoxII promoter motif	1077033	1077038	ggtaa	...
	BoxII promoter motif	1076255	1076260	ggtaa	...
	BoxII promoter motif	1074533	1074538	ggtaa	...
	GATA promoter motif [LRE]	1074784	1074789	tgatag	...
	GATA promoter motif [LRE]	1074797	1074802	agataa	...
	GATA promoter motif [LRE]	1075386	1075391	agatag	...
	GATA promoter motif [LRE]	1075725	1075730	agataa	...
	GATA promoter motif [LRE]	1075809	1075814	tgataa	...
	GATA promoter motif [LRE]	1076283	1076288	agataa	...
	GATA promoter motif [LRE]	1076930	1076935	tgatag	...
	GATA promoter motif [LRE]	1076712	1076717	agataa	...
	GATA promoter motif [LRE]	1076037	1076042	agataa	...
	GATA promoter motif [LRE]	1075241	1075246	tgataa	...
	GATA promoter motif [LRE]	1075194	1075199	agatag	...
	T-box promoter motif	1075184	1075189	actttg	...
	T-box promoter motif	1076968	1076973	actttg	...
	T-box promoter motif	1077175	1077180	actttg	...
	SORLREP3	1076442	1076450	tgtatatat	...
	SORLREP3	1076431	1076439	tgtatatat	...
	SORLIP1	1075533	1075538	agccac	...
	SORLIP1	1075467	1075472	agccac	...

Table continued -

Table continued -

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT3G07520	Bellringer/replumless/pennywise BS2 IN AG	2399029	2399036	aaattagt	Homeobox
	Bellringer/replumless/pennywise BS3 IN AG	2399028	2399035	actaattt	Homeobox
	W-box promoter motif	2398931	2398936	ttgact	WRKY
	W-box promoter motif	2398787	2398792	ttgacc	WRKY
	MYB binding site promoter	2398485	2398492	aaccaacc	MYB
	MYB4 binding site motif	2398368	2398374	aacaaac	MYB
	MYB4 binding site motif	2398486	2398492	accaacc	MYB
	RAV1-A binding site motif	2398367	2398371	caaca	ABI3VP1
	RAV1-A binding site motif	2398645	2398649	caaca	ABI3VP1
	RAV1-A binding site motif	2398744	2398748	caaca	ABI3VP1
	LFY consensus binding site motif	2398828	2398833	ccaatg	LFY
	GATA promoter motif [LRE]	2398748	2398753	tgataa	...
AT3G51480	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	19110881	19110886	actcat	bZIP
	W-box promoter motif	19110692	19110697	ttgact	WRKY
	W-box promoter motif	19111261	19111266	ttgact	WRKY
	ABFs binding site motif	19110339	19110346	cacgtggc	bZIP
	ARF1 binding site motif	19111256	19111261	tgtctc	ARF
	ARF1 binding site motif	19111026	19111031	tgtctc	ARF
	ARF1 binding site motif	19110868	19110873	tgtctc	ARF
	DPBF1&2 binding site motif	19110872	19110878	acacaag	bZIP
	DPBF1&2 binding site motif	19110341	19110347	acacgtg	bZIP
	MYB2 binding site motif	19110907	19110915	taactcgtt	MYB
	MYB4 binding site motif	19110703	19110709	aacaaac	MYB
	MYB4 binding site motif	19110562	19110568	aacaaac	MYB
	RAV1-A binding site motif	19110593	19110597	caaca	ABI3VP1
	RAV1-A binding site motif	19111496	19111500	caaca	ABI3VP1
	RAV1-A binding site motif	19111294	19111298	caaca	ABI3VP1
	RAV1-A binding site motif	19111180	19111184	caaca	ABI3VP1
	RAV1-A binding site motif	19110689	19110693	caaca	ABI3VP1
	RAV1-A binding site motif	19110565	19110569	caaca	ABI3VP1
	RAV1-A binding site motif	19110553	19110557	caaca	ABI3VP1
	LFY consensus binding site motif	19110864	19110869	ccactg	LFY
	LFY consensus binding site motif	19110958	19110963	ccattg	LFY
	LFY consensus binding site motif	19111090	19111095	ccattg	LFY
	LFY consensus binding site motif	19111207	19111212	ccattg	LFY
	LFY consensus binding site motif	19111090	19111095	ccaatg	LFY
	LFY consensus binding site motif	19111075	19111080	ccagtg	LFY
	LFY consensus binding site motif	19111069	19111074	ccaatg	LFY
	ABRE binding site motif	19110339	19110346	cacgtggc	...
	ABRE-like binding site motif	19110339	19110346	cacgtggc	...
	ARF binding site motif	19111256	19111261	tgtctc	...
	ARF binding site motif	19111026	19111031	tgtctc	...
	ARF binding site motif	19110868	19110873	tgtctc	...
	BoxII promoter motif	19111001	19111006	ggttaa	...
	GATA promoter motif [LRE]	19110675	19110680	agatag	...
	GATA promoter motif [LRE]	19110830	19110835	tgatag	...
	GATA promoter motif [LRE]	19111405	19111410	tgataa	...
	GATA promoter motif [LRE]	19110758	19110763	tgatag	...
	GATA promoter motif [LRE]	19110570	19110575	agatag	...
	G-box promoter motif [LRE]	19110342	19110347	cacgtg	...
	G-box promoter motif [LRE]	19110341	19110346	cacgtg	...
	Ibox promoter motif	19111406	19111411	gataag	...
	SORLIP5	19110884	19110890	gagttag	...

Table continued -

Table continued -

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT4G31710	MYB binding site promoter	15349085	15349092	aaccaaac	MYB
	MYB4 binding site motif	15349086	15349092	accaaac	MYB
	Ibox promoter motif	15349109	15349114	gataag	...
AT4G35290	CArG promoter motif	16789015	16789024	cctatatagg	MADS
	CArG promoter motif	16789014	16789023	cctatatagg	MADS
	ARF1 binding site motif	16788981	16788986	tgtctc	ARF
	ARF1 binding site motif	16788412	16788417	tgtctc	ARF
	RAV1-A binding site motif	16788678	16788682	caaca	ABI3VP1
	RAV1-A binding site motif	16788695	16788699	caaca	ABI3VP1
	RAV1-A binding site motif	16789026	16789030	caaca	ABI3VP1
	RAV1-A binding site motif	16789107	16789111	caaca	ABI3VP1
	RAV1-A binding site motif	16788527	16788531	caaca	ABI3VP1
	LFY consensus binding site motif	16789289	16789294	ccaatg	LFY
	ARF binding site motif	16788981	16788986	tgtctc	...
	ARF binding site motif	16788412	16788417	tgtctc	...
	BoxII promoter motif	16789418	16789423	gggttaa	...
	BoxII promoter motif	16789439	16789444	gggttaa	...
	GATA promoter motif [LRE]	16788616	16788621	tgatag	...
	GATA promoter motif [LRE]	16789423	16789428	agataa	...
	GATA promoter motif [LRE]	16788924	16788929	agataa	...
	GATA promoter motif [LRE]	16788540	16788545	agataa	...
	GATA promoter motif [LRE]	16788466	16788471	agataa	...
	Ibox promoter motif	16788923	16788928	gataag	...
	Ibox promoter motif	16788465	16788470	gataag	...
	T-box promoter motif	16789529	16789534	actttg	...
	SORLIP2	16788552	16788556	gggcc	...
	SORLIP2	16788586	16788590	gggcc	...
	SORLIP2	16788531	16788535	gggcc	...

Table continued -

Table continued -

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT5G11180	AtMYB2 BS in RD22	3564282	3564288	ctaacca	MYB
	Bellringer/replumless/pennywise BS1 IN AG	3563802	3563809	aaattaaa	Homeobox
	Bellringer/replumless/pennywise BS2 IN AG	3563367	3563374	aaattagt	Homeobox
	Bellringer/replumless/pennywise BS3 IN AG	3563368	3563375	actaattt	Homeobox
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	3561637	3561642	actcat	bZIP
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	3564015	3564020	actcat	bZIP
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	3564197	3564202	actcat	bZIP
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	3562247	3562252	actcat	bZIP
	W-box promoter motif	3561818	3561823	ttgact	WRKY
	W-box promoter motif	3563000	3563005	ttgacc	WRKY
	W-box promoter motif	3564222	3564227	ttgact	WRKY
	W-box promoter motif	3561851	3561856	ttgact	WRKY
	ARF1 binding site motif	3561810	3561815	tgtctc	ARF
	ARF1 binding site motif	3561823	3561828	tgtctc	ARF
	ARF1 binding site motif	3562422	3562427	tgtctc	ARF
	ATHB6 binding site motif	3562750	3562758	caattatta	HB
	MYB3 binding site motif	3564279	3564286	taactaac	MYB
	MYB4 binding site motif	3562390	3562396	accaacc	MYB
	MYB4 binding site motif	3563538	3563544	accaacc	MYB
	MYB4 binding site motif	3563984	3563990	aacaaac	MYB
	MYB4 binding site motif	3564214	3564220	aacaaac	MYB
	MYB4 binding site motif	3564280	3564286	aactaac	MYB
	MYB4 binding site motif	3563246	3563252	aactacc	MYB
	RAV1-A binding site motif	3561704	3561708	caaca	ABI3VP1
	RAV1-A binding site motif	3562194	3562198	caaca	ABI3VP1
	RAV1-A binding site motif	3564213	3564217	caaca	ABI3VP1
	RAV1-A binding site motif	3564479	3564483	caaca	ABI3VP1
	LFY consensus binding site motif	3561815	3561820	ccattg	LFY
	LFY consensus binding site motif	3562395	3562400	ccaatg	LFY
	LFY consensus binding site motif	3562586	3562591	ccattg	LFY
	LFY consensus binding site motif	3563172	3563177	ccattg	LFY
	LFY consensus binding site motif	3562395	3562400	ccattg	LFY
	ARF binding site motif	3561810	3561815	tgtctc	...
	ARF binding site motif	3561823	3561828	tgtctc	...
	ARF binding site motif	3562422	3562427	tgtctc	...
	BoxII promoter motif	3563571	3563576	gggttaa	...
	BoxII promoter motif	3563005	3563010	gggttaa	...
	GATA promoter motif [LRE]	3564532	3564537	agataa	...
	GATA promoter motif [LRE]	3564520	3564525	agatag	...
	GATA promoter motif [LRE]	3564426	3564431	agataa	...
	GATA promoter motif [LRE]	3564187	3564192	tgataa	...
	GATA promoter motif [LRE]	3563213	3563218	agataa	...
	GATA promoter motif [LRE]	3562745	3562750	tgataa	...
	GATA promoter motif [LRE]	3562000	3562005	agataa	...
	GATA promoter motif [LRE]	3561746	3561751	tgataa	...
	Ibox promoter motif	3564186	3564191	gataag	...
	L1-box promoter motif	3562182	3562189	taaatgta	...
	LTRE promoter motif	3562083	3562089	accgaca	...
	T-box promoter motif	3561921	3561926	actttg	...
	T-box promoter motif	3562168	3562173	actttg	...
	T-box promoter motif	3563274	3563279	actttg	...
	T-box promoter motif	3562426	3562431	actttg	...
	SORLREP1	3562942	3562951	ttatactagt	...

Table continued -

Table continued -

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT5G11210	AtMYC2 BS in RD22	3574692	3574697	cacatg	BHLH
	Bellringer/replumless/pennywise BS1 IN AG	3576798	3576805	aaattaaa	Homeobox
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	3576950	3576955	actcat	bZIP
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	3574983	3574988	actcat	bZIP
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	3574938	3574943	actcat	bZIP
	W-box promoter motif	3574886	3574891	ttgact	WRKY
	W-box promoter motif	3576702	3576707	ttgacc	WRKY
	W-box promoter motif	3576437	3576442	ttgact	WRKY
	W-box promoter motif	3576235	3576240	ttgacc	WRKY
	W-box promoter motif	3576189	3576194	ttgact	WRKY
	W-box promoter motif	3575031	3575036	ttgacc	WRKY
	W-box promoter motif	3574949	3574954	ttgacc	WRKY
	ARF1 binding site motif	3576753	3576758	tgtctc	ARF
	ARF1 binding site motif	3576774	3576779	tgtctc	ARF
	ARF1 binding site motif	3576927	3576932	tgtctc	ARF
	ATHB2 binding site motif	3575201	3575209	taatcatta	HB
	DPBF1&2 binding site motif	3575898	3575904	acacgtg	bZIP
	DPBF1&2 binding site motif	3576769	3576775	acacgag	bZIP
	DPBF1&2 binding site motif	3575898	3575904	acacgtg	bZIP
	MYB1 binding site motif	3574745	3574752	atccaacc	MYB
	MYB4 binding site motif	3575765	3575771	aacaaac	MYB
	MYB4 binding site motif	3574828	3574834	aacaaac	MYB
	MYB4 binding site motif	3574665	3574671	aacaaac	MYB
	RAV1-A binding site motif	3575254	3575258	caaca	ABI3VP1
	RAV1-A binding site motif	3575829	3575833	caaca	ABI3VP1
	RAV1-A binding site motif	3575896	3575900	caaca	ABI3VP1
	RAV1-A binding site motif	3576044	3576048	caaca	ABI3VP1
	RAV1-A binding site motif	3576450	3576454	caaca	ABI3VP1
	RAV1-A binding site motif	3575756	3575760	caaca	ABI3VP1
	RAV1-A binding site motif	3575737	3575741	caaca	ABI3VP1
	RAV1-A binding site motif	3575494	3575498	caaca	ABI3VP1
	ABRE-like binding site motif	3575905	3575912	gacgtgta	...
	ARF binding site motif	3576753	3576758	tgtctc	...
	ARF binding site motif	3576774	3576779	tgtctc	...
	ARF binding site motif	3576927	3576932	tgtctc	...
	DRE-like promoter motif	3576282	3576290	ggccgactt	...
	GATA promoter motif [LRE]	3574715	3574720	agataa	...
	GATA promoter motif [LRE]	3574769	3574774	tgatag	...
	GATA promoter motif [LRE]	3576917	3576922	tgatag	...
	GATA promoter motif [LRE]	3576823	3576828	tgataa	...
	GATA promoter motif [LRE]	3576689	3576694	tgatag	...
	GATA promoter motif [LRE]	3576674	3576679	agataa	...
	GATA promoter motif [LRE]	3574654	3574659	tgatag	...
	G-box promoter motif [LRE]	3575899	3575904	cacgtg	...
	G-box promoter motif [LRE]	3575898	3575903	cacgtg	...
	L1-box promoter motif	3574721	3574728	taaatgta	...
	T-box promoter motif	3574777	3574782	actttg	...
	SORLIP2	3575742	3575746	ggggcc	...

Table continued -

Table continued -

Gene name	Binding Site Name	Binding Site Genome Start	Binding Site Genome End	Binding Site Sequence (5'-3')	Binding Site Family/TF
AT5G27100	W-box promoter motif	9538388	9538393	ttgact	WRKY
	RAV1-A binding site motif	9538470	9538474	caaca	ABI3VP1
	RAV1-A binding site motif	9538398	9538402	caaca	ABI3VP1
	BoxII promoter motif	9538403	9538408	gggttaa	...
	T-box promoter motif	9538455	9538460	actttg	...
AT5G48400	AtMYC2 BS in RD22	19633371	19633376	cacatg	BHLH
	W-box promoter motif	19632816	19632821	ttgacc	WRKY
	W-box promoter motif	19633118	19633123	ttgacc	WRKY
	W-box promoter motif	19633264	19633269	ttgact	WRKY
	ARF1 binding site motif	19632606	19632611	tgtctc	ARF
	CCA1 binding site motif	19632823	19632830	aaaaatct	MYB-related
	DPBF1&2 binding site motif	19633371	19633377	acacatg	bZIP
	MYB4 binding site motif	19633424	19633430	aacaaac	MYB
	MYB4 binding site motif	19633095	19633101	aacaaac	MYB
	RAV1-A binding site motif	19633423	19633427	caaca	ABI3VP1
	RAV1-A binding site motif	19633035	19633039	caaca	ABI3VP1
	RAV1-A binding site motif	19632652	19632656	caaca	ABI3VP1
	ARF binding site motif	19632606	19632611	tgtctc	...
	DRE-like promoter motif	19633048	19633056	tgccgacaa	...
	GATA promoter motif [LRE]	19633207	19633212	agataa	...
	Ibox promoter motif	19633208	19633213	gataag	...
	T-box promoter motif	19633195	19633200	actttg	...
	T-box promoter motif	19633297	19633302	actttg	...
	AtMYB2 BS in RD22	19637151	19637157	ctaacca	MYB
AT5G48410	Bellringer/replumless/pennywise BS1 IN AG	19636853	19636860	aaattaaa	Homeobox
	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	19636475	19636480	actcat	bZIP
	W-box promoter motif	19636927	19636932	ttgact	WRKY
	W-box promoter motif	19637327	19637332	ttgact	WRKY
	ARF1 binding site motif	19637280	19637285	tgtctc	ARF
	DPBF1&2 binding site motif	19636987	19636993	acacaag	bZIP
	MYB binding site promoter	19637226	19637233	caccaaac	MYB
	MYB4 binding site motif	19636798	19636804	accaaac	MYB
	MYB4 binding site motif	19637227	19637233	accaaac	MYB
	MYB4 binding site motif	19637153	19637159	aactaac	MYB
	MYB4 binding site motif	19636836	19636842	aacaaac	MYB
	RAV1-A binding site motif	19636985	19636989	caaca	ABI3VP1
	RAV1-A binding site motif	19637006	19637010	caaca	ABI3VP1
	RAV1-A binding site motif	19637468	19637472	caaca	ABI3VP1
	RAV1-A binding site motif	19636967	19636971	caaca	ABI3VP1
	RAV1-A binding site motif	19636839	19636843	caaca	ABI3VP1
	RAV1-A binding site motif	19636500	19636504	caaca	ABI3VP1
	LFY consensus binding site motif	19636555	19636560	ccagtg	LFY
	ABRE-like binding site motif	19637053	19637060	tacgtgta	...
	ARF binding site motif	19637280	19637285	tgtctc	...
	BoxII promoter motif	19637022	19637027	gggttaa	...
	GATA promoter motif [LRE]	19637238	19637243	tgatag	...
	T-box promoter motif	19637372	19637377	actttg	...

2.3.6 Three-dimensional structure of AtGLRs

For a preliminary view of structural features of AtGLR proteins, we have predicted the secondary structural features of AtGLR proteins using SOPMA. Almost equal number of amino acids formed alpha helices (36.06-42.2%) and random coils (33.42-41.06%), while percentage of amino acids forming extended strands were lesser (16.37-21.6%). For getting some preliminary idea about the structural features, the three dimensional protein models were constructed by homology modeling. Search for templates by HHpred resulted the same template *i.e.* chain A of AMPA subtype ionotropic glutamate receptor from *Rattus norvegicus* (pdb: 3KG2) (Sobolevsky et al. 2009) for all the proteins. Results for 25 AtGLR proteins regarding this template as obtained from HHpred are given in Table 2.7.

Table 2.7 : Results for 25 AtGLR proteins regarding the template 3KG2 (chain A) as obtained from HHpred.

Name of the protein	Prob ^a	E-value ^b	P-value ^c	Score ^d	SS ^e	Cols ^f	Query HMM ^g	Template HMM ^h
AT1G05200.1	100	2.4E-95	7E-100	917	65.3	761	59-891	1-818(824)
AT1G05200.2	100	8.1E-79	2.2E-83	741.7	68.5	759	59-891	1-818(824)
AT1G42540.1	100	3.1E-81	8.3E-86	772.9	64.7	772	27-864	1-820(824)
AT2G17260.1	100	5.6E-78	1.5E-82	735.1	71	769	58-893	1-820(824)
AT2G24710.1	100	5.3E-79	1.4E-83	745.9	62.9	769	29-864	1-820(824)
AT2G24720.1	100	3.9E-74	1E-78	691.2	61.9	771	30-870	1-824(824)
AT2G29100.1	100	3.8E-76	1E-80	720.9	68.2	762	28-844	1-820(824)
AT2G29110.1	100	1.1E-73	2.9E-78	701.8	63.6	769	31-858	1-824(824)
AT2G29120.1	100	7.2E-74	1.9E-78	699.1	67.7	755	37-857	1-820(824)
AT2G32390.1	100	2.4E-74	6.5E-79	695.1	64	742	15-827	18-819(824)
AT2G32390.2	100	2.1E-80	5.7E-85	752.8	56.6	727	2-783	52-819(824)
AT2G32390.3	100	1.7E-65	4.5E-70	618.5	55.8	713	41-830	1-819(824)
AT2G32400.1	100	4.9E-83	1.3E-87	789.8	61.1	763	29-856	1-818(824)
AT3G04110.1	100	1.8E-77	4.8E-82	717.2	54.7	739	28-806	1-821(824)
AT3G07520.1	100	2.6E-72	7.1E-77	664.5	64.2	741	44-838	1-824(824)
AT3G51480.1	100	2.2E-84	5.9E-89	796.3	59.4	761	28-853	1-819(824)
AT4G31710.1	100	6.9E-71	1.9E-75	668.8	56.6	737	53-853	1-820(824)
AT4G35290.1	100	7.4E-84	2E-88	794.5	60.4	773	28-864	1-820(824)
AT4G35290.2	100	7.4E-84	2E-88	794.5	60.4	773	28-864	1-820(824)
AT5G11180.1	100	2E-65	5.3E-70	623.5	62.9	693	34-807	1-820(824)
AT5G11210.1	100	5.3E-74	1.4E-78	694.8	59.1	707	14-789	47-819(824)
AT5G27100.1	100	1.6E-75	4.4E-80	702.3	69.9	767	30-863	1-824(824)
AT5G48400.1	100	1.7E-67	4.7E-72	615.5	54.9	657	38-750	1-740(824)
AT5G48400.2	100	1E-74	2.8E-79	694	60	732	38-829	1-824(824)
AT5G48410.1	100	2E-71	5.4E-76	664.6	56	724	40-826	1-824(824)

^aThe 'Prob' column lists the probability in percent that the database match is a true positive, i.e. that it is homologous to the query sequence at least in some core part.

^bThe E-value for a sequence match is the expected number of false positives per database search with a score at least as good as the score of this sequence match.

^cThe P-value is equal to the E-value divided by the number of HMMs in the searched database.

^dThe 'Score' column gives the total score that includes the score from the secondary structure comparison which is listed in the next column ('SS').

^eThe 'SS' column gives the secondary structure score. This score tells how well the PSIPRED-predicted (3-state) or actual DSSP-determined (8-state) secondary structure sequence for query and template agree with each other.

^fThe 'Cols' column contains the total number of matched columns in the query–template alignment.

^gThe 'Query HMM' is the range of query match states aligned.

^hThe 'Template HMM' is the range of template match states aligned and, in parenthesis, total number of match states in template HMM.

(All these explanations have been taken from https://toolkit.tuebingen.mpg.de/hhpred/help_results.)

The C α -atoms Root Mean Square Deviation (RMSD) between the template and the corresponding models ranged from 0.44Å and 1.348 Å (Table 2.8) indicating good quality models. The structures were validated with PROCHECK. All the models showed more than 84% of amino acids in most favoured region and less than 2% in disallowed region (Table 2.8).

Table 2.8 : Quality of the predicted three-dimensional structures of AtGLR proteins by RMSD and PROCHECK

Protein name	RMSD	Results from PROCHECK				
		Residues in most favoured region (%)	Residues in allowed region (%)	Residues in generally allowed region (%)	Residues in disallowed region (%)	Residues in most favoured+allowed region (%)
AT1G05200.1	1.218	86.9	10.8	1.8	0.6	97.7
AT1G05200.2	1.218	86.9	10.8	1.8	0.6	97.7
AT1G42540.1	0.959	85.2	12.5	1.5	0.8	97.7
AT2G17260.1	1.348	86.2	11.6	1.8	0.4	97.8
AT2G24710.1	1.07	86.9	10.6	1.8	0.7	97.5
AT2G24720.1	0.703	86.8	11	1.1	1.1	97.8
AT2G29100.1	0.852	86	10.2	1.9	1.8	96.2
AT2G29110.1	0.836	85.1	12.1	2.1	0.7	97.2
AT2G29120.1	0.879	87.2	10	1.7	1.1	97.2
AT2G32390.1	0.558	84.7	11.8	1.7	1.8	96.5
AT2G32390.2	0.889	86.3	11.5	1.5	0.7	97.8
AT2G32390.3	0.681	84.3	12.2	1.6	1.9	96.5
AT2G32400.1	1.023	87.6	10	1.4	1.1	97.6
AT3G04110.1	0.628	86.2	11	1.3	1.5	97.2
AT3G07520.1	0.967	84.4	11.5	2.4	1.7	95.9
AT3G51480.1	1.11	85.9	12.1	1.3	0.7	98
AT4G31710.1	0.604	84.2	11.9	2.5	1.5	96.1
AT4G35290.1	0.854	85.1	12.2	1.8	1	97.3
AT4G35290.2	0.854	85.1	12.2	1.8	1	97.3
AT5G11180.1	0.44	86.8	11.1	1.1	1	97.9
AT5G11210.1	0.936	84.2	12.8	2.2	0.9	97
AT5G27100.1	0.625	86	11.1	1.5	1.4	97.1
AT5G48400.1	0.684	86.2	11.7	1.2	0.8	97.9
AT5G48400.2	0.549	87.5	10.2	1.4	0.9	97.7
AT5G48410.1	1.29	84.5	12.9	1.7	0.9	97.4

The total percentage of amino acids in most favoured plus allowed region ranged from 95.9 to 98%. All these showed that the predicted models were of high quality. The three-dimensional structures clearly showed the transmembrane region and the ligand-binding region (Figure 2.8).

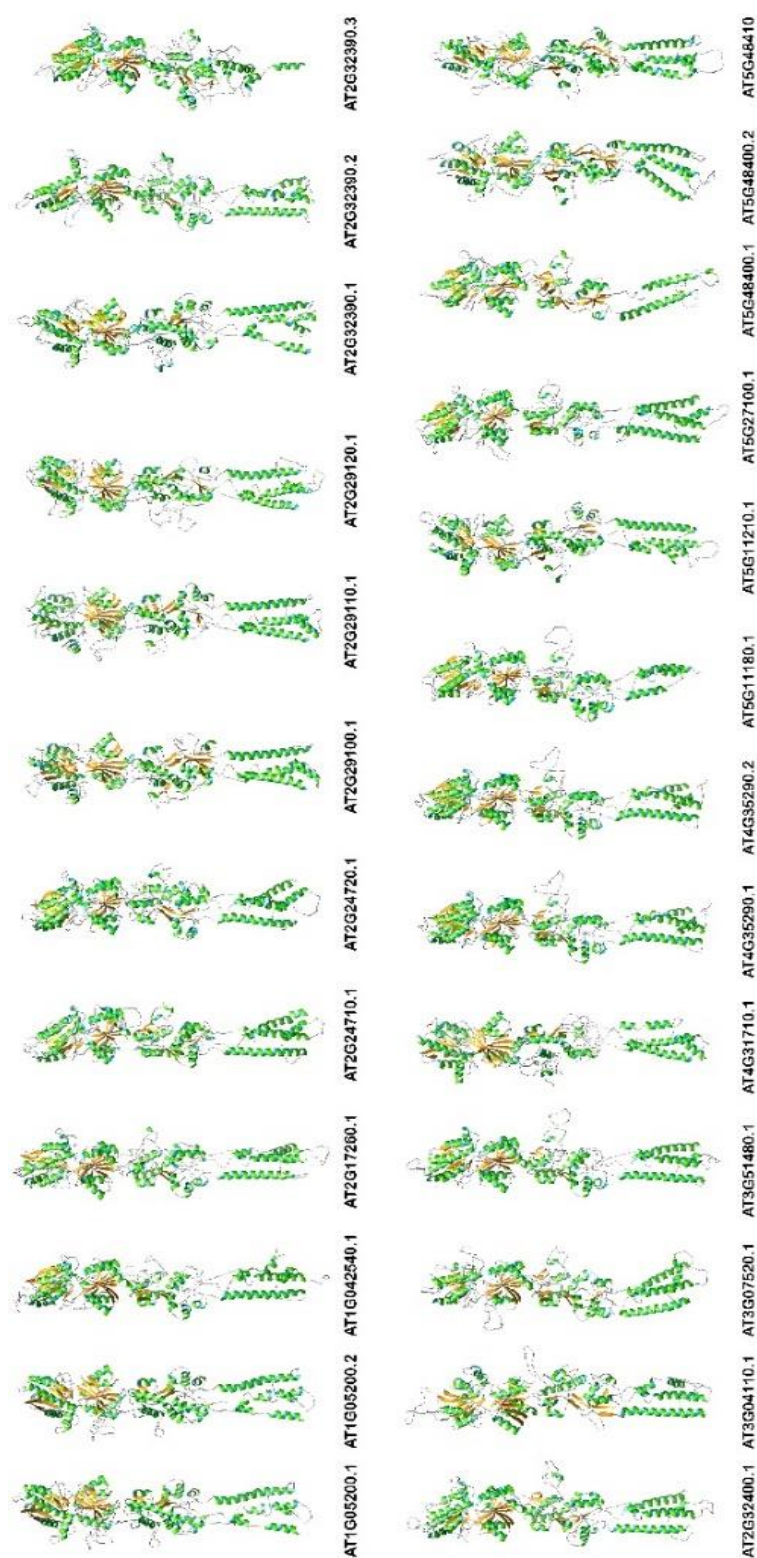


Figure 2.8 : Predicted three-dimensional structures of the AtGLR proteins. Alpha-helices and beta sheets are represented in green and orange, respectively

In all cases, as expected, the transmembrane domains were consisted of only alpha helices, but with varying number (Figure 2.8). The ligand binding domains were consisted of both alpha helices and beta sheets and the arrangement of beta sheets are similar in many of the AtGLRs. One of the splice variants (At2G32390.3) of AT2G32390 showed only one alpha helix in the transmembrane region unlike the other two splice variants. This splice variant should be studied thoroughly to know its functional aspects in *Arabidopsis*.

Evolutionary conservation of amino acid positions of the AtGLR proteins have been analysed with ConSurf. ConSurf is advantageous over other methods as it accurately compute the evolutionary rate using either an empirical Bayesian method or a maximum likelihood (ML) method (de Carvalho and De Mesquita 2013; Pupko et al. 2002). ConSurf, thus, can properly distinguish between the conservation caused by a short evolutionary time and actual sequence conservation (de Carvalho and De Mesquita 2013). The Multiple protein sequence alignment of ANF_receptor (extracellular ligand binding domain; PF01094) region (Figure 2.9a) and Lig_chan (transmembrane region of ligand-gated ion channel; PF00060) region (Figure 2.9b) of AtGLR proteins showed several conserved amino acid positions.

The start and end positions of these two domains are shown based on AT1G05200.1. Our findings revealed that transmembrane region of ligand-gated ion channel (7.86% of positions are most conserved) possess slightly more conserved positions than the extracellular ligand binding domain (7.67% of positions are most conserved). Commonly, in a protein, those residues which are positioned in active sites, involved in protein-ligand or protein-protein interactions or involved in protein structure and folding stability, faces more selective pressure and are usually more conserved than other residues (de Carvalho and De Mesquita 2013). Thus , further studies about residues in conserved positions in our analysis can throw some light on the functional role of these residues.

Computational analysis of the GLR Gene family of *Arabidopsis thaliana*

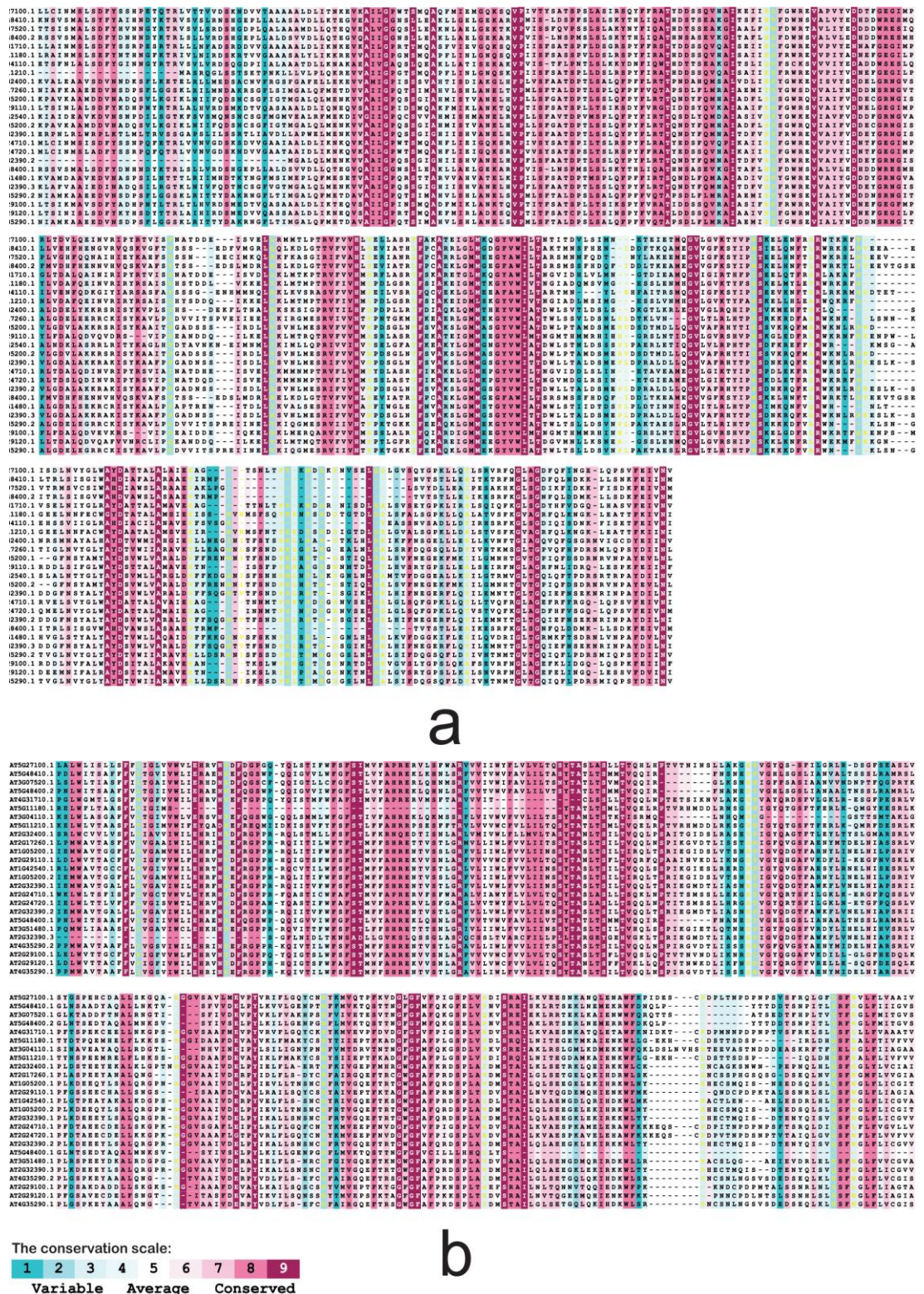


Figure 2.9 : Multiple protein sequence alignment of AtGLR proteins using ConSurf showing evolutionary conservation of amino acid residues of the (a) ANF_receptor (extracellular ligand binding domain; PF01094) and (b) Lig_chan (transmembrane region of ligand-gated ion channel; PF00060).

Note: Amino acids were
bottom left of the figure).

2.3.7 Surface properties

Surface electrostatics plays a crucial role in protein–protein interactions and protein–ligand binding (Soares et al. 2005). Clustering based on electrostatic surface has been successfully applied previously (Legrand et al. 2011; Soares et al. 2005). Based on the electrostatic surfaces, all the AtGLR proteins have been grouped into two main clusters (Figure 2.10a). The protein products of two genes in tandem arrangement (AT2G24710 and AT2G24720) on chromosome 2 grouped together based on surface electrostatics. It has been shown that proteins with similar electrostatic properties indicate similar behavior and function (Blomberg et al. 1999; Botti et al. 1998; Soares et al. 2005; Ullmann et al. 1997; Wade et al. 1998).

