Imperial College London

Department of Bioengineering

Integration of multiple environmental stresses for compound gene regulation in *Arabidopsis thaliana*

Sang Yun Lee

December, 2015

Supervisors: Dr. Reiko J. Tanaka, Dr. Jennifer H. Tweedy

Submitted in part fulfilment of the requirements for the degree of Doctor of Philosophy of Imperial College London

Abstract

Plants are known to respond to various types of environmental stresses arising from physicochemical changes and other organisms. As plants often simultaneously experience multiple stress factors due to their immobility, capacity to appropriately regulate gene expression by integrating multiple stress signals is crucial for successful adaptation to hostile environments. Although significant progress has been made in elucidating the molecular mechanisms for regulation of stress response genes under single stress, little is known about the effects of combined stress signals on gene regulation and their associated mechanisms. This thesis aimed to contribute to the understanding of plant stress response by studying the signal integration mechanisms under various perspectives: first, the thesis explored how multiple stress signals affect the choices over discretised regulatory outcomes, such as up-regulation or down-regulation. We propose that processing of multiple signals can be described as logical operations, and subsequently investigate the mechanisms for each signal integration outcome by constructing logical model of intracellular signalling network. The resulting insight was applied to analyse a transcriptomic dataset from the model plant *Arabidopsis thaliana*, leading to novel hypotheses about potential crosstalk interactions that are missing between multiple stress signalling pathways. In parallel, the thesis also explored the cases where integration of multiple stress signals modulates dynamics of gene expression. An experimental study of the expression profile of *Response-to-Dehydration 29A* (*RD29A*), a model stress response gene, was conducted to show that combination of multiple stress inputs introduces a unique qualitative effect on dynamics of gene expression. The origin of this behaviour was investigated via a dynamical model of the *RD29A* regulatory network,

which subsequently revealed potential interactions in the regulatory network that are currently unknown. Taken together, this thesis argues that systematic comparison between gene regulatory outcomes under single and combined stress inputs provides a crucial source of information for discovering functionally significant regulatory interactions in the stress signalling network.

Statement of Originality

I declare that the work presented in this thesis is my own. Any materials or ideas from the work of other people have been acknowledged and referenced.

Sang Yun Lee
Department of Bioengineering
Imperial College London

Copyright declaration

The copyright of this thesis and the material it contains is held by the author unless stated otherwise. Publication or re-use of the information or quotation from the thesis requires proper acknowledgement.

Acknowledgement

I would like to extend my sincerest gratitude to my supervisor, Dr. Reiko J. Tanaka, for providing support and motivation. This work would not have been possible without them. The years of working under her guidance has shaped my scientific interest and allowed opportunities to develop new skills, for which I am forever indebted to her.

I also would like to thank Dr. Jennifer Tweedy, my co-supervisor, for the time spent in her group. Her mentorship has been crucial for broadening my perspective on other areas of biological science.

The experimental investigation in this research was conducted in the laboratory of Professor Alex Webb (Department of Plant Sciences, University of Cambridge). I thank Prof. Webb for his deep insights in *Arabidopsis* biology and constructive criticisms. I am also grateful for logistic and technical support provided by Matthew A. Stancombe for all of the experiments carried out in Webb lab.

I would like to thank my colleagues Elisa Domínguez-Hüttinger, Alejandro Granados, Mark van Logtestijn, Panayiotis Christodoulides, Scott Lovell from Tanaka lab and Dr. Caroline Roney, Rheeda Ali, Maryam Argungu, Karla Sanchez Cazares from Tweedy lab, with whom I have shared the same passion in Systems Biology during the course of this research. The insights that emerged from various interactions I had with them have been pivotal to development of fundamental ideas presented in this thesis. Particular thanks goes to Dr. Neville J. Boon, whose theoretical expertise made significant contribution to the modelling of *RD29A* regulation in Chapter 6 of this thesis. It has been a great pleasure to have worked

with these talented individuals.

I also thank the Department of Bioengineering at Imperial College London for financial support as well as providing a wonderful environment to work throughout the course of this research.

Finally, I thank my parents, Moo-Young and Young-Min Lee: everything I have done would not have bore fruits without their enormous support and love. Special thanks goes to my sister and brother, Yerin and Ji-Yoon Lee for their companionship. It has been a true privilege to be a member of this family.

Table of Contents

Abstract **i**

Statement of Originality **iii**

Copyright declaration **iv**

Acknowledgement **v**

Table of Contents **vii**

List of Figures **xiii**

List of Tables **xv**

Abbreviations **xvi**

Terminology **xix**

1 Introduction **1**

1.1 Importance of understanding plant stress response 1

1.2 *Arabidopsis thaliana* as model organism 3

1.3 Combined stress in plants 4

1.4 Aims 6

1.5 Outline 8

2 Regulation of plant stress response genes in *Arabidopsis thaliana* **10**

2.1 Introduction . . . 10

2.2 Roles of stress-regulated genes . . . 13

2.3 ABA action and biosynthesis . . . 15

2.4 Transcription factors . . . 19

2.5 Intracellular signalling pathways for stress-induced gene regulation . . . 23

2.5.1 Abiotic stress signalling pathways . . . 23

2.5.2 Biotic stress signalling pathways . . . 27

2.5.3 Crosstalk between stress signalling pathways . . . 30

2.6 Discussion . . . 33

3 Compound gene regulation as logical operation **35**

3.1 Introduction . . . 35

3.2 Examples of signal integration from bacterial systems . . . 37

3.3 Compound logic . . . 40

3.4 Logical network models . . . 43

3.4.1 Gene regulatory network as Parallel Distributive Processing Network 43

3.4.2 Model description . . . 46

3.4.3 Genes regulated by homotypic promoters ($n = 1$) . . . 49

3.4.4 Genes regulated by heterotypic promoters ($n = 2$) . . . 49

3.5 Discussion . . . 51

4 Analysis of transcriptomic changes under combined stresses in *Arabidopsis thaliana* using compound logic **55**

4.1 Introduction . . . 55

4.2 Regulation of *Arabidopsis* transcriptome by combined environmental stresses 58

4.2.1 Regulatory Profiles Matrix . . . 58

4.2.2 Overview of transcriptomic responses to single and double stresses . . 59

4.2.3 Classification of gene regulatory profiles under single and combined stress conditions . . . 63
4.3 Prediction of signalling network topologies from the transcriptomic dataset . 68
4.3.1 Heat and biotic stress-inducible genes regulated by HSF- and WRKY-mediated signalling pathways . . . 68
4.3.2 Distribution of regulatory profiles amongst the genes regulated by HSF and WRKY pathways . . . 69
4.3.3 Gene ontology analysis . . . 72
4.3.4 Network topologies for the genes with only HSE or Wbox motifs . . . 73
4.3.5 Network topologies for the genes with both HSE and Wbox motifs . . 75
4.4 Discussion . . . 77

5 Experimental measurement of *RD29A* expression dynamics in *Arabidopsis thaliana* 81
5.1 Introduction . . . 81
5.2 Materials and Methods . . . 83
5.2.1 Stress treatment and Sample preparation . . . 83
5.2.2 Sample processing . . . 83
5.2.3 Quantitative RT-PCR Data analysis . . . 84
5.3 Results . . . 85
5.4 Discussion . . . 90

6 Mathematical modelling of *RD29A* regulatory system 93
6.1 Introduction . . . 93
6.2 Structure of *RD29A* regulatory system . . . 95
6.3 Model development . . . 98
6.3.1 Stress input dynamics . . . 98
6.3.2 Transcription factor dynamics . . . 100
6.3.3 mRNA synthesis . . . 103

6.3.4 Summary . . . 106
6.4 Results . . . 107
6.4.1 Comparison of experimental data and model outcomes . . . 107
6.4.2 Cross-input modulation of intracellular signalling processes . . . 110
6.4.3 Evidence of individual crosstalk interactions in microarray datasets . . . 114
6.4.4 Reduction of synergistic effect upon reduced input strength . . . 116
6.5 Discussion . . . 119

7 Concluding Remark 124

A Supplementary Data and Information for Chapter 5 & 6 130

A.1 Additional *RD29A* expression dynamics dataset (preliminary experiments) . . . 130
A.2 Analytical solutions for the model of the *RD29A* regulatory system . . . 133
A.2.1 Before treatment ($t \leq 0$) . . . 133
A.2.2 After treatment ($t > 0$) . . . 134
A.3 Parameter identification . . . 136
A.3.1 Monte Carlo Simulated Annealing . . . 136
A.3.2 Objective function . . . 137
A.3.3 Selection of system structure . . . 138

Bibliography 172

List of Figures

2.1 Control of stress regulon as the core signalling paradigm for stress-induced gene regulation in plants 11
2.2 Damaging effects caused by common environmental stresses in plants 14
2.3 Regulation of metabolic pathways leading to ABA biosynthesis 17
2.4 Signalling pathways for transduction of various abiotic stresses 24
2.5 Signalling pathways for transduction of various biotic stress signals 28
2.6 Crosstalk between stress-induced hormones for regulation of stress response genes 31

3.1 Examples of signal integration from bacterial systems 38
3.2 Representation of signal integration outcomes via binary compound logic . . 41
3.3 Use of compound logic functions to describe signal integration in the bacterial examples 42
3.4 Identification of compound logic functions for description of relative gene expression change based on ternary logic 44
3.5 Transcriptome regulatory network depicted as Parallel Distributive Processing network 45
3.6 Overview of a logical network model for description of signalling network topology 47
3.7 Number of possible network topologies with two signalling pathways for each regulatory profile 51

4.1 Regulatory Profiles Matrix - an illustrative example 58
4.2 Transcriptomic responses to singly applied stresses from *Col* ecotype 61
4.3 Comparison of transcriptomic responses to combinations of two stresses against the responses to singly applied stresses 62
4.4 Classification of all possible regulatory profiles via ternary compound logic . 65
4.5 Regulation of individual genes across transcriptome under different pairs of stresses 67
4.6 Identification of gene subsets controlled by Heat stress and/or Flg22 70
4.7 Distribution of regulatory profiles amongst the genes containing either HSE, Wbox or both 71
4.8 Possible Parallel Distributive Processing (PDP) network topologies for regulatory profiles of the genes with homotypic promoter 74
4.9 Proposed structure of HSF- and WRKY-mediated signalling pathways 78

5.1 Experimental measurements of *RD29A* expression level at different time and treatment conditions 86
5.2 Temporal dynamics of *RD29A* expression with intrinsic stress-independent variability removed 88

6.1 Overview of the *RD29A* regulatory system involving the DREB2 and AREB pathways 96
6.2 *RD29A* expression level upon varying NaCl and ABA concentrations (Xiong et al.) 101
6.3 The proposed mathematical model of *RD29A* regulatory system 106
6.4 Comparison of the experimental data and the outcome of the *RD29A* regulatory system model 108
6.5 Definition of cross-input modulation 111
6.6 Comparison of the system structures implementing different mechanisms of cross-input modulation 113
6.7 Effect of combining NaCl and ABA at different strengths 117

6.8 Reduction of synergistic effect upon halving the dose of one stress in a combined stress input. 118

7.1 A conceptual framework for elucidation of signal integration mechanisms from analysis of signalling network models at various levels of abstraction 127

A.1 Experimental measurements of RD29A expression level at different time and treatment conditions (2010) 131

A.2 Experimental measurements of RD29A expression level at different time and treatment conditions (2011) 132

List of Tables

2.1 Summary of six major *Arabidopsis* transcription factor families involved in stress response gene regulation 20

3.1 Logical functions describing regulation of TF activity by multiple external inputs 48

3.2 Logical functions describing the states of promoters regulated by one or two TFs 48

3.3 Logical functions reproduced by the models of regulatory network consisting of a single TF 50

4.1 Mathematical rules used for definition of the five regulatory modes 64

4.2 Enrichment of gene ontology terms amongst the genes containing either HSE or Wbox 73

4.3 Possible Parallel Distributive Processing (PDP) network topologies for the genes containing both HSE and Wbox motifs 76

5.1 List of primers used 84

5.2 Comparison of *RD29A* expression level between 3 and 5 hours of treatment (two-sample t-test) 89

6.1 Description of model parameters 102

6.2 Comparison of the system structures with cDNA microarray datasets 115

6.3 List of assumptions used in the model of the *RD29A* regulatory network 120

A.1 Parameter values used for the solutions presented in Chapter 6 figures 139

Abbreviations

ABA: Abscisic Acid

ABI: ABA-Insensitive

ABRE: ABA-Responsive Element

AP2/EREBP[i]: APETALA2/Ethylene-Responsive Element-Binding protein

AREB[ii]: ABRE-Binding protein

as-1: activation sequence 1

bHLH: Basic Helix-Loop-Helix

bZIP: Basic Leucine Zipper

CAM: Calmodulin

CAMK: Calmodulin-dependent protien kinase

CBF: CRT-Binding Factor

CDPK: Calcium-dependent Protein Kinase

CNGC: Cyclic Nucleotide-Gated Calcium channel

COI: Coroantine Insensitive

CTR: Constitutive-Triple-Response

[i]alternatively known as APETALA2/Ethylene-Responsive Factor (AP2/ERF)
[ii]also known as ABRE-Binding Factor (ABF)

DBD: DNA-Binding Domain

DRE/CRT[iii]: Dehydration-Responsive Element/C-Repeat

DREB[iv]: DRE-Binding protein

DRIP: DREB-Interacting Proteins

EIN: Ethylene Insensitive

EIL: Ethylene Insensitive 3-Like

ERD: Early Response to Dehydration

ERF: Ethylene Responsive Factor

ET: Ethylene

ETR: Ethylene Receptor

HSE: Heat Shock Element

HSP: Heat Shock Proteins

HSF[v]: Heat Shock Factor

ICE: Inducer of CBF expression

JA: Jasmonic Acid

JAZ: Jasmonate ZIM-domain

KEG: Keep-On-Going

MAPK[vi]: Mitogen-Activated Protein Kinase

MKK: MAPK Kinase

MEKK: MAPK Kinase Kinase

[iii] also simply referred as DRE
[iv] alternatively known as CRT-Binding Factor (CBF)
[v] also referred as Hsf
[vi] also referred as MPK

MYB: Myeloblastosis Related proteins

MYC: Myelocytomatosis Related proteins

NAC: No Apical Meristam (NAM)/Arabidopsis Transcription Activation Factor (ATAF)/Cup-Shaped Cotyledon (CUC)

NPR: Non-Inducer of PR expression

PAMP: Pathogen Associated Molecular Patterns

PP2C: Protein Phosphatase 2C

PR: Pathogenesis-Related

PYR/PYL/RCAR: Pyrabactin Resistance1/PYR1-like/Regulatory Components of ABA Receptors

RD29A[vii]: Response-to-Dehydration 29A

ROS: Reactive Oxygen Species

SA: Salicylic Acid

SAR: Systemic Acquired Resistance

SnRK: SNF-Related serine/threonine protein Kinase

TF: Transcription Factor

TFBS: Transcription Factor Binding Sites

TGA: TGACG Sequence-specific Binding Proteins

WRKY: WRKYGQK-containing transcription factors

ZT: Zeitgeber Time

[vii]alternatively known as Cold Regulated 78 (COR78) or Low Temperature Induced 78 (LTI78)

Terminology

Compound gene regulation: Changes in gene expression induced by more than one type of environmental signals.