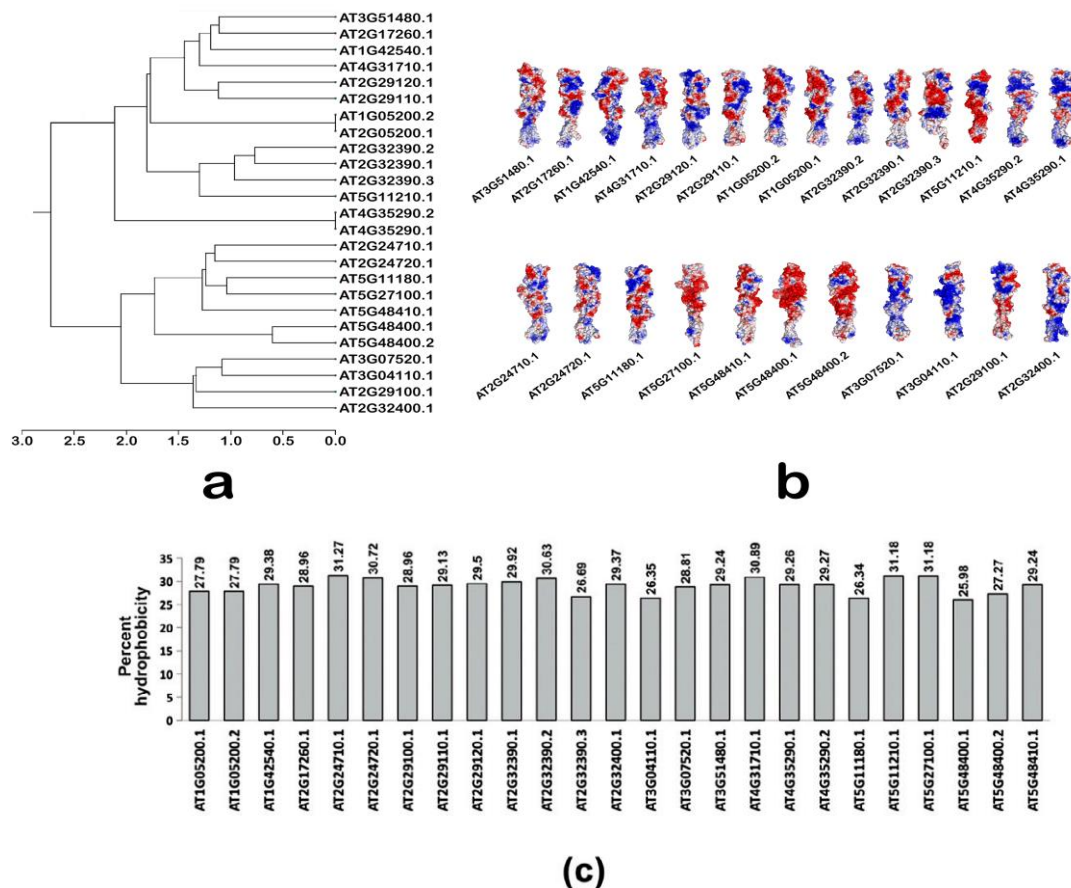


Figure 2.10 : Molecular electrostatic potential surfaces and hydrophobicity of the AtGLR models. (a) Clustering of the 25 AtGLR proteins by PIPSA. (b) Representation of the electrostatic potential projected on the solvent accessible surface of the AtGLR proteins. Each model was colored using the APBS electrostatic potential and the surface colours were clamped at -3 (red) and +3 (blue) kTe^{-1} . (c) Plot of the calculated percentage of hydrophobic surface of each AtGLR protein by PLATINUM.

Indeed, AT2G24710 and AT2G24720 showed similar expression patterns (Figure 2.4). However, among another three tandemly arranged genes (AT2G29100, AT2G29110 and AT2G29120), two (AT2G29110, AT2G29120) grouped together in the first cluster, while AT2G29100 was placed in the second cluster. Electrostatic surfaces of AtGLR proteins were clearly dissimilar, and they showed large positive and negative potential patches (Figure 2.10b). Majority of the proteins in the first cluster showed more electropositivity in the transmembrane region while in the second cluster, transmembrane regions showed comparatively more neutral regions in majority of the proteins. The splice variants of AT1G05200 as well as AT4G35290 showed no dissimilarity among them at all (Figure 2.10b) which was reflected in their grouping (Figure 2.10a). However, three splice variants of AT2G32390, although grouped together, showed some dissimilarity in their surface electrostatics.

The overall surface hydrophobicity of AtGLR proteins ranged from 25.98 - 31.27 % (Figure 2.10c) with an average ($\pm$ SE) of 29.0044 ($\pm$ 0.316) %. This shows that the surfaces of AtGLRs are mostly hydrophilic. The hydrophobic surface of the protein products of two tandemly arranged genes (AT2G24710 and AT2G24720) was very similar. However, the GRAVY score of AT2G24710.1 (0.077) was much higher than AT2G24720.1 (0.033). Three splice variants of AT2G32390 also showed difference in surface hydrophobicity. However, their GRAVY score (Table 2.2) did not show same trend as that of the surface hydrophobicity. Similar observations were made in another three tandemly arranged genes (AT2G29100, AT2G29110 and AT2G29120). Indeed, GRAVY was not found to be significantly correlated with surface hydrophobicity (Spearman's $\rho = 0.315$, $P = 0.125$). These results show that GRAVY score alone cannot be determinative of surface hydrophobicity.

2.4 Conclusion:

In the present study, we have studied twenty glutamate receptor genes from the model plant *Arabidopsis thaliana* (AtGLRs). Several phosphorylation, myristoylation and Asn-glycoylation sites have been identified which indicates their role in different functions like signal transduction, membrane targeting and protein stability. Domain organization of these proteins identified one conserved domain namely ANF_receptor (extracellular ligand binding domains) at the N-terminal domain and the C-terminal region consists of two overlapping domains namely SBP_bac_3

(bacterial extracellular solute-binding proteins and Lig_chan (transmembrane region of ligand-gated ion channel). The comparative analysis of AtGLR genes in *Arabidopsis thaliana* elucidated chromosomal location, gene structure, phylogenetic relationships and three dimensional structures. The expression analysis led to the identification of biotic, abiotic stress responsive and tissue-specific expression pattern of several of the AtGLR genes, suggesting functional diversification. Functional diversification of tandemly arranged duplicated genes has also been studied. Several *cis*-acting elements in the promoter regions of the AtGLRs indicate their role in various biological processes. Our observations revealed the biological significance of AtGLR genes which may further be validated using laboratory based functional genomics approach. Our comprehensive *in silico* analyses will also help design experiments for functional validation of their role in plant development and stress responses.

Chapter **III.**

Genome-wide analysis of glutamate
receptor gene family in allopolyploid
Brassica napus and its diploid
progenitors

Chapter III.

Genome-wide analysis of glutamate receptor gene family in allopolyploid *Brassica napus* and its diploid progenitors

3.1 Introduction

3.2 Materials and methods

3.3 Results

3.4 Discussion

3.5 Conclusion

3.1 Introduction

In mammalian central nervous systems, ionotropic glutamate receptors mediate rapid excitatory neuronal transmission (Naz et al. 2022). The plant glutamate receptor-like homologs (GLRs) are homologs of mammalian ionotropic glutamate receptors (iGluRs) and are hypothesized to be potential amino acid sensors in plants (Price et al. 2012). *Arabidopsis thaliana* possesses 20 glutamate receptor-like genes (*AtGLRs*) in its genome (Lam et al. 1998). Of these, *AtGLR1.2* and *AtGLR1.3* play vital roles in the plant response to cold stress (Zheng et al. 2018). Recent studies reporting the localization of GLRs in chloroplasts indicate an additional role of GLRs in this organ (Price et al. 2012). Indeed, Plants carrying T-DNA insertions in *AtGLR3.4* showed weak photosynthetic phenotypes (Teardo et al. 2010, 2011). Similarly, it was shown that *AtGLR3.1* was expressed preferentially in guard cells and over-expression of this gene leads to the impairment of stomatal closure that is induced by external Ca^{2+} (Cho et al. 2009). Many studies showed that plant GLRs play many roles including plant reproduction, seed germination, root growth, regulation of stomatal aperture, plant defense and others (Grenzi et al. 2022). The moss *Physcomitrium patens* contains one GLR (*PpGLR1*) which plays pivotal role in ABA-mediated growth regulation (Wang et al. 2022). In rice, *OsGLR3.4* plays vital role in plant growth as well as mechanical wound response (Yu et al. 2022). In another study, it has been shown that *OsGLR3.4* controls rice root growth and tropism towards nutrient source (Yu et al. 2023). Thus, *GLR* genes should be studied in detail in other crop species for better understanding of their roles in various physiological processes.

Several crop species of the genus *Brassica* such as *B. rapa* L. (Chinese cabbage, turnip and pak choi), *B. oleracea* L. (broccoli, cabbage, cauliflower, Brussels sprouts etc.) as well as the oilseed crop such as *B. napus* L. (rapeseed) are widely used worldwide (Wang et al. 2011; Sun et al. 2022). *Brassica oleracea* not only comprises many important vegetable crops including cauliflower, broccoli, cabbages, Brussels sprouts, Kohlrabi and kales, the species demonstrates extreme morphological diversity and crop forms, with various members grown for their leaves, flowers and stems (Mabry et al. 2021). Most *B. oleracea* crops are high in protein (Liu et al. 2014), carotenoids (Kopsell and Kopsell

2006) and contain diverse glucosinolates (GSLs) that function as unique phytochemicals for plant defence against fungal and bacterial pathogens (Halkier and Gershenzon 2006). *B. rapa* also shows morphological diversity among the cultivated types and includes both oil and vegetable crops (Sun 2015). *B. napus* is the second most important oilseed crop species in the world and there is a growing interest in rapeseed protein as human nutrition (Friedt et al. 2018).

Moreover, these three species along with *B. juncea*, *B. carinata* and *B. nigra* provide opportunities for the study of *Brassica* genome evolution as these six widely cultivated *Brassica* species are a classical example of the importance of polyploidy in botanical evolution, described by ‘U’s triangle’ (Nagaharu and Nagaharu 1935; Wang et al. 2011). Of these, *B. napus* (A and C subgenomes, $n = 19$) was formed by recent allopolyploidy between ancestors of *B. rapa* (A genome, $n = 10$) and *B. oleracea* (C genome, $n = 9$) and is probably polyphyletic (Nagaharu and Nagaharu 1935; Allender and King 2010). Comparative physical mapping studies have confirmed genome triplication in a common ancestor of *B. oleracea* (O’neill and Bancroft 2000) and *B. rapa* (Park et al. 2005) since its divergence from the *A. thaliana* lineage at least 13–17 MYA (Yang et al. 1999; Town et al. 2006; Beilstein et al. 2010). The *Brassica* genomes retains evidence of recursive whole-genome duplication (WGD) events (Bowers et al. 2003; Jiao et al. 2011) and have undergone a Brassiceae-lineage-specific whole-genome triplication (WGT) (Lysak et al. 2005; Wang et al. 2011) since their divergence from the *Arabidopsis* lineage (Liu et al. 2014).

In this study, considering the importance of plant *GLR* genes as well as the economic importance of *Brassica*, we made use of the publicly available sequenced genomes of the allotetraploid *B. napus* and its diploid progenitors *B. oleracea* and *B. rapa* to trace the evolution of *GLR* genes in these species. We have studied the phylogeny, chromosomal distribution, protein motif analysis, gene structure, orthology analysis, expression patterns and promoter analysis, subcellular localization and miRNA targeting of these *GLR* genes. This study enhances our understanding of the effect of polyploidisation events on the evolution of the *GLR* gene family in *Brassica* as well as their structural and functional aspects.

3.2 Materials and Methods

3.2.1 Identification of *GLR* genes

The genome data (gene, cds and protein data) of allopolyploid *B. napus* and its two diploid progenitors, *B. rapa* and *B. oleracea*, were obtained from the BRAD database (<http://brassicadb.org/brad/>) (Wang et al. 2015). As discussed in Chapter II, the previously analysed *Arabidopsis thaliana* GLR (AtGLR) proteins have been retrieved using the BioMart tool (Kinsella et al. 2011) of Ensembl Plants database (Kersey et al. 2014). These proteins were used as queries to perform BLASTp (Altschul et al. 1990) searches (E-value <1e-5) with all proteins from the three *Brassica* species. To identify the *GLR* genes in three *Brassica* genomes accurately, all putative protein sequences obtained from the BLASTp search, were confirmed by searching for the conserved domains using CD-search in the NCBI Conserved Domain Database (CDD; <https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) (Marchler-Bauer et al. 2017) and Pfam (<http://pfam.sanger.ac.uk/>) (El-Gebali et al. 2019) database. The conserved domain analysis helped to make the shortlisting of the best candidate to be the GLR members in the *Brassica* genomes. We have retained those genes which showed the presence of PBP1_GABAb_receptor (accession: cd06366) and GluR_Plant (cd13686) by NCBI-CD search and also showed the pfam domains ANF_receptor (extracellular ligand-binding domain; PF01094), SBP_bac_3 (bacterial extracellular solute-binding proteins, family 3; PF00497) and Lig_chan (ligand-gated ion channel; PF00060). These pfam domains were considered because these are present in *A. thaliana* GLRs (Chapter II). Moreover, the shortlisted genes have been searched in Ensembl Plants (for *B. rapa* and *B. napus*) and Phytozome 12 (Goodstein et al. 2012) (for *B. oleracea*) for confirmation of their identity as glutamate receptors. The subcellular location of *GLR* genes was predicted using the CELLO (Yu et al. 2006) web tool (<http://cello.life.nctu.edu.tw/>).

3.2.2 Phylogenetic relationship analysis

The full-length GLR protein sequences of the three *Brassica* species were aligned with MUSCLE (Edgar 2004) and subsequently, a phylogenetic tree was constructed by MEGAX (Kumar et al. 2018) based on Maximum Likelihood (ML)

method with 1000 bootstrap replicates. The tree was visualized with FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

3.2.3 Chromosomal location and gene structure analysis

The chromosomal location information of *GLR* genes in the three *Brassica* species were collected from the BRAD database and their physical positions were drafted to the corresponding chromosomes by TBTools (Chen et al. 2020). Some genes showed scaffold information only and locations of some of them were confirmed with NCBI Genome database (<https://www.ncbi.nlm.nih.gov/genome/>). Tandem duplication events were defined as adjacent homologous *GLR* genes on the same chromosome with no more than one intervening gene following Li et al. (2020). The exon/intron structures of *GLR* genes were analyzed using Gene Structure Display Server (GSDS) 2.0 (<http://gsds.gao-lab.org/>) (Hu et al. 2015) by aligning the cDNAs with their corresponding gene sequences.

3.2.4 Conserved motif analysis

Conserved motifs in *GLR* proteins were investigated by online MEME server (<https://meme-suite.org/meme/tools/meme>) (Bailey et al. 2009), with keeping the maximum motif number as 10 and the other parameters at their default values.

3.2.5 Gene expression studies using RNA-seq data

For gene expression profiling of the *GLR* genes in the three species of *Brassica*, publicly available RNA-seq data were obtained from NCBI GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) and analyzed. The following bio project accession numbers were analyzed: PRJNA530549 (He et al. 2019, 2020) for *B. napus*, PRJNA183713 (Liu et al. 2014; Yu et al. 2014) for *B. oleracea* and PRJNA185152 (Tong et al. 2013; Liu et al. 2014; Wu et al. 2022) for *B. rapa*. Gene expression patterns of each tissue were analyzed and log2 transformed FPKM values (fragments per kilobase of exon model per million mapped reads) were visualized as heat maps using TBtools (Chen et al. 2020).

3.2.6 Promoter analysis

The promoter sequences, which were the 1000 bp upstream of the transcription start site (TSS) of the *GLR* genes, were acquired from the BRAD database, and the *cis*-elements in the promoters were analyzed using the Plant Cis-Acting Regulatory Element (PlantCARE) server (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) (Lescot et al. 2002).

3.2.7 Sequence-based identification of orthologous *GLR* genes

The orthologous genes in *B. oleracea*, *B. rapa* and *B. napus* were identified using OrthoVenn2 (Xu et al. 2019) (<http://www.bioinfogenome.net/OrthoVenn>). The protein sequences of the identified *GLR* genes from all the three species of *Brassica* were used. The chromosomal position of the identified orthologous gene pairs were visualized using the TBTools program (Chen et al. 2020).for the analysis.

3.2.8 Prediction of miRNAs targeting *GLR* genes

All of the gene sequences of *GLR* family genes were submitted as candidate genes for predicting potential miRNAs by searching against the available reference of miRNA sequences of the respective species using psRNATarget (Dai et al. 2011, 2018; Dai and Zhao 2011) server with default parameters. Cytoscape software (Shannon et al. 2003) was used to visualize the interactions between the predicted miRNAs and the corresponding target *GLR* genes

3.3 Results

3.3.1 Identification of *GLR* genes in *B. napus*, *B. oleracea* and *B. rapa* and phylogenetic analysis

We have identified 11, 27 and 65 genes as *GLRs* in *B. oleracea*, *B. rapa* and *B. napus*, respectively and named them species-wise according to their distribution on chromosomes (*BoGLRs* for *B. oleracea*, *BrGLRs* for *B.rapa* and *BnGLRs* for *B.napus*) (Table 3.1). We analyzed the molecular characteristics of these *GLR* genes in *B. oleracea*, *B. rapa* and *B. napus*. In *B. napus*, the gene length of *GLR* genes ranged from 1206 bp (*BnaAnng35320D*) to 8949 bp (*BnaA04g14620D*), the peptide length ranged from 318 amino acids (*BnaC04g40280D*) to 1426 amino acids (*BnaA10g26130D*) and the GC

content of gene ranged from 30.83% (*BnaA04g14620D*) to 47.24% (*BnaA01g02120D*). In *B. oleracea*, the Gene length of *GLR* genes ranged from 2931 bp (*Bol033269*) to 7882 bp (*Bol013454*), the peptide length ranged from 739 amino acids (*Bol033267*) to 1787 amino acids (*Bol013454*) and the GC content of gene ranged from 36.02% (*Bol033267*) to 46.77% (*Bol040960*). In *B. rapa*, the gene length of *GLR* genes ranged from 2220 bp (*Bra040758*) to 9574 bp (*Bra020703*), the peptide length ranged from 397 amino acids (*Bra040759*) to 1858 amino acids (*Bra022887*) and the GC content of gene ranged from 29.88% (*Bra032063*) to 48.16% (*Bra015398*) (Table 3.1).

Table 3.1: Basic features of the identified *GLR* genes in the three species of *Brassica*. Gene IDs were obtained from BRAD database. Gene names are given in this study.

Species	Gene name	Gene ID (BRAD)	Chromosome	Start position	Stop position	Gene length (bp)	% GC content of gene	Peptide length (amino acids)	Subcellular localization
<i>B. napus</i>	<i>Bnglr1</i>	<i>Bnaa01g02120d</i>	Chra01	1046483	1048368	1886	47.24	583	Plasma membrane
	<i>Bnglr2</i>	<i>Bnaa01g33670d</i>	Chra01	22780883	22784038	3156	37.14	803	Plasma membrane
	<i>Bnglr3</i>	<i>Bnaa02g00890d</i>	Chra02	341214	345120	3907	37.11	860	Plasma membrane
	<i>Bnglr4</i>	<i>Bnaa02g30120d</i>	Chra02	21859931	21866215	6285	36.13	858	Plasma membrane
	<i>Bnglr5</i>	<i>Bnaa02g30130d</i>	Chra02	21866289	21869207	2919	40.73	730	Plasma membrane
	<i>Bnglr6</i>	<i>Bnaa02g30140d</i>	Chra02	21871483	21877918	6436	34.31	629	Plasma membrane
	<i>Bnglr7</i>	<i>Bnaa03g14950d</i>	Chra03	6882877	6886775	3899	43.19	889	Plasma membrane
	<i>Bnglr8</i>	<i>Bnaa03g14960d</i>	Chra03	6886850	6891050	4201	41.49	927	Plasma membrane
	<i>Bnglr9</i>	<i>Bnaa03g51670d</i>	Chra03	26855015	26858922	3908	42.53	960	Plasma membrane
	<i>Bnglr10</i>	<i>Bnaa04g14620d</i>	Chra04	12308301	12317249	8949	30.83	861	Plasma membrane
	<i>Bnglr11</i>	<i>Bnaa04g16910d</i>	Chra04	13790780	13795361	4582	36.58	924	Plasma membrane
	<i>Bnglr12</i>	<i>Bnaa04g16920d</i>	Chra04	13798384	13802286	3903	38.56	869	Plasma membrane
	<i>Bnglr13</i>	<i>Bnaa04g16940d</i>	Chra04	13811078	13815449	4372	35.66	640	Plasma membrane
	<i>Bnglr14</i>	<i>Bnaa04g28170d</i>	Chra04_rando m	691945	696023	4079	39.27	890	Plasma membrane
	<i>Bnglr15</i>	<i>Bnaa05g10830d</i>	Chra05	5938097	5941430	3334	43.79	740	Plasma membrane
	<i>Bnglr16</i>	<i>Bnaa05g10840d</i>	Chra05	5942119	5946545	4427	41.83	923	Plasma membrane
	<i>Bnglr17</i>	<i>Bnaa05g13030d</i>	Chra05	7682497	7687032	4536	37.26	968	Plasma membrane
	<i>Bnglr18</i>	<i>Bnaa06g30230d</i>	Chra06	20469671	20473140	3470	40.72	875	Plasma membrane
	<i>Bnglr19</i>	<i>Bnaa07g02520d</i>	Chra07	2124665	2128666	4002	42.15	913	Plasma membrane
	<i>Bnglr20</i>	<i>Bnaa08g04580d</i>	Chra08	4352179	4355111	2933	44.29	823	Plasma membrane
	<i>Bnglr21</i>	<i>Bnaa08g10690d</i>	Chra08	9905574	9911024	5451	34.21	911	Plasma membrane
	<i>Bnglr22</i>	<i>Bnaa08g29720d</i>	Chra08_rando m	498435	500465	2031	43.23	533	Plasma membrane
	<i>Bnglr23</i>	<i>Bnaa09g02950d</i>	Chra09	1502787	1505678	2892	44.16	653	Plasma membrane
	<i>Bnglr24</i>	<i>Bnaa09g02980d</i>	Chra09	1517971	1521181	3211	42.17	709	Plasma membrane
	<i>Bnglr25</i>	<i>Bnaa09g03000d</i>	Chra09	1525100	1528658	3559	39.25	855	Plasma membrane
	<i>Bnglr26</i>	<i>Bnaa09g32030d</i>	Chra09	23858647	23862328	3682	44.62	892	Plasma membrane
	<i>Bnglr27</i>	<i>Bnaa09g50380d</i>	Chra09	33337970	33340793	2824	46.6	872	Plasma membrane
	<i>Bnglr28</i>	<i>Bnaa10g03230d</i>	Chra10	1638312	1642124	3813	45.16	952	Plasma membrane
	<i>Bnglr29</i>	<i>Bnaa10g26130d</i>	Chra10	16703947	16710811	6865	40.64	1426	Plasma membrane
	<i>Bnglr30</i>	<i>Bnaa09g26240d</i>	Chrann_rando m	30168936	30173773	4838	36.56	1010	Plasma membrane
	<i>Bnglr31</i>	<i>Bnaa09g35320d</i>	Chrann_rando m	40059094	40060299	1206	43.95	402	Chloroplast
	<i>Bnglr32</i>	<i>Bnac01g03260d</i>	Chrc01	1672268	1677358	5091	44.04	913	Plasma membrane
	<i>Bnglr33</i>	<i>Bnac01g40220d</i>	Chrc01	38520477	38523439	2963	37.93	800	Plasma membrane
	<i>Bnglr34</i>	<i>Bnac02g20160d</i>	Chrc02	2189071	2193197	3746	37.61	860	Plasma membrane
	<i>Bnglr35</i>	<i>Bnac02g38540d</i>	Chrc02	41495784	41499389	3606	36.86	792	Plasma membrane
	<i>Bnglr36</i>	<i>Bnac02g38550d</i>	Chrc02	41508080	41512543	4464	39.47	885	Plasma membrane
	<i>Bnglr37</i>	<i>Bnac02g38560d</i>	Chrc02	41514008	41517361	3354	39.39	868	Plasma membrane
	<i>Bnglr38</i>	<i>Bnac03g26390d</i>	Chrc03	14977911	14981040	3130	37.51	408	Plasma membrane
	<i>Bnglr39</i>	<i>Bnac03g51940d</i>	Chrc03	18448735	18452702	3763	41.99	890	Plasma membrane
	<i>Bnglr40</i>	<i>Bnac03g51950d</i>	Chrc03	18452883	18457170	4050	41.46	927	Plasma membrane
	<i>Bnglr41</i>	<i>Bnac03g31230d</i>	Chrc03	76774361	76791138	4248	41.53	919	Plasma membrane
	<i>Bnglr42</i>	<i>Bnac04g12400d</i>	Chrc04	9661951	9665879	3929	43.32	922	Plasma membrane
	<i>Bnglr43</i>	<i>Bnac04g12410d</i>	Chrc04	9666583	9669930	3348	43.55	890	Plasma membrane
	<i>Bnglr44</i>	<i>Bnac04g15320d</i>	Chrc04	13060706	13065333	4628	38.55	972	Plasma membrane
	<i>Bnglr45</i>	<i>Bnac04g23810d</i>	Chrc04	24846056	24849197	3142	42.68	780	Plasma membrane
	<i>Bnglr46</i>	<i>Bnac04g37210d</i>	Chrc04	38667190	38670542	3353	38.47	807	Plasma membrane
	<i>Bnglr47</i>	<i>Bnac04g40280d</i>	Chrc04	41277722	41279305	1584	38.07	318	Plasma membrane
	<i>Bnglr48</i>	<i>Bnac04g40290d</i>	Chrc04	41286318	41290285	3968	39.04	590	Plasma membrane
	<i>Bnglr49</i>	<i>Bnac04g42860d</i>	Chrc04	43353531	43359263	5733	41.76	895	Plasma membrane

Table continued -

Table continued -

Species	Gene name	Gene ID (BRAD)	Chromosome	Start position	Stop position	Gene length (bp)	% GC content of gene	Peptide length (amino acids)	Subcellular localization
<i>B. napus</i>	<i>Bnglr50</i>	<i>Bnacnng66430d</i>	Chrc04	58297749	58302044	3076	40.6	810	Plasma membrane
	<i>Bnglr51</i>	<i>Bnac05g03240d</i>	Chrc05	1567784	1571031	3248	46.71	954	Plasma membrane
	<i>Bnglr52</i>	<i>Bnac06g36440d</i>	Chrc06	34974497	34978636	4140	38.48	914	Plasma membrane
	<i>Bnglr53</i>	<i>Bnacnng22760d</i>	Chrc07	4988250	4992404	3770	41.01	913	Plasma membrane
	<i>Bnglr54</i>	<i>Bnac07g26470d</i>	Chrc07	32420574	32424077	3504	40.75	874	Plasma membrane
	<i>Bnglr55</i>	<i>Bnac07g43470d</i>	Chrc07	42515456	42517887	2432	42.72	675	Plasma membrane
	<i>Bnglr56</i>	<i>Bnac07g50950d</i>	Chrc07	57599489	57635529	2424	38.94	544	Cytoplasmic/ Mitochondrial
	<i>Bnglr57</i>	<i>Bnac08g22850d</i>	Chrc08	25359634	25362884	3251	45.71	892	Plasma membrane
	<i>Bnglr58</i>	<i>Bnac08g44070d</i>	Chrc08	37504274	37508251	3978	42.38	940	Plasma membrane
	<i>Bnglr59</i>	<i>Bnac08g46160d</i>	Chrc08_random	62386	64416	2031	42.89	533	Plasma membrane
	<i>Bnglr60</i>	<i>Bnac09g02380d</i>	Chrc09	1345366	1348265	2900	44.31	671	Plasma membrane
	<i>Bnglr61</i>	<i>Bnac09g02400d</i>	Chrc09	1361958	1368324	6367	37.29	693	Plasma membrane
	<i>Bnglr62</i>	<i>Bnac09g02410d</i>	Chrc09	1370304	1373727	3424	40.68	860	Plasma membrane
	<i>Bnglr63</i>	<i>Bnacnng10410d</i>	Chrc09	51669158	51672056	2771	38.07	644	Plasma membrane
	<i>Bnglr64</i>	<i>Bnacnng32830d</i>	Chrcnn_random	31149937	31154309	4373	35.92	835	Plasma membrane
	<i>Bnglr65</i>	<i>Bnacnng48020d</i>	Chrcnn_random	47293706	47298713	5008	37.36	871	Plasma membrane
<i>B. oleracea</i>	<i>Boglr1</i>	<i>Bol029085</i>	Chrc01	1892437	1896025	3589	44.69	834	Plasma membrane
	<i>Boglr2</i>	<i>Bol034616</i>	Chrc01	34688024	34691281	3258	41.47	782	Plasma membrane
	<i>Boglr3</i>	<i>Bol008200</i>	Chrc01	38188571	38191975	3405	43.05	944	Plasma membrane
	<i>Boglr4</i>	<i>Bol008199</i>	Chrc01	38192270	38196318	4049	41.47	926	Plasma membrane
	<i>Boglr5</i>	<i>Bol013454</i>	Chrc04	19024234	19032115	7882	42.54	1787	Plasma membrane
	<i>Boglr6</i>	<i>Bol033267</i>	Chrc04	32256661	32260899	4239	36.02	739	Plasma membrane
	<i>Boglr7</i>	<i>Bol033269</i>	Chrc04	32292121	32295051	2931	40.84	758	Plasma membrane
	<i>Boglr8</i>	<i>Bol040960</i>	Chrc05	1566085	1569336	3252	46.77	951	Plasma membrane
	<i>Boglr9</i>	<i>Bol017691</i>	Chrc07	12729400	12733089	3690	41.22	912	Plasma membrane
	<i>Boglr10</i>	<i>Bol033482</i>	Chrc08	3749866	3753084	3219	44.11	929	Plasma membrane
	<i>Boglr11</i>	<i>Bol018326</i>	Chrc08	40412648	40415869	3222	45.19	852	Plasma membrane
<i>B. rapa</i>	<i>Brglr1</i>	<i>Bra011603</i>	Chra01	1449556	1453100	3545	44.77	914	Plasma membrane
	<i>Brglr2</i>	<i>Bra040147</i>	Chra01	28484030	28488142	4113	34.99	803	Plasma membrane
	<i>Brglr3</i>	<i>Bra023313</i>	Chra02	2788960	2792786	3827	36.79	860	Plasma membrane
	<i>Brglr4</i>	<i>Bra020703</i>	Chra02	23657430	23667003	9574	38.11	1616	Plasma membrane
	<i>Brglr5</i>	<i>Bra020702</i>	Chra02	23669339	23672825	3487	38.57	823	Plasma membrane
	<i>Brglr6</i>	<i>Bra020701</i>	Chra02	23674355	23682510	8156	40.11	859	Plasma membrane
	<i>Brglr7</i>	<i>Bra022887</i>	Chra03	7572439	7580285	7847	41.99	1858	Plasma membrane
	<i>Brglr8</i>	<i>Bra023989</i>	Chra03	28396850	28401973	5124	42.47	986	Plasma membrane
	<i>Brglr9</i>	<i>Bra032072</i>	Chra04	11081339	11085047	3709	40.47	897	Plasma membrane
	<i>Brglr10</i>	<i>Bra032063</i>	Chra04	11156694	11165271	8578	29.88	924	Plasma membrane
	<i>Brglr11</i>	<i>Bra040760</i>	Chra04	15679521	15684782	3831	38.16	843	Plasma membrane
	<i>Brglr12</i>	<i>Bra040761</i>	Chra04	15685945	15695322	4402	36.92	932	Plasma membrane
	<i>Brglr13</i>	<i>Bra040758</i>	Chra04	15699498	15704817	2220	40	594	Plasma membrane
	<i>Brglr14</i>	<i>Bra040759</i>	Chra04	15712883	15727082	6684	34.08	397	Plasma membrane
	<i>Brglr15</i>	<i>Bra005589</i>	Chra05	6378019	6381344	3326	43.69	946	Plasma membrane
	<i>Brglr16</i>	<i>Bra005591</i>	Chra05	6394658	6398235	3578	44.13	923	Plasma membrane
	<i>Brglr17</i>	<i>Bra018409</i>	Chra05	8057899	8062361	4463	37.28	935	Plasma membrane
	<i>Brglr18</i>	<i>Bra037487</i>	Chra06	21218646	21222111	3466	40.68	812	Plasma membrane
	<i>Brglr19</i>	<i>Bra015817</i>	Chra07	20705072	20709907	4836	36.54	1010	Plasma membrane
	<i>Brglr20</i>	<i>Bra034931</i>	Chra08	5173942	5177131	3190	44.26	930	Plasma membrane
	<i>Brglr21</i>	<i>Bra020812</i>	Chra08	11978273	11983391	5119	39.21	1015	Plasma membrane
	<i>Brglr22</i>	<i>Bra036175</i>	Chra09	2321958	2324869	2912	43.37	847	Plasma membrane
	<i>Brglr23</i>	<i>Bra036174</i>	Chra09	2328068	2331672	3605	39.08	856	Plasma membrane
	<i>Brglr24</i>	<i>Bra036795</i>	Chra09	25762588	25765963	3376	44.76	892	Plasma membrane
	<i>Brglr25</i>	<i>Bra032480</i>	Chra09	36246731	36249759	3029	46.52	962	Plasma membrane
	<i>Brglr26</i>	<i>Bra015398</i>	Chra10	2072590	2075666	3077	48.16	952	Plasma membrane
	<i>Brglr27</i>	<i>Bra009491</i>	Chra10	16898680	16905536	6857	40.78	1426	Plasma membrane

To determine the phylogenetic relationship among the GLRs in three *Brassica* genomes, we constructed the Maximum-Likelihood phylogenetic tree using the total length amino acid sequences of these proteins with 1000 bootstrap values. As shown in Figure 3.1, all GLRs were clearly divided into three groups named as A, B and C, with well-supported bootstrap values. The largest group is A (shown in red) contains forty-three GLRs from all three species. Group B (shown in blue) contains thirty four GLRs and Group C (shown in green) contains twenty six GLRs. Group C does not contain any sequences from *B. oleracea*.

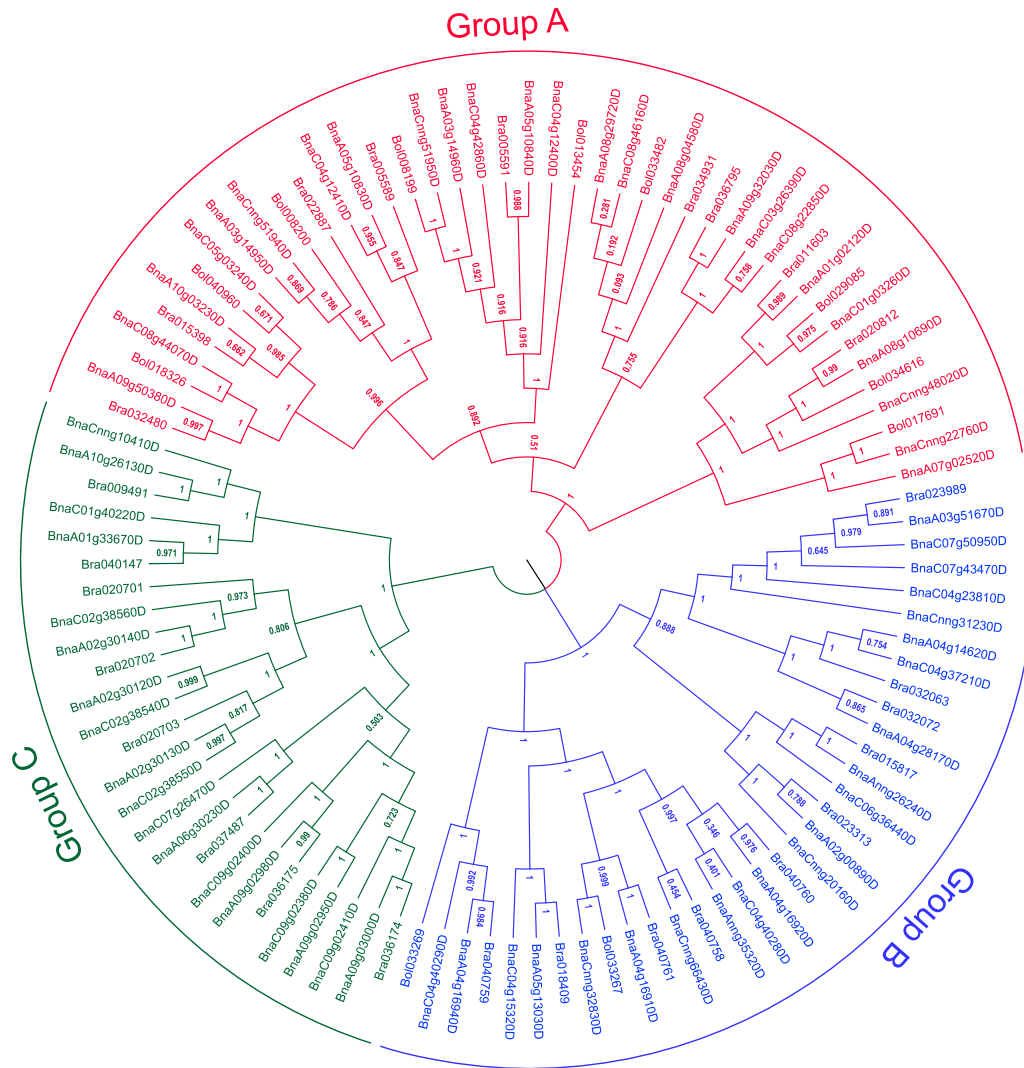


Figure 3.1: The Maximum-Likelihood (ML) phylogenetic tree of the GLR proteins of *B. rapa*, *B. oleracea* and *B. napus*. The phylogenetic tree was constructed by the MEGA X software with 1000 bootstrap replicates. Distinct colour segments represent different groups of GLRs. The numbers beside the branches suggest the decimal bootstrap values that support the adjacent node.