Compound logic: A classification framework used to identify different outcomes of signal integration based on approximation of biological processes with logical variables.

Control structure: Topological feature of a biological system that governs the behaviours of system.

Crosstalk: Convergence of biological signals facilitated by sharing of a signalling component between multiple signalling pathways, or direct interaction between components belonging to different signalling pathways.

Signal integration: A function of biological system that allows processing of multiple input signals to control a single designated outcome.

Signalling pathway / network: Molecular systems consisting of chains of biochemical events to propagate biological signals across interior of a cell. In this thesis, pathways refer to groups of molecules responsible for relaying a single type of input signal. For description of systems receiving multiple inputs, the term signalling network is used instead.

Chapter 1

Introduction

1.1 Importance of understanding plant stress response

As sessile organisms, plants possess remarkable abilities to adapt to surroundings. Fundamental to such abilities are the molecular mechanisms that allow plant cells to sense changes in their environments and subsequently commit their resources toward eliciting appropriate responses. Hostile changes in the environments leading to deviations from the physiological norm are generally referred as *stress*, which may arise from combination of suboptimal or damaging quantities of environmental factors such as water, light, and nutrient sources [221]. As responses to these changes, plants employ various molecular stress response mechanisms in order to maintain osmotic and ionic balance [227, 229], to protect and repair damaged proteins and membranes [45, 258], or to fend off incoming pathogenic attacks [49, 228].

Elucidation of the molecular stress response systems in plants holds significant social and economic interests, mainly concerning food security. The world population stands at approximately 7.349 billion individuals, which is expected to increase to 9.7 billion individuals by 2050 [249]. Such rapidly increasing population motivates searches for the ways to effectively address various types of environmental stress, which are the major factors limiting the global agricultural output [87]. For example, recent analysis has suggested that approximately 59%

of irrigated areas used for agricultural purpose is under extremely high water stress, where local water withdrawal far exceeds available water supply [82]. The climate changes introduced by global warming and changing geographical location of pathogens also create additional pressure for development of crop species that are resistant to environmental stresses [213]. Pathogen infection by fungi and bacteria is also responsible for significant yield loss in commercially important crops worldwide annually [187]: *Pseudomonas syringae* is particularly problematic for crops, due to its pathogenicity to over 180 plant species [31, 95]. The traditional methods of farm management, such as the use of pesticides and herbicides are costly, and are being met with increasing resistance from target pest species [83, 216].

The ability to produce larger quantities of food therefore depend on a variety of crops that can withstand various types of environmental stresses while having increased level of yield [87]. The traditional approach to develop stress-tolerant crop species is selective breeding, which is based on selection and enhancement of desirable phenotypic traits by controlled pollination. With the advent of post-genomic technologies, this approach has been successful in associating the genetic background in crop species with a phenotype [251]. Breeding crops for resistance to a particular type of stress, however, often cause reduced grain yield due to strong selection pressure for resistant genes. Traits associated with drought resistance such as small plant size, reduced leaf area, early maturity and prolonged stomatal closure lead to a reduced total seasonal evapotranspiration, which cause a reduced yield potential [65, 116]. Even with automated, high-throughput breeding programmes for identification of the lines with desired traits, which often involve evaluation of a million genetic variants each year [134], development of newly-bred stress resistant crops with decreased yield gap still remains as a significant challenge.

Another approach involves genetic modification, which utilises molecular genetic techniques to introduce new phenotypic traits that do not naturally occur in the targeted crop species [254, 258]. Unlike selective breeding, genetic modification permits specific choice over which gene functions to be amplified or attenuated. Such level of control has led to exploration of alternative ideas for developing stress-resistant crops, such as transformation of naturally

stress-resistant plant species as crop plants [66]. There have been some successes in engineering stress tolerance in crop: transgenic rice overexpressing the genes involved in production of trehalose that grow better in saline soil compared to control plants [81]. Developing genetically modified (GM) stress-resistant crop for widespread commercial use, however, has proven difficult because most transgenic plants enhancing the activity of a particular gene associated with stress resistance failed to thrive in field environments [79, 161, 171], or are associated with undesirable side effects such as impaired growth [149, 245].

1.2 *Arabidopsis thaliana* as model organism

The challenges in developing stress-resistant crop species, driven by socioeconomic pressures, have led to increased efforts to elucidate the molecular mechanisms of stress response gene regulation in plants. The current understanding of the molecular mechanisms responsible for responses to a variety of stress types has been pioneered by the studies of a small flowering plant, *Arabidopsis thaliana*. Since the completion of its genome sequence in 2000 [9], *A. thaliana* has been at the focal point of functional genomics research, serving as a model for other plant species and eukaryotic organisms in general [231]. *A. thaliana* has several key advantages as the *in vitro* investigative platform for various aspects of plant biology, such as rapid and inexpensive cultivation, and its ability to grow in different environmental settings, such as petri dishes, greenhouses [164], or even microfluidic devices for high-throughput screening [163]. Manipulation of its genome is facilitated via generation of transgenic lines carrying T-DNA insertions from *Agrobacterium tumefaciens* [40, 48], allowing studies of phenotypic changes from genetic perturbations.

While the species itself does not hold agronomic significance, *A. thaliana* is often used as a reference system for commercially important crop species. Direct analysis of common crop species is hindered not only by lack of molecular tools and growth logistics, but also higher complexity of genome; for example, Maize (Zea mays L.) has 2.3-gigabase genome consisting of 10 diploid chromosomes [219], which is approximately 20 times greater than that

of *A. thaliana*. Molecular understanding of *Arabidopsis* genome therefore provides a starting point for understanding the crop plants. Many *Arabidopsis* genes involved in response to environmental stresses have led to homology-based discovery of novel gene functions in crop species [200]. Given that stress tolerance is a multigenic trait [28, 254], however, many challenges remain towards developing stress-resistant crop species. More detailed elucidation of the molecular mechanisms for stress response in *A. thaliana* is therefore an important milestone towards achieving this goal.

Today, the *Arabidopsis thaliana* community possesses vast amount and variety of publicly available resources, supporting efficient design of hypotheses and models. Many of the transgenic lines are available through seed stock centres such as Arabidopsis Biological Resources Center (ABRC) in the United States, and Nottingham Arabidopsis Stock Center (NASC) in the United Kingdom. Arabidopsis researchers are also well served by comprehensive and integrated online data depositories, such as The Arabidopsis Information Resource (TAIR) [80, 100, 133]. The databases offer easy access to wide variety of information, including complete genome sequence and its genetic and physical map, gene product and ontology, and protein structure.

1.3 Combined stress in plants

As plants are often subject to multiple types of stresses at same time in field environments, the new stress-tolerant crop species must be resistant to simultanoeus presence of multiple types of stresses [28]. It has been shown, however, that development of tolerance to multiple types of stresses in crop species involves more than combining resistance to a single type of stress, because plants do not simply perceive simultaneous presence of multiple stresses as the additive sum of those stresses, but rather as a new environment [167]. As result, co-ocurrence of multiple stresses give rise to unique physiological outcome that cannot be attributed to the sum of response to individual stress: for instance, heat and dehydration stress exhibit antagonistic interaction in control of stomatal opening, where dehydration inhibits increased

transpiration caused by heat stress [211]. Effects of combined stresses are also observed at the level of gene regulation, as transcriptomic profiling studies of plants subjected to different combinations of stresses revealed significant disparity between the subsets of genes regulated by combination of two stresses and those regulated by individually applied stresses [110, 206].

Surprisingly, the theme of combined stresses received little attention in the field of plant stress biology to recent dates, with only around 180 original articles that study combination of two or more stresses in plants [156]. These studies mainly focused on characterising the effects of combining multiple stresses on various plant processes, comparing their influences on physiological characteristics such as growth and molecular changes such as gene regulation. In a simplest term, the outcomes of simultaneously applying multiple stress signals can be either synergistic or antagonistic. Such binate approach, mainly advocated by Mittler [167] and Suzuki et al. [238], has been adopted to describe the modes of non-additive interaction between various stress pairs. Mode of interaction for a specific pair of stress is determined by examining the broad effects on various physiological traits such as growth and yield upon presence of the pair stress. So far, most stress combinations have been shown to interact positively, inflicting greater damage to plants in comparison to singly applied stresses [238]. Drought and Heat, which is one of the most commonly observed stress combination, exacerbate the detrimental effect on photosynthetic capacity and growth [41, 253]. On the other hand, several stress combinations are known to exhibit negative interaction by either mitigating the damage or enhancing tolerance to the other stress. For example, decreased stomatal conductance caused by drought stress can reduce the amount of O_3 uptake and its associated damages [26, 191].

The presence of interactions between multiple stresses in regulating cellular and physiological processes indicate that combined stress is likely to be perceived as a unique environmental condition, rather than a sum of two types of singular stress. The interactions also suggest for connectivity and interdependence of the molecular mechanisms regulating the responses to each type of stress. It has also been argued, however, such approach may be an oversimpli-

fication of the nature of interactions, as it only presents the average of numerous responses that may exhibit wide variability. It has also been suggested that the responses to stress combinations may depend on intra- and inter-specific variation, the order in which the pair of stresses are applied, and the developmental stage [156].

1.4 Aims

The studies of combined stresses in plants are at their beginning, and their significance in development of stress-tolerant crops has started to gain wider recognition across the community [201]. Efforts are being directed to discovery of unknown interactions between different stress pairs in modulating molecular and physiological processes, and under various experimental and developmental conditions. Tangential to these goals, we identify two additional research areas that need be addressed in order to achieve system-level understanding of how plants address stress in realistic field environments:

- **Systematic classification of interaction between pairs of stresses**

  Given the complexity in the outcomes of simultaneous presence of multiple stresses, describing the nature of interactions between two stresses only in terms of synergism and antagonism may be misleading. This is because interaction between pairs of stresses may bring different qualitative changes in the targeted process that cannot simply be described as positive or negative relationship. In the case of gene regulation, for instance, a positive relationship between two stresses may be used to describe the three independent situations interchangeably: i) greater-than-additive increase in the level of expression by a gene that is inducible by individual application of either stress, ii) greater level of expression by a gene that is normally inducible by only a single type of stress, or iii) induction of a gene that is not inducible by any of the singly applied stress upon treatment with combination of two stresses. Such ambiguity undermines assessment of the potential biological consequence of the observed interaction between the given stress pairs. In addition, interaction

between two stresses may also lead to quantitative changes that are specific to intensities, time and order of applied stress inputs, which introduces further complexity in the behaviours of affected genes [14]. Thus, a systematic classification of different modes of interaction between a stress pair would simplify the understanding of potential outcomes of combined stress, and allow investigation of their origin and biological significance.

- **Elucidation of the molecular mechanisms coordinating the responses to combined stress**

  Despite the increasing number of observations regarding the molecular and physiological changes under different stress combinations, the knowledge of the molecular mechanisms responsible for those observed behaviours is lacking. Most of the studies with the aims of explaining various molecular changes in plants, such as regulation of stress response genes, have relied on bottom-up assembly of the observations made for individual components into a larger system. The resulting models are depicted as networks of interactions between various signalling molecules, which often resemble a input-output system connecting the external stress and the molecular/physiological changes in question. Such models provide detailed understanding of the signalling components for relaying different types of stress and have decent predictive power over the qualitative behaviours of the system such as phenotypic changes upon genetic perturbations [105, 280]. Those models are, however, insufficient to provide explanations for the behaviours observed under combined stress settings: for example, transcriptomic analysis of various plant species including *Arabidopsis* [206], tobacco [211], rice [182] and *Sorghum bicolar* [110] under combination of two or more stresses showed that numerous genes are activated specifically under stress combinations, not their individual application. Because many of the traditional models of stress gene regulatory network have been developed upon the observations under single stress conditions, those models cannot explain how the genes specifically activate in response to combined stresses. In order to translate the current knowledege of stress signalling pathways into understanding of how plants respond to multiple simultaneous stress, it is necessary to elucidate the signal integration mechanisms that facilitate crosstalk between

signalling networks for different stress types.

In the light of these needs, we investigate how plants integrate multiple types of stress stimuli present in their surroundings to improve their survival. Here, we specifically investigate control of gene expression by multiple simultaneous stress inputs, which we call *compound gene regulation*. Because changes in gene expression profiles are relatively easy to measure via existing experimental tools [16, 150] and measurement data on gene expression changes upon both single and combined stress settings is abundantly available, there are ample opportunities to explore various outcomes of combining multiple types of stresses. To narrow down the scope of research further, we focus on stress re *Arabidopsis thaliana*. Utility of *Arabidopsis thaliana* as the model system is discussed in more detail in the next chapter.

1.5 Outline

The thesis is organised into two parts based on the aspect of gene regulation affected by integration of multiple stress signals. The first part of thesis examines how combining multiple stress inputs modulate the cellular decision over types of regulation, such as up-regulation or down-regulation. To achieve this, we extrapolate a classification framework developed by Tanaka and Kimura [239] called compound logic, which allows reduction of possible gene regulatory behaviours into finite number of signal integration outcomes. The mechanisms behind each of the signal integration outcomes identified under this framework were then investigated by developing logical network models, which we used to describe the topologies of signalling network. To further investigate utility of the classification framework and the logical network models, a publicly available transcriptomic dataset measuring gene expression responses to single and combined stresses in the model organism *Arabidopsis thaliana* [206] was examined, which led to elucidation of diverse behaviours amongst the genes induced by combined stress. By analysing the behaviours of a small subset of genes known to be regulated by common transcription factors, the logical network models could be implemented to generate novel hypothesis for the structure of the associated signalling network.

The second focus of this thesis is to assess the effect of multiple stress stimuli on temporal dynamics of gene expression. For this aim, we reduced our scope of study by selecting a single gene, *Response-to-Dehydration 29A* (*RD29A*) [269, 270]. The gene is chosen for its inducibility by multiple types of stress stimuli including salt stress (NaCl) and abscisic acid (ABA), which is a principal plant hormone for dehydration signalling. We conduct experimental investigation on *RD29A* expression time course profiles under various combination of NaCl and ABA treatments to examine presence of interaction between the two stress inputs in regulating *RD29A* expression. To explain the observed behaviours of *RD29A* upon various treatment conditions, we construct a dynamical network model for transduction of NaCl and ABA signals based on the current understanding of *RD29A* regulatory system. Comparison between the experimental data and the model outcome reveals that the model is insufficient to reproduce the synergistic effect on *RD29A* expression by combined NaCl and ABA inputs. By systematic investigation of different network topologies, we identify the possible mechanisms that may be responsible for the observed synergistic effect. We show that the new model implementing the proposed mechanisms can correctly predict the qualitative change in the dynamics of *RD29A* expression upon change in NaCl and ABA input conditions.

The chapter details are as follows: Chapter 2 provides a general introduction of plant stress biology. Development of classification framework based on compound logic, and analysis of logical network models for investigation of signal integration mechanisms are presented in Chapter 3, the results of which are applied to understand the transcriptomic profiles upon combined stresses in Chapter 4. The next two chapters explore the latter theme: Chapter 5 presents and analyses the experimental data for *RD29A* expression profiles upon NaCl and ABA treatments, and Chapter 6 describes development and analysis of the mathematical model of *RD29A* regulatory system. Chapter 7 provides general conclusions and suggests further work.

Chapter 2

Regulation of plant stress response genes in *Arabidopsis thaliana*

2.1 Introduction

The researches in plant stress biology has shown extraordinary progress to this date, resulting in detailed knowledge of molecular and physiological mechanisms in response to a wide variety of stresses [96]. Before investigating the molecular basis of plant gene regulation in response to combined stress, a review of the current understanding of the gene regulatory mechanisms involved with singular stress is required. An extensive summary of plant stress gene regulation, however, is beyond the purpose of this chapter, as it is provided by numerous comprehensive texts [97, 208]. Instead, this chapter aims to develop and present our view regarding the current knowledge of stress response gene regulation by focusing on key subsidiary themes, and highlight the gaps that must be addressed in order to integrate the availabile information into understanding of the mechanisms behind compound gene regulation.

As a whole, the researches in plant stress gene regulation can be broadly divided into two separate themes: i) identification and functional characterisation of transcription factors (TFs), and ii) elucidation of the intracellular signalling network. The studies investigat-

ing the former theme employ various techniques ranging from comparative genomics approaches to identify loci in genome coding for TFs [209], molecular characterisation of TF-DNA binding [226, 270, 278], and assessment of phenotypic effects from perturbing activities of those TF genes [75, 214, 250]. The latter cohorts of studies focus on the molecular mechanisms for perception of stress signals [168, 190, 195, 257], identification of intracellular signalling components for transduction of the stress signals, and their functional characterisation [102, 165, 274]. Convergence of the insights from these researches led to models of signalling pathways for transduction of various types of stresses, which have been traditionally depicted as sequential cascades [180, 267]. Figure 2.1 provides a summary of the core signalling paradigm for stress response gene regulation in plants that emerges from examining the models of various stress signalling pathways.

Identification of TF genes and their corresponding target DNA sequences (cis-elements) has

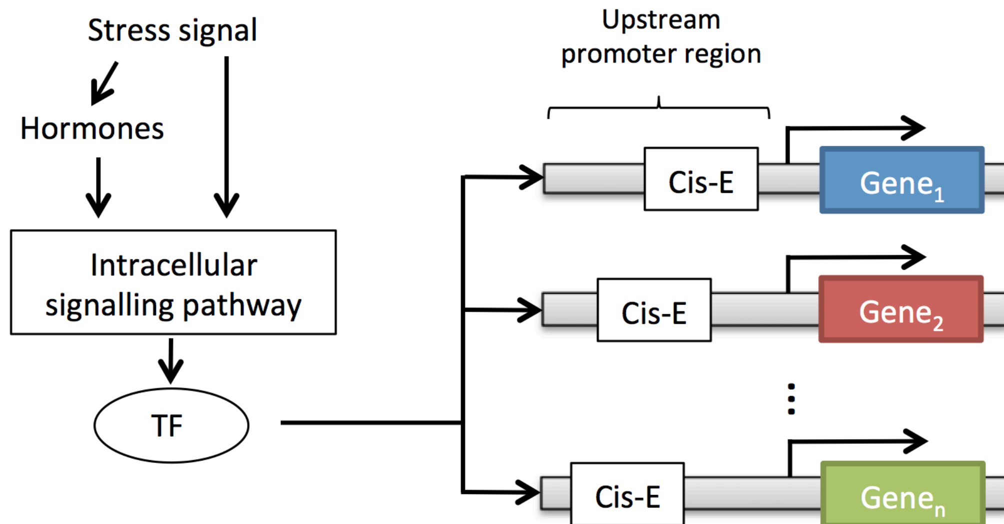


Figure 2.1: **Control of stress regulon as the core signalling paradigm for stress-induced gene regulation in plants**. A stress regulon consists of a group of stress response genes ($Gene_1$, $Gene_2$, ..., $Gene_n$) with various molecular functions that are regulated together due to presence of a common type of cis-element (Cis-E) in their upstream promoter regions. Intracellular signalling pathways for regulation of stress response gene expression have been typically classified into two groups, depending on involvement of corresponding stress hormones in transduction of stress signals for induction of TF activities.

revealed the fundamental basis for transcriptional control of stress response genes. In plants, stress response genes that contain common cis-elements are regulated together as a single functional unit called *regulon* [181]. Organisation of stress response genes into regulons serve a similar purpose to having bacterial operons by facilitating tighter spatial and temporal control over expression of the member genes [136]. Plant stress regulons, however, differ to bacterial operons in that they do not involve synthesis of polycistronic messenger RNAs, and exert control over genes spatially distributed across the plant genome. Because each gene in a stress regulon possesses its own unique promoter architecture, finer control of expression is also possible via combining different types of cis-elements. The cis-elements that orchestrate the responses to abiotic and biotic stresses have been identified in *Arabidopsis*: the best-studied examples of stress-inducible cis-elements include the Dehydration-Responsive-Element / C-Repeat (DRE/CRT) for drought-, salt-, and cold-induced gene expression [270], the ABA-Responsive-Element (ABRE) [159, 175] for drought-, salt- and ABA-induced gene expression, and as-1 [112], W-box elements [276] for biotic-stress-induced gene expression.

Detailed knowledge of TFs has also helped to concentrate the efforts in characterising the intracellular signalling pathways that induce the activities of the TFs upon presence of stress signals. Traditionally, studies of plant stress signalling have focused on elucidating the roles of hormones that are synthesised in response to environmental stresses. The major plant stress hormones include abscisic acid (ABA), which has long been regarded as the principal transducer of various abiotic stress signals [248], and jasmonic acid (JA), salicylic acid (SA) and ethylene (ET) for their roles in transduction of biotic stress signals [22]. Although the activities of the TFs in stress response gene expression are regulated by a highly complex network of molecular interactions, understanding of the relationship between the TFs and the stress-induced hormones has greatly simplified the reconstruction of intracellular signalling network architecture. For instance, TFs that are known to regulate subset of stress-inducible genes have been often assigned into hormone-dependent and hormone-independent pathways. Notable examples include identification of the TFs involved in ABA-independent and ABA-dependent signalling pathways for regulation of drought and salt-induced genes [271], and

the TFs in SA-dependent and SA-independent regulation of PR regulon [199].

This chapter first briefly discusses the physiological, cellular and molecular aspects of stress response in plants, with focus on *Arabidopsis thaliana*. The chapter then reviews the studies on molecular mechanisms of stress responses in three subsidiary themes: action and synthesis of stress-induced hormones (with focus on ABA - Section 2.3), transcriptional regulation (with focus on TFs and cis-elements - Section 2.4), and intracellular signal transduction (with focus on abiotic and biotic signalling pathways - Section 2.5).

2.2 Roles of stress-regulated genes

To understand how plants respond to hostile changes in the environment, it is useful to distinguish different types of stresses. In plants, environmental stresses have been broadly categorised into two groups: *abiotic stress* represents suboptimal environmental conditions imposed by physical and chemical factors, such as high salinity, drought, high and low temperature, strong light, UV, heavy metals and hypoxia. *Biotic stress*, on the other hand, denotes damages inflicted by other organisms such as bacterial and fungal infection, and physical wounding caused by insects and other herbivorous animals. Each type of stress imposes different effects on plant physiology. Figure 2.2 summarises the key damaging effects introduced by each type of stress.

Abiotic and biotic stress factors detrimentally affect various cellular aspects such as osmotic, ionic and metabolic balances, photosynthesis, membrane integrity, proteome stability, and redox status [97]. While each stress may arise from distinct environmental conditions, the type of cellular and physiological damages caused by different stresses show significant overlaps (Fig. 2.2). This is exemplified from the effects of salt, drought and cold stress, which all lead to reduction of plasma membrane integrity: in the case of drought stress, intracellular water potential decreases due to removal of water in the aqueous environment surrounding the membrane, which subsequently leads to reduced hydrostatic pressure [32]. In the case

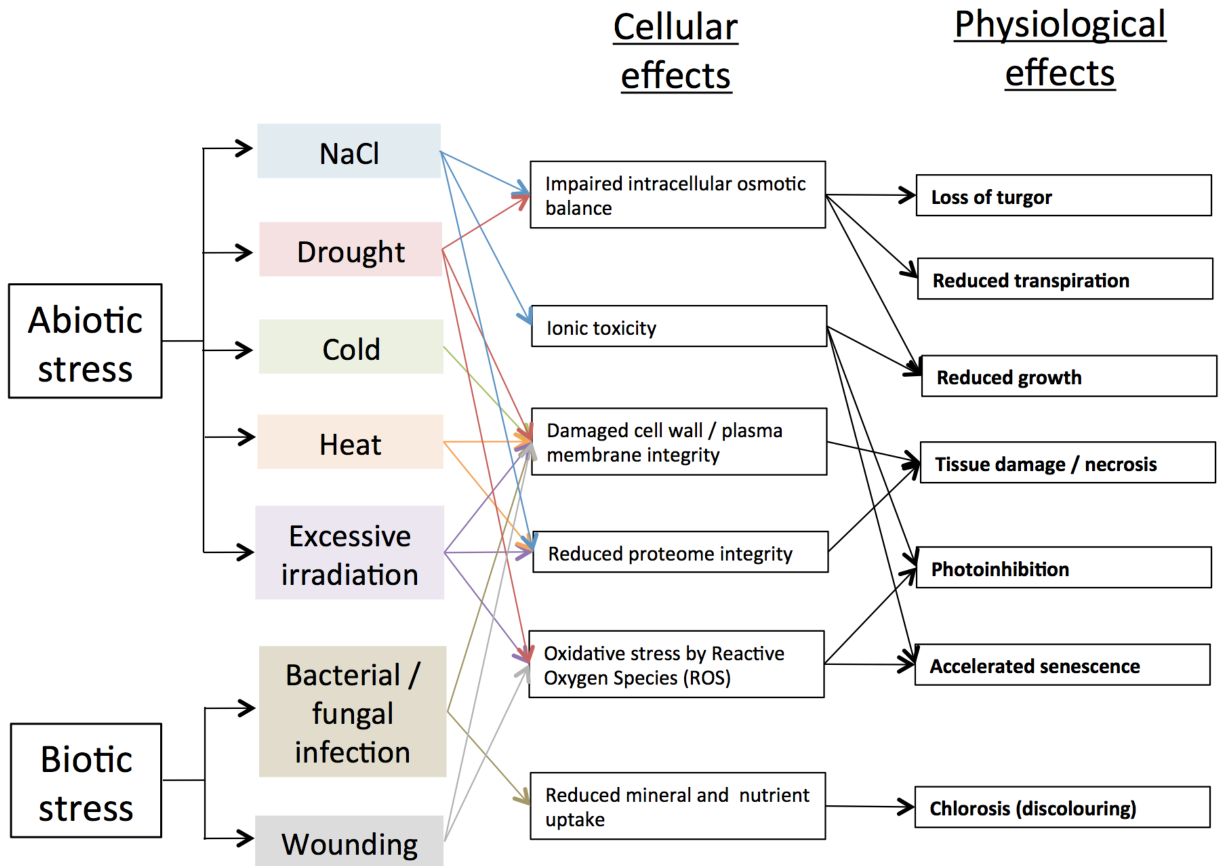


Figure 2.2: **Damaging effects caused by common environmental stresses in plants**. Each type of stress elicits both distinct and common cellular effects, which subsequently causes various types of impairments in the physiological processes of plants. Notably, all stress types lead to reduced cell wall and plasma membrane integrity, suggesting that ubiquitous responses may be involved in mitigating the damages incurred by different stress types. The information regarding cellular and physiological effects of each stress type has been obtained from [97].

of salt stress, membrane structure becomes destabilised via changes in static charge balance across the membrane surface [237]. Cold stress, on the other hand, also directly affects the mechanical properties of plasma membrane via perturbation of hydrophobic interaction within lipid bilayer, causing the membrane to become less fluid [45, 257].