3.3.2 Chromosomal localization of *GLR* genes in three species of *Brassica*

The chromosomal locations of *GLR* gene family members in *B. rapa*, *B. oleracea* and *B. napus* genomes were analysed. In *B. rapa*, genes are distributed in all the 10 chromosomes while in *B. oleracea*, *GLR* genes are distributed in only 5 out of 9 chromosomes. In *B. rapa*, we have found five gene clusters with genes arranged in tandem arrangements indicating tandem duplication events. Of these, chromosome A02 has one cluster with three genes; chromosome A04 has two clusters with two and four genes, respectively; chromosome A05 has one cluster with two genes and chromosome A09 has one cluster with two genes. In *B. oleracea*, we have found only two clusters: one in chromosome C01 and another in chromosome C04 with two genes in each cluster. In *B. napus*, exact chromosomal locations of 7 *GLR* genes could not be obtained and thus, 58 *GLR* genes could be properly located on the chromosomes. By comparing the *B. napus* A subgenome with *B. rapa* genome, we have found 27 genes in each genome with similar relative positions for majority of the genes. However, while chromosome A04 of *B. rapa* possess two clusters of two and four tandemly arranged genes, respectively, chromosome A04 of *B. napus* possess only one cluster with three genes. Other changes include one cluster in chromosome A03 of *B. napus* which is absent in *B. rapa* and one extra gene in the cluster present in chromosome A09 in *B. napus* compared to *B. rapa* (Figure 3.2). However, when comparing *B. napus* C subgenome with *B. oleracea* genome, we have found five gene clusters in *B. napus* C subgenome compared to only two in *B. oleracea* genome indicating gene amplification in the C subgenome of *B. napus* after allopolyploidization. More importantly, *B. napus* C subgenome contains *GLR* genes in all the chromosomes unlike *B. oleracea* in which four chromosomes do not possess any *GLR* genes.

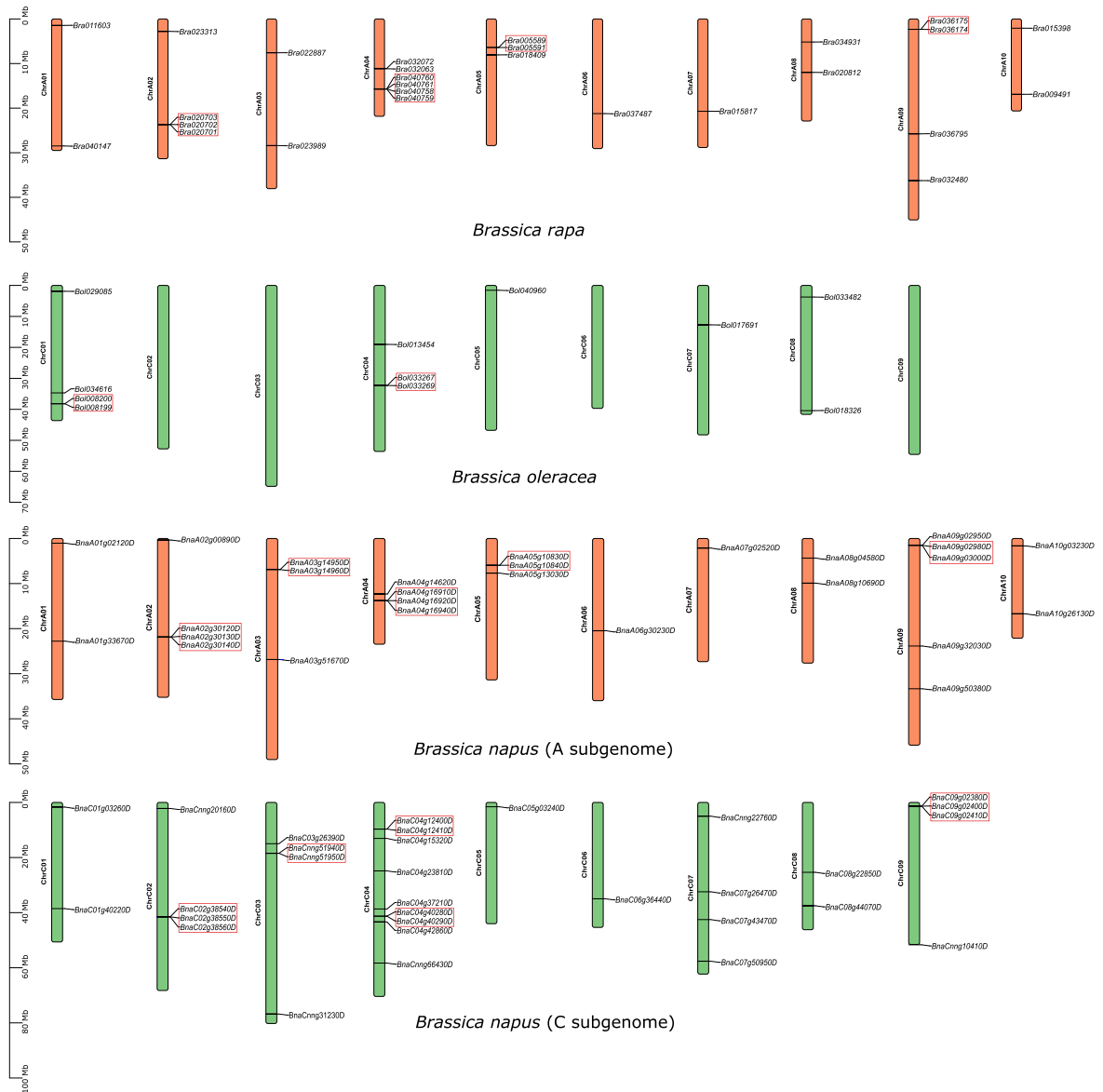


Figure 3.2: Distribution of *GLR* gene family members on chromosomes of *B. rapa*, *B. oleracea* and *B. napus*. *B. napus* A subgenome and C subgenome chromosomes are coloured in accordance with the *B. rapa* (A genome) and *B. oleracea* (C genome), respectively for better understanding of the subgenome specific *GLR* gene family expansion in *B. napus*. For *B. napus*, 58 *GLR* genes for which exact chromosomal information were available in the database were mapped. Genes inside the red boxes represent tandem arrangement. Scales at the left side indicate chromosome sizes. Note that in *B. napus*, the *GLR* gene family expanded mainly in the C subgenome.

3.3.3 Subcellular localization of GLR proteins in the three species of *Brassica*

The CELLO webtool was used to predict the subcellular localization of the GLR proteins in *B. oleracea*, *B. rapa* and *B. napus*. All of the GLRs of *B. oleracea* and *B. rapa* are predicted to be localized in plasma membrane. In *B. napus*, BnaAnng35320D and BnaC07g50950D were predicted to be localized in Chloroplast and Cytoplasm/Mitochondria, respectively. All the other proteins of this species were predicted to be localized in plasma membrane (Table 3.1). It indicates that the GLR family was highly conserved in subcellular localization during the allopolyploidization.

3.3.4 Gene structure and protein conserved motifs of *GLR* genes in the three species of *Brassica*

To study the *GLR* gene structure, we analyzed the distribution of the exons and introns in them on the basis of phylogenetic similarities of GLR proteins (Figure 3.3a, 3.3b). The number of introns ranged from zero (*BnaAnng35320D*) to 12 (*Bra022887*). Average introns per gene were 4.86, 4.59 and 4.73 for group A, B and C, respectively. Notably, genes with tandem arrangements did not show much conservation in exon-intron structures. We have also searched 10 conserved motifs in all GLR proteins of from *B. oleracea*, *B. rapa* and *B. napus* by using MEME (Figure 3.3c). As shown in Figure 3.3c, all the motifs were present in majority of GLR proteins, mainly in the same sequence. However, some motifs were present in duplicate in some of the genes, while some genes lack some of these motifs. Notably, motif 8 were absent in several members mainly in the members of the group C (Figure 3.1 and 3.3a). In contrast to the conservation in short amino acid sequence motifs, the exon-intron structures in the same clade were not highly conserved. Notably, all the short amino acid sequence motifs are present in duplicate in *Bra022887* protein. The *Bra022887* gene contains the highest number of introns (12), as mentioned earlier. Figure 3.3: exon–intron structure of *GLR* genes and conserved motif distribution of GLR proteins. (a) The phylogenetic tree of GLR proteins as shown in Figure 3.1 (coloured similarly for different groups); (b) Exon-intron structures of the 103 *GLR* genes; (c) Ten conserved motifs identified by MEME are indicated by different colors in all the 103 GLR proteins.

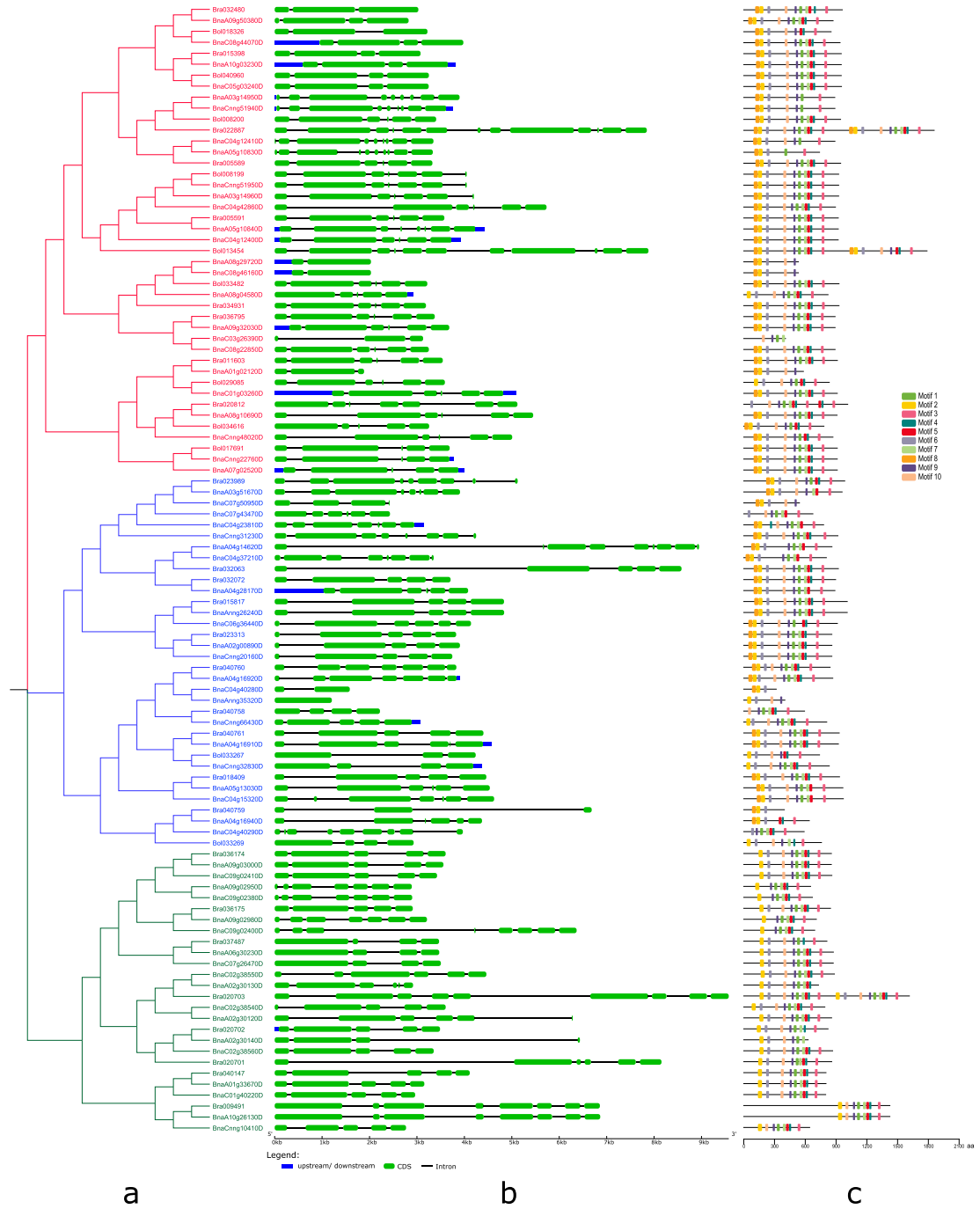


Figure 3.3: Exon–intron structure of *GLR* genes and conserved motif distribution of *GLR* proteins. (a) The phylogenetic tree of *GLR* proteins as shown in Figure 3.1 (coloured similarly for different groups); (b) Exon-intron structures of the 103 *GLR* genes; (c) Ten conserved motifs identified by MEME are indicated by different colors in all the 103 *GLR* proteins.

3.3.5 Identification of orthologous gene clusters in the three species of *Brassica*

The protein sequences of the identified *GLR* genes from all the three species were used for the orthology analysis. This analysis was performed to evaluate the effect of polyploidisation events, which shaped the evolution of the *Brassica* genomes, on the *GLR* gene family. We have considered only those genes for which exact chromosomal locations were available. In this way, several orthologous genes were identified of which the highest number (118 gene pairs) was identified between *B. rapa* and *B. napus*. However, only 52 gene pairs were identified between *B. oleracea* and *B. napus* which is much lesser than the earlier case. Only 27 gene pairs were identified between *B. rapa* and *B. oleracea* (Figure 3.4). Notably, orthology were found between the *GLR* genes of *B. rapa* and *B. oleracea* with both A and C subgenomes of *B. napus*.

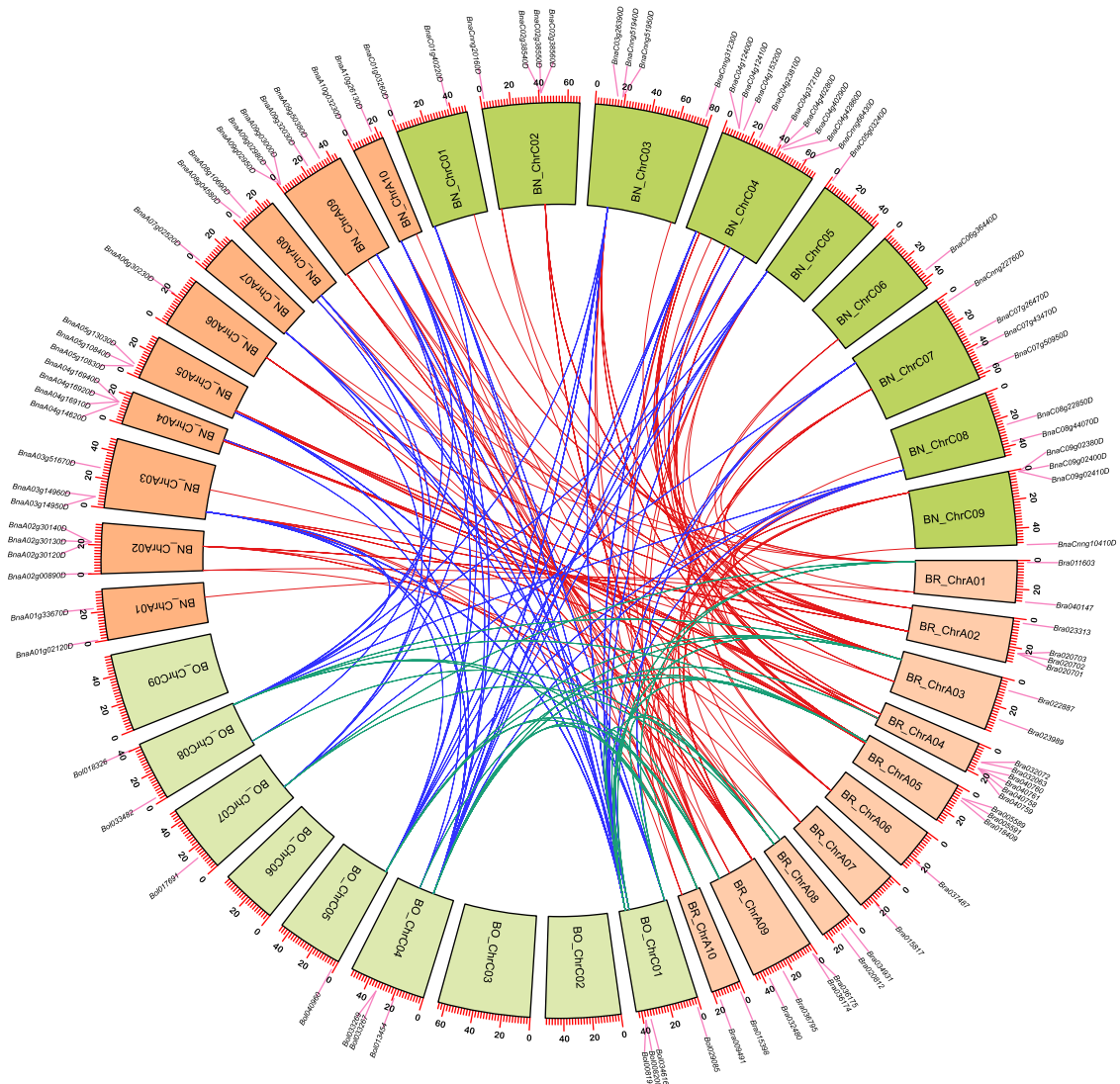


Figure 3.4: Orthologous *GLR* genes among *B. napus*, *B. rapa* and *B. oleracea* as found with OrthoVenn2. Nine *B. oleracea* chromosomes are coloured in light green, ten *B. rapa* chromosomes are coloured in light orange, nine *B. napus* chromosomes belonging to C subgenome are coloured in darker green and ten *B. napus* chromosomes belonging to A subgenome are coloured in darker orange. The respective chromosome numbers are written inside the coloured boxes. The gene names are written outside the chromosomes linked with pink lines. For *B. napus*, 58 *GLR* genes for which exact chromosomal information was available in the database were mapped only. The scale for the chromosome sizes (Mb) is rendered in red lines outside the chromosomes. The different orthologous genes are linked with different coloured lines; Blue, green and red lines indicate orthology between *B. oleracea* with *B. napus*, *B. oleracea* with *B. rapa* and *B. napus* with *B. rapa*, respectively. The figure was created using the software TBTools.

3.3.6 Analysis of *GLR* gene expression in the three species of *Brassica*

Based on the RNA-seq data, we investigated the expression patterns of *GLR* genes in the three species of *Brassica*. For *B. oleracea*, dataset of seven tissues (bud, callus, flower, leaf, root, silique and stem) were obtained. For *B. rapa*, eight tissues (callus, flower, leaf1, leaf2, root1, root2, silique and stem) were obtained. However, for *B. napus*, expression profile of genes of two early-maturing rapeseed varieties (HX17 and HX58) treated with cold shock at chilling (4°C) and freezing (−4°C) temperatures were obtained (He et al. 2019, 2020). For both the varieties, the treatments are as follows: Control (MA and MB), CA (4°C 12 h following cold acclimation), FA (−4°C 12 h following cold acclimation), CB (4°C) and FB (−4°C). *GLR* genes of *B. oleracea* were differentially expressed among the seven tissues (Figure 3.5a), reflecting the diversity of *GLR* gene functions during growth and development. Of these genes, *Bol034616* showed very low or no expression in all the tissues. Majority of the other genes expressed in all the tissues with varying expression level. Of these, *Bol013454* expressed at a much higher level in flower and silique than bud, leaf, root and stem indicating its primary role in reproductive growth. However, it was also highly expressed in callus which is noteworthy and need further study. *Bol033267* and *Bol033269* did not express in roots. In *B. rapa*, several genes showed distinct tissue specific expression patterns while others were expressed in all tissues but the expression levels varied in different tissues. Three genes namely *Bra009491*, *Bra040758* and *Bra040760* showed very high expression levels in callus tissues. Notably, they were also highly expressed in flowers (Figure 3.5b). In *B. napus*, we have studied differential expression of *GLR* genes in abiotic stress response. Some genes did not express at all in treated as well as control. There is a possibility that these genes lost the ability to be transcribed when regulatory regions became impaired, which presumably occurred later than the formation of the characteristic protein structural features. Similar observations were made by Xie and Shang (2018) while working with purple acid phosphatase (PAP) family of enzymes in plants. Other genes showed differential expression pattern in response to different cold stress. *BnaA10g03230D* was found to be downregulated in response to cold stress while *BnaA10g26130D* and

BnaC04g12410D were upregulated. Some other genes were also differentially expressed but their expression level varied in response to chilling and freezing stress compared to control. Also, different expression patterns were found among the two varieties (HX17 and HX58). For example, *BnaC04g12400D* showed downregulation of expression in response to chilling (CB) as well as freezing (FB) compared to the control (MB) in HX58. While, in HX17, the gene showed downregulation in response to freezing (FB) but showed similar expression in response to chilling (CB) compared to the control (MB) (Figure 3.5c). From these observations, it is clear that *GLR* genes take part in growth and development of vegetative and reproductive organs and also plays important role in abiotic stress response in different species of *Brassica*.

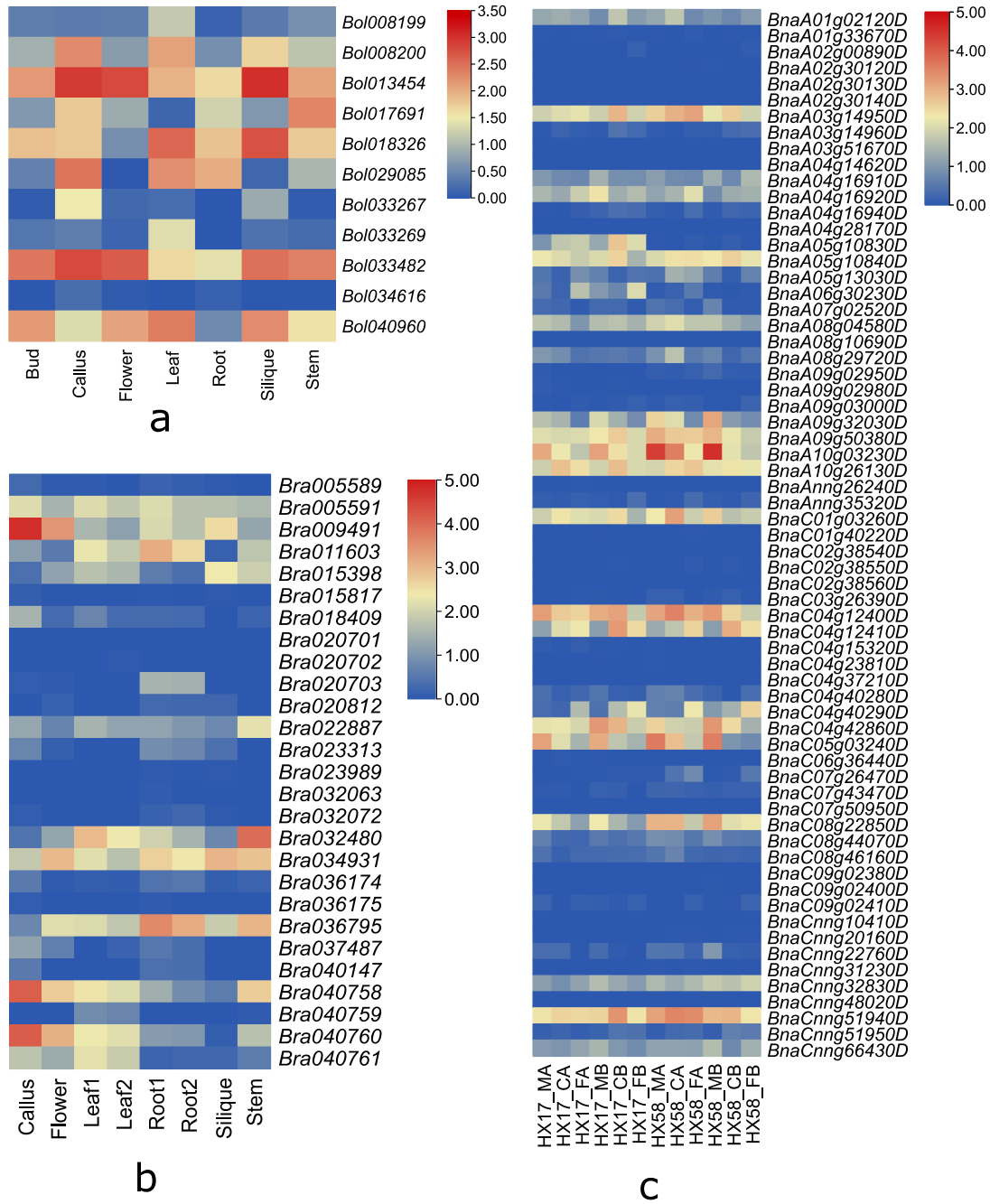


Figure 3.5: Heat map representation of expression patterns of *GLR* genes analyzed by RNA-Seq. a) *B. oleracea*, b) *B. rapa*, c) *B. napus*. Expression levels of the *GLR* genes are shown as the Log2 transformed FPKM values obtained from the respective RNA-Seq data.

3.3.7 Analysis of *cis*-acting elements in the promoters of *GLR* genes in the three species of *Brassica*

The PlantCARE database was used to predict the *cis*-acting elements present in the upstream 1000 bp of *GLR* genes of the three species of *Brassica* to investigate the potential function of these genes. We have found many *cis*-acting elements which can be broadly categorized into six types including promoters/enhancers (6), cell cycle related (4), development related (18), light responsive (26), plant hormone responsive (17), stress related (14) and other types (8) of *cis*-acting elements (Figure 3.6). Of these, almost all the *GLR* genes possess Myb and MYC binding sites indicating their roles in plant development. Majority of These genes also contain some light responsive elements like Box 4, G-box and GT1-motif. Plant hormone response elements and stress response elements play important roles for plant defense. Among the plant hormone responsive elements, ABRE elements are present in majority of genes. Several stress responsive elements were also found indicating that *GLR* genes play important role in plant stress response. Some of the frequently found stress responsive elements are LTR, MBS, STRE, W-box etc. It is noteworthy that there are some *cis*-elements which were found at least in one gene of *B. napus*, but not found at all in its two diploid progenitors. These elements are dOCT, E2Fb, OCT, re2f-1, AACA_motif, GRA, 4cl-CMA2b, AAAC-motif, LS7, Sp1, JERE, TATC-box and GC-motif. However, pattern of *cis*-element found is overall similar in these three species indicating that as a whole; the functional aspects of these genes remained unaltered after allopolyploidization.

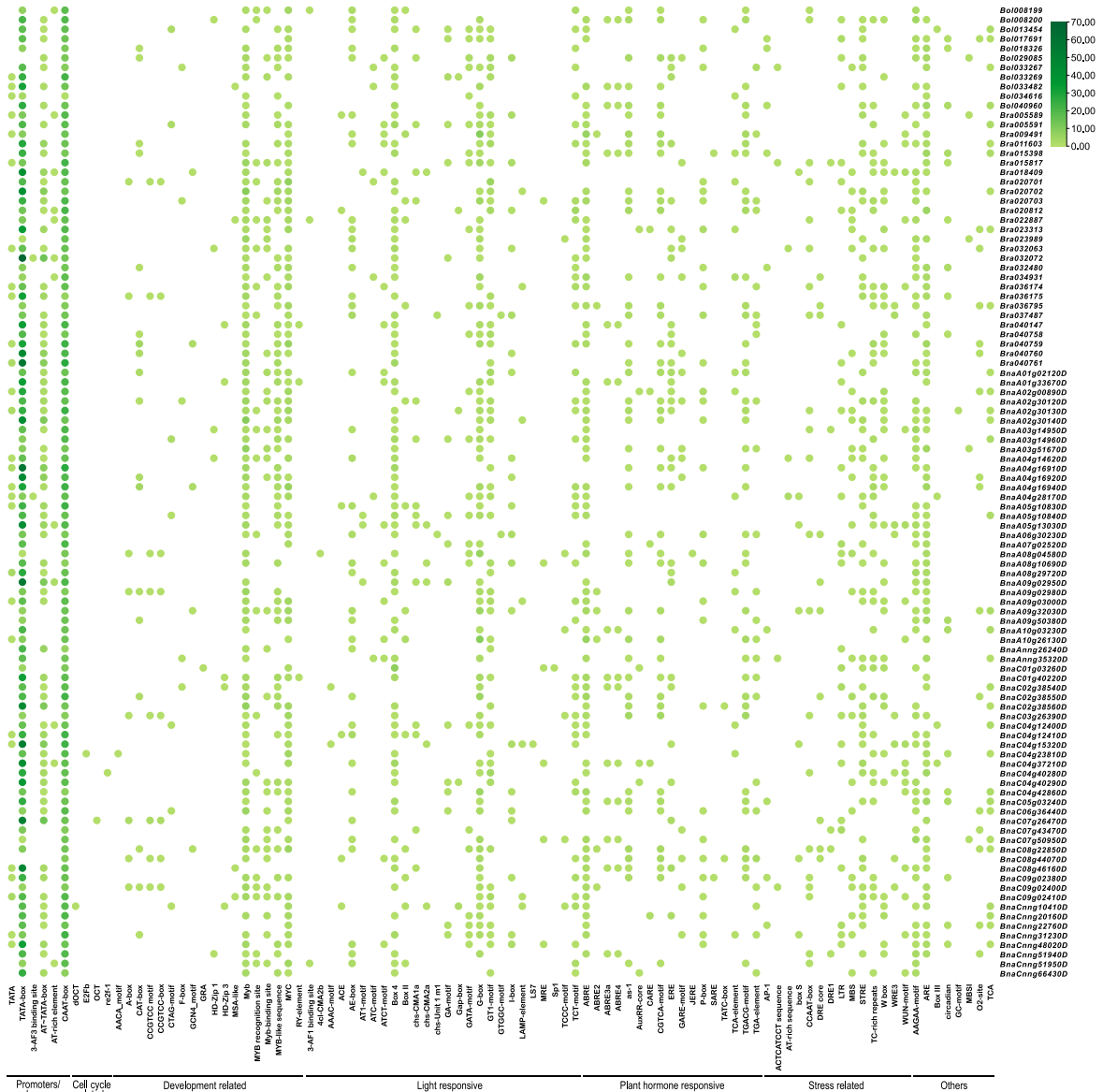


Figure 3.6: Main *cis*-elements in *GLR* gene promoters in *B. rapa*, *B. oleracea* and *B. napus*. Seven types of *cis*-elements viz. promoters/enhancers, cell cycle related, light responsive, development-related, plant hormone responsive, stress related and other elements were identified in the *GLR* gene promoter regions. The colour scale bar represents the number of a particular *cis*-element.

3.3.8 Comprehensive analysis of microRNAs targeting *GLR* Genes in the three species of *Brassica*

We have found three miRNAs targeting three different *GLR* genes in *B. oleracea* of which two showed inhibition by cleavage and one showed inhibition by translation (Figure 3.7a). In *B. rapa*, we have identified sixty one miRNAs targeting twenty five *GLR* genes (Figure 3.7b). Of these, the gene *Bra009491* is the most targeted gene (targeted by 12 miRNAs). On the other hand, *Bra015817* and *Bra040758* are targeted by only one miRNA each. Majority of these miRNAs inhibit the genes by cleavage. Furthermore, *bra-miR156a-3p* was identified as the highest targeting miRNA targeting six genes (*Bra040760*, *Bra011603*, *Bra020703*, *Bra036174*, *Bra020701* and *Bra020702*) all by cleavage. In *B. napus*, thirty six miRNAs target twenty seven *GLR* genes (Figure 3.7c). Of these, twelve belong to A subgenome and fifteen belong to the C subgenome. *BnaC08g22850D* is the most targeted gene (targeted by nine miRNAs). On the other hand, *BnaA04g28170D*, *BnaC04g40290D* and *BnaC05g03240D* are targeted by a single miRNA each. In *B. napus*, the members of miR172 family (*bn-miR172a*, *bn-miR172b*, *bn-miR172c* and *bn-miR172d*) are the highest targeting miRNAs (targeting eight genes each, all by cleavage). These genes are *BnaA01g33670D*, *BnaA03g14950D*, *BnaA05g10830D*, *BnaA10g26130D*, *BnaC01g40220D*, *BnaC04g12410D*, *BnaCnng10410D* and *BnaCnng51940D* and all of them are targeted by *bn-miR172a-d*.

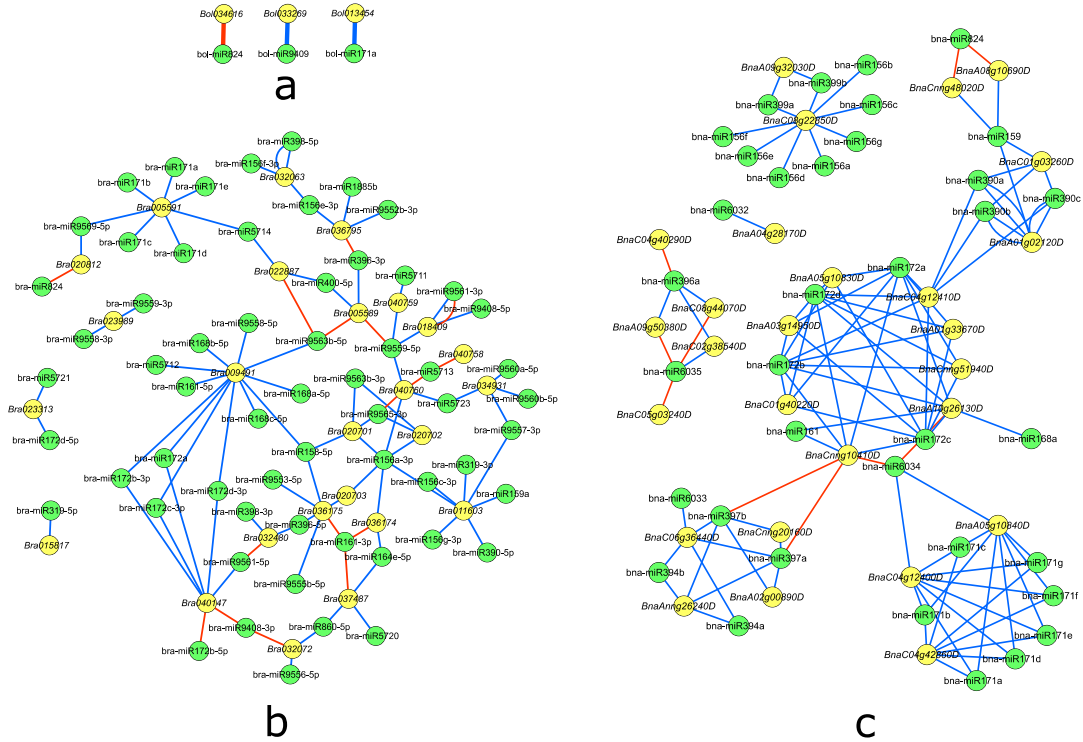


Figure 3.7: A schematic representation of the regulatory network relationships between the miRNAs and their targeted genes in a) *B. oleracea*, b) *B. rapa* and c) *B. napus*. Yellow circles represent regulated genes, green circles represent miRNAs that regulates the genes expression; blue and red lines represent inhibition by cleavage and by inhibition of translation, respectively.

3.4 Discussion

GLRs are very important in plant signaling and hence, have been studied extensively in past few years. Bortoli et al (2016) studied extensively on this gene family in different photosynthetic organisms including *Brassica rapa*. However, considering the economic importance of the genus *Brassica*, it is important to study the *GLR* gene family in different members of this genus. The recently sequenced genome of allopolyploid *B. napus* as well as its diploid progenitors *B. rapa* and *B. oleracea* gave us the opportunity to identify and study the *GLR* gene family in these species. Moreover, it is important to study the fate of *GLR* genes after polyploidization in this genus. So, this study can also be important in the context of genome-wide study of gene families after polyploidization. In this study, we have found that the *GLR* gene family in *Brassica napus* has expanded after the allopolyploidization as the total number of *GLR* genes in *B. napus* is greater than the

combined number of *GLR* genes in *B. rapa* and *B. oleracea*. Similar results were found regarding *RC12* gene family in these three species (Sun et al. 2022). However, this expansion of *GLR* gene family was not uniform in the A and C subgenomes, as revealed by the present study. While the *GLR* gene number remains constant in the A subgenome (with some alterations in chromosomal positions within the genome), it expanded considerably in the C subgenome. We have found several tandem duplications of *GLR* genes after allopolyploidization especially in the C subgenome. It shows that polyploidization has different effect on different subgenomes, at least in *Brassica*. *Brassica* has undergone a WGT (whole genome triplication) event followed by gene duplications, deletions and rearrangements (Lysak et al. 2005). Moreover, gene loss event after triploidisation has also been suggested previously (Song et al. 2014; Lohani et al. 2019). Our study indicate that these events may have shaped the expansion of *GLR* genes in *B. napus* differently in A and C subgenomes. However, the analysis of orthologous gene relationship using the protein sequences shows that *B. napus* *GLR* genes from both A and C subgenomes are orthologous to *B. rapa* and *B. oleracea*. In fact, phylogenetic analysis of *GLR* proteins showed that these sequences can be subdivided into three groups but these groups are not species-specific. This shows that protein sequence level conservation is high among the members of this group. Moreover, study of conserved motifs in the *GLR* proteins with the help of MEME also supports this as the arrangements of these motifs are more or less conserved. Thus, comparative analysis of the *GLR* proteins highlighted that the protein sequence and domain compositions of these genes have been relatively conserved during evolution of the *Brassica* species. However, we have found diverse intron/exon structures in *Brassica* *GLR* genes. Introns and RNA splicing offer great genetic plasticity of the eukaryotic cells (Patthy 1999). So, it has been found that protein level conservation is more in these *GLRs* in *Brassica* than gene level structural conservation.

The *cis*-acting elements also showed tremendous diversity in the promoter region which shows they are involved in diverse physiological processes like development, response to different plant hormones, light response, stress responses etc. However, overall, the pattern of *cis*-elements in the promoters found is overall similar in *B. napus* and its two diploid progenitors indicating that as a whole; the functional aspects of these genes

remained unaltered after allopolyploidization. Expression analysis showed difference in expression pattern, expression level and tissue specificity in *B. rapa* and *B. oleracea*. Thus, these genes play important roles in plant growth and development. Previous studies showed that in plants, *GLR* genes are expressed in response to wounds, pathogen attack etc. (Grenzi et al. 2021). We have shown that in *B. napus*, *GLRs* are up or downregulated in response to cold stress indicating their role in abiotic stress response. This is also evident from the presence of *cis*-acting elements related to stress response.