The apparent similarity in some cellular effects induced by different types of stresses have led to a hypothesis that responses to different stress factors might involve common molecular response mechanisms to address the cellular damages caused by the stress. Indeed, stress response genes inducible by different types of stress show extensive overlap. From the microarray profiles under different stress conditions in *Arabidopsis* and bioinformatic analyses of the molecular functions, it has been suggested that the abiotic-stress-inducible genes typically encode proteins responsible for protecting cells from stress such as molecular chaperones, LEA (late embryogenesis abundant) proteins, osmotin, antifreeze proteins, mRNA-binding proteins, key enzymes for osmolyte biosynthesis such as proline, water channel proteins, sugar and proline transporters, detoxification enzymes, enzymes for fatty acid metabolism, proteinase inhibitors, ferritin, and lipid-transfer proteins [129, 222, 272]. Similar microarray analyses have shown that the genes inducible by biotic stresses encode proteins with antibacterial, antifungal and insecticidal effects, various catalytic enzymes such as chitinases, peroxidases, ribonucleases and lysozyme, and the enzymes leading to enhanced integrity of cell wall such as those involved in lignin-biosynthesis [53, 157].

2.3 ABA action and biosynthesis

The similarity of molecular changes in responses to different stresses is at least partially contributed by the actions of common hormonal regulators at upstream of those processes, synthesisesd in response to multiple types of environmental stress. All of the hormones mentioned above (ABA, JA, SA, and ET) are known to induce diverse yet designated sets of genes. The stress-induced hormones can be transported across long distances via xylem or phloem, diffuse locally between cells, or are maintained in their site of synthesis to exert

their influence on target cells [69]. Before discussing how the molecular responses to counter stresses are triggered, it is necessary to understand how the stress-induced hormones are synthesised in responses to environmental stresses. Since ABA has been most intensively studied amongst the stress-induced hormones so far with detailed information regarding synthesis, signalling and molecular actions, we mainly focus on ABA in this section.

ABA is a small compound ($C_{15}H_{20}O_4$) with weak acidic properties, first identified 50 years ago as a growth inhibitor accumulating in abscising cotton fruits [3]. It is a ubiquitously found in plants and also produced in some phytopathogenic fungi, bacteria and protozoans [260]. Since its initial discovery, ABA has been found to be involved in many cellular and physiological processes in various stages of development: ABA plays a crucial role in seed maturation and maintenance of dormancy by mediating transition from growth via cell division to growth via cell enlargement and accumulation of storage reserves [144]. Maintenance of seed dormancy is attributed to continued production of ABA [141], whereas transition from dormancy to germination involves decreased ABA content relative to gibberellin via ABA catabolism [189]. ABA is also an inhibitor of floral transition, as shown by early and delayed flowering in ABA-deficient and ABA-hypersensitive mutants compared to wild type plants, respectively [131, 152]. While the effects of ABA on many growth-related processes appear to be inhibitory, however, some low levels of ABA may be required for normal growth of plant. This is shown through stunted growth exhibited by ABA-deficient plants in comparison to wild type plants owing to reduced transpiration and turgor [64].

More importantly for us, plants use ABA as a molecular signal for various types of environmental stress during vegetative growth stage. Its concentration substantially increases under environmental stresses such as drought, salinity, cold, pathogen attack and UV radiation [63]. For example, the baseline ABA level in maize measured under normal condition was found to be in range of 300 to 400 pmol g^{-1} fresh weight[i] (FW), which increases up to 6000 pmol g^{-1} FW in response to dehydration stress within 24 hours [207]. Such increase in ABA content leads to numerous changes that are beneficial to the plants under stress: at

[i]Fresh weight denotes biomass that includes water content, as opposed to dry weight.

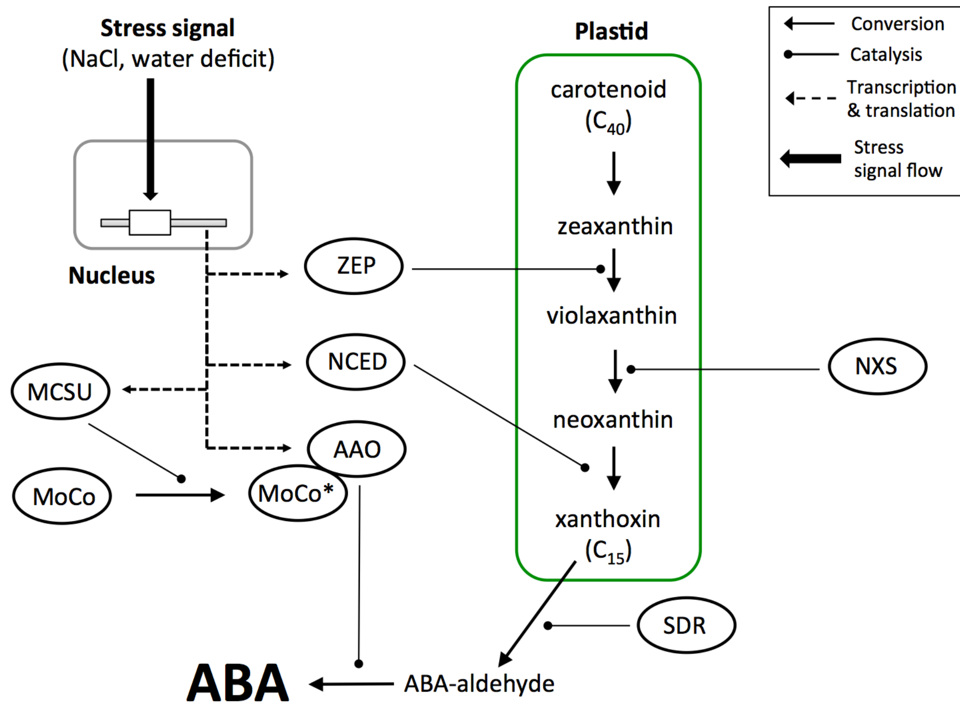


Figure 2.3: **Regulation of metabolic pathways leading to ABA biosynthesis**. The ABA biosynthetic pathway starts from zeaxanthin, which is a non-specific precursor to ABA generated from the 40-carbon epoxycarotenoid reserve. Conversion of zeanxanthin to the next intermediate, violaxanthin, is mediated by zeaxanthin epoxidase (ZEP). Violaxanthin is subsequently converted to neoxanthin, a structural isomer of violaxanthin, through the action of neoxanthin synthase (NXS). Next, 9-cis-epoxycarotenoid dioxygenase (NCED) cleaves neoxanthin into 15-carbon xanthoxin, the ABA-specific precursor. Xanthoxin translocates to cytoplasm, and is converted to ABA-aldehyde by short-chain alcohol reductase (SDR). The final step of the pathway involves oxidation of ABA-aldehyde into ABA, catalysed by ABA-aldehyde oxidase (AAO). The activity of AAO requires presence of molybdenum cofactor (MoCo), which is activated (MoCo*) by sulfurylation catalysed by MoCo sulfurase (MCSU). Stress induces ABA biosynthesis by induction of the key regulatory genes, namely ZEP, NCED, AAO and MCSU. The information from this figure has been obtained from [63] and [268].

the physiological level, ABA induces stomatal closure to minimise water loss through transpiration [123]. It also mitigates cellular damages incurred by stress through up-regulation of stress-responsive genes, which encode enzymes for the production osmolytes and proteins to enhance cellular integrity [64]. Thus, the changes triggered by ABA collectively lead to increased tolerance of plants to the specific environmental stress they are exposed to.

Synthesis of ABA is a complex process with numerous regulatory steps, and understanding its mechanism is crucial for understanding how plants respond to various types of environmental stress. Figure 2.3 outlines the metabolic processes leading to *de novo* synthesis of ABA, generalisable to different plant species. In summary, the current understanding suggests that presence of stress results in induction of several genes, which encode enzymes or cofactor proteins involved in different steps of ABA biosynthesis. Increased rates of reaction from greater availability of those gene products consequently drive the whole pathway forward. Deletion or loss-of-function mutants of those genes lead to ABA-deficient phenotype showing severely impaired resistance to drought and salt stresses [63]. Among those genes, 9-cis-epoxycarotenoid dioxygenase (NCED) is regarded as the key regulator of ABA biosynthesis controlled by stress, as the reaction it catalyses is mostly the rate-limiting step for the whole pathway [268]. The extent at which expression levels of other ABA biosynthetic genes are regulated from stress varies between different organs and plant species: for example, salt and drought stress have little effect on expression of zeaxanthin epoxidase (ZEP) in tobacco [15] and tomato [243], whereas ZEP is strongly induced by those stresses in Arabidopsis [266].

Synthesis of other phytohormones also involve mechanisms similar to the example of the ABA biosynthetic pathway, where perception of stress in the environment leads to induction of numerous biosynthetic enzymes and other regulatory genes involved in production of the target compound. The detailed mechanisms of how environmental stresses trigger hormone biosynthesis in plants have been extensively reviewed elsewhere (SA synthesis - [43], JA synthesis - [247], and ET synthesis - [256]). Taken together, understanding of stress-induced hormone actions and biosynthesis provided important insights regarding the molecular changes occuring from initial perception of stress to early stages of stress response.

2.4 Transcription factors

At the other end of plant stress research, numerous studies focused on identifying and characterising the transcription factors that regulate stress response genes. Since sequencing of *Arabidopsis* genome, 1533 genes were initially identified as TFs (5.9% of genome) [209]. These numbers are particularly high compared to that of other eukaryotic organisms such as *Drosophila menlanogaster* (635 genes, 3.5% of genome) and *Caenorhabditis elegans* (669 genes, 4.5% of genome) [209], suggesting that regulation of transcription in *Arabidopsis* is highly complex and diversified. There are currently four representative online databases of *Arabidopsis* TFs: AGRIS[ii] [51], RARTF[iii] [103], DATF[iv] [90] and PlnTFDB [v][196].

Each transcription factor is classified into different families according to their characteristic DNA-Binding Domains (DBD), which recognise unique base pair sequences in the upstream cis-regulatory region of the targeted genes [153], or their ability to form a transcriptional complex with a particular interacting TF partner. Due to different criteria used for definition of TF families, the number of genes assigned to each family of TF slightly varies amongst the databases. For example, AGRIS ignores auxiliary transcriptional regulator proteins that interact with TFS but do not directly bind DNA, while the others classify them as TFs [166]. Merging the information from each of the four databases described above, there are now 72 *Arabidopsis* TF families discovered so far [98]. While many of the TF families identified in *Arabidopsis* are conserved across other species, there are also TF families that are exclusively found in plants. A summary of six major TF families (three general, three plant-specific) that are known to play important role in regulation of stress response gene expression is shown in Table 2.1.

TF proteins in the Basic region/leucine zipper (bZIP) family are characterised by distinct structure consisting of Basic region (BR) that facilitates DNA-binding, and leucine zipper

[ii] Arabidopsis Gene Regulatory Information Server (http://arabidopsis.med.ohio-state.edu/AtTFDB/)
[iii] RIKEN Arabidopsis Transcription Factor database (http://rarge.gsc.riken.jp/rartf/)
[iv] Database for Arabidopsis Transcription Factors (http://datf.cbi.pku.edu.cn/)
[v] Plant Transcription Factor Database (http://plntfdb.bio.uni-potsdam.de/v3.0/)

Table 2.1: Summary of six major *Arabidopsis* transcription factor families involved in stress response gene regulation

Specificity	**Family name**	**Cis-element**	**Consensus Sequence**	**Notable TFs**	**Stress factor**
General	bZIP	as-1 (G-box)	CCACGTGG	TGA1-7	Pathogen infection [218]
		ABRE	(C/T)ACGTGGC	AREB1-3	Drought, Salt, ABA [250]
	HSF	HRE	AGAANNTTCT	HsfA1	Heat [220]
	bHLH	G-box	CACGTG	MYC2	Drought, Salt, ABA [1]
Plant	AP2/ERF	GCC-box	GCCGCC	ERF1-5	Wounding, Salt, Drought, ABA [72]
		DRE	(A/G)CCGAC	DREB1,2	Drought, Salt, Cold [149]
	NAC	NACR	CATGTG	RD26	Drought, Salt, ABA [73]
	WRKY	W-box	(T)TGAC(C/T)	WRKY22,29	Pathogen infection [10]

region (ZR) that allows dimerisation [118]. In *Arabidopsis*, there are 76 putative bZIP TFs identified to this date [98], which play diverse roles not only related to abiotic and biotic stress responses, but also in light signalling, seed maturation and flower development too [108]. A notable bZIP TF orchestrating biotic stress response is TGACG Sequence-specific Binding Proteins (TGAs), which are principal regulators of Systemic Acquired Resistance [67]. TGA binds to activation sequence-1 (as-1) motif, a variant of G-box element commonly found in upstream promoter regions of Pathogenesis-Related (PR) genes [218]. Deletion of various TGAs lead to loss of basal resistance [120], suggesting their significance in pathogenic defence coordination. Another class of TF in bZIP family with a major role in abiotic stress signalling consists ABRE-Binding Proteins (AREBs) [250]. Suppressing expression of the AREB genes lead to reduced ABA sensitivity and drought resistance [275], implying their significance in mediating responses to drought stress.

The Heat-Shock Factors (HSFs) family forms another generic TF family found across different organisms, which play central role in heat stress response. The main downstream targets of HSF proteins are Heat Shock Proteins (HSP) involved in orchestration of folding, intracellular distribution and degradation of proteins [185, 259]. Specificity to HSF binding is conferred by Heat-Shock Element (HSE) that is found in the promoters of HSP genes [186]. A notable example in this family is HsfA1, which is the master regulator of thermotolerance in *Arabidopsis* [220].

The Basic helix-loop-helx (bHLH) family is the second largest TF families in Eukaryotes [209], with 162 putative members identified in *Arabidopsis* [18]. The bHLH proteins are characterised by two functional domains, with the basic domain for contacts with recognised DNA sequences and HLH domain for oligomerisation. The bHLH proteins bind to various types of cis-elements, but G-box is known to be their most common target [11]. Due to diversity in structure and function, the bHLH TFs are classified into several classes and subfamilies [244]. A notable member of the bHLH family that facilitates regulation of abiotic stress response in *Arabidopsis* is MYC2, which is expressed in response to dehydration stress and ABA [2]. Overexpression of MYC2 leads to higher sensitivity to ABA and osmotic stress

tolerance in transgenic *Arabidopsis* [1]. Interestingly, MYC2 is also known to regulate defense genes in response to JA and ET [29, 151], suggesting that it facilitates crosstalk between ABA-induced abiotic stress response and JA/ET-induced biotic stress response.

The APETALA2/Ethylene-Responsive Element Binding Factor (AP2/ERF) family constitutes one of the major groups of transcription factors that are specific to plants. The members of this family possess a highly conserved region of 60 to 70 amino acids (AP2 domain) originally identified from APETALA2, a homeotic gene involved in *Arabidopsis* flower development [210]. Within the AP2/ERF family, two classes of TFs are of interest with respect to regulation of stress response gene expression: Ethylene-Response Element Binding Factors (ERF) recognises GCC-box [188], which is commonly found in the promoters of biotic stress response genes. Another important class of AP2/ERF TFs consists of Dehydration-Responsive-Element Binding (DREB) TFs, which target DRE to regulate stress response genes. The DREB TFs are further classified into into two functional subgroups: DREB1/C-Repeat Binding Factor (DREB1/CBF) for cold stress responses, and DREB2 for dehydration stress responses [149].

Numerous members of the NAM/ATAF1,2/CUC2 (NAC) family of TFs are also recognised as abiotic and biotic stress response regulators [204]. NAC proteins contain highly conserved region of approximately 150 amino acids and bind to NAC recognition (NACR) sites as multimers [246]. An important member of NAC family is RD26, whose role in ABA-dependent signalling of salt and dehydration stress has been established via differing ABA sensitivities in transgenic plants with changes in RD26 expression [73]. Microarray analysis showed that ABA- and stress-inducible genes were upregulated in RD26-overexpressing plants and repressed in RD26-repressed plants, indicating that a cis-regulatory element, the NAC recognition site [73], may function in ABA-dependent gene expression under stress conditions.

Finally, the WRKY TFs form another plant-specific family of TFs with established roles in biotic stress signalling [193]. The WRKY TFs are named after its highly conserved 60

amino acid long WRKY domain, which contains WRKYGQK at the N-terminus and novel zinc-finger-like motif at the C-terminus [47]. The WRKY TFs bind to consensus cis-element called W-box found in many *Arabidopsis* defence genes [59]. The best-known TFs in the WRKY family are WRKY22 and WRKY29, which play a central role in regulation of innate immune response gene expression and disease resistance [60]. They are important components of MAPK-cascade for Pathogen-associated Molecular Patterns (PAMP) signalling, and enhancement of their activities via overexpression of WRKY22 and AtWRKY29 increases resistance to both bacterial and fungal pathogens [10].

2.5 Intracellular signalling pathways for stress-induced gene regulation

Connecting the insights gained regarding the action and biosynthesis of plant stress hormones and transcription factors acting at downstream led to elucidation of key intracellular signalling pathways. Here, we selectively review the current knowledge of the signalling pathways known to play important roles in regulation of genes in response to various abiotic and biotic stresses.

2.5.1 Abiotic stress signalling pathways

Many abiotic-stress-inducible genes are controlled by ABA, but some are not, suggesting that both ABA-dependent and ABA-independent regulatory systems are involved in stress-responsive gene expression (Fig. 2.4). Resposes to drought and salinity stresses are known to be regulated via both ABA-dependent and ABA-independent pathways: the AREBs represent the major components of ABA-dependent signalling pathways for transduction of drought and salt signals along with several members of MYC and NAC family of TFs also known for regulating in ABA-responsive gene expression, whereas DREB2s mediate

the drought and salt-induced gene expression in absence of ABA. In contrast, heat and cold stress response appears to be regulated essentially by ABA-independent mechanisms [4], with HsfA1 and DREB1 as the major TFs for heat- and cold-indued gene expression, respectively.

The intracellular signalling pathway leading to activation of AREBs is the best-studied example for ABA-dependent regulation of gene expression, and the only complete sig-

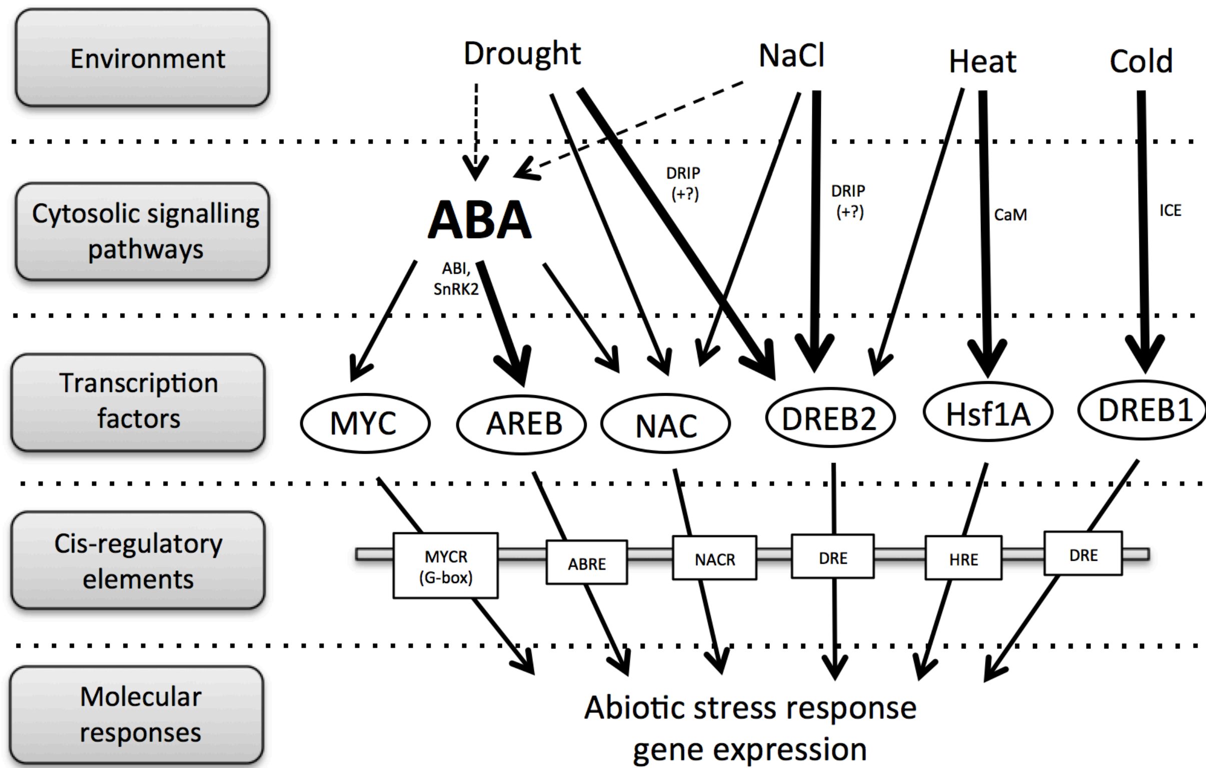


Figure 2.4: **Signalling pathways for transduction of various abiotic stress signals**. The stress signals flow across different signalling layers, which are separated by dotted lines. Genes for drought and salt responses are largely controlled by ABA-dependent and ABA-independent pathways, each comprising of TFs (ellipses) that recognise specific cis-elements in the targeted genes (Boxes). Responses to Heat and Cold largely occur in a ABA-independent manner. Bold arrows represent the signalling pathways that are characterised in greater detail compared to other pathways, with highlights on the major signalling components. Dashed arrows indicate the pathways for ABA biosynthesis. The figure is based on the information mainly from [271] and other references mentioned in the text.

nalling pathway with detailed knowledge regarding initial signal perception and transduction. Pyrabactin resistance 1 / Pyr-likes / Regulatory Component of ABA Receptors (PYR1/PYLs/RCARs) were identified as a putative receptor for ABA perception, which triggers downstream phosphorylation cascade for activation of AREB. PYR1/PYLs/RCARs interact with type 2C protein phosphatases (PP2C) such as ABA Insensitive 1 and 2 (ABI1 and ABI2) [165], which blocks downstream signalling by dephosphorylating SNF1-related kinase 2 (SnRK2)-type protein kinases (SnRK2) in the absence of ABA. Upon ABA accumulation, PYRs/PYLs/RCARs bind to inactivate PP2Cs, eliminating the inhibitory effect on SnRK2 kinases. SnRK2 kinases switch to the phosphorylated state, which in turn activate AREBs by the phosphorylation of multiple Ser/Thr residues [77]. There are also evidences for involvement of Ca^{2+}-dependent protein kinase (CDPK) in phosphorylating AREBs in response to increased intracellular Ca^{2+} levels induced by various stimuli [115, 281], suggesting that AREBs may integrate multiple signal inputs under abiotic stress conditions.

The DREB2 proteins have been identified as the ABA-independent regulators of gene expression, as the DREB2 genes were shown to be induced by both drought and salt stresses, but not ABA [149]. Although the complete information regarding signalling pathway architecture for activation of DREB2 is still missing, the DREB2 regulatory system is relatively well characterised compared to other stress-mediating TFs. Overexpression of DREB2 genes did not result in any phenotypic changes in transgenic plants [149], suggesting that DREB2 transcriptional activity requires post-translational activation. Domain analysis of DREB2 revealed that the central region of DREB2 contains a Negative Regulatory Domain (NRD) and that deletion of this region makes DREB2 constitutively active (DREB2A-CA) [214]. The NRD in DREB2 has been proposed to regulate *in vivo* stability of DREB2 proteins, as recombinant DREB2A-CA proteins without NRD is stable in the nucleus, while the native DREB2A proteins are not. Regulation of DREB2 stability is thought to occur via protein degradation mediated by DREB2-interacting proteins (DRIPs), which negatively regulates DREB2 abundance via ubquitin-mediated proteolysis [205]. A recent finding, however, suggests that inhibition of DRIP protein activity by either deletion of their genes or addition

of proteasome inhibitor is not sufficient to induce expression of the genes at downstream of DREB2, suggesting that there may be additional mechanism for activation of DREB2 activity [172]. Overexpression of DREB2A-CA also induced expression of heat shock (HS)-related genes and improved thermotolerance in transgenic plants [214], suggesting the role of DREB2 proteins in regulating Heat stress response. The details of signalling pathway transducing heat signals for activation of DREB2 proteins is still at large.

The signalling mechanisms for regulation of various HSF proteins including HsfA1 are starting to become clarified. Heat stress is initially recognised by various membrane sensors, and it has been proposed that the plasma membrane cyclic nucleotide gated calcium channel (CNGC2) acts as the primary thermosensor [62]. The CNGC2 proteins convert heat stress to intracellular Ca^{2+} signals, which are subsequently perceived by calmodulin (CaM), a ubiquitous calcium-binding messenger protein in plants and other eukaryotic organisms [230]. CaM is an essential component in the heat signal transduction as calmodulin-binding protein kinase 3 (CaMK3) interacts with the HsfA1 [147]. The role of calmodulin in heat signalling has been further established by genetic perturbation of the calmodulin genes, where CaM3 knockout mutant plants in *Arabidopsis* were more sensitive to heat stress, whereas overexpression of CaM3 showed enhanced thermotolerance [279]. HsfA1 exists as transcriptionally inactive monomer in a complex with Hsp40/Hsp70 and Hsp90 in absence of stress, and its activity is only triggered when HsfA1 proteins are released from the complex to form trimers after CaMK-dependent phosphorylation [220]. Notably, the negative regulators Hsp70 and Hsp90 are up-regulated by HsfA1, suggesting that the TF is subject to negative feedback control [91].

Genes inducible by cold stress are regulated by DREB1/CBF, which are DRE binding proteins initially identified with DREB2 [149]. Transgenic *Arabidopsis* plants overexpressing DREB1/CBF genes showed strong tolerance to cold stress [107, 117, 149], suggesting that DREB1/CBF is the major regulator of the cold stress response genes. This contrasts DREB2, which does not lead to stress tolerance when overexpressed. Thus, transduction of cold stress signals via DREB1/CBF has been proposed to occur via transcriptional cascade rather than

enzymatic cascades involving post-translational modification of TFs [250]. Expression of DREB1/CBF is regulated at the transcriptional level by Inducer of CBF Expression 1 (ICE1) [45]. ICE1 is a MYC-type transcription factor that has also been shown to regulate stomata formation [113]. ICE1 proteins are subject to post-translational control via ubiquitination and sumoylation, which exert negative and positive effect on transcriptional activity of the TFs. High expression of Osmotically Responsive gene 1 (HOS1), which a RING finger ubiquitin ligase similar to DRIP [55]. In contrast SUMO E3 ligase (SIZ1) sumoylates ICE1 and enhances its activation of the DREB1/CBF genes [169].

2.5.2 Biotic stress signalling pathways

Transduction of biotic stress signals are also broadly classified as hormone-independent and hormone-dependent signalling pathways. While multiple hormones including SA, JA, and ET play central roles in biotic stress signaling upon pathogen infection and mechanical wounding, there are also signalling pathways that do not require production of any of these hormones (Fig. 2.5). It has been suggested that the apparent multiplicity in number of signalling pathways contribute towards combining generic defence responses with the responses that are specific to the invading pathogens [198]. While conserved Mitogen-Activated Protein Kinase (MAPK) signalling pathways that recognise Pathogen-Associated Molecular Patterns (PAMP) allow responses to a broad range of pathogenic organisms, responses specific to invading pathogen can occur via production of defensive compounds with different composition from prioritisation of either SA-, JA-, or ET-inducible responses. For example, infection with biotrophic pathogens that establish long-term feeding relationship with hosts predominantly results in the activation of the SA-dependent systemic acquired resistance response (SAR) leading to the accumulation of SA-inducible PRs, while JA- and ET-inducible defence responses are triggered by necrotrophic pathogens that kill the host cells to thrive [85, 128, 240]. Wounding also results in the activation of JA- and ET-inducible defence responses, but not SA-inducible defence responses [33].