The expression of some of these genes in all the species is controlled by micro RNAs as revealed in the present study. In *B. rapa*, the gene *Bra009491* was found to be targeted by 12 miRNAs. Interestingly, this gene showed the highest expression in callus compared to other tissues. So, in normal conditions, its expression is lower. This may be contributed by higher targeting by miRNAs. In *B. napus*, the members of miR172 family are the highest targeting miRNAs. MiR172 was first reported from *Arabidopsis* and then in many plants because of its high conservation (Park et al. 2002; Wang et al. 2019). The members of this miRNA family and their target genes play key roles in floral organ differentiation and flowering time (Wang et al. 2019). In *B. napus*, four putative miR172 family members (Bna-miR172a-d) were predicted in miRbase. These four mature miR172 sequences are highly conserved except for the first base and divergence was observed in the precursor sequences (Wang et al. 2019). It is interesting to observe that these miRNAs also target several *GLR* genes.

3.5 Conclusion

In conclusion, we comprehensively analysed the evolutionary pattern, gene structure, conserved motifs, orthology, *cis*-acting elements and expression profiles of *GLR* genes in three *Brassica* species. This effort can serve as a preliminary step in comprehensive functional characterization of *GLR* genes by reverse genetic approaches and molecular genetics research. In addition, this study may also facilitate our understanding of the evolution of the economically important species of *Brassica* during the evolution of polyploidy.

Chapter IV.

Homology modeling and *in silico*
structural characterization of glutamate
receptors of allopolyploid *Brassica*
napus and its diploid progenitors

Chapter IV.

Homology modeling and *in silico* structural characterization of glutamate receptors of allopolyploid *Brassica napus* and its diploid progenitors

4.1 Introduction

4.2 Methodology

4.3 Results

4.4 Discussion

4.5 Conclusion

4.1 Introduction

Plant glutamate-receptors (GLRs) are homologs of animal ionotropic glutamate receptors (iGluRs) and are known to be associated with various physiological processes in plants including carbon and nitrogen metabolism, defense response, stomatal aperture regulation, pollen tube growth, root development, and stress responses (Simon et al. 2023); abscisic acid (ABA) biosynthesis and signaling, water loss, seed germination (Kong et al. 2015); root gravitropism (Miller et al. 2010), lateral root initiation (Vincill et al. 2013), root development (Singh et al. 2016), innate immune responses (Forde and Roberts 2014) and so on. Since their discovery in 1998 in *Arabidopsis thaliana* (Lam et al. 1998), many such ion channels have been discovered from other plants. However, study of their structural details is available only from a few GLR proteins of *A. thaliana*. It is now known that they are multimeric (homo or heteromeric) membrane proteins with structural similarity with animal iGluRs.

However, with the discovery of many GLRs from other plants, it is necessary to study the structural features of these GLRs for better understanding the structure, function, and evolution of these proteins. Considering their role in diverse physiological processes in plants, study of GLRs is even more important for crop plants. In chapter III, we have identified and studied several GLR genes from *Brassica napus*, *B. oleracea* and *B. rapa*. However, their structural properties are yet to be analyzed.

Protein characterization by experimental methods involve high cost and much time. Thus, these methods are not amenable for high throughput techniques. In these situations, *in silico* methods can be used for getting an overall picture of properties of proteins that can further be studied by experimental methods. Thus, molecular approach linked to bioinformatics is essential (Shukla et al. 2021). Several computational tools are helpful to understand the physico-chemical and structural properties of proteins. As the three-dimensional structure of proteins can answer many queries regarding their functional aspects, three-dimensional structure prediction is an essential aspect of bioinformatics and homology modeling is used for prediction of the three-dimensional structure of proteins (Shukla et al. 2021). This approach is based on the fact that evolutionary related proteins possess similar structures (Altschul et al. 1994; Shukla et al. 2021). We have modeled the GLRs of *Arabidopsis thaliana* earlier based on AMPA subtype ionotropic glutamate receptor from *Rattus norvegicus*

(Chapter II). However, crystal structure of *A. thaliana* GLR is available now (Green et al. 2021). So, it is expected that more accurate prediction is possible now with the advent of new experimental data. So, in this work, we have studied the structural properties of GLRs from three species of *Brassica* with the help of some bioinformatic tools.

4.2 Methodology

4.2.1 Sequence retrieval

We have used 65, 27 and 11 GLR proteins from *B. napus*, *B. rapa* and *B. oleracea* as described in Chapter III. The protein sequences of the identified genes were retrieved in FASTA format and used for further analysis.

4.2.2 Physicochemical characterization and secondary structure prediction

For physicochemical characterization, theoretical pI, molecular weight, instability index (Guruprasad et al. 1990) and grand average hydropathy (GRAVY) (Kyte and Doolittle 1982) were computed using the Expasy's ProtParam (Gasteiger et al. 2005) server (<https://web.expasy.org/protparam/>). Secondary structure prediction was carried out with SOPMA (Geourjon and Deleage 1995).

4.2.3 Homology modeling and model validation

SWISS-MODEL workspace (<https://swissmodel.expasy.org/>) (Waterhouse et al. 2018) was used to predict the three-dimensional structure of the GLRs. The most suitable template was selected for modeling. All the modeled structures were evaluated by different quality evaluation tools such as QMEAN (Benkert et al. 2011) and QMEANDisCo (Studer et al. 2020) available with the Swiss Model workspace. Further quality assessment was performed by Ramachandran plot analysis in PROCHECK (Laskowski et al. 1993) and ERRAT (Colovos and Yeates 1993) by using SAVES server (<https://saves.mbi.ucla.edu/>) as well as by ProSA (Sippl 1993; Wiederstein and Sippl 2007).

Using Swiss-Model, the most suitable template for homology modeling was chosen from the protein data bank (PDB) (Berman et al. 2000) (<http://www.rcsb.org/>) which contains all publicly available 3D structures of proteins. On the basis of the results obtained from template search, the crystal structure of the glutamate receptor-like

channel AtGLR3.4 (PDB id: 7LZH, chain A) at a resolution of 3.57Å (Green et al. 2021) was used as the template for homology modeling of the 3D models of GLRs of three *Brassica* species. Modeled proteins were visualized with Chimera (Pettersen et al. 2004). Energy minimization was carried out with Gromos force field (van Gunsteren et al. 1996) in Swiss-PdbViewer (Guex and Peitsch 1997).

4.2.4 Surface- electrostatics property analysis

Electrostatic potentials of the proteins were compared with the web-based version of PIPSA (Protein Interaction Property Similarity Analysis) (Blomberg et al. 1999; Wade et al. 2001; Gabdoulline et al. 2007; Richter et al. 2008). For this, monomers of the GLRs were considered as the tetramers were huge and resulted in some errors while performing PIPSA. The models were superposed with Chimera and submitted to the PIPSA web server and Adaptive Poisson-Boltzmann Solver (APBS) software (Baker et al. 2001) was chosen to calculate the electrostatic potentials. For pair wise comparisons of electrostatic potentials of two proteins, Hodgkin (Hodgkin and Richards 1987; Good et al. 1992) similarity index (SI) for two proteins (labeled 1 and 2) was used as follows:

$$SI_{12} = \frac{2(\mathbf{p}_1, \mathbf{p}_2)}{(\mathbf{p}_1, \mathbf{p}_1) + (\mathbf{p}_2, \mathbf{p}_2)},$$

where $(\mathbf{p}_1, \mathbf{p}_1)$, $(\mathbf{p}_1, \mathbf{p}_2)$ and $(\mathbf{p}_2, \mathbf{p}_2)$ are the scalar products of the electrostatic potentials over the region where the potentials are compared (Blomberg et al. 1999). If the electrostatic potentials (ϕ) are computed on a three-dimensional grid over the two superimposed molecules, the scalar product is

$$(\mathbf{p}_1, \mathbf{p}_2) = \sum_{i,j,k} \phi_1(i, j, k) \phi_2(i, j, k),$$

Where summation is carried out over the grid points (i,j,k) that are within the region of interest (Blomberg et al. 1999). The resulting similarity indices have been converted by PIPSA into distances as follows:

$$D_{a,b} = \sqrt{2 - 2SI_{a,b}}$$

where $D_{a,b}$ and $SI_{a,b}$ are the respective distance and similarity index of two proteins (labeled a and b) (Richter et al. 2008). Finally, a cluster analysis has been performed with PIPSA.

4.3 Results

4.3.1 Physico-chemical properties of GLRs of three species of *Brassica*

The physico-chemical properties of the GLRs were studied with ExPASy's ProtParam tool. In *B. napus*, the molecular weight of the proteins ranged from 36.01 kDa (BnaC04g40280D) to 159.78 kDa (BnaA10g26130D) with an average ($\pm$ SE) of 91.21 ($\pm$ 2.36) kDa. In *B. rapa*, the molecular weight of the proteins ranged from 44.19 kDa (Bra040759) to 207.96 kDa (Bra022887) with an average ($\pm$ SE) of 107.06 ($\pm$ 6.07) kDa. In *B. oleracea*, the molecular weight of the proteins ranged from 83.35 kDa (Bol033267) to 199.24 kDa (Bol013454) with an average ($\pm$ SE) of 105.80 ($\pm$ 9.68) kDa.

The GRAVY value for a peptide is the sum of hydropathy values (Kyte and Doolittle 1982). In *B. napus*, Maximum and minimum GRAVY score were recorded as 0.097 (BnaC04g23810D) and -0.276 (BnaC04g40280D), respectively. Twenty-five of these proteins showed positive GRAVY scores while the rest forty showed negative values. In *B. rapa*, the GRAVY score ranged from -0.19 to 0.078. Twenty of these GLRs showed negative GRAVY score while the rest seven showed positive GRAVY score. In *B. oleracea*, the GRAVY score ranged from -0.259 to 0.078. Eight of these proteins showed negative GRAVY score while only three showed positive GRAVY score. In *B. napus*, the value of pI ranged from 5.73 (BnaCnng20160D) to 9.26 (BnaCnng10410D) with an average ($\pm$ SE) of 7.52 ($\pm$ 0.14). In *B. rapa*, pI ranged from 6.1 (Bra023313) to 9.47 (Bra020701) with an average ($\pm$ SE) of 7.82 ($\pm$ 0.17). In *B. oleracea*, pI ranged from 6.11 to 8.86 with an average ($\pm$ SE) of 7.67 ($\pm$ 0.30). Regarding instability index, *B. napus* showed that 80% (52 out of 65) GLRs are stable *i.e.* their instability index is lower than 40. In *B. rapa* and *B. oleracea*, 74% (20 out of 27) and 63% (7 out of 11) GLRs are stable, respectively (Table 4.1).

Table 4.1: Physico-chemical properties of GLRs of *Brassica* species

Protein name	Gene ID	Molecular weight (kD)	pi	GRAVY	Instability index
BnGLR1	BnaA01g02120D	64.45	9.11	-0.02	33.36
BnGLR2	BnaA01g33670D	89.79	7.69	0.038	34.68
BnGLR3	BnaA02g00890D	97.05	6.05	-0.143	41.31
BnGLR4	BnaA02g30120D	96.20	6.79	-0.005	30.71
BnGLR5	BnaA02g30130D	82.37	6.19	0.005	39.5
BnGLR6	BnaA02g30140D	70.47	6.39	-0.041	36.89
BnGLR7	BnaA03g14950D	99.01	6.83	-0.201	41.7
BnGLR8	BnaA03g14960D	103.52	8.58	0.063	35.39
BnGLR9	BnaA03g51670D	108.11	8.64	-0.061	32.72
BnGLR10	BnaA04g14620D	96.83	6.82	-0.047	34.6
BnGLR11	BnaA04g16910D	104.10	7.04	-0.186	39.9
BnGLR12	BnaA04g16920D	97.34	8.14	-0.103	33.14
BnGLR13	BnaA04g16940D	71.97	6.46	-0.132	36.95
BnGLR14	BnaA04g28170D	100.51	8.77	-0.058	36.84
BnGLR15	BnaA05g10830D	81.89	6.05	-0.055	42.53
BnGLR16	BnaA05g10840D	103.08	7.3	0.066	39.75
BnGLR17	BnaA05g13030D	108.55	8.82	-0.133	31.58
BnGLR18	BnaA06g30230D	97.87	8.41	-0.022	29.59
BnGLR19	BnaA07g02520D	101.59	7.27	0.003	41.67
BnGLR20	BnaA08g04580D	92.33	8.7	-0.055	35.46
BnGLR21	BnaA08g10690D	100.96	8.23	0.075	39.24
BnGLR22	BnaA08g29720D	59.27	8.92	0.032	31.19
BnGLR23	BnaA09g02950D	73.55	6.35	0.033	34.17
BnGLR24	BnaA09g02980D	79.55	5.94	0.013	34.29
BnGLR25	BnaA09g03000D	96.25	6.42	-0.027	33.75
BnGLR26	BnaA09g32030D	99.35	8.65	0.075	40.62
BnGLR27	BnaA09g50380D	97.23	6.47	-0.084	34.59
BnGLR28	BnaA10g03230D	106.16	8.34	-0.088	38.47
BnGLR29	BnaA10g26130D	159.78	8.63	-0.187	45.82

Table continued -

Table continued -

Protein name	Gene ID	Molecular weight (kD)	pi	GRAVY	Instability index
BnGLR30	BnaAnng26240D	113.54	8.64	-0.155	40.1
BnGLR31	BnaAnng35320D	45.11	5.98	-0.226	35.16
BnGLR32	BnaC01g03260D	101.59	8.55	0.023	38.94
BnGLR33	BnaC01g40220D	89.65	7.31	0.031	38.59
BnGLR34	BnaCnng20160D	97.04	5.73	-0.143	41.51
BnGLR35	BnaC02g38540D	89.12	8.92	-0.023	34.72
BnGLR36	BnaC02g38550D	98.63	5.94	-0.004	35.04
BnGLR37	BnaC02g38560D	97.32	7.65	0.024	36.72
BnGLR38	BnaC03g26390D	44.92	6.64	-0.036	36.97
BnGLR39	BnaCnng51940D	99.18	5.9	-0.204	40.79
BnGLR40	BnaCnng51950D	103.50	8.57	0.078	35.02
BnGLR41	BnaCnng31230D	103.48	7.11	0.019	36.64
BnGLR42	BnaC04g12400D	102.82	7.95	0.074	39.08
BnGLR43	BnaC04g12410D	99.51	6.22	-0.094	36.6
BnGLR44	BnaC04g15320D	108.66	8.8	-0.112	33.45
BnGLR45	BnaC04g23810D	87.07	6.15	0.097	32.22
BnGLR46	BnaC04g37210D	91.11	6.56	-0.002	33.43
BnGLR47	BnaC04g40280D	36.01	8.64	-0.276	40.21
BnGLR48	BnaC04g40290D	67.40	6.05	-0.144	37.82
BnGLR49	BnaC04g42860D	100.45	8.17	0.033	35.9
BnGLR50	BnaCnng66430D	90.96	8.69	-0.164	35.64
BnGLR51	BnaC05g03240D	106.30	7.94	-0.051	39.09
BnGLR52	BnaC06g36440D	102.78	8.9	-0.227	37.43
BnGLR53	BnaCnng22760D	101.54	7.26	0.005	41.95
BnGLR54	BnaC07g26470D	97.83	8.37	0.011	30.46
BnGLR55	BnaC07g43470D	76.12	6.54	0.029	35.49
BnGLR56	BnaC07g50950D	61.49	8.66	-0.262	33.46
BnGLR57	BnaC08g22850D	99.36	8.77	0.052	40.54
BnGLR58	BnaC08g44070D	104.64	7.59	-0.055	34.58
BnGLR59	BnaC08g46160D	59.33	8.9	0.041	31.53

Table continued -

Table continued -

Protein name	Gene ID	Molecular weight (kD)	pi	GRAVY	Instability index
BnGLR60	BnaC09g02380D	75.77	6.02	0.039	33.46
BnGLR61	BnaC09g02400D	78.05	5.74	-0.066	36.46
BnGLR62	BnaC09g02410D	96.55	6.21	-0.025	36.59
BnGLR63	BnaCnng10410D	72.06	9.26	-0.017	37.55
BnGLR64	BnaCnng32830D	94.30	8.49	-0.218	38.22
BnGLR65	BnaCnng48020D	96.45	7.97	0.022	42.8
BoGLR1	Bol029085	92.76	8.86	-0.031	40.02
BoGLR2	Bol034616	87.39	7.63	-0.032	41.1
BoGLR3	Bol008200	105.80	6.45	-0.098	40.22
BoGLR4	Bol008199	103.50	8.57	0.078	35.02
BoGLR5	Bol013454	199.24	7.75	-0.022	38.88
BoGLR6	Bol033267	83.35	8.63	-0.259	39.44
BoGLR7	Bol033269	85.88	6.11	-0.077	30.78
BoGLR8	Bol040960	106.14	8.17	-0.046	38.79
BoGLR9	Bol017691	101.62	7.26	0.001	42.07
BoGLR10	Bol033482	103.95	8.57	0.01	34
BoGLR11	Bol018326	94.16	6.36	-0.077	34.84
BrGLR1	Bra011603	101.40	8.23	0.059	37.58
BrGLR2	Bra040147	89.97	7.34	0.026	35.82
BrGLR3	Bra023313	97.07	6.1	-0.145	41.92
BrGLR4	Bra020703	181.66	8.09	-0.031	34.13
BrGLR5	Bra020702	91.97	6.36	-0.019	35.39
BrGLR6	Bra020701	96.85	9.47	-0.066	40.42
BrGLR7	Bra022887	207.96	7.59	-0.044	38.38
BrGLR8	Bra023989	111.00	8.18	-0.037	33.38
BrGLR9	Bra032072	101.38	8.46	-0.027	36.99
BrGLR10	Bra032063	103.99	6.83	-0.05	33.88
BrGLR11	Bra040760	93.78	8.13	-0.028	30.76
BrGLR12	Bra040761	104.92	6.85	-0.176	40.18
BrGLR13	Bra040758	66.68	8.77	-0.098	32.63

Table continued -

Table continued -

Protein name	Gene ID	Molecular weight (kD)	pi	GRAVY	Instability index
BrGLR14	Bra040759	44.19	7.06	0.012	28.22
BrGLR15	Bra005589	105.86	7.16	-0.101	39.98
BrGLR16	Bra005591	102.92	7.3	0.078	39.1
BrGLR17	Bra018409	104.78	8.45	-0.15	32.4
BrGLR18	Bra037487	90.31	8.73	-0.043	28.98
BrGLR19	Bra015817	113.54	8.64	-0.155	40.1
BrGLR20	Bra034931	103.95	8.66	0.005	33.24
BrGLR21	Bra020812	113.31	8.4	-0.009	41.15
BrGLR22	Bra036175	94.71	6.57	0.01	32.54
BrGLR23	Bra036174	96.35	6.63	-0.037	32.46
BrGLR24	Bra036795	99.35	8.65	0.075	40.62
BrGLR25	Bra032480	106.69	7.59	-0.061	34.78
BrGLR26	Bra015398	106.14	8.34	-0.084	38.84
BrGLR27	Bra009491	159.84	8.65	-0.19	45.86

Regarding instability index, 52 of 65 GLR proteins of *B. napus* were found to have instability index less than 40 while the other 13 showed instability index more than 40. In *B. rapa*, 20 GLRs showed instability index less than 40 while the rest 7 showed instability index of more than 40. In *B. oleracea*, 4 GLRs showed instability index more than 40 while the other 7 showed instability index of less than 40.

4.3.2 Secondary structural features of GLRs of three *Brassica* species

The secondary structural features were analyzed by SOPMA. In All the GLRs, predominance of alpha helices and random coils were observed (Figure 4.1, 4.2 and 4.3). In *B. oleracea*, average percentage of random coil (37.29 ± 0.89) was slightly higher than alpha helix (36.41 ± 0.72). However, in *B. rapa*, average percentage of alpha helix (37.57 ± 0.69) was slightly higher than random coil (36.58 ± 0.55). Similarly, in *B. napus*, average percentage of alpha helix was slightly higher (37.59 ± 0.37) than random coil (36.75 ± 0.32).

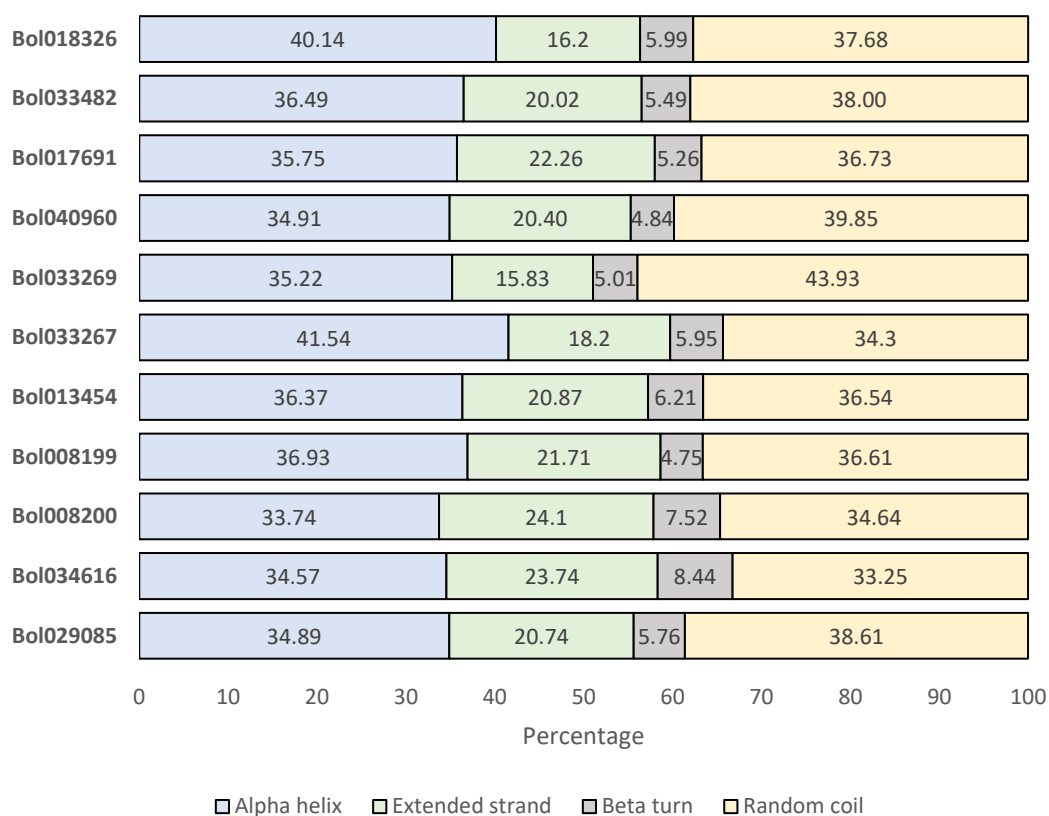


Figure 4.1: Secondary structural features of GLRs of *B. oleracea*

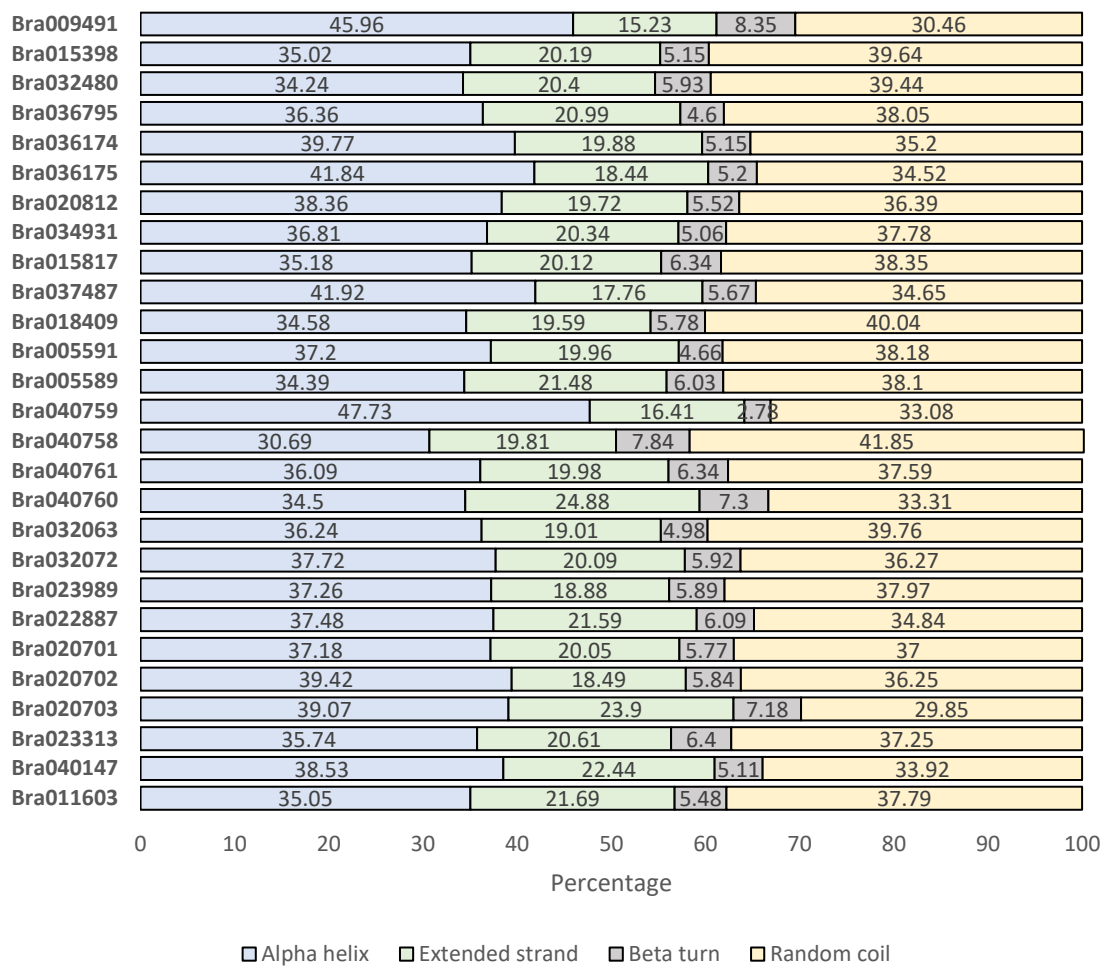


Figure 4.2: Secondary structural features of GLRs of *B. rapa*

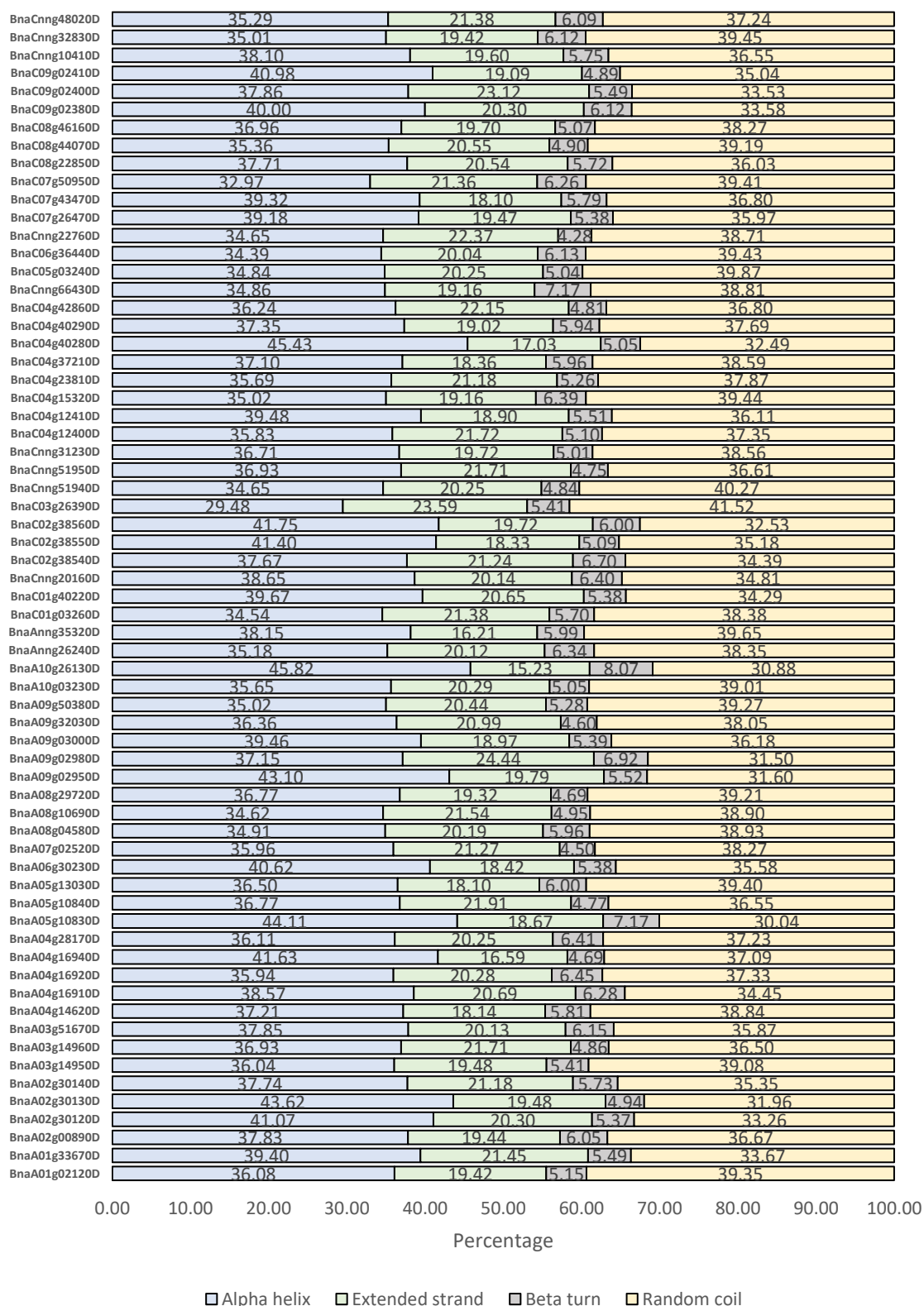


Figure 4.3: Secondary structural features of GLRs of *B. napus*

4.3.3 Three-dimensional structures of GLRs

Homology modeling of the GLRs were performed with SWISS-MODEL and the Structure of the glutamate receptor-like channel AtGLR3.4 (PDB id: 7LZH, chain A) at a resolution of 3.57Å (Green et al. 2021) was used as the template for homology modeling. All the modeled structures were carefully and manually checked. Only those structures were selected which showed the presence of ATD, LBD and TMD. Finally, 74 models (7 from *B. oleracea*, 21 from *B. rapa* and 46 from *B. napus*) were selected for further study. Sequence similarity with the template ranged from 29-58 percent. Regarding sequence coverage, two sequences showed less than 50% sequence coverage (47% for Bra020703 and 48% for Bra022887). However, these are very long proteins (peptide length of 1616 and 1858 amino acids for Bra020703 and Bra022887, respectively) while the average peptide length for all the 65 GLRs are 865.68±22.25 amino acids. Moreover, the domain search with InterPro (Paysan-Lafosse et al. 2023) showed the presence of Ionotropic glutamate receptor (IPR015683) family domain in the first half of Bra020703 (amino acid position 48-808) and in the second half of Bra022887 (amino acid position 961-1810). The target-template coverage was at per with this result. For Bra020703, the template coverage was from amino acid position 43-810 of the target and for Bra022887, the template coverage was from amino acid position 956-1791 of the target. Rest of the proteins showed sequence coverage ranging from 72% (Bra020812) to 100% (BnaA09g50380D). The SWISS-MODEL workspace provided several quality parameters including QMEAN Z-scores and QMEANDisCo Global (Table 4.2). The QMEAN score ranged from -5.4 to -2.66. The QMEANDisCo Global score ranged from 0.5 to 0.77. Ramachandran plot showed that residues in core region ranges from 83.8 to 90% and 0.1 to 1.6% residues are in the disallowed region, indicating acceptable model quality. ERRAT value ranged from 85.08 to 95.27 which confirms the acceptable quality of the modeled structures (Table 4.2). The ProSA Z-score ranged from -6.32 to -10.21. The Z-score of the template is -8.5. Thus, the modeled structures showed almost similar Z-scores compared to the template. The three-dimensional structures of GLRs of *B. oleracea* and *B. rapa* are shown in Figure 4.4 and Figure 4.5, respectively. Of the 46 modeled GLRs of *B. napus*, 20 belong to A subgenome (Figure 4.6) and 26 belong to C subgenome (Figure 4.7). Sequence identity They all share the typical structures of GLRs, but some structural differences were noted. These differences were mainly prominent in the ATD region while LBD and

TMD were found to be structurally more conserved. It is noteworthy that in some GLRs, ATD structures seemed to be slightly incomplete (Bra040760, BnaC07g43470D, BnaC09g02400D and BnaCnng10410D). However, we have not discarded them as they can also be studied in detail in future studies which may open new avenues of the understanding of structure-function relationship of plant GLRs.

Surface electrostatic analysis with PIPSA showed the relationship of different GLRs based on surface electrostatic potential. The global electrostatic potential in every GLR structure was compared to each other by calculating the individual electrostatic distances, and the numeric matrix was used for the construction of a heatmap (Figure 4.8). The analysis clearly showed two clusters based on the electrostatic potential although the clusters were not species specific.

Table 4.2: Evaluation of modeled structures of GLRs

Gene ID	Sequence Similarity	Sequence Coverage	QMEAN	c β	All atom	Solvation	Torsion	QMEAN DisCoGlobal	Ramachandran plot				ERRAT	ProSA
									Core	Allow	Gener	Disall		
BnaA01g33670D	0.34	0.93	-4.46	-2.97	-1.74	-0.61	-3.65	0.57	85.50%	12.40%	1.10%	1.00%	91.41	-8.10
BnaA02g00890D	0.37	0.91	-3.49	-2.36	-1.98	0.15	-3.03	0.65	86.80%	11.80%	1.00%	0.40%	88.46	-8.84
BnaA02g30120D	0.35	0.88	-3.78	-3.08	-1.23	-0.22	-3.11	0.62	87.70%	10.30%	1.30%	0.70%	93.11	-9.10
BnaA03g14950D	0.53	0.94	-3.12	-1.32	-1.26	-0.65	-2.6	0.73	87.40%	11.30%	0.90%	0.40%	91.28	-9.65
BnaA03g14960D	0.42	0.95	-3.05	-0.96	-1.21	0.11	-2.83	0.69	88.70%	10.10%	0.90%	0.30%	91.14	-9.22
BnaA04g14620D	0.37	0.87	-4.17	-2.58	-1.91	-1.08	-3.27	0.61	86.50%	11.60%	1.50%	0.40%	87.38	-9.10
BnaA04g16910D	0.38	0.86	-3.15	-2	-1.37	0.14	-2.77	0.65	85.60%	12.50%	1.10%	0.70%	90.97	-8.75
BnaA04g16920D	0.39	0.85	-3.42	-1.81	-1.53	-0.51	-2.85	0.62	85.50%	11.50%	1.60%	0.40%	91.64	-7.49
BnaA04g28170D	0.37	0.9	-4.07	-2.53	-1.61	-0.01	-3.52	0.65	86.60%	10.70%	2.20%	0.50%	93.00	-9.51
BnaA05g10840D	0.42	0.95	-3.58	-1.35	-0.85	-0.17	-3.21	0.69	87.90%	10.60%	1.00%	0.50%	94.92	-9.51
BnaA05g13030D	0.38	0.82	-4.08	-2.34	-2.24	-0.66	-3.34	0.62	85.50%	12.50%	1.50%	0.40%	90.29	-9.18
BnaA06g30230D	0.35	0.88	-3.9	-2.49	-1.31	0.05	-3.4	0.6	88.30%	10.00%	1.00%	0.70%	91.33	-8.94
BnaA07g02520D	0.45	0.96	-2.79	-1.56	-1.11	0.2	-2.51	0.71	89.00%	9.40%	1.40%	0.20%	92.68	-9.44
BnaA08g04580D	0.46	0.96	-3.2	-0.99	-1.12	-0.18	-2.88	0.7	88.40%	10.10%	1.10%	0.40%	95.10	-8.53
BnaA08g10690D	0.45	0.95	-3.16	-1.69	-1.16	0.1	-2.82	0.72	89.00%	9.50%	1.20%	0.20%	92.83	-9.13
BnaA09g03000D	0.35	0.89	-4.07	-3.31	-1.92	0.29	-3.3	0.6	88.20%	10.10%	1.00%	0.70%	88.40	-8.38
BnaA09g32030D	0.45	0.97	-2.66	-1.29	-1.12	-0.1	-2.33	0.71	89.30%	9.70%	0.90%	0.20%	91.46	-9.33
BnaA09g50380D	0.57	1	-3.74	-0.63	-1.05	-0.39	-3.4	0.75	88.80%	9.60%	1.20%	0.40%	93.50	-8.48
BnaA10g03230D	0.58	0.97	-3.19	-1.12	-1.01	0.35	-3.02	0.77	88.20%	10.70%	0.80%	0.30%	92.78	-9.24
BnaAnng26240D	0.37	0.81	-3.11	-2.26	-1.48	0.38	-2.75	0.65	87.40%	11.30%	1.00%	0.30%	88.74	-10.21
BnaC01g03260D	0.45	0.96	-2.66	-1.78	-1.07	0.24	-2.37	0.71	89.40%	9.20%	1.20%	0.20%	92.46	-9.35
BnaC01g40220D	0.34	0.92	-4.66	-3.01	-1.64	-0.3	-3.94	0.57	86.30%	11.30%	1.70%	0.70%	90.94	-7.85
BnaC02g38540D	0.35	0.91	-4.01	-3.18	-1.79	-0.68	-3.15	0.61	87.70%	10.40%	1.20%	0.70%	90.96	-7.94
BnaC02g38550D	0.35	0.87	-3.86	-3.09	-1.34	-0.3	-3.15	0.6	87.60%	10.20%	1.40%	0.80%	94.78	-9.50
BnaC02g38560D	0.35	0.88	-4.39	-2.52	-1.17	-0.5	-3.71	0.58	86.80%	10.90%	1.30%	1%	92.39	-9.22
BnaC04g12400D	0.42	0.95	-3.41	-1.42	-1.04	-0.12	-3.06	0.69	88.60%	10.10%	0.80%	0.50%	93.67	-9.44
BnaC04g15320D	0.38	0.82	-4.52	-2.51	-1.95	-1	-3.64	0.6	84.10%	14.00%	1.40%	0.50%	90.22	-9.19
BnaC04g3810D	0.36	0.94	-4.05	-2.48	-2.44	-0.94	-3.19	0.6	85.70%	12.30%	1.40%	0.60%	87.64	-8.00
BnaC04g37210D	0.37	0.88	-3.94	-2.32	-1.54	-0.84	-3.17	0.6	87.70%	10.40%	1.30%	0.60%	92.01	-8.05
BnaC04g42860D	0.42	0.94	-3.55	-1.25	-0.76	0.03	-3.27	0.68	88.20%	10.00%	1.30%	0.50%	93.30	-9.66
BnaC05g03240D	0.57	0.97	-3.2	-1.03	-1.07	0.32	-3.03	0.77	87.80%	11.10%	0.80%	0.30%	92.12	-9.41
BnaC06g36440D	0.37	0.87	-3.34	-2.2	-1.55	0.14	-2.91	0.64	87.10%	11.20%	1.20%	0.50%	88.54	-9.49
BnaC07g26470D	0.35	0.88	-4.05	-3.32	-1.6	-0.24	-3.31	0.59	87.60%	10.30%	1.40%	0.60%	94.65	-8.57
BnaC07g43470D	0.36	0.9	-4.03	-2.68	-1.87	-0.84	-3.19	0.57	86.40%	12.00%	0.80%	0.80%	89.84	-7.05
BnaC08g22850D	0.45	0.97	-2.99	-1.51	-0.91	-0.05	-2.65	0.7	89.40%	8.90%	1.60%	0.10%	92.69	-9.67
BnaC08g44070D	0.56	0.98	-3.14	-0.65	-1.03	0.09	-2.96	0.75	88.70%	9.90%	0.90%	0.50%	94.12	-9.45
BnaC09g02400D	0.35	0.86	-4.22	-3.35	-2.2	-1	-3.21	0.53	87.30%	11.20%	1.00%	0.60%	87.87	-6.78
BnaC09g02410D	0.35	0.89	-4.07	-3.25	-1.44	-0.3	-3.33	0.61	87.20%	10.70%	1.20%	0.90%	92.79	-8.83
BnaCnng10410D	0.34	0.91	-5.4	-3.97	-2.13	-1.46	-4.12	0.5	85.90%	11.50%	1.70%	0.90%	88.23	-6.32
BnaCnng20160D	0.37	0.92	-3.34	-1.93	-1.55	0.11	-2.94	0.64	87.50%	11.20%	0.80%	0.50%	90.61	-9.02
BnaCnng22760D	0.45	0.96	-2.84	-1.5	-1.07	0.13	-2.55	0.71	88.80%	9.60%	1.40%	0.20%	92.30	-9.31
BnaCnng31230D	0.36	0.88	-4.77	-2.51	-1.57	-0.49	-4.06	0.6	83.80%	12.60%	2.00%	1.60%	89.70	-9.11
BnaCnng32830D	0.39	0.84	-3.15	-2.44	-1.73	-0.12	-2.61	0.62	86.60%	12.10%	0.90%	0.40%	87.65	-6.82
BnaCnng48020D	0.45	0.96	-2.85	-1.46	-0.72	0.05	-2.55	0.72	90.00%	8.60%	1.00%	0.40%	95.27	-9.90
BnaCnng51950D	0.42	0.94	-3.4	-0.91	-0.86	-0.16	-3.11	0.69	89.20%	9.10%	1.30%	0.30%	94.21	-9.41
BnaCnng66430D	0.39	0.84	-3.09	-1.81	-1.84	-0.57	-2.49	0.62	85.90%	12.90%	0.70%	0.50%	89.98	-7.28

Table continued -

Table continued -

Gene ID	Sequence Similarity	Sequence Coverage	QMEAN	c β	All atom	Solvation	Torsion	QMEAN DisCo Global	Ramachandran plot				ERRAT	ProSA
									Core	Allow	Gener	Disall		
Bol008199	0.42	0.94	-3.4	-0.91	0.86	-0.16	-3.11	0.69	89.20%	9.10%	1.30%	0.30%	94.21	-9.41
Bol008200	0.55	0.95	-2.77	-0.96	-0.69	0.25	-2.62	0.74	89.30%	9.20%	1.00%	0.50%	93.58	-9.21
Bol017691	0.45	0.96	-3.02	-1.8	-1.26	0.16	-2.68	0.71	87.90%	10.70%	1.30%	0.10%	91.50	-9.12
Bol029085	0.44	0.96	-3	-1.57	-0.71	-0.58	-2.49	0.69	89.50%	8.90%	1.20%	0.30%	94.92	-8.53
Bol033482	0.46	0.94	-2.94	-0.62	-0.89	0.15	-2.79	0.71	89.00%	9.40%	1.00%	0.50%	93.57	-10.05
Bol034616	0.45	0.98	-2.89	-1.2	-1.01	-0.18	-2.54	0.7	88.90%	9.60%	1.30%	0.30%	92.71	-8.77
Bol040960	0.57	0.97	-3.19	-0.93	-0.91	0.34	-3.05	0.77	88.20%	10.50%	1.10%	0.20%	92.88	-9.45
Bra005589	0.54	0.97	-3.08	-1.22	-1.25	0.26	-2.86	0.75	87.90%	11.00%	0.80%	0.30%	92.29	-9.42
Bra005591	0.42	0.95	-3.6	-1.69	-0.87	-0.19	-3.17	0.69	88.60%	9.90%	1.10%	0.40%	95.22	-9.49
Bra011603	0.45	0.96	-2.97	-1.88	-1.19	0.05	-2.59	0.75	89.10%	9.50%	1.20%	0.20%	92.15	-9.21
Bra015398	0.58	0.97	-3.14	-0.87	-0.87	0.29	-3	0.77	88.10%	10.70%	0.90%	0.20%	93.02	-9.30
Bra015817	0.29	0.81	-3.11	-2.26	-1.48	0.38	-2.75	0.65	87.40%	11.30%	1.00%	0.30%	88.74	-10.21
Bra018409	0.3	0.87	-4.01	-2.33	-1.38	-0.25	-3.43	0.63	86.40%	11.80%	1.30%	0.60%	92.55	-9.38
Bra020702	0.35	0.91	-4.15	-2.31	-1.18	-0.77	-3.42	0.58	87.90%	10.10%	1.10%	0.80%	92.65	-9.74
Bra020703	0.35	0.47	-3.95	-3.28	-1.61	-0.51	-3.14	0.59	88.40%	9.90%	1.00%	0.80%	92.82	-8.91
Bra020812	0.45	0.72	-3.22	-1.39	-1	-0.12	-2.86	0.66	89.10%	9.80%	1.00%	0.20%	94.00	-7.19
Bra022887	0.55	0.48	-2.79	-0.83	-0.57	0.2	-2.64	0.73	89.40%	9.00%	1.00%	0.60%	93.79	-9.14
Bra023313	0.37	0.92	-3.39	-1.63	-1.09	0.18	-3.08	0.64	87.70%	10.40%	1.30%	0.60%	91.88	-8.92
Bra032063	0.37	0.88	-3.91	-2.22	-1.56	0.19	-3.48	0.65	87.30%	10.70%	1.50%	0.50%	89.81	-9.72
Bra032072	0.37	0.91	-3.6	-2.44	-2.07	0.14	-3.11	0.67	86.80%	12.20%	0.90%	0.20%	86.94	-9.43
Bra032480	0.56	0.96	-3.38	-0.67	-1.14	-0.11	-3.13	0.74	87.40%	11.00%	0.90%	0.80%	93.73	-9.24
Bra034931	0.46	0.94	-2.85	-0.7	-0.65	0.14	-2.7	0.71	89.00%	9.30%	1.10%	0.60%	94.32	-9.84
Bra036174	0.35	0.89	-4.09	-3.21	-1.76	-0.25	-3.36	0.6	87.80%	10.20%	1.00%	1.00%	90.56	-8.50
Bra036175	0.52	0.89	-3.92	-3	-1.66	-0.66	-3.1	0.62	87.40%	11.20%	1.10%	0.30%	90.72	-8.68
Bra036795	0.45	0.97	-2.66	-1.29	-1.12	-0.1	-2.33	0.71	89.30%	9.70%	0.90%	0.20%	91.46	-9.33
Bra040147	0.34	0.93	-4.46	-3.02	-1.79	-0.33	-3.73	0.58	85.60%	12.50%	1.10%	0.90%	91.47	-8.13
Bra040760	0.38	0.85	-3.49	-2.46	-2.18	-0.7	-2.73	0.61	87.20%	11.20%	1.20%	0.50%	85.08	-6.87
Bra040761	0.38	0.85	-3.13	-1.98	-1.65	0.07	-2.72	0.64	85.80%	12.60%	1.00%	0.50%	89.00	-8.56

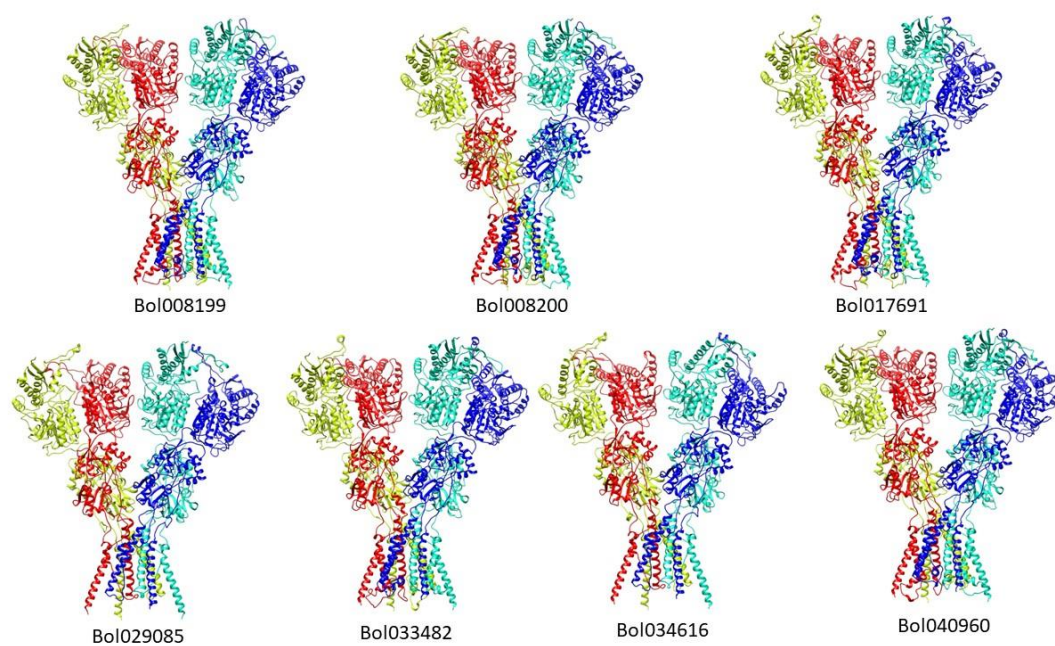


Figure 4.4: Predicted homotetrameric three-dimensional structures of the GLR proteins from *B. oleracea*. The four chains are coloured in yellow, red, cyan and blue.

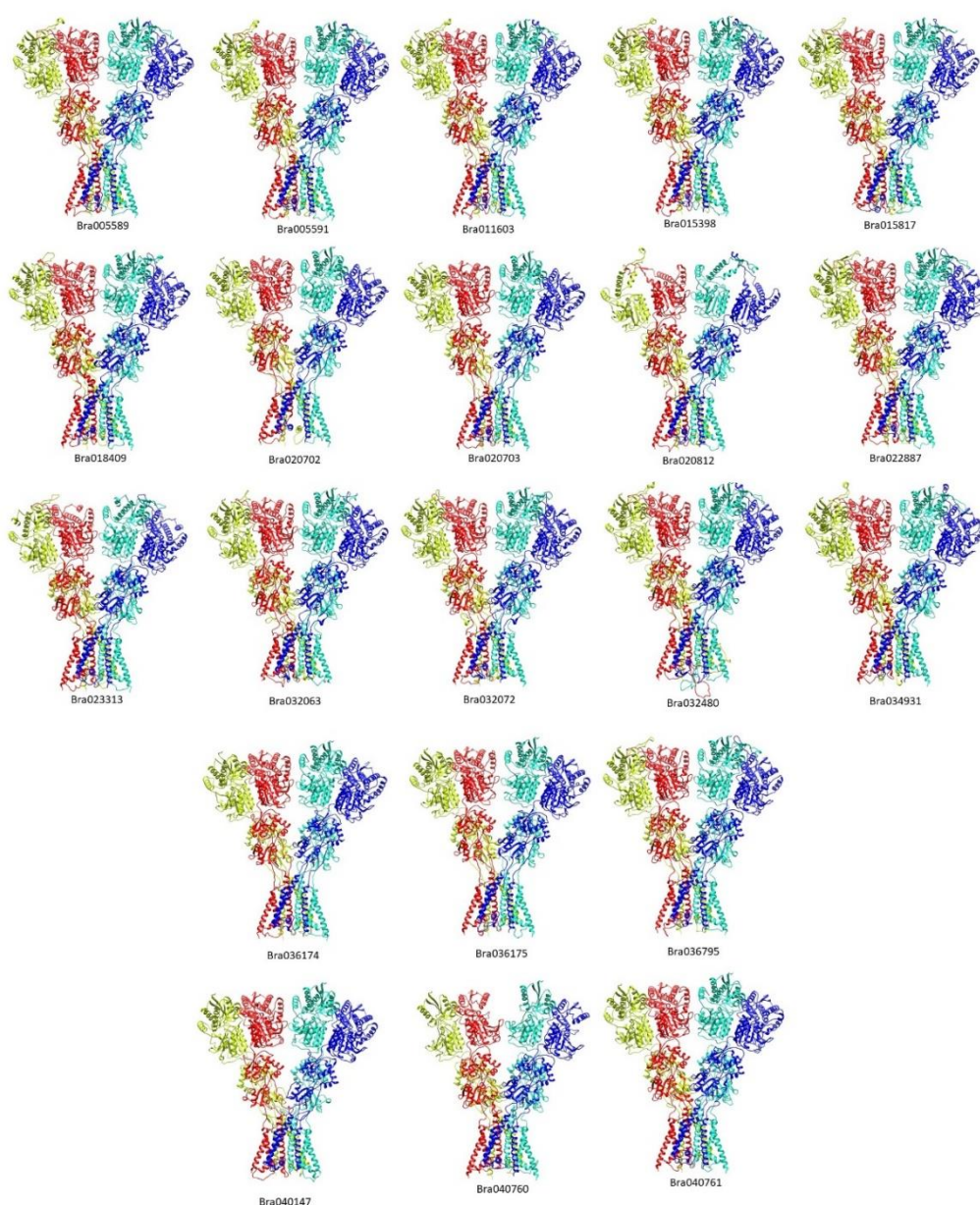


Figure 4.5: Predicted homotetrameric three-dimensional structures of the GLR proteins from *B. rapa*. The four chains are coloured in yellow, red, cyan and blue.

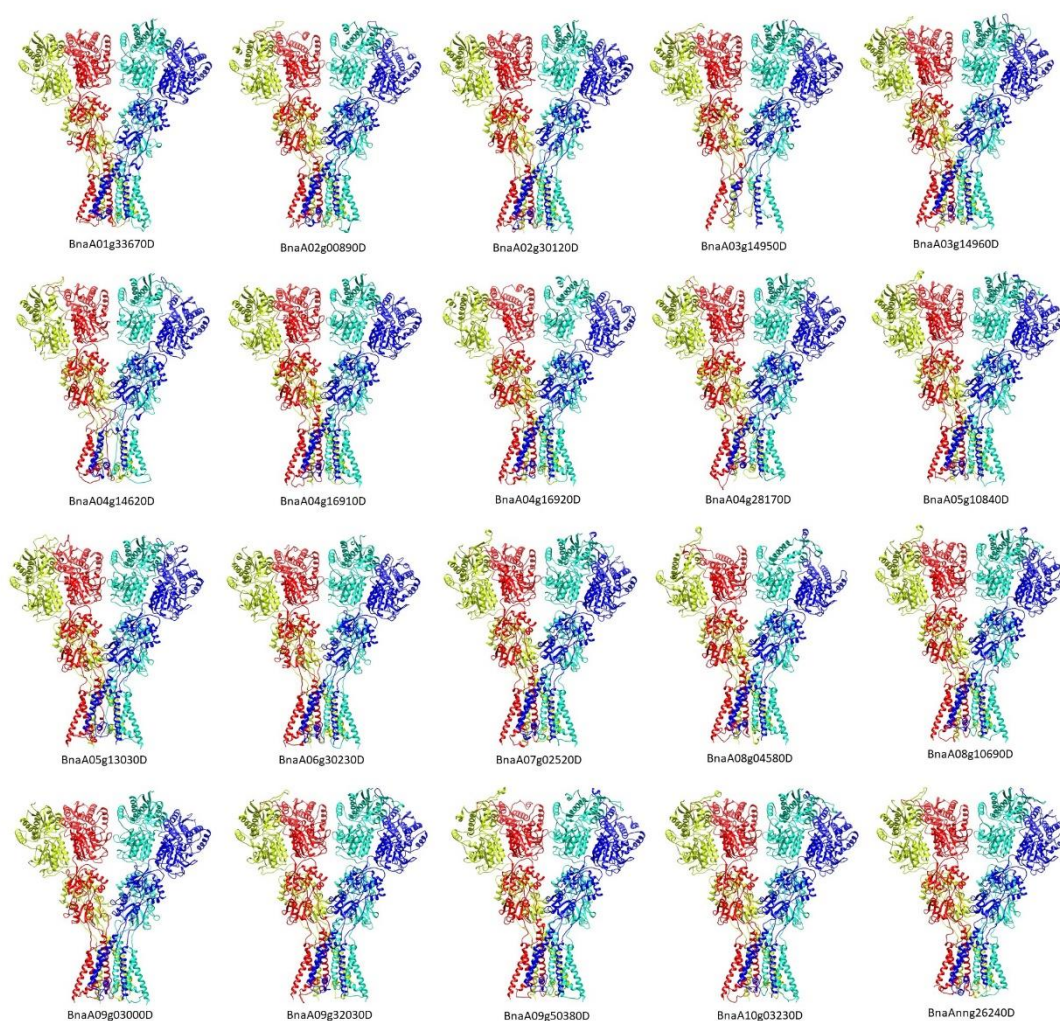


Figure 4.6: Predicted homotetrameric three-dimensional structures of the GLR proteins from A subgenome of *B. napus*. The four chains are coloured in yellow, red, cyan and blue.

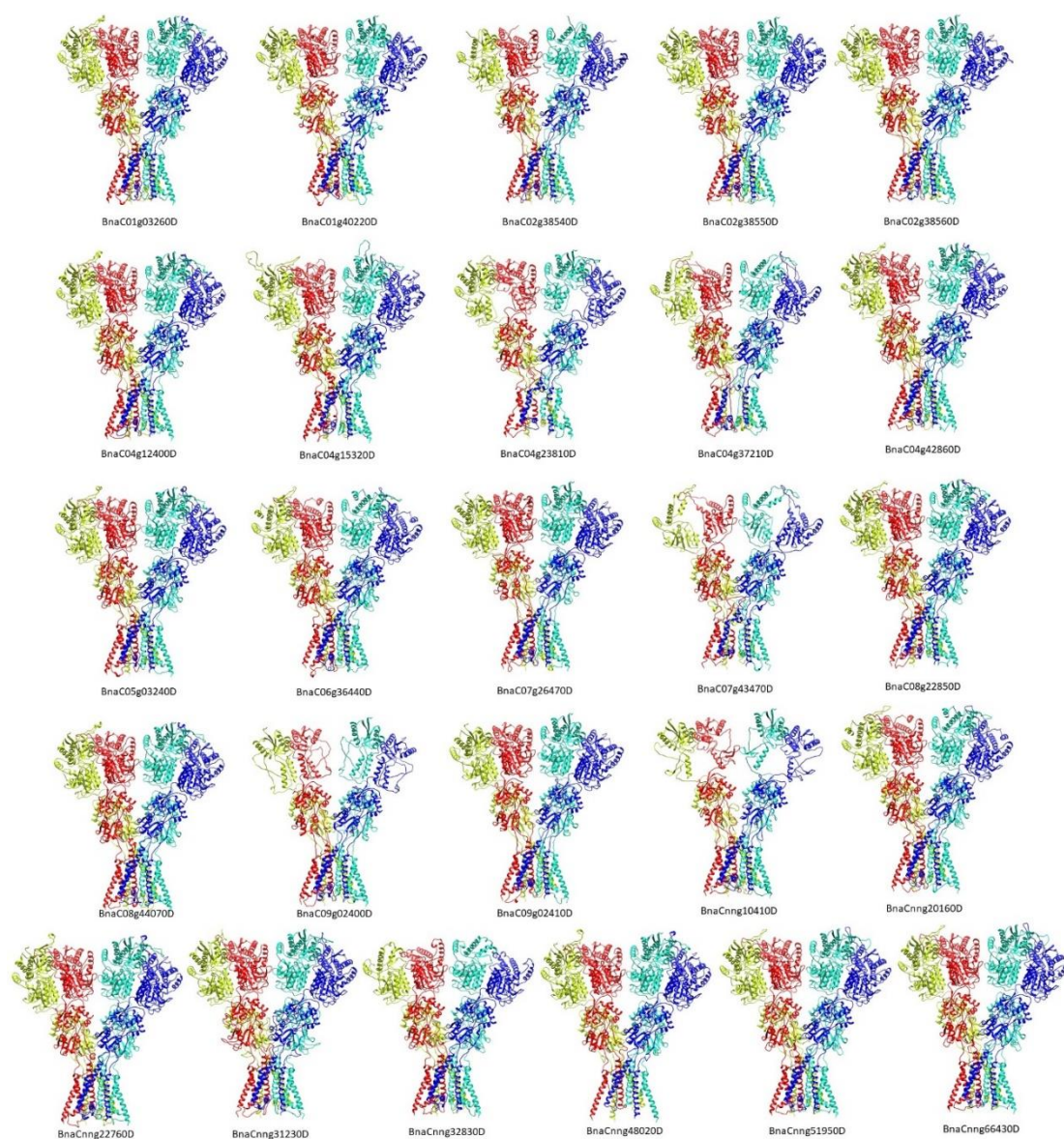


Figure 4.7: Predicted homotetrameric three-dimensional structures of the GLR proteins from C subgenome of *B. napus*. The four chains are coloured in yellow, red, cyan and blue.

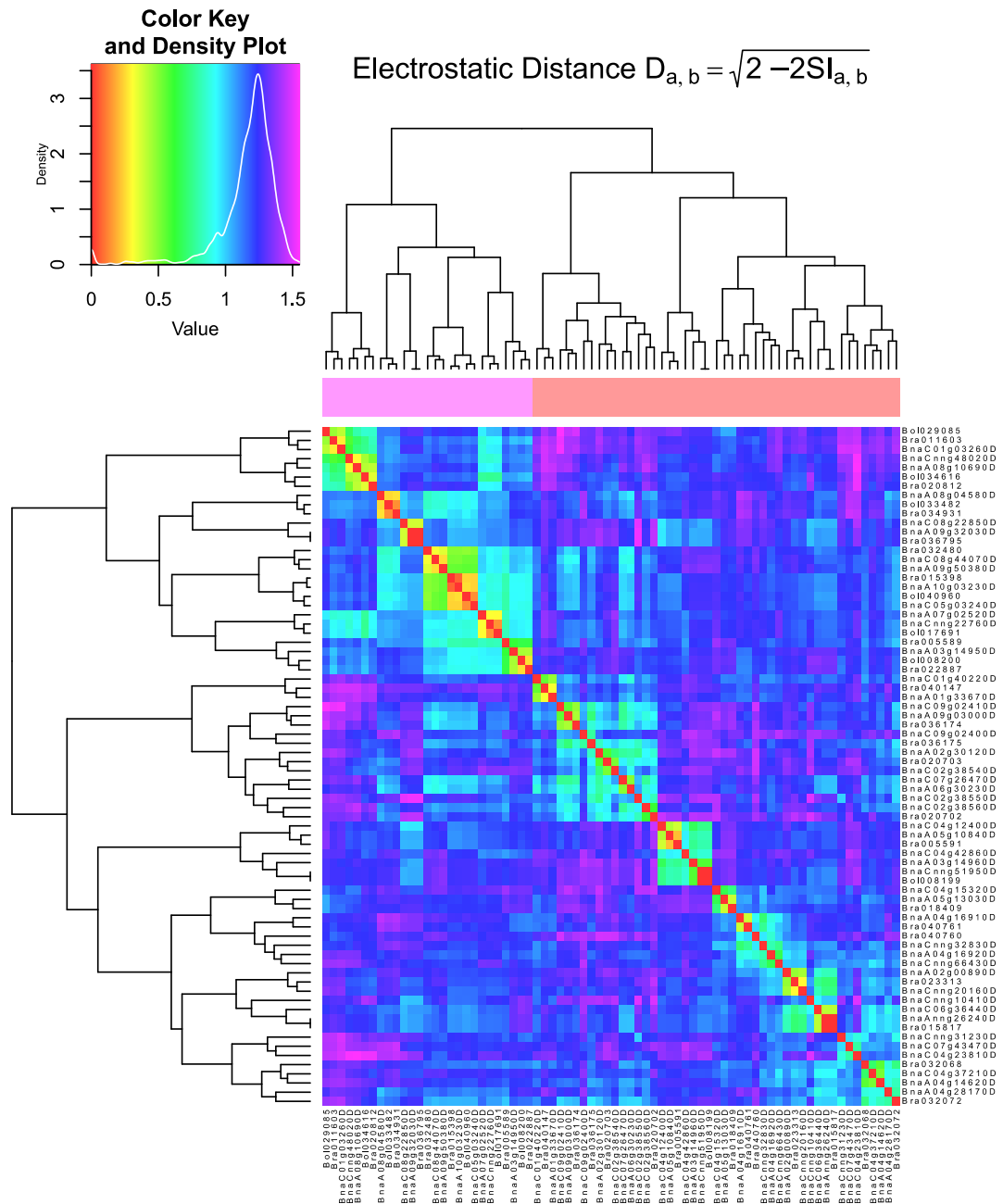


Figure 4.8: Epogram showing surface electrostatic potential of GLRs of three species of *Brassica* with PIPSA analysis.

4.4 Discussion

Plant GLRs are gaining importance as many new members are being discovered from several plant species (De Bortoli et al. 2016; Grenzi et al. 2022). However, information regarding their three-dimensional structures is scarce. However, their architecture and stoichiometry are thought to be similar to those of animal iGluRs as they are homologs of each other (Lam et al. 1998). Thus, presumably the plant GLRs are homo- or hetero-tetrameric membrane proteins (Wudick et al. 2018a). The four subunits are consisted of ATD (amino-terminal domain), LBD (ligand-binding domain), TMD (transmembrane domains) and CTD (carboxy-terminal domain). Crystal structures of a GLR from *Arabidopsis thaliana* (GLR 3.4) has been studied in detail recently and it showed all these features. Besides this, the LBD structure of other two GLRs (GLR3.2 and GLR3.3) from *A. thaliana* have also been determined and provided the first glimpse of the structural features of LBDs of GLRs and their ligand-binding features (Alfieri et al. 2020; Gangwar et al. 2021). However, structural features of other GLRs from different plant species is needed for better understanding of the diverse role of this group of proteins. For providing a preliminary idea of structural features of GLRs of economically important species of the genus *Brassica*, we have modeled several *Brassica* GLRs by homology modeling approach. We have also studied primary and secondary structural features of these proteins. The computed isoelectric point (pi) is useful for developing buffer systems for protein purification by isoelectric focusing (Sahay et al. 2020). Majority of the GLRs from all the three species was found to be basic. The instability index of a protein is a measure of its stability in a test tube and a value above 40 indicates its unstable nature (Guruprasad et al. 1990; Sahay et al. 2020). The majority of the *Brassica* GLRs showed that they are stable. The Grand Average Hydropathy (GRAVY) of a protein is calculated as the sum of hydropathy values of all amino acids, divided by the number of residues in the sequence and a negative value of GRAVY indicates hydrophilic nature (Wang et al. 2021). Although GLRs are membrane proteins, they have shown both positive and negative GRAVY scores. Kyte and Doolittle (1982) proposed that GRAVY score of membrane bound proteins are higher than the mean value of sequenced soluble proteins (-0.4). Marmagne et al. (2004) reported that The GRAVY score of the *Arabidopsis* plasma membrane proteins analyzed ranged from -1.6 to +1. Almost half of the GLRs of *A. thaliana* showed positive GRAVY scores (Chapter II). However, in the present study, majority of the

GLRs of *Brassica* species showed negative GRAVY scores. The negative values indicate the possibility of better interaction with water for those GLRs. Regarding secondary structures, random coil and alpha helices are predominant in the *Brassica* GLRs. As membrane proteins, presence of alpha helices is expected. Alpha helices also give thermostability to the proteins (Kabiraj et al. 2023). On the other hand, random coils give flexibility to a protein (Dutta et al. 2021; Kabiraj et al. 2023). As plant GLRs has a broader ligand specificity than animal iGluRs (Vincill et al. 2012; Tapken et al. 2013; Forde and Roberts 2014), presence of random coils may favour the promiscuity needed to accommodate different ligands. However, further study is needed to gather more information regarding this aspect. We have used homology modeling approach as it is known that two homologous proteins with similar sequences possess similar structures (Babic and Kovacic 2021). However, in the present study, the sequence similarity between the targets and the template was often found to be low (29-58%) despite the fact that the coverage was good in majority of cases. As structure of a protein is more conserved than the sequence (Illergård et al. 2009), structural modeling is a reliable approach to detect homology between proteins with low sequence similarity (<20%) (Kelley et al. 2015; Babic and Kovacic 2021). Thus, homology modeling of GLRs of *Brassica* species can be a starting point of analysis of structural features of this group of proteins. The *Brassica* GLRs showed typical GLR structures composed of TMD, LBD and ATD. The structures are similar in majority of the GLRs. However, ATD structures of a few GLRs seemed to be slightly different. These should be studied in detailed in future studies as this structural difference can be linked to novel functions. The surface electrostatic potential based epogram broadly showed two groups but these groups are not species specific. Thus, it may be result of functional diversity of GLRs.

4.5 Conclusion

For better understanding of the structural properties of GLR proteins from three species of *Brassica*, several computational tools were used to analyze the physico-chemical properties, secondary structure and to generate three-dimensional structure by homology modeling. This study, for the first time, analyzed the structural aspects of these proteins and also showed that these proteins are structurally conserved. The results obtained this study will serve as the foundation of detailed structural analysis for any future research.

Summary & Conclusions

Summary & Conclusions

Summary of the research work carried out in the present study entitled
**“BIOINFORMATIC ANALYSIS OF SOME PLANT GLUTAMATE
 RECEPTOR LIKE GENES FROM FAMILY BRASSICACEAE”**

Glutamate Receptors are synaptic and non-synaptic receptors situated primarily on the membranes of neuronal and glial cells. Glutamate is abundant in the human body and functions as a major excitatory neurotransmitter. It activates two receptor families *i.e.* metabotropic glutamate receptor proteins (mGluR) and ionotropic glutamate receptors (iGluR). The iGluRs are glutamate-gated ion channels involved in the movement of Na^+ and Ca^{2+} ions across the post-synaptic plasma membrane. The discovery of a family of glutamate receptor-like (GLR) genes in the model plant *Arabidopsis thaliana* (Brassicaceae) was the first indication that glutamate receptors were present in organisms lacking a nervous system. Glutamate is a ubiquitous plant cell ligand recognized by glutamate receptors located in the plasma membrane. Numerous studies have shown the glutamate usually exerts its signaling role by these receptors. *A. thaliana* possess 20 glutamate receptor-like genes (*AtGLRs*) in its genome which are involved in many functions including light signal transduction and calcium homeostasis. However, little is known about the physicochemical, functional and structural properties of *AtGLRs*. Moreover, study of GLRs from different crop species can lead to better understanding of their evolution, structure and function. This will, ultimately benefit the different crop improvement programmes. As GLR genes are known from *A. thaliana*, we have studied these genes in detail and also studied other GLR genes from some of the crop plants of the family Brassicaceae. We have studied GLRs from *Brassica napus*, *B. oleracea* and *B. rapa* for this study as sequenced genomes are available for these plants. Moreover, the fact that allotetraploid *Brassica napus* originated from the hybridization between the diploid *Brassica oleracea* and *Brassica rapa*, this study was carried out to study the effect of polyploidization on the evolution of the GLRs. Thus, in this study, glutamate receptor-like genes of *A. thaliana* and three *Brassica* species have been studied *in silico*. In *A. thaliana*, exon-intron structure revealed a common origin for most of these genes. The presence of several phosphorylation and myristoylation sites indicate the involvement of *AtGLRs* in various signaling processes. Gene ontology analysis revealed that *AtGLRs* are involved in

various biological processes, including different stress responses. In two genes namely AT2G17260 and AT4G35290, presence of RAV1-A binding site motif in the promoter coupled with results from gene ontology annotation indicate their role in stomatal movement through abscisic acid signaling. Expression analysis of these genes showed differential expression of several tandemly arranged genes which indicates neo or sub-functionalization. Two genes namely AT5G48400 and AT5G48410 showed significantly more expression in response to *Botrytis cinerea* infection. Five of these genes have shown G-protein-coupled γ -aminobutyric acid (GABA) receptor activity indicating a possible interaction between AtGLRs and GABA. Structurally, all of them were similar while differences were found regarding electrostatic surfaces as well as surface hydrophobicity. Results of this study provide a comprehensive reference regarding AtGLRs for further analysis regarding the structure, function, and evolution of the glutamate receptors in plants.

Till now, a comprehensive genome-wide analysis of the *GLR* gene family members in different economically important species of *Brassica* is missing. Considering the origin of allotetraploid *B. napus* from the hybridization between the diploid *B. oleracea* and *B. rapa*, we have identified 11, 27 and 65 *GLR* genes in *B. oleracea*, *B. rapa* and *B. napus*, respectively showing expansion of this gene family in *B. napus*. Chromosomal locations revealed several tandemly duplicated *GLR* genes in all the three species. Moreover, the gene family expanded in *B. napus* after allopolyploidization. The phylogenetic analysis showed that the 103 GLRs are classified into three main groups. The exon-intron structures of these genes are not very conserved and showed wide variation in intron numbers. However, protein sequences are much conserved as shown by the presence of ten short amino acid sequence motifs. Predicted *cis*-acting elements in 1 kb promoters of *GLR* genes are mainly involved in light, stress and hormone responses. Expression analysis using RNA-seq data showed that in *B. oleracea* and *B. rapa*, some GLRs are more tissue specific than others. In *B. napus*, some GLRs are downregulated under cold stress, while others are upregulated. Some GLR genes have also been found to be controlled by microRNAs in the three species of *Brassica*. Majority of these miRNAs control the expression of the target genes by cleavage and some by translation. This data may help in improving these *Brassica* crop species by miRNA-based gene regulation. In summary, this bioinformatic study of the *GLR* gene family of the three *Brassica* species provides evidence for the expansion of this gene

family in *B. napus* and also provided useful information for in-depth studies of their biological functions in *Brassica*. Different physio-chemical Properties of *Brassica* GLRs has also been Studied. Majority of the GLRs from all the three species of *Brassica* were found to be basic. Although almost half of the GLRs of *A. thaliana* showed positive GRAVY (Grand average hydropathy) scores, majority of the *Brassica* GLRs showed negative GRAVY score. This indicates the possibility of better interaction with water for those GLRs. *Brassica* GLRs comprised of mainly random coils and alpha helices. Plant GLRs are known for accepting several molecules as ligands. As random coils confer flexibility in proteins, presence of random coils in the GLRs may favour the multiple ligand binding. Study of three-dimensional structure by homology modeling approach of several GLRs of *Brassica* has also been studied. The structure showed typical plant GLR structure and composed of transmembrane domain (TMD), ligand binding domain (LBD) and amino terminal domain (ATD). The TMD region is structurally more conserved. However, the ATD region showed some differences among the different GLRs. The surface electrostatic potential of different *Brassica* GLRs has been studied. Based on the surface electrostatic potential, *Brassica* GLRs has been divided into two groups, but this groups were not species specific.