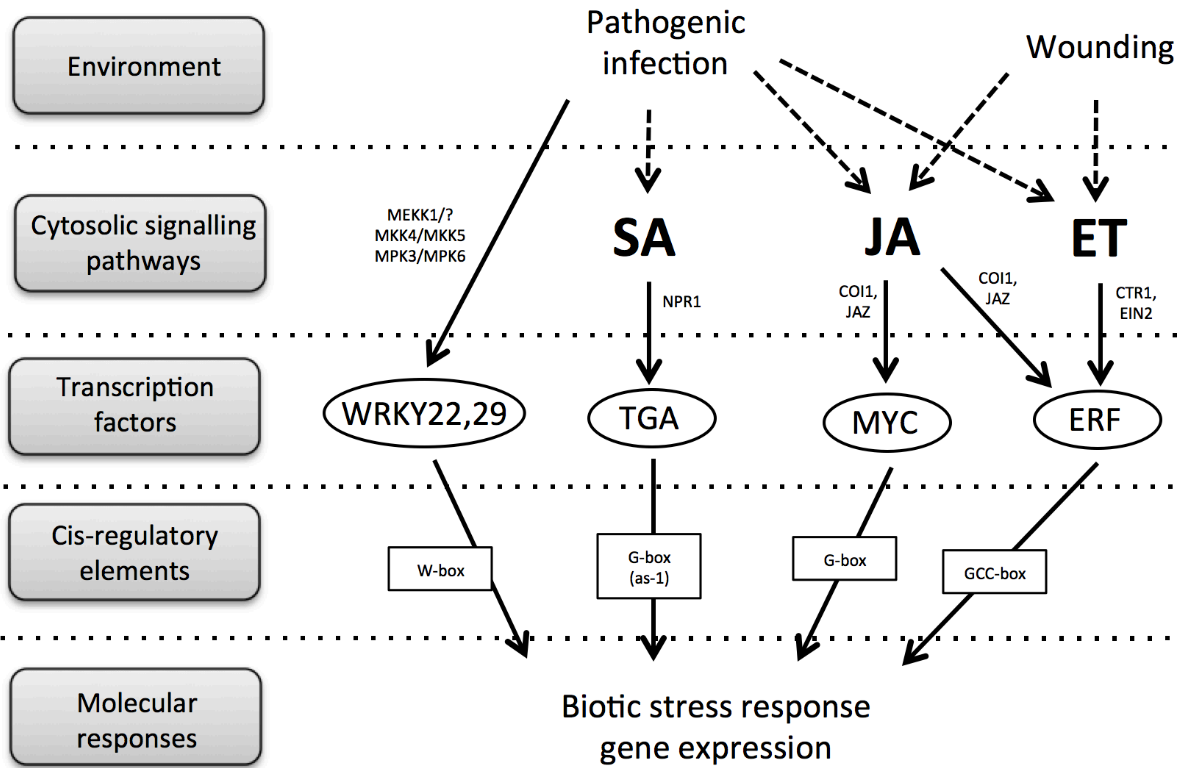


Figure 2.5: **Signalling pathways for transduction of various biotic stress signals**. As before, the pathogenic stress signals flow across different signalling layers, which are separated by dotted lines. Perception of pathogenic infection triggers production of phytohormones SA, JA and ET, which are induce biotic stress response gene expression via TFs (ellipses) that bind to corresponding cis-elements in the targeted genes (Boxes). In addition, a signalling pathway via MAPK cascade is activated in absence of hormones. Wounding stress is primarily transduced by JA and ET, not SA. Dashed arrows indicate the pathways for SA, JA and ET biosynthesis. The information presented in this figure was taken from [22] and other references cited in the text.

The PAMP-induced MAPK signalling of biotic stress signals typically coordinate to the early stage of stress response due to its rapid activation. The MAPK cascades in plants, similar to those found in other eukaryotic organisms, consist of three types of protein kinases that are sequentially activated from MAPK Kinase Kinase (MAPKKK, also known as MEKK), MAPK Kinase (MAPKK, also known as MKK) to MAPK. The best studied example is the flagellin-induced MAPK module, with complete steps of signal transduction from initial perception of stress to regulation of gene expression: the flagellin derived peptide flg22 is detected by FLS2 receptor kinase [88], which subsequently triggers a rapid and strong ac-

tivation of MEKK1-MKK4/MKK5-MPK3/MPK6 cascade [10]. The activated MPK3 and MPK6 proteins induce expression of WRKY22 and WRKY29 TFs, which control expression of defence genes. The mechanism of WRKY22/WRKY29 gene activation by MPK3/MPK6 is not currently understood, but it has been proposed that MPK3/MPK6 phosphorylates to attenuate the activity of transcriptional repressor for WRKY22/WRKY29 genes [60]. Transgenic *Arabidopsis* constitutively expressing of MEKK1 or WRKY29 showed reduced symptoms under *P. syringae* infection, suggesting enhanced activity of this flagellin-induced MAPK module confers greater resistance to pathogenic invasion [10]. However, it is unclear whether MEKK1 is the sole mediator of MPK3/MPK6 activation because deletion of MEKK1 gene does not lead to reduced activity of MPK3/MPK6 [236].

The SA-dependent pathways for defence gene regulation play key roles in plant defense against pathogens, especially regarding whole-plant adaptive responses to pathogens termed systemic acquired resistance (SAR) [128, 130]. SAR helps to restrict pathogen growth and the enhancement of disease resistance in the whole plant after local infection [130, 137]. Both endogenous increase in SA levels or exogenous SA treatments lead to rapid activation of defense responses during which defense related genes such as PR genes are activated. The significance of SA in disease resistance has been demonstrated via SA-deficient mutants, which show reduced resistance to biotrophic pathogens [35, 86, 130]. Intracellular transduction of the SA signal occurs via Nonexpressor of PR1 (NPR1), where increase in the SA level leads to breakdown of inactive NPR1 oligomers to transcriptionally active monomers via redox changes [173]. The NPR1 proteins do not act as a TF, but rather transcriptional co-regulator to induce the activity of other TFs. The primary partner for NPR1 is TGA proteins, which belongs to bZIP family of TFs discussed in the previous section (2.4).

In contrast to the SA-dependent pathway, the JA-dependent pathways coordinate resistance to a series of necrotrophic pathogens [130, 184, 242]. The key components of the JA signaling pathway are the Coronatine Insensitive 1 (COI1) and the Jasmonate ZIM-domain (JAZ) proteins, which act as activator and repressor of JA signalling, respectively. In the absence of JA, JAZ proteins associates with various TFs including MYC2, a key inducer of JA-

dependent defence gene expression, and tightly repress their transcriptional activities [232]. The repression by JAZ is alleviated when COI1, the intracellular receptor for JA, targets the JAZ repressor proteins for degradation by the proteasome [241]. Interestingly, JA signalling via the COI1/JAZ system also regulates the ET-inducible regulons, because JAZ proteins also repress ET-responsive TFs such as ERF [197, 261, 283].

The ET-dependent pathway show significant overlap with the JA-dependent pathway in terms of physiological function, as impaired ET signalling also results in reduced resistance to necrotrophic pathogens [184, 242]. ET signals are perceived transmembrane receptors that are homologous to bacterial two-component histidine kinases [39, 78]. In the absence of ET, the receptors activate Constitutive Triple Response 1 (CTR1), a Raf-like kinase [121]. CTR1 phosphorylates another type of transmembrane proteins found in Endoplasmic Reticulum (ER) called Ethylene Insensitive 2 (EIN2), which in their phosphorylated form are targeted for proteolytic degradation. Inactivation of the receptors from ET binding prevents EIN2 phosphorylation by CTR1, which causes cleavage and nuclear localisation of EIN2 C-terminus and subsequent activation of TFs including ERFs [111].

2.5.3 Crosstalk between stress signalling pathways

With growing knowledge of the stress signalling network architecture in plants, the mechanisms of interaction amongst different types of stresses recently have became the subjects of significant interest. Studies of interaction between abiotic and biotic stress hormones are the prime examples of such initiative (Fig. 2.6). While ABA is defined as a key hormone controlling abiotic stress responses, it also plays important role in biotic stress responses by interacting with biotic stress signals [14, 74]. In general, ABA treatments reduce SAR increases the susceptibility to pathogens by inhibiting SA-induced expression of defence genes, where the two hormones interact antagonistically at both the biosynthetic and signaling levels [273]. Some pathogens exploit such antagonism to suppress SA-induced host defence mechanisms by enhancing production of ABA [27]. For example, a bacterial toxin coronatine

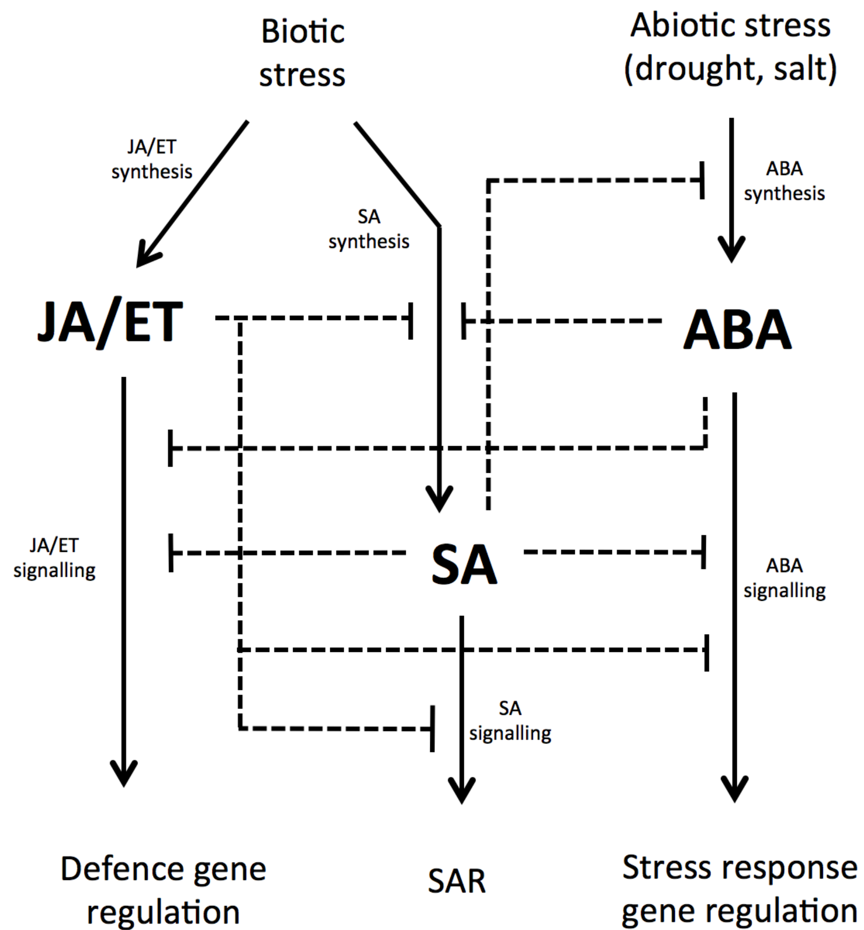


Figure 2.6: **Crosstalk between stress-induced hormones for regulation of stress response genes**. Interactions between biotic and abiotic stress signals (ABA vs. SA, JA/ET), and interactions amongst different biotic stress signals (SA vs. JA/ET) for regulation of stress response genes are characterised by complex network of regulatory cues targeting either synthesis or signalling of other hormones. The information presented in this figure was adapted from [12].

triggers ABA accumulation, resulting in the suppression of SA synthesis [52]. SA also exerts negative effects on drought tolerance mediated by ABA, as the expression of ABA biosynthesis and ABA responsive genes was repressed by signaling downstream of NPR1, the principal regulator of SA signalling [76]. ABA also inhibits JA/ET signaling pathways in a similar manner, as disruption of either ABA-signalling genes or JA/ET-signalling genes resulted in up-regulation of JA/ET-responsive genes and ABA-responsive genes, respectively [7]. These results collectively suggest that there is mutual antagonism between ABA-mediated abiotic stress signalling and SA/JA/ET-mediated biotic stress signalling.

Biotic stress hormones are also known to interact amongst each other (Fig. 2.6). SA antagonises JA/ET-inducible gene transcription by inducing the expression of genes encoding transcriptional regulators that interfere with JA/ET-signaling [130]. These SA-induced regulators could inhibit a positive regulator of JA-inducible gene expression by interacting with it. SA could induce transcription of suppressive transcription factors that directly bind to the promoter of JA/ET responsive genes to repress their expression. Alternatively, JA also suppresses SA production because it induces production of ET-signaling gene EIN3/EIL1, which in turn encodes suppressors of SA biosynthesis [119]. It has been suggested that JA and ET signaling pathways take the precedence over SA mediated defense pathways in plant defense, because in the mutants where JA and ET responses are constitutively activated, the JA responsive genes were not suppressed by SA [143]. Pharmacological assays and mutant studies also illustrated that ET signaling could render the JA signaling insensitive to the suppression of SA [143].

Other than stress hormones, several signalling molecules have been proposed to facilitate crosstalk between the abiotic stress signalling pathways. Intracellular Ca^{2+} regulate a variety of TFs involved in abiotic stress response [127]. For instance, FRY1 locus of *Arabidopsis* encoding a Inositol 1,4,5-*tris* phosphate (IP3)-gated Ca^{2+} channel protein leads to an increase in cytosolic Ca^{2+} in response to ABA, salt, and cold stress signals [265]. Considering activation of HsfA1 against heat stress also involves a CDPK cascade, the role of Ca^{2+} in abiotic stress signalling appears prominent. Another source of crosstalk between abiotic

stress signals is multiple substrate specificty exhibited by some members of MAPK cascades: activities of MAPK4 and MAPK6 proteins are triggered by both cold and salt stresses [102], providing the basis for convergence of multiple stress signals [44]. However, the behaviours of those potential mediators of crosstalk are still poorly characterised in presence of multiple abiotic stress signals, which prevents further mechanistic understanding of their functions in facilitating interactions between various abiotic stress signals.