References

References:

- Abdallah, M., Dubousset, L., Meuriot, F., Etienne, P., Avice, J. C., & Ourry, A. (2010). Effect of mineral sulphur availability on nitrogen and sulphur uptake and remobilization during the vegetative growth of *Brassica napus* L. *Journal of Experimental Botany*, 61(10), 2635–2646.
- Acher, F. C., & Bertrand, H.-O. (2005). Amino acid recognition by Venus flytrap domains is encoded in an 8-residue motif. *Peptide Science: Original Research on Biomolecules*, 80(2-3), 357–366.
- Ahuja, I., Rohloff, J., & Bones, A. M. (2011). Defence mechanisms of *Brassicaceae*: Implications for plant-insect interactions and potential for integrated pest management. *Sustainable Agriculture Volume 2*, 623–670.
- Airoidi, C. A., & Davies, B. (2012). Gene duplication and the evolution of plant MADS-box transcription factors. *Journal of Genetics and Genomics*, 39(4), 157–165.
- Alfieri, A., Doccia, F. G., Pederzoli, R., Grenzi, M., Bonza, M. C., Luoni, L., Candeo, A., Romano Armada, N., Barbiroli, A., & Valentini, G. (2020). The structural bases for agonist diversity in an *Arabidopsis thaliana* glutamate receptor-like channel. *Proceedings of the National Academy of Sciences*, 117(1), 752–760.
- Allan, A. C., Hellens, R. P., & Laing, W. A. (2008). MYB transcription factors that colour our fruit. *Trends in Plant Science*, 13(3), 99–102.
- Allender, C. J., & King, G. J. (2010). Origins of the amphiploid species *Brassica napus* L. investigated by chloroplast and nuclear molecular markers. *BMC Plant Biology* 10, 54.
- Al-Shehbaz, & Ihsan, A. (1973). The biosystematics of the genus *Thelypodium* (Cruciferae). *Contributions from the Gray Herbarium of Harvard University*, 204, 3–148.
- Al-Shehbaz, I. A., O’Kane Jr, S. L., & Price, R. A. (1999). Generic placement of species excluded from *Arabidopsis* (Brassicaceae). *Novon*, 9(3), 296–307.
- Altschul, S. F., Boguski, M. S., Gish, W., & Wootton, J. C. (1994). Issues in searching molecular sequence databases. *Nature Genetics*, 6, 119–129.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410.
- Aouini, A., Matsukura, C., Ezura, H., & Asamizu, E. (2012). Characterisation of 13 glutamate receptor-like genes encoded in the tomato genome by structure, phylogeny and expression profiles. *Gene*, 493(1), 36–43.
- Appel, O., & Al-Shehbaz, I. A. (2003). Cruciferae. In: Kubitzki, K. & Bayer, C. (Eds.) *Flowering Plants: Dicotyledons: Malvales, Capparales and Non-betain Caryophyllales* (pp. 75–174). Springer Berlin, Heidelberg.

- Ashkenazy, H., Erez, E., Martz, E., Pupko, T., & Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Research*, 38(suppl 2), W529–W533.
- Asthir, B. (2015). Mechanisms of heat tolerance in crop plants. *Biologia Plantarum*, 59, 620–628.
- Ayalon, G., & Stern-Bach, Y. (2001). Functional Assembly of AMPA and Kainate Receptors Is Mediated by Several Discrete Protein-Protein Interactions. *Neuron*, 31(1), 103–113.
- Babic, N., & Kovacic, F. (2021). Predicting drug targets by homology modelling of *Pseudomonas aeruginosa* proteins of unknown function. *PLoS ONE*, 16(10), e0258385. <https://doi.org/10.1371/journal.pone.0258385>
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., & Nobel, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* 37, W202–W208. <https://doi.org/10.1093/nar/gkp335>
- Baker, N. A., Sept, D., Joseph, S., Holst, M. J., & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, 98(18), 10037–10041.
- Beilstein, M. A., Nagalingum, N. S., Clements, M. D., Manchester, S. R., & Mathews, S. (2010). Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, 107(43), 18724–18728. <http://doi.org/10.1073/pnas.0909766107>
- Benkert P, Biasini M, & Schwede T (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics Oxford Academic*, England, 27,343–350. <https://doi.org/10.1093/bioinformatics/btq662>
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, L. N., & Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, 28(1), 235–242.
- Blanc, G., & Wolfe, K. H. (2004). Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *The Plant Cell*, 16(7), 1679–1691.
- Blomberg, N., Gabdoulline, R. R., Nilges, M., & Wade, R. C. (1999). Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. *Proteins: Structure, Function, and Bioinformatics*, 37(3), 379–387.
- Bortoli, S. D., Teardo, E., Szabò, I., Morosinotto, T., & Alboresi, A. (2016). Evolutionary insight into the ionotropic glutamate receptor superfamily of photosynthetic organisms. *Biophysical Chemistry*, 218, 14–26. <https://doi.org/10.1016/j.bpc.2016.07.004>
- Botti, S. A., Felder, C. E., Sussman, J. L., & Silman, I. (1998). Electrotactins: a class of adhesion proteins with conserved electrostatic and structural motifs. *Protein Engineering*, 11(6), 415–420.
- Bouché, N., & Fromm, H. (2004). GABA in plants: just a metabolite? *Trends in Plant Science*, 9(3), 110–115.

- Bowers, J. E., Chapman, B. A., Rong, J., & Paterson, A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, 422, 433–438.
- Brassai, A., Suvanjev, R.-G., Bán, E.-G., & Lakatos, M. (2015). Role of synaptic and nonsynaptic glutamate receptors in ischaemia induced neurotoxicity. *Brain Research Bulletin*, Elsevier, 112, 1–6.
- Brenner, E. D., Martinez-Barboza, N., Clark, A. P., Liang, Q. S., Stevenson, D. W., & Coruzzi, G. M. (2000). *Arabidopsis* mutants resistant to S (+)- β -methyl- α , β -diaminopropionic acid, a cycad-derived glutamate receptor agonist. *Plant Physiology*, 124(4), 1615–1624.
- Brosnan, J. T., & Brosnan, M. E. (2013). Glutamate: A truly functional amino acid. *Amino Acids*, 45, 413–418.
- Buell, C. R., & Last, R. L. (2010). Twenty-first century plant biology: Impacts of the *Arabidopsis* genome on plant biology and agriculture. *Plant Physiology*, 154(2), 497–500.
- Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics*, 129(3), 897–907.
- Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., Pupko, T., & Ben-Tal, N. (2013). ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Israel Journal of Chemistry*, 53(3–4), 199–206. <http://doi.org/10.1002/ijch.201200096>
- Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., & Samans, B. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*, 345(6199), 950–953.
- Chattopadhyay, S., Ang, L.-H., Puente, P., Deng, X.-W., & Wei, N. (1998). *Arabidopsis* bZIP protein HY5 directly interacts with light-responsive promoters in mediating light control of gene expression. *The Plant Cell*, 10(5), 673–683.
- Chen C, Chen H, Zhang Y, Thomas, H. R., Frank, M. H., He, Y., & Xia, R. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant* 13, 1194–1202.
- Chen, G. Q., Cui, C., Mayer, M. L., & Gouaux, E. (1999). Functional characterization of a potassium-selective prokaryotic glutamate receptor. *Nature*, 402(6763), 817–821. <https://doi.org/10.1038/45568>
- Chen, H., Zhang, J., Neff, M. M., Hong, S.-W., Zhang, H., Deng, X.-W., & Xiong, L. (2008). Integration of light and abscisic acid signaling during seed germination and early seedling development. *Proceedings of the National Academy of Sciences*, 105(11), 4495–4500.
- Chen, W., Provart, N. J., Glazebrook, J., Katagiri, F., Chang, H.-S., Eulgem, T., Mauch, F., Luan, S., Zou, G., Whitham, S. A., Budworth, P. R., Tao, Y., Xie, Z., Chen, X., Lam, S., Kreps, J. A., Harper, J. F., Si-Ammour, A., Mauch-Mani, B., Heinlein, M., Kobayashi, K., Hohn, T., Dangl, J. L., Wang, X., & Zhu, T. (2002). Expression

- profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. *The Plant Cell*, 14(3), 559–574.
- Chen, Y.-H., Chao, Y.-Y., Hsu, Y. Y., Hong, C.-Y., & Kao, C. H. (2012). Heme oxygenase is involved in nitric oxide-and auxin-induced lateral root formation in rice. *Plant Cell Reports*, 31, 1085–1091.
- Cheng, H., Deng, W., Wang, Y., Ren, J., Liu, Z., & Xue, Y. (2014). dbPPT: a comprehensive database of protein phosphorylation in plants. *Database*, 2014, bau121.
- Cheng, Y., Zhang, X., Sun, T., Tian, Q., & Zhang, W.-H. (2018). Glutamate receptor homolog3. 4 is involved in regulation of seed germination under salt stress in *Arabidopsis*. *Plant and Cell Physiology*, 59(5), 978–988.
- Cheong, Y. H., Chang, H.-S., Gupta, R., Wang, X., Zhu, T., & Luan, S. (2002). Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in *Arabidopsis*. *Plant Physiology*, 129(2), 661–677.
- Chiu, J. C., Brenner, E. D., DeSalle, R., Nitabach, M. N., Holmes, T. C., & Coruzzi, G. M. (2002). Phylogenetic and expression analysis of the glutamate-receptor-like gene family in *Arabidopsis thaliana*. *Molecular Biology and Evolution*, 19(7), 1066–1082.
- Chiu, J., DeSalle, R., Lam, H.-M., Meisel, L., & Coruzzi, G. (1999). Molecular evolution of glutamate receptors: a primitive signaling mechanism that existed before plants and animals diverged. *Molecular Biology and Evolution*, 16(6), 826–838.
- Cho, D., Kim, S. A., Murata, Y., Lee, S., Jae, S.-K., Nam, H. G., & Kwak, J. M. (2009). De-regulated expression of the plant glutamate receptor homolog *AtGLR3.1* impairs long-term Ca^{2+} -programmed stomatal closure. *The Plant Journal*, 58(3), 437–449. <http://doi.org/10.1111/j.1365-313X.2009.03789.x>
- Choi, W.-G., Miller, G., Wallace, I., Harper, J., Mittler, R., & Gilroy, S. (2017). Orchestrating rapid long-distance signaling in plants with Ca^{2+} , ROS and electrical signals. *The Plant Journal*, 90(4), 698–707. <https://doi.org/10.1111/tpj.13492>
- Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science*, 2(9), 1511–1519. <https://doi.org/10.1002/pro.5560020916>.
- Cominelli, E., & Tonelli, C. (2009). A new role for plant R2R3-MYB transcription factors in cell cycle regulation. *Cell Research*, 19(11), 1231.
- Conesa, A., & Götz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics*, 2008, 619832. <https://doi.org/10.1155/2008/619832>
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18), 3674–3676. <http://doi.org/10.1093/bioinformatics/bti610>
- Dai, X., Zhao, & P. X. (2011). psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Research* 39, W155-159. <https://doi.org/10.1093/nar/gkr319>

- Dai, X., Zhuang, Z., & Zhao, P. X. (2011). Computational analysis of miRNA targets in plants: current status and challenges. *Briefings in Bioinformatics*, 12, 115–121. <https://doi.org/10.1093/bib/bbq065>
- Dai, X., Zhuang, Z., & Zhao, P. X. (2018). psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Research* 46, W49–W54. <https://doi.org/10.1093/nar/gky316>
- Dash, S., Van Hemert, J., Hong, L., Wise, R. P., & Dickerson, J. A. (2012). PLEXdb: gene expression resources for plants and plant pathogens. *Nucleic Acids Research*, 40(D1), D1194–D1201.
- Davenport, R. (2002). Glutamate receptors in plants. *Annals of Botany*, 90(5), 549–557.
- Davis, A. R., Pylatuik, J. D., Paradis, J. C., & Low, N. H. (1998). Nectar-carbohydrate production and composition vary in relation to nectary anatomy and location within individual flowers of several species of *Brassicaceae*. *Planta*, 205, 305–318.
- De Bortoli, S., Teardo, E., Szabò, I., Morosinotto, T., & Alboresi, A. (2016) Evolutionary insight into the ionotropic glutamate receptor superfamily of photosynthetic organisms. *Biophysical Chemistry*, 218, 14–26
- de Carvalho, M. D. C., & De Mesquita, J. F. (2013). Structural modeling and *in silico* analysis of human superoxide dismutase 2. *PLoS ONE*, 8(6), e65558. <https://doi.org/10.1371/journal.pone.0065558>
- De Vos, M., Van Oosten, V. R., Van Poecke, R. M., Van Pelt, J. A., Pozo, M. J., Mueller, M. J., Buchala, A. J., Metraux, Jean-Pierre., Van Loon, L. C., Dicke, M., & Pieterse, M. J. C. (2005). Signal signature and transcriptome changes of *Arabidopsis* during pathogen and insect attack. *Molecular Plant-Microbe Interactions*, 18(9), 923–937. <https://doi.org/10.1094/MPMI-18-0923>
- Deinlein, U., Stephan, A. B., Horie, T., Luo, W., Xu, G., & Schroeder, J. I. (2014). Plant salt-tolerance mechanisms. *Trends in Plant Science*, 19(6), 371–379.
- Demidchik, V., Davenport, R. J., & Tester, M. (2002). Nonselective cation channels in plants. *Annual Review of Plant Biology*, 53(1), 67–107.
- Demidchik, V., Essah, P. A., & Tester, M. (2004). Glutamate activates cation currents in the plasma membrane of *Arabidopsis* root cells. *Planta*, 219(1), 167–175. <http://doi.org/10.1007/s00425-004-1207-8>
- Dennison, K. L., & Spalding, E. P. (2000). Glutamate-gated calcium fluxes in *Arabidopsis*. *Plant Physiology*, 124(4), 1511–1514.
- Diaz, C., Purdy, S., Christ, A., Morot-Gaudry, J.-F., Wingler, A., & Masclaux-Daubresse, C. (2005). Characterization of markers to determine the extent and variability of leaf senescence in *Arabidopsis*. A metabolic profiling approach. *Plant Physiology*, 138(2), 898–908.
- Ding, Y., Shi, Y., & Yang, S. (2019). Advances and challenges in uncovering cold tolerance regulatory mechanisms in plants. *New Phytologist*, 222(4), 1690–1704.
- Dingledine, R., Borges, K., Bowie, D., & Traynelis, S. F. (1999). The glutamate receptor ion channels. *Pharmacological Reviews*, 51(1), 7–62.

- Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., & Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Research*, 32(suppl 2), W665–W667.
- Dong, J., Chen, C., & Chen, Z. (2003). Expression profiles of the *Arabidopsis* WRKY gene superfamily during plant defense response. *Plant Molecular Biology*, 51(1), 21–37.
- Dosztányi, Z., Csizmok, V., Tompa, P., & Simon, I. (2005a). The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *Journal of molecular biology*, 347(4), 827–839.
- Dosztányi, Z., Csizmok, V., Tompa, P., & Simon, I. (2005b). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, 21(16), 3433–3434.
- Dubos, C., Huggins, D., Grant, G. H., Knight, M. R., & Campbell, M. M. (2003). A role for glycine in the gating of plant NMDA-like receptors. *The Plant Journal*, 35(6), 800–810.
- Duermeyer, L., Khodapanahi, E., Yan, D., Krapp, A., Rothstein, S. J., & Nambara, E. (2018). Regulation of seed dormancy and germination by nitrate. *Seed Science Research*, 28(3), 150–157.
- Duret, L. (2002). Evolution of synonymous codon usage in metazoans. *Current Opinion in Genetics & Development*, 12(6), 640–649.
- Duret, L., & Mouchiroud, D. (1999). Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 96(8), 4482–4487.
- Dutta, B., Deska, J., Bandopadhyay, R., & Shamekh, S. (2021). In silico characterization of bacterial chitinase: illuminating its relationship with archaeal and eukaryotic cousins. *Journal of Genetic Engineering and Biotechnology*, 19, 19. <https://doi.org/10.1186/s43141-021-00121-6>
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797.
- Efremov, R. G., Chugunov, A. O., Pyrkov, T. V., Priestle, J. P., Arseniev, A. S., & Jacoby, E. (2007). Molecular lipophilicity in protein modeling and drug design. *Current Medicinal Chemistry*, 14(4), 393–415.
- 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., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research* 47(D1), D427–D432.
- Elhaik, E., & Tatarinova, T. (2012). GC3 Biology in Eukaryotes and Prokaryotes. In T. Tatarinova (Ed.), *DNA Methylation - From Genomics to Technology*. InTech. pp. 55–68. Retrieved from <http://www.intechopen.com/books/dna-methylation-from-genomics-to-technology/gc3-biology-in-eukaryotes-and-prokaryotes>
- Feng, C.-Z., Chen, Y., Wang, C., Kong, Y.-H., Wu, W.-H., & Chen, Y.-F. (2014). *Arabidopsis* RAV1 transcription factor, phosphorylated by SnRK2 kinases,

- regulates the expression of *ABI3*, *ABI4*, and *ABI5* during seed germination and early seedling development. *The Plant Journal*, 80(4), 654–668.
- Feng, S., Pan, C., Ding, S., Ma, Q., Hu, C., Wang, P., & Shi, K. (2021). The Glutamate Receptor Plays a Role in Defense against *Botrytis cinerea* through Electrical Signaling in Tomato. *Applied Sciences*, 11(23), Article 23. <https://doi.org/10.3390/app112311217>
- Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., & Bateman, A. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, 44(D1), D279–D285. <http://doi.org/10.1093/nar/gkv1344>
- Forde B. G., & Roberts M. R. (2014). Glutamate receptor-like channels in plants: a role as amino acid sensors in plant defence? *F1000prime Reports* 6, 37.
- Forde, B. G. (2014). Glutamate signalling in roots. *Journal of Experimental Botany*, 65(3), 779–787.
- Forde, B. G., & Lea, P. J. (2007). Glutamate in plants: metabolism, regulation, and signalling. *Journal of Experimental Botany*, 58(9), 2339–2358.
- Fowler, S., & Thomashow, M. F. (2002). *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *The Plant Cell*, 14(8), 1675–1690.
- Franzke, A., Lysak, M. A., Al-Shehbaz, I. A., Koch, M. A., & Mummenhoff, K. (2011). Cabbage family affairs: The evolutionary history of *Brassicaceae*. *Trends in Plant Science*, 16(2), 108–116.
- Friedrich, T., Faivre, L., Bäurle, I., & Schubert, D. (2019). Chromatin-based mechanisms of temperature memory in plants. *Plant, Cell & Environment*, 42(3), 762–770.
- Friedt, W., Tu, J., & Fu, T. (2018). Academic and Economic Importance of *Brassica napus* Rapeseed. In: Liu S, Snowdon R, Chalhoub B (Eds.) *The Brassica napus Genome*. Springer International Publishing, Cham, pp 1–20.
- Gabdoulline, R. R., Stein, M., & Wade, R. C. (2007). qPIPSA: relating enzymatic kinetic parameters and interaction fields. *BMC Bioinformatics*, 8(1), 373.
- Gangappa, S. N., & Chattopadhyay, S. (2013). MYC2 differentially regulates GATA-box containing promoters during seedling development in *Arabidopsis*. *Plant Signaling & Behavior*, 8(10), e25679.
- Gangwar, S. P., Green, M. N., Michard, E., Simon, A. A., Feijó, J. A., & Sobolevsky, A. I. (2021). Structure of the *Arabidopsis* glutamate receptor-like channel GLR3.2 ligand-binding domain. *Structure*, 29(2), 161–169. e4.
- Garcia-Servin, M. A., Mendoza-Sanchez, M., & Contreras-Medina, L. M. (2021). Electrical signals as an option of communication with plants: A review. *Theoretical and Experimental Plant Physiology*, 33(2), 125–139.
- Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). *Protein identification and analysis tools on the ExPASy server*. In: Walker,

- J. M. (Ed.) *The Proteomics Protocols Handbook*. Springer Protocols Handbooks. Humana Press. pp 571–607. <https://doi.org/10.1385/1-59259-890-0:571>
- Gene Ontology Consortium. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, 32(suppl 1), D258–D261.
- Geourjon, C., & Deleage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Computer Applications in the Biosciences: CABIOS*, 11(6), 681–684.
- Giuliano, G., Pichersky, E., Malik, V. S., Timko, M. P., Scolnik, P. A., & Cashmore, A. R. (1988). An evolutionarily conserved protein binding sequence upstream of a plant light-regulated gene. *Proceedings of the National Academy of Sciences*, 85(19), 7089–7093.
- Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., & Ben-Tal, N. (2003). ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. *Bioinformatics*, 19(1), 163–164.
- Glez-Peña, D., Gómez-Blanco, D., Reboiro-Jato, M., Fdez-Riverola, F., & Posada, D. (2010). ALTER: program-oriented conversion of DNA and protein alignments. *Nucleic Acids Research*, 38(suppl 2), W14–W18.
- Goetz, R. M., & Fuglsang, A. (2005). Correlation of codon bias measures with mRNA levels: analysis of transcriptome data from *Escherichia coli*. *Biochemical and Biophysical Research Communications*, 327(1), 4–7.
- Golicz, A. A., Bayer, P. E., Barker, G. C., Edger, P. P., Kim, H., Martinez, P. A., Chan, C. K. K., Severn-Ellis, A., McCombie, W. R., & Parkin, I. A. (2016). The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nature Communications*, 7(1), 13390.
- Gómez-Campo, C. (1980). Morphology and morpho-taxonomy of the tribe *Brassicaceae*. In: Tsunoda, S., Hinata, K., Gómez-Campo, C. (Eds.) *Brassica Crops and Wild Allies.[I]*, Japan Scientific Societies Press, Tokyo, pp: 3–31.
- Gong, M., Chen, S.-N., Song, Y.-Q., & Li, Z.-G. (1997). Effect of calcium and calmodulin on intrinsic heat tolerance in relation to antioxidant systems in *maize* seedlings. *Functional Plant Biology*, 24(3), 371–379.
- González-Pérez, S., Gutiérrez, J., García-García, F., Osuna, D., Dopazo, J., Lorenzo, Ó., Revuelta, J. L., & Arellano, J. B. (2011). Early transcriptional defense responses in *Arabidopsis* cell suspension culture under high-light conditions. *Plant Physiology*, 156(3), 1439–1456. <https://doi.org/10.1104/pp.111.177766>
- Good, A. C., Hodgkin, E. E., & Richards, W. G. (1992). Utilization of Gaussian functions for the rapid evaluation of molecular similarity. *Journal of the American Chemical Society*, ACS Publications, Information Computer Science 32, 188–191
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* 40, D1178–D1186.

- Green, M. N., Gangwar, S. P., Michard, E., Simon, A. A., Portes, M. T., Barbosa-Caro, J., Wudick, M. M., Lizzio, M. A., Klykov, O., Yelshanskaya, M. V., Feijó, J. A., & Sobolevsky, A. I. (2021). Structure of the *Arabidopsis thaliana* glutamate receptor-like channel GLR3.4. *Molecular Cell*, 81(15), 3216–3226.e8. <https://doi.org/10.1016/j.molcel.2021.05.025>
- Green, P. J., Kay, S. A., & Chua, N.-H. (1987). Sequence-specific interactions of a pea nuclear factor with light-responsive elements upstream of the *rbcS-3A* gene. *The EMBO Journal*, 6(9), 2543.
- Grenzi, M., Bonza M. C., Alfieri A, & Costa, A. (2021). Structural insights into long-distance signal transduction pathways mediated by plant glutamate receptor-like channels. *New Phytologist* 229, 1261–1267. <https://doi.org/10.1111/nph.17034>
- Grenzi, M., Bonza, M. C., & Costa, A. (2022). Signaling by plant glutamate receptor-like channels: What else! *Current Opinion in Plant Biology* 68, 102253. <https://doi.org/10.1016/j.pbi.2022.102253>
- Guex, N., & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis*, 18(15), 2714–2723.
- Guo, A. Y., Zhu, Q. H., Chen, X., & Luo, J. C. (2007). [GSDS: a gene structure display server]. *Yi Chuan= Hereditas/Zhongguo Yi Chuan Xue Hui Bian Ji*, 29(8), 1023–1026.
- Guruprasad, K., Reddy B. B., & Pandit, M.W. (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engineering, Design and Selection*, 4(2), 155–161.
- Halkier, B.A., & Gershenzon, J., (2006) Biology and biochemistry of glucosinolates. *Annual Review Plant Biology* 57, 303–333.
- Hanf, V., & Gonder, U. (2005). Nutrition and primary prevention of breast cancer: Foods, nutrients and breast cancer risk. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 123(2), 139–149.
- Hansen, K. B., Wollmuth, L. P., Bowie, D., Furukawa, H., Menniti, F. S., Sobolevsky, A. I., Swanson, G. T., Swanger, S. A., Greger, I. H., Nakagawa, T., McBain, C. J., Jayaraman, V., Low, C. M., Dell'Acqua, M. L., Diamond, J. S., Camp, C. R., Perszyk, R. E., Yuan, H., & Traynelis, S. F. (2021). Structure, Function, and Pharmacology of Glutamate Receptor Ion Channels. *Pharmacological Reviews*, 73(4), 298–487.
- Hara, K., Yagi, M., Kusano, T., & Sano, H. (2000). Rapid systemic accumulation of transcripts encoding a tobacco WRKY transcription factor upon wounding. *Molecular and General Genetics*, 263(1), 30–37.
- Hasanuzzaman, M. (Ed.) (2020). *The plant family Brassicaceae: Biology and Physiological Responses to Environmental Stresses*. Springer Nature Singapore Pte Ltd.

- Hasanuzzaman, M., Shabala, L., Brodribb, T. J., Zhou, M., & Shabala, S. (2019). Understanding physiological and morphological traits contributing to drought tolerance in barley. *Journal of Agronomy and Crop Science*, 205(2), 129–140.
- He, X., Kang, Y., Li, W., Liu, W., Xie, P., Liao, L., Huang, L., Yao, M., Qian, L., Liu, Z., Guan, C., Guan, M., & Hua, W. (2020). Genome-wide identification and functional analysis of the TIFY gene family in the response to multiple stresses in *Brassica napus* L. *BMC Genomics* 21, 736. <https://doi.org/10.1186/s12864-020-07128-2>
- Xin, H., Xianchao, N., Pan, X., Wei, L., Min, Y., Yu, K., Lunwen, Q., & Wei, H. (2019). Comparative Transcriptome Analyses Revealed Conserved and Novel Responses to Cold and Freezing Stress in *Brassica napus* L. *G3: Genes Genomes Genetics*. 9, 2723–2737. <https://doi.org/10.1534/g3.119.400229>
- Hebda, A., Liszka, A., Lewandowska, A., Lyczakowski, J. J., Gabryś, H., & Krzeszowiec, W. (2022). Upregulation of GLRs expression by light in *Arabidopsis* leaves. *BMC Plant Biology*, 22(1), 197. <https://doi.org/10.1186/s12870-022-03535-7>
- Hernandez-Garcia, C. M., & Finer, J. J. (2014). Identification and validation of promoters and cis-acting regulatory elements. *Plant Science*, 217, 109–119.
- Hershberg, R., & Petrov, D. A. (2008). Selection on codon bias. *Annual Review of Genetics*, 42, 287–299.
- Hershberg, R., & Petrov, D. A. (2009). General rules for optimal codon choice. *PLoS Genetics*, 5(7), e1000556.
- Hodgkin, E. E., & Richards, W. G. (1987). Molecular similarity based on electrostatic potential and electric field. *International Journal of Quantum Chemistry*, 32(S14), 105–110. <http://doi.org/10.1002/qua.560320814>
- Hu, B., Jin, J., Guo, A-Y., Zhang, H., Luo, J., & Gao, G. (2015) GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297.
- Huang, T., & Duman, J. G. (2002). Cloning and characterization of a thermal hysteresis (antifreeze) protein with DNA-binding activity from winter bittersweet nightshade, *Solanum dulcamara*. *Plant Molecular Biology*, 48(4), 339–350.
- Huang, X.-C., German, D. A., & Koch, M. A. (2020). Temporal patterns of diversification in *Brassicaceae* demonstrate decoupling of rate shifts and mesopolyploidization events. *Annals of Botany*, 125(1), 29–47.
- Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Biology and Evolution*, 2(1), 13–34.
- Illergård, K., Ardell, D. H., & Elofsson, A. (2009). Structure is three to ten times more conserved than sequence—A study of structural response in protein cores. *Proteins: Structure, Function, and Bioinformatics*, 77, 499–508. <https://doi.org/10.1002/prot.22458>
- Imperiali, B., & O'Connor, S. E. (1999). Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Current Opinion in Chemical Biology*, 3(6), 643–649.

- Iwano, M., Ito, K., Fujii, S., Kakita, M., Asano-Shimosato, H., Igarashi, M., Kaothien-Nakayama, P., Entani, T., Kanatani, A., & Takehisa, M. (2015). Calcium signalling mediates self-incompatibility response in the *Brassicaceae*. *Nature Plants*, 1(9), 1–9.
- Jahangir, M., Abdel-Farid, I. B., Kim, H. K., Choi, Y. H., & Verpoorte, R. (2009). Healthy and unhealthy plants: The effect of stress on the metabolism of *Brassicaceae*. *Environmental and Experimental Botany*, 67(1), 23–33.
- Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., Tomsho, L. P., Hu, Y., Liang, H., Soltis, P. S., Soltis, D. E., Clifton, S. W., Schlarbaum, S. E., Schuster, S. C., Ma, H., Leebens-Mack, J., & dePamphilis, C. W. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100.
- Jin, L., Cai, Y., Sun, C., Huang, Y., & Yu, T. (2019). Exogenous l-glutamate treatment could induce resistance against *Penicillium expansum* in pear fruit by activating defense-related proteins and amino acids metabolism. *Postharvest Biology and Technology*, 150, 148–157.
- Kabiraj, A., Laha, A., Panja, A. S., & Bandopadhyay, R. (2023). *In silico* comparative structural and functional analysis of arsenite methyltransferase from bacteria, fungi, fishes, birds, and mammals. *Journal of Genetic Engineering and Biotechnology*, 21, 64. <https://doi.org/10.1186/s43141-023-00522-9>
- Kang, J., & Turano, F. J. (2003). The putative glutamate receptor 1.1 (AtGLR1. 1) functions as a regulator of carbon and nitrogen metabolism in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, 100(11), 6872–6877.
- Kang, J., Mehta, S., & Turano, F. J. (2004). The putative *glutamate receptor 1.1* (AtGLR1.1) in *Arabidopsis thaliana* regulates abscisic acid biosynthesis and signaling to control development and water loss. *Plant and Cell Physiology*, 45(10), 1380–1389.
- Kang, S., Kim, H. B., Lee, H., Choi, J. Y., Heu, S., Oh, C. J., Kwon, S. I., & An, C. S. (2006). Overexpression in *Arabidopsis* of a Plasma Membrane-targeting Glutamate Receptor from Small Radish Increases Glutamate-mediated Ca²⁺ Influx and Delays Fungal Infection. *Molecules and Cells*, 21(3), 418–427. [https://doi.org/10.1016/S1016-8478\(23\)12872-X](https://doi.org/10.1016/S1016-8478(23)12872-X)
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, 10, 845–858.
- Kersey, P. J., Allen, J. E., Christensen, M., Davis, P., Falin, L. J., Grabmueller, C., Hughes, D. S. T., Humphrey, J., Kerhornou, A., Khobova, J., Langridge, N., McDowall, M. D., Maheswari, U., Maslen, G., Nuhn, M., Ong, C. K., Paulini, M., Pedro, H., Toneva, I., Tuli, M. A., Walts, B., Williams, G., Wilson, D., Youens-Clark, K., Monaco, M. K., Stein, J., Wei, X., Ware, D., Bolser, D. M., Howe, K. L., Kulesha, E., Lawson, D., & Staines, D. M. (2014). Ensembl Genomes 2013: scaling up access to genome-wide data. *Nucleic Acids Research* 42(D1), D546–D552. <https://doi.org/10.1093/nar/gkt979>

- Kim, S. A., Kwak, J. M., Jae, S.-K., Wang, M.-H., & Nam, H. G. (2001). Overexpression of the *AtGluR2* Gene Encoding an *Arabidopsis* Homolog of Mammalian Glutamate Receptors Impairs Calcium Utilization and Sensitivity to Ionic Stress in Transgenic Plants. *Plant and Cell Physiology*, 42(1), 74–84.
- Kinsella, R. J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., Kersey, P., & Flicek, P. (2011). Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database: The Journal of Biological Databases and Curation*, 2011, bar030. <http://doi.org/10.1093/database/bar030>
- Koch, M. A., Wernisch, M., & Schmickl, R. (2008). *Arabidopsis thaliana*'s wild relatives: An updated overview on systematics, taxonomy and evolution. *Taxon*, 57(3), 933–3E.
- Kong, D., Hu, H.-C., Okuma, E., Lee, Y., Lee, H. S., Munemasa, S., Cho, D., Ju, C., Pedoeim, L., & Rodriguez, B. (2016). L-Met activates *Arabidopsis* GLR Ca^{2+} channels upstream of ROS production and regulates stomatal movement. *Cell Reports*, 17(10), 2553–2561.
- Kong, D., Ju, C., Parihar, A., Kim, S., Cho, D., & Kwak, J. M. (2015). *Arabidopsis* glutamate receptor homolog3. 5 modulates cytosolic Ca^{2+} level to counteract effect of abscisic acid in seed germination. *Plant Physiology*, 167(4), 1630–1642.
- Kopsell, D. A., & Kopsell, D. E., (2006). Accumulation and bioavailability of dietary carotenoids in vegetable crops. *Trends in Plant Science* 11, 499–507.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35, 1547–1549.
- Kumar, J., & Mayer, M. L. (2013). Functional insights from glutamate receptor ion channel structures. *Annual Review of Physiology*, 75, 313–337.
- Kusakina, J., & Dodd, A. N. (2012). Phosphorylation in the plant circadian system. *Trends in Plant Science*, 17(10), 575–583.
- Kwaaitaal, M., Huisman, R., Maintz, J., Reinstädler, A., & Ralph, P. (2011). Ionotropic glutamate receptor (iGluR)-like channels mediate MAMP-induced calcium influx in *Arabidopsis thaliana*. *Biochemical Journal*, 440(3), 355–365.
- Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157(1), 105–132.
- La, V. H., Lee, B.-R., Islam, M. T., Mamun, M. A., Park, S.-H., Bae, D.-W., & Kim, T.-H. (2020). Characterization of Glutamate-Mediated Hormonal Regulatory Pathway of the Drought Responses in Relation to Proline Metabolism in *Brassica napus* L. *Plants*, 9(4), 512.
- Lagercrantz, U., Putterill, J., Coupland, G., & Lydiat, D. (1996). Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time. *The Plant Journal*, 9(1), 13–20.
- Lam, H.-M., Chiu, J., Hsieh, M.-H., Meisel, L., Oliveira, I. C., Shin, M., & Coruzzi, G. (1998). Glutamate-receptor genes in plants. *Nature*, 396(6707), 125–126.

- Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., & Ben-Tal, N. (2005). ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Research*, 33(suppl 2), W299–W302.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), 2947–2948.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2), 283–291.
- Laxa, M., Liebthal, M., Telman, W., Chibani, K., & Dietz, K.-J. (2019). The role of the plant antioxidant system in drought tolerance. *Antioxidants*, 8(4), 94.
- Lazarova, G., Zeng, Y., & Kermode, A. R. (2002). Cloning and expression of an *ABSCISIC ACID-INSENSITIVE 3 (ABI3)* gene homologue of yellow-cedar (*Chamaecyparis nootkatensis*). *Journal of Experimental Botany*, 53(371), 1219–1221.
- Lee, H.-L., & Irish, V. F. (2011). Gene duplication and loss in a MADS box gene transcription factor circuit. *Molecular Biology and Evolution*, 28(12), 3367–3380.
- Legrand, B., Giudice, E., Nicolas, A., Delalande, O., & Le Rumeur, E. (2011). Computational study of the human dystrophin repeats: interaction properties and molecular dynamics. *PloS ONE*, 6(8), e23819. <https://doi.org/10.1371/journal.pone.0023819>
- Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., de-Peer, Y. V., Rouze, P., & Rombauts, S. (2002). PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Research* 30, 325–327.
- Li, Y., Song, Q., Zhang, Y., Li, Z., Guo, J., Chen, X., & Zhang, G. (2020). Genome-wide identification, characterization, and expression patterns analysis of the SBP-box gene family in wheat (*Triticum aestivum* L.). *Scientific Reports* 10, 17250.
- Li, D., Dreher, K., Knee, E., Brkljacic, J., Grotewold, E., Berardini, T. Z., Lamesch, P., Garcia-Hernandez, M., Reiser, L. & Huala, E. (2014). *Arabidopsis* database and stock resources. In Sanchez-Serrano, J., Salinas, J. (Eds.) *Arabidopsis Protocols: Methods in Molecular Biology*, (Vol. 1062), pp. 65–96. Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-62703-580-4_4
- Li, F., Wang, J., Ma, C., Zhao, Y., Wang, Y., Hasi, A., & Qi, Z. (2013). Glutamate receptor-like channel3. 3 is involved in mediating glutathione-triggered cytosolic calcium transients, transcriptional changes, and innate immunity responses in *Arabidopsis*. *Plant Physiology*, 162(3), 1497–1509.
- Li, H., Jiang, X., Lv, X., Ahammed, G. J., Guo, Z., Qi, Z., Yu, J., & Zhou, Y. (2019a). Tomato GLR3. 3 and GLR3. 5 mediate cold acclimation-induced chilling tolerance by regulating apoplastic H₂O₂ production and redox homeostasis. *Plant, Cell & Environment*, 42(12), 3326–3339.