2.6 Discussion

In this chapter, we have provided a broad introduction of plant stress biology by explaining the role of *Arabidopsis thaliana* as the model plant organism, cellular and physiological damages incurred by different types of stresses, and the functions of the genes that are triggered in response to each stress. The chapter also reviewed the past studies involving the molecular mechanisms of stress response regulation by focusing on the key signalling pathways for transduction of different types of abiotic and biotic stress. The signalling pathways have been reconstructed by combining the insights from the studies on the stress response TFs and their corresponding cis-elements (stress regulons), the stress signalling hormones, and the intermediate signalling genes such as protein kinases and phosphatases, enzymes involved in phospholipids metabolism, and calmodulin-binding proteins. The resulting models of stress signalling pathways are often depicted as sequential cascades of biochemical processes. This reductionistic approach to reconstruct stress signalling pathway has led to detailed mechanistic understanding of signal transduction from initial perception of stress to transcriptional regulation: one of the remarkable examples of such success is full reconstitution of the ABA signalling pathway *in vitro* [71].

Integrating the insights regarding individual signalling pathways has revealed common elements shared by multiple signalling pathways, which are likely to serve as potential nodes for crosstalk. Examples of such crosstalk are observed from all levels of signal transduction, including the TFs (Fig. 2.4, 2.5), stress signalling hormones (Fig. 2.6), and other signalling

components such as calcium and ubiquitous protein kinases shared by multiple signalling pathways. Numerous evidences suggests that the control of these shared nodes may add on to the diversity in gene expression outcomes. For example, it has been observed that DREB2, a potential integrator of drought and heat stress, exhibit signature expression profiles that are specific to the stress input [215]. Interactions between stress hormones are unlikely to be simple antagonistic relationships, as the microarray profiling of *Arabidopsis* transcriptome upon combination of SA and JA demonstrated simultaneous presence of both synergistically and antagonistically regulated genes [217]. Moreover, it has been shown that the interaction between SA and JA may be dynamically coordinated depending on time and quantity of hormone treatments [177]. More detailed functional characterisation of the known crosstalk connections is therfore necessary, with focus on the activities of multifunctional TFs and calcium under various stress conditions. In addition to the map of interaction between hormones, the origin of observed interaction between stress signalling hormones also awaits further investigation.

With the availability of -omics datasets and bioinformatic methods, transcriptional networks have also been studied in large-scale. Unlike the signalling pathways that have been reconstructed via bottom-up assembly of individual signalling components, these transcriptional networks are typically inferred directly from large-scale datasets based on pairwise correlation [37], TF-gene interaction [192], or combination of multiple datasets and manual curation [109]. While the large-scale transcriptional networks do not offer specific insights regarding regulation of stress response genes or transduction of stress signals, the apparent density of connection between the genes suggests that there may be numerous undiscovered interactions with crucial roles in shaping gene regulatory outcomes under various stress conditions and facilitating interaction between multiple stress. How does one then identify those functionally important interactions lacking in the current model of stress signalling network? Answering this question is the focus of the upcoming chapters.

Chapter 3

Compound gene regulation as logical operation

3.1 Introduction

As our understanding of the plant stress signalling pathways accumulates, it is becoming increasingly clear that transduction of environmental stress signals occurs via highly complex network. From the previous chapter, it was shown that interactions between multiple stress inputs have effects on gene regulation and the resulting physiological responses, as exemplified by apparent non-additive behaviours observed between the stress signalling hormones (Fig. 2.6). Several studies have investigated the potential mechanisms behind such interaction between multiple hormonal inputs in relatively well-characterised plant systems such as guard cell signalling [23] and root development [148]. Notably, these studies have utilised mechanistic models to determine the role of crosstalk between multiple hormones in controlling given processes, such as stomatal aperture and auxin biosynthesis. Unfortunately, similar detailed mechanistic modelling approaches are not yet available to the forthcoming sets of data on gene regulatory outcomes in response to combined stresses, owing to incomplete characterisation of the upstream stress signalling pathways regulating many of those genes. While further experimental characterisation of individual signalling components un-

der a variety of environmental settings is one approach to shed more light on how multiple external inputs are processed to regulate expression of stress response genes, such efforts are likely to be expensive and time-consuming.

An alternative approach to elucidate the mechanisms for compound gene regulation is to systematically characterise different outcomes of signal integration, and investigate each of the observed outcomes in the light of a particular design of the regulating network. One of the key concepts in Systems Biology is that design principles of a biological network, such as network structure and kinetic parameters that govern network dynamics, determine the behaviour of the network as a whole [125]. Numerous studies attempted to functionally characterise recurring structural patterns called network motifs [140, 225] by associating them with specific behaviours. For example, coherent feedforward loops commonly serve as so-called persistence detector [158], where output is not produced by transient activation, but only upon presence of a persistent input. Although it has been shown that behaviours observed from biological network cannot be solely attributed to topological structure of the network [104], it still leads to an interesting question whether a specific gene regulatory outcomes observed under combined inputs can be mapped on to a specific type of network structure.

The main difficulty in assigning the outcomes of signal integration into specific control structures, however, stems from large number of possible signal integration outcomes. The previous chapters discussed that the effects of combining multiple inputs may not sometimes be adequately described by simple synergistic or antagonistic interaction. Thus, a conceptual framework that identifies and limits the space of possible signal integration outcomes is an important step towards elucidation of the molecular mechanisms behind compound gene regulation. In this chapter, we examined compound logic, which was originally proposed by [239], as candidate for such classification framework. The mechanisms behind the possible outcomes of signal integration classified under compound logic were then investigated by constructing simple models of signalling networks based on Parallel Distributive Processing network [160]. The simplicity of the proposed model allowed systematic investigation of vari-

ous signalling network topologies, which provide insights regarding the potential mechanisms responsible for the classified signal integration outcomes.

3.2 Examples of signal integration from bacterial systems

To better understand the variety of signal integration outcomes, it is useful to consider examples from other organisms that experience decision-making problems upon presence of multiple environmental signals. Bacterial organisms are well known for their abilities to integrate wide variety of environmental signals to increase their chance of survival [25, 38]. While there are large differences in the exact signalling mechanisms employed by bacterial and plant systems, similarities can also be found amongst those examples with respect to the biological outcomes triggered by multiple environmental inputs, such as level of gene expression or protein activity.

One of the best studied examples of biological networks integrating multiple external inputs is the *lac* operon in *Escherichia coli* [106]. The *lac* operon contains the genes required for lactose utilisation, including LacY for lactose import, LacZ and LacA for lactose metabolism. The activity of the *lac* operon is regulated by two carbon sources, namely lactose and glucose, and the key function of *lac* operon is to induce expression of the three genes only in presence of lactose. Figure 3.1a outlines the core control structure involved with regulation of lac expression. There are two signalling pathways, one involving the *lac* repressor (LacI) as the lactose sensor, and catabolite activator protein (CAP) as the glucose sensor. LacI represses expression of the *lac* genes, whereas CAP activates expression of the *lac* genes. Presence of lactose and glucose inhibits the DNA-binding capability of LacI and CAP, respectively. The ability of CAP to trigger expression of the lac genes is prevented by LacI, as LacI blocks the transcription initiation sites immediately downstream of CAP. This cross-regulation from LacI to CAP plays a key role in establishing glucose as the preferred carbon source over

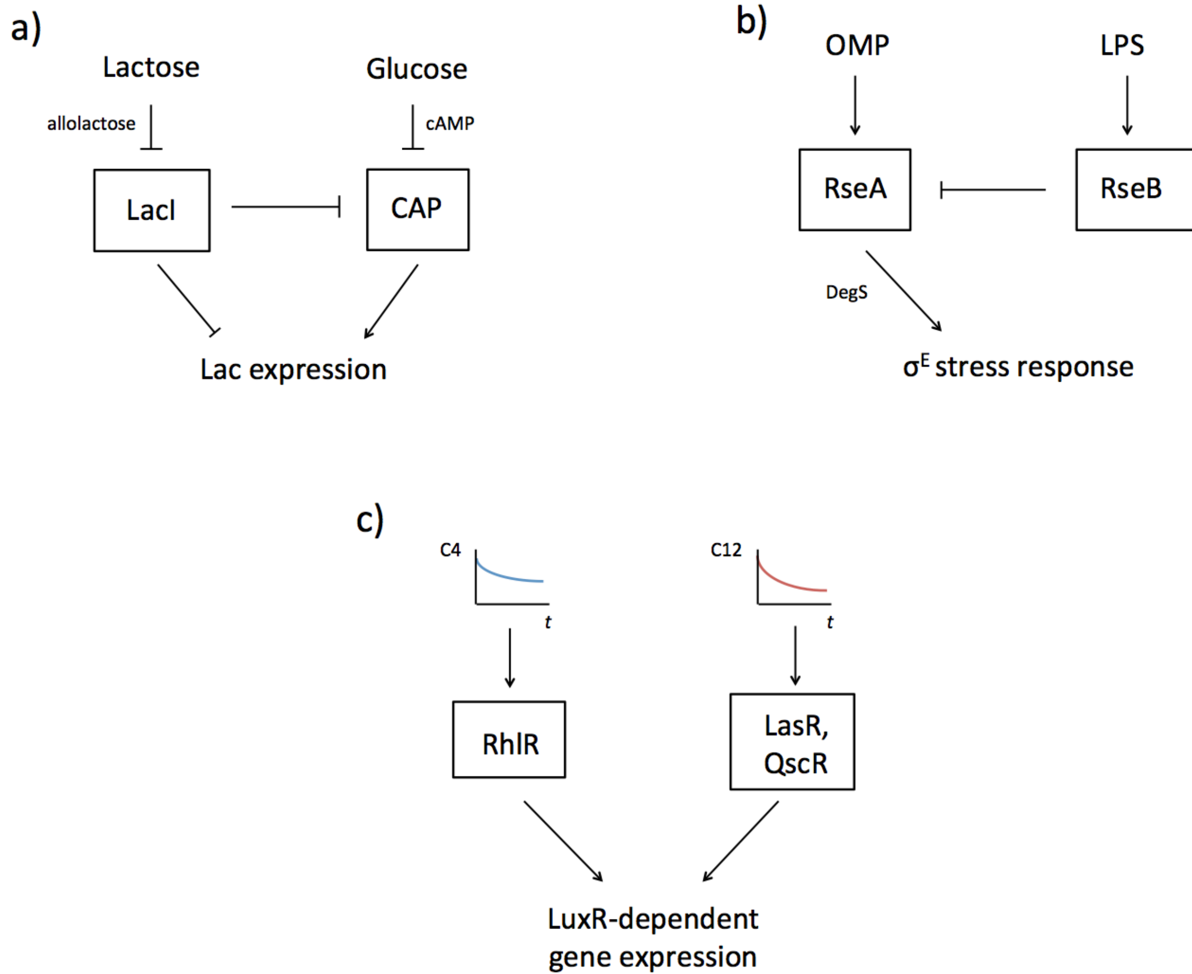


Figure 3.1: **Examples of signal integration from bacterial systems**. a) The core control structure of the *lac* operon outlines signalling pathways for lactose and glucose consisting of LacI and CAP, which are the repressor and the activator of *lac* expression. b) Regulation of σ^{E} stress response is facilitated by two periplasmic proteins RseA and RseB that are responsive to OMP and LPS signals, respectively. The latter binds the former, rendering it insensitive to OMP. c) Two quorum inducers, C4-HSL (C4) and 3-oxo-C12-HSL (C12) in *Pseudomonas aeruginosa* activate LuxR-dependent expression. C4 and C12 have different half-life.

lactose because expression of the lac genes specifically occurs when only CAP is active. Thus, this example demonstrates that *E. coli* is able to integrate simultaneous lactose and glucose signals to make decisions over induction of its genes, preventing committment to production of the proteins involved in lactose metabolism when both carbon sources are present.

Another example of signal integration in bacterial gene regulation is control of σ^{E}-dependent regulation of stress response genes in gram-negative bacteria such as *E. coli*, which plays a crucial role in repair of damaged envelope. Biochemical signals characterising damages in outer membranes is accumulation of unfolded outer membrane proteins (OMPs) and Lipopolysaccharides (LPS) in periplasmic space [5]. The pathways for OMP and LPS signalling is known to interact with eacther other (Fig. 3.1b). In the OMP signalling pathway, degradation of RseA mediated by DegS is responsible for production of σ^{E} factors. OMP, however, is unable to initiate RseA degradation on its own because RseA is tightly associated with RseB, the LPS sensor [146]. Simultaneous presence of OMP and LPS is therefore required to trigger σ^{E}-dependent expression of stress response. This example demonstrates that integrating multiple signals can be used to reduce waste of cellular resources by preventing unnecessary expression of downstream genes upon transient increase in the stress signals, which are not likely to represent severe damage to the envelope.

Signal integration is also observed in bacterial quorum sensing, which involves intercellular signalling between bacterial cells to modify collective behaviours of bacterial cell population based on population density [262]. To achieve this, individual bacterial cells produce and detect chemical signaling molecules called autoinducers (AIs) to modulate expression of their own genes and across neighbouring cells. Notably, many quorum-sensing bacteria utilise multiple types of AI, each with designated receptors at the membrane level but regulating a common signalling target. For example, *Pseudomonas aeruginosa* biofilm formation is triggered by two AIs, namely C4-HSL and 3-oxo-C12-HSL [46], which lead to activation of master quorum-sensing transcriptional regulator, LuxR (Fig. 3.1c). The utility of integrating multiple AI signals, however, has been a controversial topic. Having two types of input

signals that activate a common downstream target may lead to loss of information, as noisy fluctuation associated with one signal undermines signalling of the other [162]. A relatively recent study conducted by Cornforth et al. [50] suggested that combination of multiple AIs allow detection and response to different environmental states at greater resolution, particularly when the AIs have different chemical properties such as decay rate. From the changes in *Pseudomonas aeruginosa* transcriptome in response to two inducer molecules with distinct decay rates, the same study identified several gene clusters displaying constrasting behaviours. Although the detailed mechanisms leading to specific behaviours are not yet understood, the bacterial quorum-sensing example demonstrates that signal integration helps to fine-tune downstream responses to the precise environmental state by harnessing the information about dynamics of multiple inputs.