- Li, J., Zhu, S., Song, X., Shen, Y., Chen, H., Yu, J., Yi, K., Liu, Y., Karplus, V. J., & Wu, P. (2006). A rice glutamate receptor-like gene is critical for the division and survival of individual cells in the root apical meristem. *The Plant Cell*, 18(2), 340–349. <https://doi.org/10.1105/tpc.105.037713>
- Li, M., Wang, F., Li, S., Yu, G., Wang, L., Li, Q., Zhu, X., Li, Z., Yuan, L., & Liu, P. (2020). Importers drive leaf-to-leaf jasmonic acid transmission in wound-induced systemic immunity. *Molecular Plant*, 13(10), 1485–1498.
- Li, W.-H., Yang, J., & Gu, X. (2005). Expression divergence between duplicate genes. *Trends in Genetics*, 21(11), 602–607.
- Li, Z.-G., Gong, M., Xie, H., Yang, L., & Li, J. (2012). Hydrogen sulfide donor sodium hydrosulfide-induced heat tolerance in tobacco (*Nicotiana tabacum* L) suspension cultured cells and involvement of Ca²⁺ and calmodulin. *Plant Science*, 185, 185–189.
- Li, Z.-G., Ye, X.-Y., & Qiu, X.-M. (2019b). Glutamate signaling enhances the heat tolerance of *maize* seedlings by plant glutamate receptor-like channels-mediated calcium signaling. *Protoplasma*, 256(4), 1165–1169. <https://doi.org/10.1007/s00709-019-01351-9>
- Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I. A. P., Zhao, M., Ma, J., Yu, J., Huang, S., Wang, X., Wang, J., Lu, K., Fang, Z., Bancroft, I., Yang, T.-J., Hu, Q., Wang, X., Yue, Z., Li, H., Yang, L., Wu, J., Zhou, Q., Wang, W., King, G. J., Pires, J. C., Lu, C., Wu, Z., Sampath, P., Wang, Z., Guo, H., Pan, S., Yang, L., Min, J., Zhang, D., Jin, D., Li, W., Belcram, H., Tu, J., Guan, M., Qi, C., Du, D., Li, J., Jiang, L., Batley, J., Sharpe, A. G., Park, B.-S., Ruperao, P., Cheng, F., Waminal, N. E., Huang, Y., Dong, C., Wang, L., Li, J., Hu, Z., Zhuang, M., Huang, Y., Huang, J., Shi, J., Mei, D., Liu, J., Lee, T.-H., Wang, J., Jin, H., Li, Z., Li, X., Zhang, J., Xiao, L., Zhou, Y., Liu, Z., Liu, X., Qin, R., Tang, X., Liu, W., Wang, Y., Zhang, Y., Lee, J., Kim, H. H., Denoeud, F., Xu, X., Liang, X., Hua, W., Wang, X., Wang, J., Chalhoub, B., & Paterson, A. H. (2014). The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nature Communications* 5, 3930.
- Liu, C., Thong, Z., & Yu, H. (2009). Coming into bloom: the specification of floral meristems. *Development*, 136(20), 3379–3391.
- Locher, K. P., Lee, A. T., & Rees, D. C. (2002). The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science*, 296(5570), 1091–1098.
- Lohani N, Golicz, A. A., Singh, M. B., & Bhalla, P. L. (2019). Genome-wide analysis of the *Hsf* gene family in *Brassica oleracea* and a comparative analysis of the *Hsf* gene family in *B. oleracea*, *B. rapa* and *B. napus*. *Functional and Integrative Genomics* 19, 515–531. <https://doi.org/10.1007/s10142-018-0649-1>
- López-Bucio, J. S., de la Cruz, H. R., & Guevara-García, A. A. (2018). Glutamate sensing in plants. In Ramakrishna, A., & Roshchina, V. V. (Eds.) *Neurotransmitters in Plants: Perspectives and Applications* (pp. 231–240). CRC Press. <https://doi.org/10.1201/b22467>

- Lu, G., Wang, X., Liu, J., Yu, K., Gao, Y., Liu, H., Wang, C., Wang, W., Wang, G., & Liu, M. (2014). Application of T-DNA activation tagging to identify glutamate receptor-like genes that enhance drought tolerance in plants. *Plant Cell Reports*, 33, 617–631.
- Lynch, M., & Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science*, 290(5494), 1151–1155.
- Lysak, M. A., Koch, M. A., Pecinka, A., & Schubert, I. (2005). Chromosome triplication found across the tribe *Brassicaceae*. *Genome Research* 15, 516–525.
- Mabry, M. E., Turner-Hissong, S. D., Gallagher, E. Y., McAlvay, A. C., An, H., Edger, P. P., Moore, J. D., Pink, D. A. C., Teakle, G. R., Stevens, C. J., Barker, G., Labate, J., Fuller, D. Q., Allaby, R. G., Beissinger, T., Decker, J. E., Gore, M. A., & Pires, J. C. (2021). The Evolutionary History of Wild, Domesticated, and Feral *Brassica oleracea* (*Brassicaceae*). *Molecular Biology and Evolution* 38, 4419–4434. <https://doi.org/10.1093/molbev/msab183>
- Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, 18(1), 50–60. <http://doi.org/10.1214/aoms/1177730491>
- Manuka, R., Saddhe, A. A., & Kumar, K. (2015). Genome-wide identification and expression analysis of WNK kinase gene family in rice. *Computational Biology and Chemistry*, 59, 56–66.
- Manzoor, H., Kelloniemi, J., Chiltz, A., Wendehenne, D., Pugin, A., Poinssot, B., & Garcia-Brugger, A. (2013). Involvement of the glutamate receptor A t GLR 3.3 in plant defense signaling and resistance to *Hyaloperonospora arabidopsidis*. *The Plant Journal*, 76(3), 466–480.
- Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., Geer, L. Y., & Bryant, S. H. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research* 45, D200–D203.
- Marmagne, A., Rouet, M.-A., Ferro, M., Rolland, N., Alcon, C., Joyard, J., Garin, J., Barbier-Brygoo, H., & Ephritikhine, G. (2004). Identification of new intrinsic proteins in *Arabidopsis* plasma membrane proteome. *Molecular & Cellular Proteomics*, 3(7), 675–691. <https://doi.org/10.1074/mcp.M400001-MCP200>
- Mayer, M. L. (2006). Glutamate receptors at atomic resolution. *Nature*, 440(7083), 456–462.
- Meyerhoff, O., Müller, K., Roelfsema, M. R. G., Latz, A., Lacombe, B., Hedrich, R., Dietrich, P., & Becker, D. (2005). AtGLR3. 4, a glutamate receptor channel-like gene is sensitive to touch and cold. *Planta*, 222, 418–427.
- Michard, E., Lima, P. T., Borges, F., Silva, A. C., Portes, M. T., Carvalho, J. E., Gilliam, M., Liu, L.-H., Obermeyer, G., & Feijó, J. A. (2011). Glutamate receptor-like genes form Ca²⁺ channels in pollen tubes and are regulated by pistil D-serine. *Science*, 332(6028), 434–437.

- Michard, F. (2011). The burden of high-risk surgery and the potential benefit of goal-directed strategies. *Critical Care*, 15(5), 447.
- Miller, N. D., Brooks, T. L. D., Assadi, A. H., & Spalding, E. P. (2010). Detection of a gravitropism phenotype in *glutamate receptor-like 3.3* mutants of *Arabidopsis thaliana* using machine vision and computation. *Genetics*, 186(2), 585–593.
- Miura, K., & Furumoto, T. (2013). Cold signaling and cold response in plants. *International Journal of Molecular Sciences*, 14(3), 5312–5337.
- Mousavi, S. A., Chauvin, A., Pascaud, F., Kellenberger, S., & Farmer, E. E. (2013). GLUTAMATE RECEPTOR-LIKE genes mediate leaf-to-leaf wound signalling. *Nature*, 500(7463), 422–426.
- Moyroud, E., Kusters, E., Monniaux, M., Koes, R., & Parcy, F. (2010). LEAFY blossoms. *Trends in Plant Science*, 15(6), 346–352.
- Muchate, N. S., Nikalje, G. C., Rajurkar, N. S., Suprasanna, P., & Nikam, T. D. (2016). Plant salt stress: Adaptive responses, tolerance mechanism and bioengineering for salt tolerance. *The Botanical Review*, 82, 371–406.
- Mukherjee, A. (2014). Effect of Gamma-Aminobutyric acid (GABA) on root growth of *Allium sativum*. *Indian Journal of Fundamental and Applied Life Sciences*, 4(1), 171–175.
- Mukherjee, D., Mukherjee, A., & Ghosh, T. C. (2016). Evolutionary Rate Heterogeneity of Primary and Secondary Metabolic Pathway Genes in *Arabidopsis thaliana*. *Genome Biology and Evolution*, 8(1), 17–28.
- Mun, J.-H., Yu, H.-J., Park, S., & Park, B.-S. (2009). Genome-wide identification of NBS-encoding resistance genes in *Brassica rapa*. *Molecular Genetics and Genomics*, 282, 617–631.
- Nagaharu, U., & Nagaharu, N. (1935). Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japanese Journal of Botany*, 7, 389–452.
- Naz, R., Khan, A., Alghamdi, B. S., Ashraf, Md. G., Alghanmi, M., Ahmad, A., Bashir, S. S., & Haq, Q. M. R. (2022). An Insight into Animal Glutamate Receptors Homolog of *Arabidopsis thaliana* and Their Potential Applications—A Review. *Plants* 11, 2580. <https://doi.org/10.3390/plants11192580>
- Nguyen, C. T., Kurenda, A., Stolz, S., Chételat, A., & Farmer, E. E. (2018). Identification of cell populations necessary for leaf-to-leaf electrical signaling in a wounded plant. *Proceedings of the National Academy of Sciences*, 115(40), 10178–10183.
- Ni, J., Yu, Z., Du, G., Zhang, Y., Taylor, J. L., Shen, C., Xu, J., Liu, X., Wang, Y., & Wu, Y. (2016). Heterologous expression and functional analysis of rice GLUTAMATE RECEPTOR-LIKE family indicates its role in glutamate triggered calcium flux in rice roots. *Rice*, 9, 1–14.
- O’neill, C. M., & Bancroft, I. (2000). Comparative physical mapping of segments of the genome of *Brassica oleracea* var. *alboglabra* that are homoeologous to sequenced

- regions of chromosomes 4 and 5 of *Arabidopsis thaliana*. *The Plant Journal* 23, 233–243.
- Olsen, H. B., & Kaarsholm, N. C. (2000). Structural effects of protein lipidation as revealed by LysB29-myristoyl, des (B30) insulin. *Biochemistry*, 39(39), 11893–11900.
- Ortiz-Ramírez, C., Michard, E., Simon, A. A., Damineli, D. S., Hernández-Coronado, M., Becker, J. D., & Feijó, J. A. (2017). GLUTAMATE RECEPTOR-LIKE channels are essential for chemotaxis and reproduction in mosses. *Nature*, 549(7670), 91–95.
- Osborn, T. C., Kole, C., Parkin, I. A. P., Sharpe, A. G., Kuiper, M., Lydiate, D. J., & Trick, M. (1997). Comparison of flowering time genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana*. *Genetics*, 146(3), 1123–1129.
- Pagni, M., Ioannidis, V., Cerutti, L., Zahn-Zabal, M., Jongeneel, C. V., Hau, J., Martin, O., Kuznetsov, D., & Falquet, L. (2007). MyHits: improvements to an interactive resource for analyzing protein sequences. *Nucleic Acids Research*, 35(suppl 2), W433–W437.
- Pandey, S., & Assmann, S. M. (2004). The *Arabidopsis* putative G protein-coupled receptor GCR1 interacts with the G protein α subunit GPA1 and regulates abscisic acid signaling. *The Plant Cell*, 16(6), 1616–1632.
- Park, J. Y., Koo, D. H., Hong, C. P., Lee, S. J., Jeon, J. W., Lee, S. H., Yun, P. Y., Park, B. S., Kim, H. R., Bang, J. W., Plaha, P., Bancroft, I., & Lim, Y. P. (2005). Physical mapping and microsynteny of *Brassica rapa* ssp. *pekinensis* genome corresponding to a 222 kbp gene-rich region of *Arabidopsis* chromosome 4 and partially duplicated on chromosome 5. *Molecular Genetics and Genomics* 274, 579–588.
- Park, W., Li, J., Song, R., Messing, J., & Chen, X. (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Current Biology* 12, 1484–1495. [https://doi.org/10.1016/s0960-9822\(02\)01017-5](https://doi.org/10.1016/s0960-9822(02)01017-5)
- Parkin, I. A. P., Sharpe, A. G., Keith, D. J., & Lydiate, D. J. (1995). Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). *Genome*, 38(6), 1122–1131.
- Parkin, I. A., Gulden, S. M., Sharpe, A. G., Lukens, L., Trick, M., Osborn, T. C., & Lydiate, D. J. (2005). Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics*, 171(2), 765–781.
- Parkin, I. A., Koh, C., Tang, H., Robinson, S. J., Kagale, S., Clarke, W. E., Town, C. D., Nixon, J., Krishnakumar, V., & Bidwell, S. L. (2014). Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. *Genome Biology*, 15(6), 1–18.
- Patthy, L. (1999). Genome evolution and the evolution of exon-shuffling — a review. *Gene* 238, 103–114. [https://doi.org/10.1016/s0378-1119\(99\)00228-0](https://doi.org/10.1016/s0378-1119(99)00228-0)
- Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunić

- I, Marchler-Bauer A, Mi H, Natale DA, Orengo CA, Pandurangan AP, Rivoire, C., Sigrist, C.J.A., Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., & Bateman, A. (2023). InterPro in 2022. *Nucleic acids research*, 51, D418–D427. <https://doi.org/10.1093/nar/gkac993>
- Peden, J. F. (1999). *Analysis of Codon Usage*. University of Nottingham, United Kingdom.
- Petrey, D., & Honig, B. (2005). Protein structure prediction: inroads to biology. *Molecular Cell*, 20(6), 811–819.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612.
- Pnueli, L., Hallak-Herr, E., Rozenberg, M., Cohen, M., Goloubinoff, P., Kaplan, A., & Mittler, R. (2002). Molecular and biochemical mechanisms associated with dormancy and drought tolerance in the desert legume *Retama raetam*. *The Plant Journal*, 31(3), 319–330.
- Podell, S., & Gribskov, M. (2004). Predicting N-terminal myristoylation sites in plant proteins. *Bmc Genomics*, 5(1), 37. <http://doi.org/doi:10.1186/1471-2164-5-37>
- Prakash, V., Singh, V. P., Tripathi, D. K., Sharma, S., & Corpas, F. J. (2019). Crosstalk between nitric oxide (NO) and abscisic acid (ABA) signalling molecules in higher plants. *Environmental and Experimental Botany*, 161, 41–49.
- Price, M. B., Jelesko, J., & Okumoto, S. (2012). Glutamate receptor homologs in plants: functions and evolutionary origins. *Frontiers in Plant Science*, 3, 235.
- Price, M. B., Kong, D., & Okumoto, S. (2013). Inter-subunit interactions between glutamate-like receptors in *Arabidopsis*. *Plant Signaling & Behavior*, 8(12), e27034.
- Pupko, T., Bell, R. E., Mayrose, I., Glaser, F., & Ben-Tal, N. (2002). Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics*, 18(suppl 1), S71–S77.
- Puranik, S., Sahu, P. P., Mandal, S. N., Parida, S. K., & Prasad, M. (2013). Comprehensive genome-wide survey, genomic constitution and expression profiling of the NAC transcription factor family in foxtail millet (*Setaria italica* L.). *Plos ONE*, 8(5), e64594. <https://doi.org/10.1371/journal.pone.0064594>
- Pyrkov, T. V., Chugunov, A. O., Krylov, N. A., Nolde, D. E., & Efremov, R. G. (2009). PLATINUM: a web tool for analysis of hydrophobic/hydrophilic organization of biomolecular complexes. *Bioinformatics*, 25(9), 1201–1202.
- Qi, Z., Stephens, N. R., & Spalding, E. P. (2006). Calcium entry mediated by GLR3.3, an *Arabidopsis* glutamate receptor with a broad agonist profile. *Plant Physiology*, 142(3), 963–971.
- Qiu, X.-M., Sun, Y.-Y., Ye, X.-Y., & Li, Z.-G. (2020). Signaling Role of Glutamate in Plants. *Frontiers in Plant Science*, 10, 1743. <https://doi.org/10.3389/fpls.2019.01743>

- R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Raju, S. K. K., Barnes, A. C., Schnable, J. C., & Roston, R. L. (2018). Low-temperature tolerance in land plants: Are transcript and membrane responses conserved? *Plant Science*, 276, 73–86.
- Rakow, G. (2004). Species origin and economic importance of *Brassica*. In E. C. Pua & C. J. Douglas (Eds.) *Brassica* (pp. 3–11). Springer-Verlag Berlin Heidelberg.
- Rampitsch, C., & Bykova, N. V. (2012). The beginnings of crop phosphoproteomics: exploring early warning systems of stress. *Frontiers in Plant Science*, 3, 144. <https://doi.org/10.3389/fpls.2012.00144>.
- Reiner, A., & Levitz, J. (2018). Glutamatergic signaling in the central nervous system: Ionotropic and metabotropic receptors in concert. *Neuron*, 98(6), 1080–1098.
- Resh, M. D. (1999). Fatty acylation of proteins: new insights into membrane targeting of myristoylated and palmitoylated proteins. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1451(1), 1–16.
- Reyes, J. C., Muro-Pastor, M. I., & Florencio, F. J. (2004). The GATA family of transcription factors in *Arabidopsis* and rice. *Plant Physiology*, 134(4), 1718–1732.
- Richter, S., Wenzel, A., Stein, M., Gabdoulline, R. R., & Wade, R. C. (2008). webPIPSA: a web server for the comparison of protein interaction properties. *Nucleic Acids Research*, 36(suppl 2), W276–W280.
- Rizhsky, L., Liang, H., & Mittler, R. (2002). The combined effect of drought stress and heat shock on gene expression in tobacco. *Plant Physiology*, 130(3), 1143–1151.
- Robert, L. S., Robson, F., Sharpe, A., Lydiate, D., & Coupland, G. (1998). Conserved structure and function of the *Arabidopsis* flowering time gene *CONSTANS* in *Brassica napus*. *Plant Molecular Biology*, 37, 763–772.
- Román-Palacios, C., Molina-Henao, Y. F., & Barker, M. S. (2020). *Polyploidy increases overall diversity despite higher turnover than diploids in the Brassicaceae*. *Proceedings of the Royal Society B: Biological Sciences*, 287, 20200962. <https://doi.org/10.1098/rspb.2020.0962>
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology*, 61(3), 539–542.
- Roy, S. J., Gilliam, M., Berger, B., Essah, P. A., Cheffings, C., Miller, A. J., Davenport, R. J., Liu, L.-H., Skynner, M. J., & Davies, J. M. (2008). Investigating glutamate receptor-like gene co-expression in *Arabidopsis thaliana*. *Plant, Cell & Environment*, 31(6), 861–871.
- Saddhe, A. A., Malvankar, M. R., Karle, S. B., & Kumar, K. (2019). Reactive nitrogen species: Paradigms of cellular signaling and regulation of salt stress in plants. *Environmental and Experimental Botany*, 161, 86–97.

- Sahay, A., Piprodhe, A., & Pise, M. (2020). *In silico* analysis and homology modeling of strictosidine synthase involved in alkaloid biosynthesis in *Catharanthus roseus*. *Journal of Genetic Engineering and Biotechnology*, 18(1), 44.
- Šali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3), 779–815.
- Salvador-Recatalà, V., Tjallingii, W. F., & Farmer, E. E. (2014). Real-time, *in vivo* intracellular recordings of caterpillar-induced depolarization waves in sieve elements using aphid electrodes. *New Phytologist*, 203(2), 674–684.
- Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D. & Lohmann, J. U. (2005). A gene expression map of *Arabidopsis thaliana* development. *Nature Genetics*, 37(5), 501–506.
- Schrödinger, L. L. C. (2010, August). *The PyMOL Molecular Graphics System, Version 1.3r1*.
- Seifi, H. S., Van Bockhaven, J., Angenon, G., & Höfte, M. (2013). Glutamate metabolism in plant disease and defense: Friend or foe? *Molecular Plant-Microbe Interactions*, 26(5), 475–485.
- Serrano, N., Ling, Y., Bahieldin, A., & Mahfouz, M. M. (2019). Thermoprimer reprograms metabolic homeostasis to confer heat tolerance. *Scientific Reports*, 9(1), 1–14.
- Shankar, S., Segaran, G., Sundar, R. D. V., Settu, S., & Sathivelu, M. (2019). *Brassicaceae* - A classical review on its pharmacological activities. *International Journal of Pharmaceutical Sciences Review and Research*, 55(1), 107–113.
- Shanmukhan, A. P., Mathew, M. M., Radhakrishnan, D., Aiyaz, M., & Prasad, K. (2020). Regrowing the damaged or lost body parts. *Current Opinion in Plant Biology*, 53, 117–127.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13, 2498–2504. <https://doi.org/10.1101/gr.1239303>
- Sharp, P. A. (1981). Speculations on RNA splicing (minireview). *Cell*, 23, 643.
- Shields, D. C., & Sharp, P. M. (1987). Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases. *Nucleic Acids Research*, 15(19), 8023–8040.
- Shields, D. C., Sharp, P. M., Higgins, D. G., & Wright, F. (1988). “Silent” sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Molecular Biology and Evolution*, 5(6), 704–716.
- Shiu, S.-H., & Bleecker, A. B. (2003). Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in *Arabidopsis*. *Plant Physiology*, 132(2), 530–543.
- Shukla, E., Singh, P., Singh, V. K., Singh, S. S., & Mishra, A. K., (2021). Homology modeling in combination of phylogenetic assortment, a new approach to resolve the phylogeny of selected heterocystous cyanobacteria based on phycocyanin

- encoding cpcBA-IGS locus. *Vegetos* 34, 339–354. <https://doi.org/10.1007/s42535-021-00207-z>
- Sillito, D., Parkin, I. A., Mayerhofer, R., Lydiate, D. J., & Good, A. G. (2000). *Arabidopsis thaliana*: A source of candidate disease-resistance genes for *Brassica napus*. *Genome*, 43(3), 452–460.
- Simon, A. A., Navarro-Retamal, C., & Feijó, J. A. (2023). Merging Signaling with Structure: Functions and Mechanisms of Plant Glutamate Receptor Ion Channels. *Annual Review of Plant Biology*, 74, 415–452. <https://doi.org/10.1146/annurev-arplant-070522-033255>
- Singh, S. K., & Chang, I.-F. (2018). Pharmacological studies with specific agonist and antagonist of animal iGluR on root growth in *Arabidopsis thaliana*. In J. Samardzic (Ed.), *GABA and Glutamate-New Developments in Neurotransmission Research* (pp. 47–63). *InTech*. <http://dx.doi.org/10.5772/intechopen.72121>.
- Singh, S. K., Chien, C.-T., & Chang, I.-F. (2016). The *Arabidopsis* glutamate receptor-like gene GLR3.6 controls root development by repressing the Kip-related protein gene KRP4. *Journal of Experimental Botany*, 67(6), 1853–1869.
- Sippl, M. J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Bioinformatics*, 17, 355–362
- Soares, D. C., Gerloff, D. L., Syme, N. R., Coulson, A. F. W., Parkinson, J., & Barlow, P. N. (2005). Large-scale modelling as a route to multiple surface comparisons of the CCP module family. *Protein Engineering Design and Selection*, 18(8), 379–388. <http://doi.org/10.1093/protein/gzi039>
- Sobolevsky, A. I. (2015). Structure and gating of tetrameric glutamate receptors. *The Journal of Physiology*, 593(1), 29–38.
- Sobolevsky, A. I., Rosconi, M. P., & Gouaux, E. (2009). X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature*, 462(7274), 745–756.
- Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, 33(suppl 2), W244–W248.
- Song, X., Liu, G., Duan, W., Liu, T., Huang, Z., Ren, J., Li, Y., & Hou, X. (2014). Genome-wide identification, classification and expression analysis of the heat shock transcription factor family in Chinese cabbage. *Molecular Genetics and Genomics* 289, 541–551. <https://doi.org/10.1007/s00438-014-0833-5>
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1), 72–101. <https://doi.org/10.2307/1412159>
- Staub, R. E., Feng, C., Onisko, B., Bailey, G. S., Firestone, G. L., & Bjeldanes, L. F. (2002). Fate of indole-3-carbinol in cultured human breast tumor cells. *Chemical Research in Toxicology*, 15(2), 101–109.
- Straussman, R., Nejman, D., Roberts, D., Steinfeld, I., Blum, B., Benvenisty, N., Simon, I., Yakhini, Z., & Cedar, H. (2009). Developmental programming of CpG island

- methylation profiles in the human genome. *Nature Structural & Molecular Biology*, 16(5), 564–571.
- Stroebel, D., & Paoletti, P. (2021). Architecture and function of NMDA receptors: An evolutionary perspective. *The Journal of Physiology*, 599(10), 2615–2638. <https://doi.org/10.1113/JP279028>
- Studer, G., Rempfer, C., Waterhouse, A. M., Gumienny, R., Haas, J., & Schwede, T. (2020). QMEANDisCo-distance constraints applied on model quality estimation. *Bioinformatics* Oxford England 36, 1765–1771. <https://doi.org/10.1093/bioinformatics/btz828>
- Sukhov, V., Sukhova, E., & Vodeneev, V. (2019). Long-distance electrical signals as a link between the local action of stressors and the systemic physiological responses in higher plants. *Progress in Biophysics and Molecular Biology*, 146, 63–84.
- Sukhova, E., Akinchits, E., & Sukhov, V. (2017). Mathematical models of electrical activity in plants. *The Journal of Membrane Biology*, 250, 407–423.
- Sun, R. (2015). Economic/Academic Importance of *Brassica rapa*. In X. Wang & C. Kole (Eds.), *The Brassica rapa Genome* (pp 1–15), Springer, Berlin, Heidelberg.
- Sun, W., Li, M., & Wang, J. (2022). Genome-Wide Identification and Characterization of the *RCI2* Gene Family in Allotetraploid *Brassica napus* Compared with Its Diploid Progenitors. *International Journal of Molecular Science* 23, 614. <https://doi.org/10.3390/ijms23020614>
- Sun, C., Jin, L., Cai, Y., Huang, Y., Zheng, X., & Yu, T. (2019). L-Glutamate treatment enhances disease resistance of tomato fruit by inducing the expression of glutamate receptors and the accumulation of amino acids. *Food Chemistry*, 293, 263–270.
- Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, 34(suppl 2), W609–W612.
- Sverdlov, A. V., Rogozin, I. B., Babenko, V. N., & Koonin, E. V. (2003). Evidence of splice signal migration from exon to intron during intron evolution. *Current Biology*, 13(24), 2170–2174.
- Tadege, M., Sheldon, C. C., Helliwell, C. A., Stoutjesdijk, P., Dennis, E. S., & Peacock, W. J. (2001). Control of flowering time by FLC orthologues in *Brassica napus*. *The Plant Journal*, 28(5), 545–553.
- Takahashi, F., & Shinozaki, K. (2019). Long-distance signaling in plant stress response. *Current Opinion in Plant Biology*, 47, 106–111.
- Tam, R., & Saier, M. H. (1993). Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiological Reviews*, 57(2), 320–346.
- Tapken, D., Anschütz, U., Liu, L.-H., Huelsken, T., Seebohm, G., Becker, D., & Hollmann, M. (2013). A plant homolog of animal glutamate receptors is an ion channel gated by multiple hydrophobic amino acids. *Science Signaling*, 6(279), ra47.

- Tardieu, F., Simonneau, T., & Muller, B. (2018). The physiological basis of drought tolerance in crop plants: A scenario-dependent probabilistic approach. *Annual Review of Plant Biology*, 69, 733–759.
- Teardo, E., Carraretto, L., Bortoli, S. D., Costa, A., Behera, S., Wagner, R., Lo Schiavo, F., Formentin, E., & Szabo, I. (2015). Alternative splicing-mediated targeting of the *Arabidopsis* GLUTAMATE RECEPTOR3.5 to mitochondria affects organelle morphology. *Plant Physiology*, 167(1), 216–227.
- Teardo, E., Formentin, E., Segalla, A., Giacometti, G. M., Marin, O., Zanetti, M., Schiavo, F. L., Zoratti, M., & Szabò, I. (2011). Dual localization of plant glutamate receptor AtGLR3.4 to plastids and plasmamembrane. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1807(3), Article 3.
- Teardo, E., Segalla, A., Formentin, E., Zanetti, M., Marin, O., Giacometti, G. M., Lo Schiavo, F., Zoratti, M., & Szabò, I. (2010). Characterization of a plant glutamate receptor activity. *Cellular Physiology and Biochemistry*, 26(2), Article 2.
- Terzaghi, W. B., & Cashmore, A. R. (1995). Light-regulated transcription. *Annual Review of Plant Biology*, 46(1), 445–474.
- The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408(6814), 796–815. <https://doi.org/10.1038/35048692>
- Thomas, B. C., Pedersen, B., & Freeling, M. (2006). Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Research*, 16(7), 934–946.
- Tichelaar, W., Safferling, M., Keinänen, K., Stark, H., & Madden, D. R. (2004). The Three-dimensional Structure of an Ionotropic Glutamate Receptor Reveals a Dimer-of-dimers Assembly. *Journal of Molecular Biology*, 344(2), 435–442.
- Tong, C., Wang, X., Yu, J., Wu, J., Li, W., Huang, J., Dong, C., Hua, W., & Liu, S. (2013). Comprehensive analysis of RNA-seq data reveals the complexity of the transcriptome in *Brassica rapa*. *BMC Genomics* 14, 689. <https://doi.org/10.1186/1471-2164-14-689>
- Town, C. D., Cheung, F., Maiti, R., Crabtree, J., Haas, B. J., Wortman, J. R., Hine, E. E., Althoff, R., Arbogast, T. S., Tallon, L. J., Vigouroux, M., Trick, M., & Bancroft, I. (2006). Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. *The Plant Cell* 18, 1348–1359.
- Toyota, M., Spencer, D., Sawai-Toyota, S., Jiaqi, W., Zhang, T., Koo, A. J., Howe, G. A., & Gilroy, S. (2018). Glutamate triggers long-distance, calcium-based plant defense signaling. *Science*, 361(6407), 1112–1115.
- Traynelis, S. F., Wollmuth, L. P., McBain, C. J., Menniti, F. S., Vance, K. M., Ogden, K. K., Hansen, K. B., Yuan, H., Myers, S. J., & Dingledine, R. (2010). Glutamate Receptor Ion Channels: Structure, Regulation, and Function. *Pharmacological Reviews*, 62(3), 405–496. <https://doi.org/10.1124/pr.109.002451>

- Turano, F. J., Panta, G. R., Allard, M. W., & van Berkum, P. (2001). The putative glutamate receptors from plants are related to two superfamilies of animal neurotransmitter receptors via distinct evolutionary mechanisms. *Molecular Biology and Evolution*, 18(7), 1417–1420.
- Twomey, E. C., & Sobolevsky, A. I. (2018). Structural mechanisms of gating in ionotropic glutamate receptors. *Biochemistry*, 57(3), 267–276.
- Ullmann, G. M., Hauswald, M., Jensen, A., Kostic, N. M., & Knapp, E.-W. (1997). Comparison of the physiologically equivalent proteins cytochrome c_6 and plastocyanin on the basis of their electrostatic potentials. Tryptophan 63 in cytochrome c_6 may be isofunctional with tyrosine 83 in plastocyanin. *Biochemistry*, 36(51), 16187–16196.
- Van de Peer, Y., Mizrachi, E., & Marchal, K. (2017). The evolutionary significance of polyploidy. *Nature Reviews Genetics*, 18(7), 411–424.
- van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hünenberger, P. H., Krüger, P., Mark, A. E., Scott, W. R. P., Tironi, I. G. (1996). *Biomolecular simulation: The {GROMOS96} manual and user guide*. Hochschulverlag AG an der ETH Zürich, Zürich.
- Venkatakrishnan, A. J., Deupi, X., Lebon, G., Tate, C. G., Schertler, G. F. & Babu, M. M. (2013). Molecular signatures of G-protein-coupled receptors. *Nature*, 494(7436), 185–194.
- Vincill, E. D., Clarin, A. E., Molenda, J. N., & Spalding, E. P. (2013). Interacting glutamate receptor-like proteins in phloem regulate lateral root initiation in *Arabidopsis*. *The Plant Cell* 25, 1304–1313
- Vincill, E. D., Bieck, A. M., & Spalding, E. P. (2012). Ca^{2+} conduction by an amino acid-gated ion channel related to glutamate receptors. *Plant Physiology*, 159(1), 40–46.
- Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity*, 93(1), 77–78.
- Wade, R. C., Gabdoulline, R. R., & De Rienzo, F. (2001). Protein interaction property similarity analysis. *International Journal of Quantum Chemistry*, 83(3-4), 122–127.
- Wade, R. C., Gabdoulline, R. R., & Luty, B. A. (1998). Species dependence of enzyme-substrate encounter rates for triose phosphate isomerases. *Proteins: Structure, Function, and Bioinformatics*, 31(4), 406–416.
- Wahid, A., Gelani, S., Ashraf, M., & Foolad, M. R. (2007). Heat tolerance in plants: An overview. *Environmental and Experimental Botany*, 61(3), 199–223.
- Walch-Liu, P., & Forde, B. G. (2007). L-Glutamate as a novel modifier of root growth and branching: What's the sensor? *Plant Signaling & Behavior*, 2(4), 284–286.
- Wang, H., Zhong, H., Gao, C., Zang, J., & Yang, D. (2021). The Distinct Properties of the Consecutive Disordered Regions Inside or Outside Protein Domains and Their Functional Significance. *International Journal of Molecular Sciences*, 22, 10677. <https://doi.org/10.3390/ijms221910677>

- Wang, T., Ping, X., Cao, Y., Jian, H., Gao, Y., Wang, J., Tan, Y., Xu, X., Lu, K., Li, J., & Liu, L. (2019). Genome-wide exploration and characterization of miR172/euAP2 genes in *Brassica napus* L. for likely role in flower organ development. *BMC Plant Biology* 19, 336. <https://doi.org/10.1186/s12870-019-1936-2>
- Wang, X., Wu, J., Liang, J., Cheng, F., & Wang, X. (2015). *Brassica* database (BRAD) version 2.0: integrating and mining Brassicaceae species genomic resources. *Database: The Journal of Biological Databases and Curation* 2015, bav093.
- Wang, Y., Yu, D., Zhao, H., Jiang, L., Gao, L., Song, Y., Liu, Z., Bao, F., Hou, C., He, Y., Ju, C., Li, L., & Kong, D. (2022). A glutamate receptor-like gene is involved in ABA-mediated growth control in *Physcomitrium* (*Physcomitrella*) *patens*. *Plant Signaling and Behavior* 17(1), 2145057. <https://doi.org/10.1080/15592324.2022.2145057>
- Wang, P.-H., Lee, C.-E., Lin, Y.-S., Lee, M.-H., Chen, P.-Y., Chang, H.-C., & Chang, I.-F. (2019). The Glutamate Receptor-Like Protein GLR3.7 Interacts With 14-3-3 ω and Participates in Salt Stress Response in *Arabidopsis thaliana*. *Frontiers in Plant Science*, 10, 1169. <https://doi.org/10.3389/fpls.2019.01169>
- Wang, W., Vinocur, B., Shoseyov, O., & Altman, A. (2004). Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in Plant Science*, 9(5), 244–252.
- Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.-H., Bancroft, I., & Cheng, F. (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics*, 43(10), 1035–1039.
- Warwick, S. I., Francis, A., & Al-Shehbaz, I. A. (2006). *Brassicaceae*: Species checklist and database on CD-Rom. *Plant Systematics and Evolution*, 259(2), 249–258.
- Wasternack, C. (2019). New light on local and systemic wound signaling. *Trends in Plant Science*, 24(2), 102–105.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, TAP., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*, 46, W296–W303. <https://doi.org/10.1093/nar/gky427>
- Weiland, M., Mancuso, S., & Baluska, F. (2015). Signalling via glutamate and GLRs in *Arabidopsis thaliana*. *Functional Plant Biology*, 43(1), 1–25.
- Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*, 35, W407–W410.
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6), 80–83.
- Wink, M., & Van Wyk, B.-E. (2008). *Mind-altering and poisonous plants of the world*. Timber Press.

- Wo, Z. G., & Oswald, R. E. (1995). Unraveling the modular design of glutamate-gated ion channels. *Trends in Neurosciences*, 18(4), 161–168.
- Wright, F. (1990). The ‘effective number of codons’ used in a gene. *Gene*, 87(1), 23–29.
- Wu X, Pan L, Guo X, Li, T., Li, J., Duan, Q., & Huang, J. (2022). Genome-wide identification and expression analysis of *BrAGC* genes in *Brassica rapa* reveal their potential roles in sexual reproduction and abiotic stress tolerance. *Frontiers in Genetics* 13, 1044853. <https://doi.org/10.3389/fgene.2022.1044853>
- Wudick, M. M., Michard, E., Oliveira Nunes, C., & Feijó, J. A. (2018a). Comparing plant and animal glutamate receptors: Common traits but different fates? *Journal of Experimental Botany*, 69(17), 4151–4163.
- Wudick, M. M., Portes, M. T., Michard, E., Rosas-Santiago, P., Lizzio, M. A., Nunes, C. O., Campos, C., Santa Cruz Damineli, D., Carvalho, J. C., & Lima, P. T. (2018b). CORNICHON sorting and regulation of GLR channels underlie pollen tube Ca^{2+} homeostasis. *Science*, 360(6388), 533–536.
- Xie, L., & Shang, Q. (2018). Genome-wide analysis of purple acid phosphatase structure and expression in ten vegetable species. *BMC Genomics* 19, 646. <https://doi.org/10.1186/s12864-018-5022-1>
- Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., Zhang, G., Gu, Y. Q., Coleman-Derr, D., Xia, Q. & Wang, Y. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research* 47, W52–W58. <https://doi.org/10.1093/nar/gkz333>
- Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E., Lancet, D., & Shmueli, O. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*, 21(5), 650–659. <http://doi.org/10.1093/bioinformatics/bti042>
- Yang, L., Zhao, Y., Wu, X., Zhang, Y., Fu, Y., Duan, Q., Ma, W., & Huang, J. (2022). Genome-Wide Identification and Expression Analysis of *BraGLRs* Reveal Their Potential Roles in Abiotic Stress Tolerance and Sexual Reproduction. *Cells*, 11(23), 3729. <https://doi.org/10.3390/cells11233729>
- Yang, Y.-W., Lai, K.-N., Tai, P.-Y., & Li, W.-H. (1999). Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. *Journal of Molecular Evolution*, 48(5), 597–604.
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8), 1586–1591.
- Yao, Y., Carretero-Paulet, L., & Van de Peer, Y. (2019). Using digital organisms to study the evolutionary consequences of whole genome duplication and polyploidy. *PLoS ONE*, 14(7): e0220257. <https://doi.org/10.1371/journal.pone.0220257>
- Yilmaz, A., Mejia-Guerra, M. K., Kurz, K., Liang, X., Welch, L., & Grotewold, E. (2011). AGRIS: the *Arabidopsis* gene regulatory information server, an update. *Nucleic Acids Research*, 39(suppl 1), D1118–D1122.

- Yoshida, R., Mori, I. C., Kamizono, N., Shichiri, Y., Shimatani, T., Miyata, F., Honda, K., & Iwai, S. (2016). Glutamate functions in stomatal closure in *Arabidopsis* and fava bean. *Journal of Plant Research*, 129, 39–49.
- Yu, B., Sun, Y., Jin, X., Xie, Z., Li, X., & Huang, J. (2023). Rice glutamate receptor-like channel OsGLR3.4 modulates the root tropism growth towards amino acids via plasma membrane depolarization and ROS generation. *Environmental and Experimental Botany* 205, 105146. <https://doi.org/10.1016/j.envexpbot.2022.105146>
- Yu, B., Wu, Q., Li, X., Zeng, R., Min, Q., & Huang, J. (2022). GLUTAMATE RECEPTOR-like gene *OsGLR3.4* is required for plant growth and systemic wound signaling in rice (*Oryza sativa*). *New Phytologist* 233(3), 1238–1256. <https://doi.org/10.1111/nph.17859>
- Yu, C-S., Chen, Y-C., Lu, C-H., & Hwang J-K (2006). Prediction of protein subcellular localization. *Proteins: Structure, Function, and Bioinformatics* 64, 643–651.
- Yu, J., Tehrim, S., Zhang, F., Tong, C., Huang, J., Cheng, X., Dong, C., Zhou, Y., Qin, R., Hua, W. & Liu, S. (2014). Genome-wide comparative analysis of NBS-encoding genes between *Brassica* species and *Arabidopsis thaliana*. *BMC Genomics* 15, 3. <https://doi.org/10.1186/1471-2164-15-3>
- Zhang, J., Cui, T., Su, Y., Zang, S., Zhao, Z., Zhang, C., Zou, W., Chen, Y., Cao, Y., Chen, Y., Que, Y., Chen, N., & Luo, J. (2022). Genome-Wide Identification, Characterization, and Expression Analysis of Glutamate Receptor-like Gene (GLR) Family in Sugarcane. *Plants*, 11(18), Article 18. <https://doi.org/10.3390/plants11182440>
- Zhang, Y. (2009). Protein structure prediction: when is it useful? *Current Opinion in Structural Biology*, 19(2), 145–155.
- Zheng, Y., Luo, L., Wei, J., Chen, Q., Yang, Y., Hu, X., & Kong, X. (2018). The glutamate receptors *AtGLR1.2* and *AtGLR1.3* increase cold tolerance by regulating jasmonate signaling in *Arabidopsis thaliana*. *Biochemical and Biophysical Research Communications* 506(4), 895–900.
- Zheng, F., Erreger, K., Low, C.-M., Banke, T., Lee, C. J., Conn, P. J., & Traynelis, S. F. (2001). Allosteric interaction between the amino terminal domain and the ligand binding domain of NR2A. *Nature Neuroscience*, 4(9), 894–901.
- Zheng, J., Knighton, D. R., Taylor, S. S., Xuong, N.-H., Sowadski, J. M., & Eyck, L. F. T. (1993). Crystal structures of the myristylated catalytic subunit of cAMP-dependent protein kinase reveal open and closed conformations. *Protein Science*, 2(10), 1559–1573.
- Zukalová, H., & Vasak, J. (2002). The role and effects of glucosinolates of *Brassica* species-a review. *Rostlinná Výroba*, 48(4), 175–180.

List of Publications

List of Publications:

A. Full Research Articles:

1. Computational analysis of the glutamate receptor gene family of *Arabidopsis thaliana*

- **Bidhan Chandra Roy** & Ashutosh Mukherjee
Pages 2454-2474 | Received 03 May 2016, Accepted 03 Aug 2016, Published online: 15 Sep 2016, <https://doi.org/10.1080/07391102.2016.1222968>, (*Journal of Biomolecular Structure and Dynamics*, Volume 35, 2017-Issue 11, Taylor & Francis).

2. Genome-wide analysis of glutamate receptor gene family in allopolyploid *Brassica napus* and its diploid progenitors

- **Bidhan Chandra Roy**, Nikita Shukla, Ratan Gachhui, Ashutosh Mukherjee
Received: 16 April 2023 / Accepted: 10 August 2023, Published Online: 25 August 2023,
<https://doi.org/10.1007/s10709-023-00192-y>, (*Genetica*, Volume 151, Pages 293-310, Springer Nature).

B. Abstracts in Proceedings of Conferences and Seminars:

1. An *in silico* Study of microRNAs targeting Glutamate receptor genes in three species of *Brassica*

- **Bidhan Chandra Roy**, Ratan Gachhui, Ashutosh Mukherjee,
Abstract Book of “2nd International Conference On INTEGRATIVE BIOLOGY & APPLIED GENETICS”, held in 20th -22nd, July, 2022, Organized by Department of Genetics & Biotechnology, Osmania University, Hyderabad-500007, Telangana, India.

2. Homology modelling and *in silico* analysis of three-dimensional structure of a glutamate receptor of “*Brassica oleracea*”

- **Bidhan Chandra Roy**, Ratan Gachhui, Ashutosh Mukherjee,
Abstract Book of “State level Seminar On Biotechnological Interventions Towards Health and Agriculture”, held in 24th September, 2022, Organized by Department of Molecular Biology & Biotechnology, Kanchrapara College, Affiliated to University of Kalyani, West Bengal, India.

3. An *in silico* Analysis of Structural Features of A Glutamate Receptor of “*Brassica oleracea*”

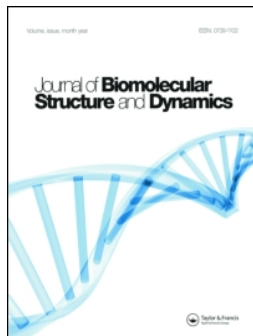
- **Bidhan Chandra Roy**, Ratan Gachhui, Ashutosh Mukherjee, Abstract Book of “Fifth Regional Science and Technology Congress” ,*Volume 6*, ISBN: 978-81-959389-6-4, held in 19th-20th January, 2023, Organized by Department of Science and Technology and Biotechnology, Government of West Bengal and West Bengal State University, West Bengal, India.

C. Awards Received:

- Securing best award in Oral Presentation during 2nd International Conference “Integrative Biology and Applied Genetics” held during July 20th- 22nd, 2022 at Osmania University, Hyderabad-500007, Telangana, India.

D. Events attended/participated:

- Participated in a 2 days National Open Online Workshop on “BIOINFORMATICS: CONCEPTS, TOOLS AND DATABASE” Organized by Research Development Committee, K J College of Science & Commerce, Mumbai-77 on 27th-28th, July, 2020.
- Participated in the E-conference entitled “International E-conference on Biotechnology, Bioinformatics, and Biomedicine 2020” Organized by Faculty of Applied Sciences, AIMST University, Malaysia, held on 24th- 26th August 2020
- Participated in the International Webinar on “Advances in Biological Science-ABS 2020” held on 17th-19th, August, 2020, Organized by the Department of Botany & Biotechnology, Bharathiar University, Coimbatore, Tamil Nadu, India.
- Participated in the National Virtual conference on “Genomics to Phenomics: A New Horizon in Plant Science Research” held on 28th February and 1st March, 2021, Organized by Department of Botany, University of Calcutta, West Bengal, India.
- Participated in the “1st BOTANICAL CONGRESS 2023(An International Meet)” held on 23th -25th, March, 2023, Organized by Botanical Society of Bengal In collaboration with Department of Botany, University of Calcutta, West Bengal, India.
- Participated in the 2 days National Workshop on “Basics of NGS Data Analysis” held at University of Kalyani, Kalyani, West Bengal, India, During August 24th and 25th, 2023.



Computational analysis of the glutamate receptor gene family of *Arabidopsis thaliana*

Bidhan Chandra Roy & Ashutosh Mukherjee

To cite this article: Bidhan Chandra Roy & Ashutosh Mukherjee (2017) Computational analysis of the glutamate receptor gene family of *Arabidopsis thaliana*, Journal of Biomolecular Structure and Dynamics, 35:11, 2454-2474, DOI: [10.1080/07391102.2016.1222968](https://doi.org/10.1080/07391102.2016.1222968)

To link to this article: <http://dx.doi.org/10.1080/07391102.2016.1222968>



View supplementary material [↗](#)



Accepted author version posted online: 10 Aug 2016.
Published online: 15 Sep 2016.



Submit your article to this journal [↗](#)



Article views: 163



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

Computational analysis of the glutamate receptor gene family of *Arabidopsis thaliana*

Bidhan Chandra Roy^a and Ashutosh Mukherjee^{b*} 

^aDepartment of Botany, Dinabandhu Mahavidyalaya, North 24 Parganas, Bongaon, West Bengal 743235, India; ^bDepartment of Botany, Vivekananda College, 269, Diamond Harbour Road, Thakurpukur, Kolkata, West Bengal 700063, India

Communicated by Ramaswamy H. Sarma

(Received 3 May 2016; accepted 3 August 2016)

Ionotropic glutamate receptors (iGluRs) function as glutamate-activated ion channels in rapid synaptic transmission in animals. *Arabidopsis thaliana* possess 20 glutamate receptor-like genes (AtGLRs) in its genome which are involved in many functions including light signal transduction and calcium homeostasis. However, little is known about the physico-chemical, functional and structural properties of AtGLRs. In this study, glutamate receptor-like genes of *A. thaliana* have been studied *in silico*. Exon–intron structures revealed common origin of majority of these genes. The presence of several phosphorylation and myristoylation sites indicate the involvement of AtGLRs in various signaling processes. Gene ontology analysis showed the participation of AtGLRs in various biological processes including different stress responses. In two genes namely AT2G17260 and AT4G35290, presence of RAV1-A binding site motif in the promoter coupled with results from gene ontology annotation indicate their role in stomatal movement through abscisic acid signaling. Expression analysis showed differential expression of several tandemly arranged genes which indicates neo or sub-functionalization. Two genes namely AT5G48400 and AT5G48410 showed significantly more expression in response to *Botrytis cinerea* infection. Five of these genes have shown G-protein-coupled γ -aminobutyric acid (GABA) receptor activity indicating a possible interaction between AtGLRs and GABA. Structurally, all of them were similar while differences were found regarding electrostatic surfaces as well as surface hydrophobicity. Results of this study provide a comprehensive reference regarding AtGLRs for further analysis regarding the structure, function, and evolution of the glutamate receptors in plants.

Keywords: ionotropic glutamate receptor; gene ontology; *cis*-acting elements; tandem duplication; stress response; protein surface properties

Introduction

Membranes of plant cells possess a variety of ion channels which play vital roles in transport of ions used as signaling molecules, nutrients, metabolites, and osmotica (Davenport, 2002). Membrane proteins at the surface of the cell serve as the communication interface between the cell's internal and external environments (Venkatakrishnan et al., 2013). Although many works have focused on highly selective ion channels in plants, a variety of cations are transported via non-selective cation channels (NSCC) (Davenport, 2002). NSCCs are ion channels that are selectively permeable for cations over anions, but do not strongly distinguish between monovalent cations (Demidchik, Davenport, & Tester, 2002). Some of these putative NSCC genes of plant origin are related to animal glutamate receptors (Demidchik et al., 2002). Using the genome sequence and expressed sequence tag database of the model plant *Arabidopsis thaliana* (L.) Heynh., (Lam et al., 1998) identified two complementary DNA clones known as *GLR1* and *GLR2* as well as two genomic sequences known as *GLR3* and *GLR4*. In addition to

primary sequence similarity, the *Arabidopsis* GLR (AtGLR) genes are also predicted to be similar to animal iGluRs (Lam et al., 1998). To date, 20 GLRs have been identified in the *Arabidopsis* genome (Davenport, 2002). Study of glutamate-like receptor homologs from sequenced genomes of *A. thaliana*, lycophytes and bryophytes revealed that these plant genes share homology with mammalian iGluRs, which are involved in neurotransmission (Price, Jelesko, & Okumoto, 2012). iGluRs are NSCC that function predominantly as glutamate-gated Na^+ and Ca^{2+} influx pathways at neuronal (in vertebrates) or neuromuscular (in invertebrates) junctions (Dingledine, Borges, Bowie, & Traynelis, 1999). GLR-like genes have also been identified in other dicotyledonous and monocotyledonous species as well as in gymnosperms indicating that these genes are of general importance in plant physiology (Davenport, 2002).

Following the discovery of the *Arabidopsis* AtGLR family, (Dennison & Spalding, 2000) demonstrated that addition of glutamate (1 mM) to the medium around *Arabidopsis* seedlings stimulated a rapid increase in

*Corresponding author. Email: ashutoshcaluniv@gmail.com

cytosolic Ca^{2+} and caused membrane depolarization of root tip cells. These effects were attributed to activation of plasma membrane Ca^{2+} permeable channels, even though the possibility of secondary activation of Ca^{2+} influx by glutamate uptake was not excluded (Davenport, 2002). To date, GLRs are found to be involved in many biological processes in plants (Price et al., 2012). These include Carbon/Nitrogen balance (Kang & Turano, 2003), photosynthesis (Teardo, Formentin et al., 2010; Teardo, Segalla et al., 2011), abiotic stress response (Kang, Mehta, & Turano, 2004; Meyerhoff et al., 2005), growth of pollen tube (Michard et al., 2011), morphogenesis of root (Li et al., 2006; Miller, Brooks, Assadi, & Spalding, 2010), plant–pathogen interaction (Kang et al., 2006; Kwaaitaal, Huisman, Maintz, Reinstädler, & Panstruga, 2011), and regulation of cellular Ca^{2+} kinetics (Cho et al., 2009; Dubos, Huggins, Grant, Knight, & Campbell, 2003; Kang et al., 2006; Kim, Kwak, Jae, Wang, & Nam, 2001; Qi, Stephens, & Spalding, 2006; Vincill, Bieck, & Spalding, 2012).

Similar to the animal GluRs, plant GLRs are predicted to have a long N-terminal domain of unidentified function, with homology to the ligand-binding domain of mGluRs (metabotropic glutamate receptors), GABA_B receptors (receptors of gamma-aminobutyric acid [GABA]), and extracellular calcium sensors (Davenport, 2002). This domain can be deleted from vertebrate iGluRs without loss of function, although it has been implicated in desensitization membrane targeting and allosteric modulation by Zn^{2+} (Zheng et al., 2001). Moreover, plant GLRs preserve the predicted channel structure of animal iGluRs and show high amino acid sequence identity in the ligand-binding domains and in the M1 and M2 transmembrane domains (Chiu, DeSalle, Lam, Meisel, & Coruzzi, 1999).

In spite of several studies concerning *Arabidopsis thaliana* glutamate receptor-like genes (AtGLRs), there are several areas to explore. For example, analysis of *cis*-acting elements can throw light on the expression and possible role of these proteins in response to different environmental conditions. Moreover, evolutionary history and functional divergence can be inferred from the study of gene structure, chromosomal location, and evolutionary rates along with expression analysis. Additionally, studies of protein three-dimensional (3D) structures are crucial for deciphering the actual process of their functioning.

In the post genomic era, several areas of *Arabidopsis* research are easy to perform because of the enormous information and databases associated with *Arabidopsis* (Li et al., 2014). Because of the lack of experimental 3D structural data for AtGLR proteins, it is necessary to use comparative modeling and structural prediction to study their molecular properties. The utility of such approaches in designing experiments and interpreting experimental results is now widely recognized (Petrey & Honig, 2005;

Zhang, 2009). Therefore, in the present study, we have carried out detailed *in silico* analysis of the AtGLR genes and proteins to gain insights into their structural and functional attributes. We have analyzed physico-chemical properties, gene structure, and gene ontology of AtGLRs. Protein 3D structures were predicted with homology modeling, which would facilitate deciphering the molecular function of these proteins. These results may be helpful in designing of experiments concerning GLR proteins of *A. thaliana* for their further characterization. Subsequently, we analyzed their expression profiles in different tissues and developmental stages, which will further help future studies for elucidating the precise roles of AtGLRs in growth, development, and environmental interactions.

Materials and methods

Identification, sequence retrieval, and phylogenetic analysis

The previously reported (Chiu et al., 2002) AtGLR genes have been retrieved using the BioMart tool (Kinsella et al., 2011) of EnsemblPlants database (Kersey et al., 2014). Additionally, the *A. thaliana* genes data-set of EnsemblPlants database was searched for the Pfam (Finn et al., 2016) domain PF00060 (Ligand-gated ion channel) which is characteristic of iGluRs to find any new member. However, no new member could be found and thus, the final data-set was comprised of the previously reported 20 AtGLR genes. The gene and protein sequences were retrieved using the BioMart tool of EnsemblPlants database. Some genes showed more than one protein products and they were also studied. Multiple sequence alignment was carried out with MUSCLE (Edgar, 2004). The resulting multiple alignment was converted into nexus format with ALTER (Glez-Peña, Gómez-Blanco, Reboiro-Jato, Fdez-Riverola, & Posada, 2010). Phylogeny reconstruction using the Bayesian inference was conducted with MrBayes 3.2.6 (Ronquist et al., 2012).

Physicochemical characterization, motif, and domain analysis of the AtGLR proteins

Different physico-chemical parameters viz. amino acid composition, theoretical pI (isoelectric point) and grand average of hydropathicity (GRAVY) were measured using ProtParam (Gasteiger et al., 2005) server (<http://web.expasy.org/protparam/>). Different phosphorylation, Asn-glycosylation, and myristoylation sites have been predicted using motifscan tool of MyHits (Pagni et al., 2007). Protein domains were identified with Pfam (Finn et al., 2016). Protein disorder content has been studied using IUPred (Dosztányi, Csizsmok, Tompa, & Simon, 2005a, 2005b). The different *cis*-elements of promoter

regions of the AtGLR genes and corresponding transcription factors (TFs) of these binding sites have been predicted from AGRIS, the *Arabidopsis* gene regulatory information server (Yilmaz et al., 2011).

Chromosomal distribution and gene structure prediction

Using the BioMart tool (Kinsella et al., 2011) of EnsemblPlants database (Kersey et al., 2014), the chromosomal locations of each gene were retrieved. Subsequently, the genes were displayed onto the five chromosomes of *A. thaliana* according to their ascending order of physical position (Mbp) using MapChart (Voorrips, 2002). Tandem duplications were characterized on the basis of the criteria that tandem duplicated genes are placed within 10 predicted open reading frames or within 30 kb of each other as described previously (Shiu & Bleecker, 2003). The exon–intron structures of the genes were determined using Gene structure display server (<http://gsds.cbi.pku.edu.cn/>) (Guo, Zhu, Chen, & Luo, 2007) by comparing the full-length CDSs (predicted coding sequence) of AtGLRs with their corresponding genomic sequences.

Estimation of synonymous and non-synonymous substitution rates and divergence time

For the calculation of the synonymous (d_N) and non-synonymous (d_S) substitution rates between the tandemly duplicated genes, ClustalW-based (Larkin et al., 2007) multiple sequence alignment was performed using their amino acid sequences. The aligned amino acid sequences and their corresponding original cDNA sequences were analyzed using PAL2NAL (Suyama, Torrents, & Bork, 2006), which used the CODEML program of PAML (Yang, 2007), to estimate d_N and d_S . The orthologous gene pairs (with 1:1 orthology) among *A. thaliana*, *Arabidopsis lyrata* (L.) O’Kane & Al-Shehbaz and *Brassica rapa* L. as well as their corresponding d_N and d_S values have been retrieved using the BioMart tool (Kinsella et al., 2011) of EnsemblPlants database (Kersey et al., 2014). Time (million years ago, Mya) of duplication and divergence of each gene were calculated using a synonymous mutation rate of λ substitutions per synonymous site per year as $T = K_S/2\lambda$ (Puranik et al., 2013) where $\lambda = 6.1 \times 10^{-9}$ (Lynch & Conery, 2000).

Expression analysis and functional annotation

The normalized expression values of the 20 AtGLR genes have been obtained from PLEXdb (Dash, Van Hemert, Hong, Wise, & Dickerson, 2012). For obtaining the expression profiles of these genes in different plant parts and developmental stages, normalized expression data of ‘Expression Atlas of *Arabidopsis* Development

(AtGenExpress)’ (accession No. AT40) (Schmid et al., 2005) have been downloaded from PLEXdb. To examine the expression pattern of AtGLRs under different biotic stresses, we have downloaded the data regarding ‘Transcriptome changes of *Arabidopsis* during pathogen and insect attack’ (accession No. AT49) (De Vos et al., 2005) and ‘Response to *Botrytis cinerea* infection’ (accession No. AT51). In AT49, effect of gene expression of *A. thaliana* has been studied in response to different pathogens and insects namely *Pseudomonas syringae* pv. *tomato* (a pathogenic leaf bacterium), *Alternaria brassicicola* (a pathogenic leaf fungus), *Pieris rapae* (a tissue-chewing caterpillar), *Frankliniella occidentalis* (a cell-content-feeding thrip), and *Myzus persicae* (a phloem-feeding aphid). To examine the light-dependent expression pattern, we have downloaded expression data of ‘Gene expression from *Arabidopsis* under high light conditions’ (accession No. AT123) (González-Pérez et al., 2011). In this experiment, *Arabidopsis* cell suspension cultures were treated under control light (50 $\mu\text{E}/\text{m}^2/\text{s}$), high light (1800 $\mu\text{E}/\text{m}^2/\text{s}$) and dark conditions.

The tissue specificity index τ of the AtGLR genes was calculated following Yanai et al. (2005) as follows:

$$\tau = \frac{\sum_{i=1}^N (1 - x_i)}{(N - 1)},$$

where N is the number of tissues and x_i is the expression profile component normalized by the maximal component value. The index τ ranges from 0 to 1, with a higher value suggesting higher tissue specificity (Mukherjee, Mukherjee, & Ghosh, 2016). EN_c (effective number of codons) (Wright, 1990) and GC_3 (Guanine and Cytosine content of the third codon position) were calculated from the coding sequences of the corresponding transcripts of the 20 AtGLR genes using CodonW (Peden, 1999) obtained from <http://codonw.sourceforge.net/>.

The functional annotation of AtGLR gene sequences and subsequent analyses were performed using Blast2GO (Conesa & Götz, 2008; Conesa et al., 2005). The gene sequences of AtGLRs were submitted into Blast2GO program to perform three subsequent steps namely (i) BLASTx search (Altschul, Gish, Miller, Myers, & Lipman, 1990) against the nucleotide database of NCBI, (ii) mapping of GO (Gene Ontology Consortium, 2004) terms associated with the BLAST results, and (iii) annotation of GO terms associated with each query to relate the sequences to known function. Three categories of GO classification namely biological processes, molecular functions, and cellular components were then analyzed and graphically represented.

Homology modeling and structure validation

Secondary structure prediction was carried out with SOPMA (Geourjon & Deléage, 1995). 3D structures of



Genome-wide analysis of glutamate receptor gene family in allopolyploid *Brassica napus* and its diploid progenitors

Bidhan Chandra Roy^{1,2} · Nikita Shukla^{2,4} · Ratan Gachhui² · Ashutosh Mukherjee³

Received: 16 April 2023 / Accepted: 10 August 2023
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract

Ionotropic glutamate receptors are ligand-gated nonselective cation channels that mediate neurotransmission in the central nervous system of animals. Plants possess homologous proteins called glutamate receptor-like channels (GLRs) which are involved in vital physiological processes including seed germination, long-distance signaling, chemotaxis, Ca^{2+} signaling etc. Till now, a comprehensive genome-wide analysis of the *GLR* gene family members in different economically important species of *Brassica* is missing. Considering the origin of allotetraploid *Brassica napus* from the hybridization between the diploid *Brassica oleracea* and *Brassica rapa*, we have identified 11, 27 and 65 *GLR* genes in *B. oleracea*, *B. rapa* and *B. napus*, respectively showing an expansion of this gene family in *B. napus*. Chromosomal locations revealed several tandemly duplicated *GLR* genes in all the three species. Moreover, the gene family expanded in *B. napus* after allopolyploidization. The phylogenetic analysis showed that the 103 GLRs are classified into three main groups. The exon-intron structures of these genes are not very conserved and showed wide variation in intron numbers. However, protein sequences are much conserved as shown by the presence of ten short amino acid sequence motifs. Predicted *cis*-acting elements in 1 kb promoters of *GLR* genes are mainly involved in light, stress and hormone responses. RNA-seq analysis showed that in *B. oleracea* and *B. rapa*, some GLRs are more tissue specific than others. In *B. napus*, some GLRs are downregulated under cold stress, while others are upregulated. In summary, this bioinformatic study of the *GLR* gene family of the three *Brassica* species provides evidence for the expansion of this gene family in *B. napus* and also provided useful information for in-depth studies of their biological functions in *Brassica*.

Keywords Allopolyploidy · *cis*-acting elements · Expression · Orthology · Phylogeny

Introduction

In mammalian central nervous systems, ionotropic glutamate receptors mediate rapid excitatory neuronal transmission (Naz et al. 2022). The plant glutamate receptor-like

homologs (GLRs) are homologs of mammalian ionotropic glutamate receptors (iGluRs) and are hypothesized to be potential amino acid sensors in plants (Price et al. 2012). *Arabidopsis thaliana* possesses 20 glutamate receptor-like genes (*AtGLRs*) in its genome which are involved in many functions including light signal transduction and calcium homeostasis (Roy and Mukherjee 2017). In *Arabidopsis*, *AtGLR1.2* and *AtGLR1.3* play vital roles in the plant response to cold stress (Zheng et al. 2018). Recent studies reporting the localization of GLRs in chloroplasts indicate an additional role of GLRs in this organelle (Price et al. 2012). Indeed, plants carrying T-DNA insertions in *AtGLR3.4* showed weak photosynthetic phenotypes (Teardo et al. 2010, 2011). Similarly, it was shown that *AtGLR3.1* was expressed preferentially in guard cells and over-expression of this gene leads to the impairment of stomatal closure that is induced by external Ca^{2+} (Cho et al. 2009). Many studies showed that plant GLRs play many roles including

✉ Ashutosh Mukherjee
ashutoshcaluniv@gmail.com

¹ Department of Botany, Dinabandhu Mahavidyalaya, North 24 Parganas, Bongaon, West Bengal 743235, India

² Department of Life Science & Biotechnology, Jadavpur University, 188 Raja S.C. Mullick Road, Kolkata, West Bengal 700032, India

³ Department of Botany, Vivekananda College, 269, Diamond Harbour Road, Thakurpukur, Kolkata, West Bengal 700063, India

⁴ Present Address: CSIR–Centre for Cellular and Molecular Biology (CCMB), Hyderabad 500007, India

plant reproduction, seed germination, root growth, regulation of stomatal aperture, plant defense and others (Grenzi et al. 2022). The moss *Physcomitrium patens* contains one GLR (PpGLR1) which plays a pivotal role in ABA-mediated growth regulation (Wang et al. 2022). In rice, *OsGLR3.4* plays a vital role in plant growth as well as mechanical wound response (Yu et al. 2022). In another study, it has been shown that *OsGLR3.4* controls rice root growth and tropism towards nutrient source (Yu et al. 2023). Thus, *GLR* genes should be studied in detail in other crop species for better understanding of their roles in various physiological processes.

Several crop species of the genus *Brassica* such as *B. rapa* L. (Chinese cabbage, turnip and pak choi), *B. oleracea* L. (broccoli, cabbage, cauliflower, Brussels sprouts etc.) as well as the oilseed crop such as *B. napus* L. (rapeseed) are widely used worldwide (Wang et al. 2011; Sun et al. 2022). *Brassica oleracea* not only comprises many important vegetable crops including cauliflower, broccoli, cabbages, Brussels sprouts, Kohlrabi and kales; the species demonstrates extreme morphological diversity and crop forms, with various members grown for their leaves, flowers and stems (Mabry et al. 2021). Most *B. oleracea* crops are high in protein (Liu et al. 2014), carotenoids (Kopsell and Kopsell 2006) and contain diverse glucosinolates (GSLs) that function as unique phytochemicals for plant defence against fungal and bacterial pathogens (Halkier and Gershenzon 2006). *B. rapa* also shows morphological diversity among the cultivated types and includes both oil and vegetable crops (Sun 2015). *B. napus* is the second most important oilseed crop species in the world and there is a growing interest in rapeseed protein as human nutrition (Friedt et al. 2018).

Moreover, these three species along with *B. juncea*, *B. carinata* and *B. nigra* provide opportunities for the study of *Brassica* genome evolution as these six widely cultivated *Brassica* species are a classical example of the importance of polyploidy in botanical evolution, described by the ‘U’s triangle’ (Nagaharu and Nagaharu 1935; Wang et al. 2011). Of these, *B. napus* (A and C subgenomes, $n=19$) was formed by recent allopolyploidy between ancestors of *B. rapa* (A genome, $n=10$) and *B. oleracea* (C genome, $n=9$) and is probably polyphyletic (Nagaharu and Nagaharu 1935; Allender and King 2010). Comparative physical mapping studies have confirmed genome triplication in a common ancestor of *B. oleracea* (O’neill and Bancroft 2000) and *B. rapa* (Park et al. 2005) since its divergence from the *A. thaliana* lineage at least 13–17 MYA (Yang et al. 1999; Town et al. 2006; Beilstein et al. 2010). The *Brassica* genomes retain evidence of recursive whole-genome duplication (WGD) events (Bowers et al. 2003; Jiao et al. 2011) and have undergone a Brassicaceae-lineage-specific whole-genome triplication (WGT) (Lysak et al. 2005; Wang et al. 2011) since their divergence from the *Arabidopsis* lineage (Liu et al. 2014).

In this study, considering the importance of plant *GLR* genes as well as the economic importance of *Brassica*, we made use of the publicly available sequenced genomes of the allotetraploid *B. napus* and its diploid progenitors *B. oleracea* and *B. rapa* to trace the evolution of the *GLR* genes in these species. We have studied the phylogeny, chromosomal distribution, protein motif analysis, gene structure, orthology analysis, expression patterns and promoter analysis, subcellular localization and miRNA targeting of these *GLR* genes. This study enhances our understanding of the effect of polyploidisation events on the evolution of the *GLR* gene family in *Brassica* as well as their structural and functional aspects.

Materials and methods

Identification of *GLR* genes

The genome data (gene, cds and protein data) of allopolyploid *B. napus* and its two diploid progenitors, *B. rapa* and *B. oleracea*, were obtained from the BRAD database (<http://brassicadb.org/brad/>) (Wang et al. 2015). The previously analysed (Roy and Mukherjee 2017) *Arabidopsis thaliana* GLR (AtGLR) proteins have been retrieved using the BioMart tool (Kinsella et al. 2011) of Ensembl Plants database (Kersey et al. 2014). These proteins were used as queries to perform BLASTp (Altschul et al. 1990) searches ($E\text{-value} < 1e-5$) with all proteins from the three *Brassica* species. To identify the *GLR* genes in three *Brassica* genomes accurately, all putative protein sequences obtained from the BLASTp search, were confirmed by searching for the conserved domains using CD-search in the NCBI Conserved Domain Database (CDD; <https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) (Marchler-Bauer et al. 2017) and Pfam (<http://pfam.sanger.ac.uk/>) (El-Gebali et al. 2019) database. The conserved domain analysis helped to make the shortlisting of the best candidate to be the GLR members in the *Brassica* genomes. We have retained those genes which showed the presence of PBP1_GABAb_receptor (accession: cd06366) and GluR_Plant (cd13686) by NCBI-CD search and also showed the pfam domains ANF_receptor (extracellular ligand-binding domain; PF01094), SBP_bac_3 (bacterial extracellular solute-binding proteins, family 3; PF00497) and Lig_chan (ligand-gated ion channel; PF00060). These pfam domains were considered because these are present in *A. thaliana* GLRs (Roy and Mukherjee 2017). Moreover, the shortlisted genes have been searched in Ensembl Plants (for *B. rapa* and *B. napus*) and Phytozome 12 (Goodstein et al. 2012) (for *B. oleracea*) for confirmation of their identity as glutamate receptors. The subcellular location of *GLR* genes was predicted using the CELLO (Yu et al. 2006) web tool (<http://cello.life.nctu.edu.tw/>).

Phylogenetic relationship analysis

The full-length GLR protein sequences of the three *Brassica* species were aligned with MUSCLE (Edgar 2004) and subsequently, a phylogenetic tree was constructed by MEGAX (Kumar et al. 2018) based on Maximum Likelihood (ML) method with 1000 bootstrap replicates. The tree was visualized with FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

Chromosomal location and gene structure analysis

The chromosomal location information of *GLR* genes in the three *Brassica* species were collected from the BRAD database and their physical positions were drafted to the corresponding chromosomes by TBTools (Chen et al. 2020). Some genes showed scaffold information only and locations of some of them were confirmed with NCBI Genome database (<https://www.ncbi.nlm.nih.gov/genome/>). Tandem duplication events were defined as adjacent homologous *GLR* genes on the same chromosome with no more than one intervening gene following Li et al. (2020). The exon/intron structures of *GLR* genes were analyzed using Gene Structure Display Server (GSDS) 2.0 (<http://gsds.gao-lab.org/>) (Hu et al. 2015) by aligning the cDNAs with their corresponding gene sequences.

Conserved motif analysis

Conserved motifs in GLR proteins were investigated by online MEME server (<https://meme-suite.org/meme/tools/meme>) (Bailey et al. 2009), with keeping the maximum motif number as 10 and the other parameters at their default values.

Gene expression studies using RNA-seq data

For gene expression profiling of the *GLR* genes in the three species of *Brassica*, publicly available RNA-seq data were obtained from NCBI GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) and analyzed. The following bio project accession numbers were analyzed: PRJNA530549 (He et al. 2019, 2020) for *B. napus*, PRJNA183713 (Liu et al. 2014; Yu et al. 2014) for *B. oleracea* and PRJNA185152 (Tong et al. 2013; Liu et al. 2014; Wu et al. 2022) for *B. rapa*. Gene expression patterns of each tissue were analyzed and log2 transformed FPKM values (fragments per kilobase of exon model per million mapped reads) were visualized as heat maps using TBtools (Chen et al. 2020).

Promoter analysis

The promoter sequences, which were the 1000 bp upstream of the transcription start site (TSS) of the *GLR* genes, were

acquired from the BRAD database, and the *cis*-elements in the promoters were analyzed using the Plant Cis-Acting Regulatory Element (PlantCARE) server (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) (Lescot et al. 2002).

Sequence-based identification of orthologous GLR genes

The orthologous genes in *B. oleracea*, *B. rapa* and *B. napus* were identified using OrthoVenn2 (Xu et al. 2019) (<http://www.biioinfo genome.net/OrthoVenn>). The protein sequences of the identified *GLR* genes from all the three species of *Brassica* were used. The chromosomal position of the identified orthologous gene pairs were visualized using the TBTools program (Chen et al. 2020) for the analysis.

Prediction of miRNAs targeting GLR genes

All of the gene sequences of *GLR* family genes were submitted as candidate genes for predicting potential miRNAs by searching against the available reference of miRNA sequences of the respective species using psRNATarget (Dai et al. 2011, 2018; Dai and Zhao 2011) server with default parameters. Cytoscape software (Shannon et al. 2003) was used to visualize the interactions between the predicted miRNAs and the corresponding target *GLR* genes.

Results

Identification of *GLR* genes in *B. napus*, *B. oleracea* and *B. rapa* and phylogenetic analysis

We have identified 65, 11 and 27 genes as *GLRs* in *B. napus*, *B. oleracea*, and *B. rapa*, respectively and named them species-wise according to their distribution on chromosomes (*BnGLRs* for *B. napus*, *BoGLRs* for *B. oleracea* and *BrGLRs* for *B. rapa*) (Table 1). We analyzed the molecular characteristics of these *GLR* genes in *B. oleracea*, *B. rapa* and *B. napus*. In *B. napus*, the gene length of *BnGLR* genes ranged from 1206 bp (*BnaAnng35320D*) to 8949 bp (*BnaA04g14620D*), the peptide length ranged from 318 amino acids (*BnaC04g40280D*) to 1426 amino acids (*BnaA10g26130D*) and the GC content of gene ranged from 30.83% (*BnaA04g14620D*) to 47.24% (*BnaA01g02120D*). In *B. oleracea*, the Gene length of *BoGLR* genes ranged from 2931 bp (*Bol033269*) to 7882 bp (*Bol013454*), the peptide length ranged from 739 amino acids (*Bol033267*) to 1787 amino acids (*Bol013454*) and the GC content of gene ranged from 36.02% (*Bol033267*) to 46.77% (*Bol040960*). In *B. rapa*, the gene length of *BrGLR* genes ranged from 2220 bp (*Bra040758*) to 9574 bp (*Bra020703*), the peptide length