3.3 Compound logic

The above examples shows that integration of multiple environmental signals governs cellular decision regarding regulation of gene expression. Specifically, those examples showed that the outcomes of gene regulation from integrating multiple environmental signals can be described as logical operation. Such interpretation of biological processes integrating multiple environmental signals led to development of compound logic as a systematic framework for classifying different signal integration outcomes by Tanaka and Kimura [239].

The basic idea of compound logic is simple (Fig. 3.2): given that logical variables can be used to describe the states of input signals, S_1 and S_2, and the output R that is jointly controlled by the two inputs, the input-output relationship can be described with a logical function, K, which is defined by the states of R at all possible input combinations. Systematic identification of signal integration outcomes is possible, as the finite number of states for the inputs and the output limits the number of unique K functions. For example, binary description of the inputs and the output states leads to 2^4 unique K functions. Thus, a signal integration outcome, which describes how different combinations of the input signals

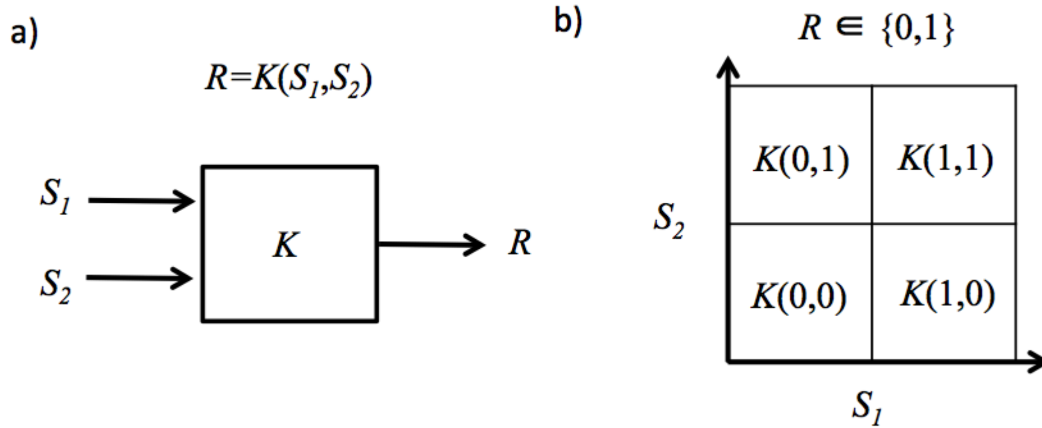


Figure 3.2: **Representation of signal integration outcomes via compound logic**. a) Signal integration is a process where two input signals lead to regulation of a shared output. b) The outcomes of signal integration can be defined by a logical function K describing the state of output at different combinations of inputs. If the states of the inputs are approximated with binaries, the space of possible output can be divided into four quadrants.

affect the output, can be succinctly described with a single function K. Figure 3.3 shows representation of signal integration outcomes presented in the three examples discussed above based on binary compound logic.

The compound logic functions identified based on binary approximation of input and output states can be described in terms of Boolean algebra. However, one of the most frequently encountered questions regarding gene regulation is not whether a gene simply switches on or off, but how the external signals induce change in the activity of the gene with respect to control. Differential cDNA microarray, for example, detects the genes showing relative difference in behaviours under two experimental conditions. This means that transcriptional regulation of a gene, which can either be positive or negative, cannot be adequately captured with binaries. Compound logic offers a solution to this problem by allowing a larger number of states for description of the inputs and the output. More adequate description of gene regulatory outcomes can be achieved by considering three states of gene activity that include

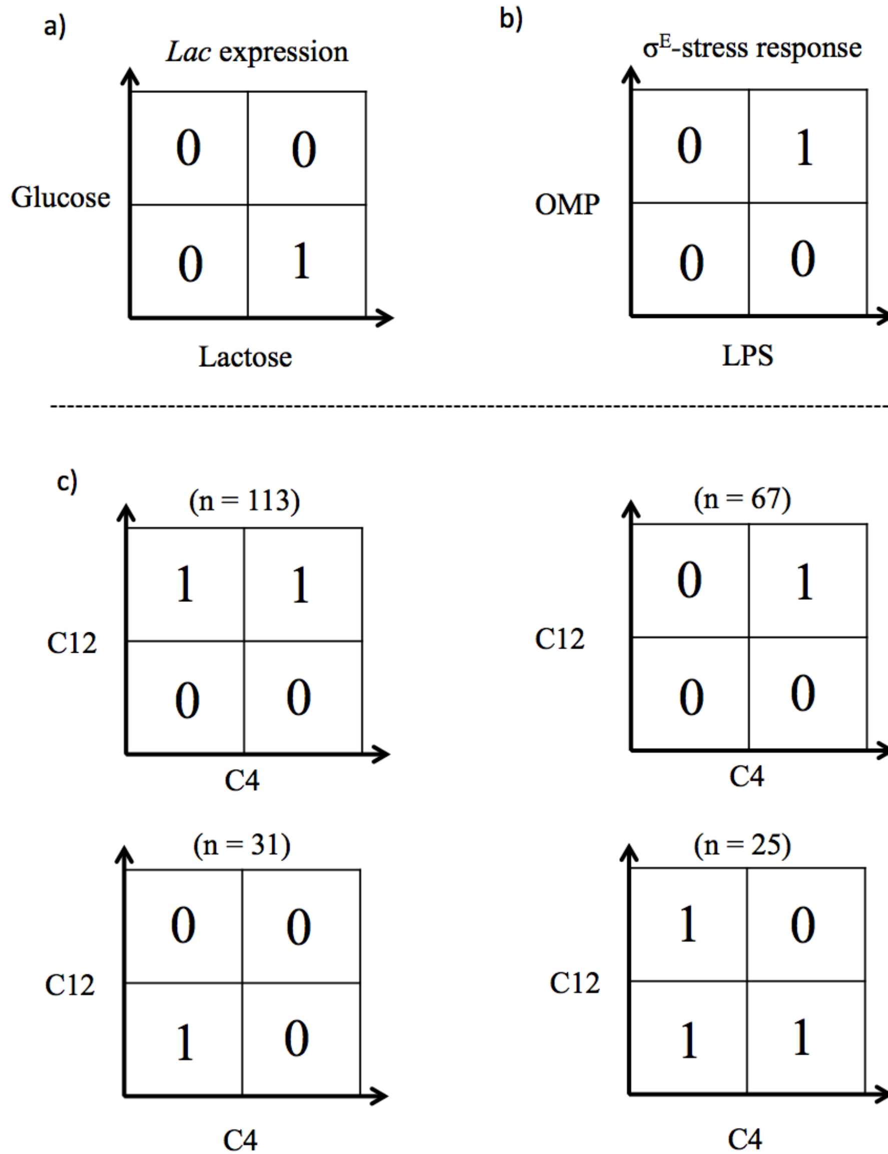


Figure 3.3: **Use of compound logic functions to describe signal integration in the bacterial examples**. The effects of integrating multiple environmental signals from the examples described in Figure 3.1 are shown. Numbers in each quadrant indicates the state of the genes subject to regulation, described in binaries (0,1). a) Integration of lactose and glucose signals by interaction between LacI and CAP leads to selective expression of lac operon under single lactose condition. b) Both OMP and LPS signals are required to trigger σ^{E} stress response due to the interaction between RseA and RseB. This is equivalent to logical AND operation. c) Combined treatments of C4 and C12 lead to various regulatory outcomes amongst LuxR-induced transcriptome [50]. Four most frequently observed logical functions are shown (n = number of genes).

up-regulation (1), no change (0), and down-regulation (-1).

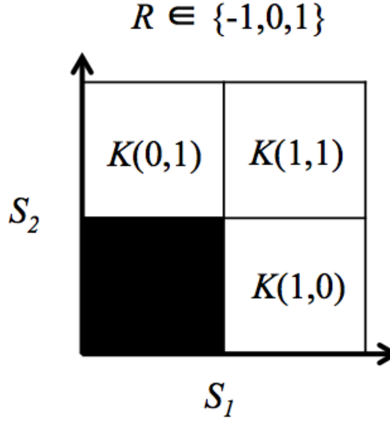
The resulting compound logic functions that describe the effects of signal integration on direction of gene expression change are shown in Figure 3.4. Here, we consider possible changes in gene expression under three input conditions (S_1 only, S_2 only, and $S_1 + S_2$) because the outcomes of signal integration describe relative change in gene expression with respect to control. Given that there are three possible states of gene expression for each of three input conditions, there are 3^3 possible regulatory profiles.

Compound logic can also be generalised further to account for larger numbers of possible states for the output and input conditions. Given that the numbers of possible states for the output and input conditions are represented by r and s, respectively, the possible number of signal integration outcomes would correspond to r^s. While characterisation of signal integration outcomes for arbitrary number of states is possible, it introduces combinatorial complexity that quickly becomes overwhelming. We argue that classification based on ternary gene regulatory outcomes (Fig. 3.4) provides sufficiently detailed and comprehensive coverage of diverse behaviours observed from compound gene regulation.

3.4 Logical network models

3.4.1 Gene regulatory network as Parallel Distributive Processing Network

Having identified the possible signal integration outcomes in terms of compound logic functions, we then investigate whether each function can be assigned to particular network structure. As discussed in chapter 2, environmental stress generates multiple biochemical signals in a variety of forms, such as damages to membrane structure and changes in intracellular osmotic balance. Those signals are subsequently processed by multiple signalling pathways acting in parallel, each regulating a subset of genes. Such aspect of of stress-induced gene



Function	S_1	S_2	$S_1 + S_2$	Function	S_1	S_2	$S_1 + S_2$	Function	S_1	S_2	$S_1 + S_2$
K_1	-1	-1	-1	K_{10}	0	-1	-1	K_{19}	1	-1	-1
K_2	-1	-1	0	K_{11}	0	-1	0	K_{20}	1	-1	0
K_3	-1	-1	1	K_{12}	0	-1	1	K_{21}	1	-1	1
K_4	-1	0	-1	K_{13}	0	0	-1	K_{22}	1	0	-1
K_5	-1	0	0	K_{14}	0	0	0	K_{23}	1	0	0
K_6	-1	0	1	K_{15}	0	0	1	K_{24}	1	0	1
K_7	-1	1	-1	K_{16}	0	1	-1	K_{25}	1	1	-1
K_8	-1	1	0	K_{17}	0	1	0	K_{26}	1	1	0
K_9	-1	1	1	K_{18}	0	1	1	K_{27}	1	1	1

Figure 3.4: **Identification of compound logic functions for description of relative gene expression change based on ternary logic**. Relative changes in transcriptional activity of a gene with respect to the basal level (black quadrant) can occur in three directions: increase (1), no change (0) and decrease (-1). This leads to identification of 27 unique logical functions highlighting various effects of signal integration on gene regulatory outcomes.

regulatory system bears remarkable resemblence to Parallel Distributive Processing (PDP) networks [160].

An example of stress-induced gene regulatory system expressed as a PDP network is shown in Fig. 3.5. A PDP network consists of three layers of nodes (or units); the nodes in the 'input' layer denote the state of environment, which in our case, indicate presence of stress. The state of input nodes determine the state of the nodes in the intermediate layer, which correspond to the activities of intracellular signalling pathways. This layer is referred as 'hidden' because states of the nodes belonging to this layer are not observable directly from experiments. Finally, the 'output' nodes represent the states of the genes regulated by the 'hidden' nodes. Collective states of the nodes in the output layer is equivalent to the overall transcriptomic response induced by stress inputs.

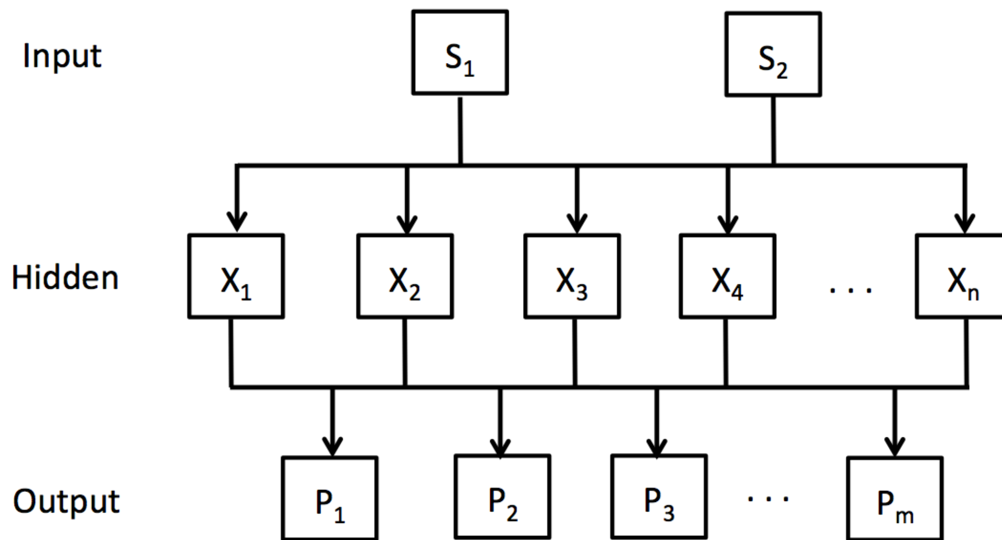


Figure 3.5: **Transcriptome regulatory network depicted as Parallel Distributive Processing network**. Different combinations of two stresses, S_1, S_2, induces changes in promoter states of m genes, denoted by P, via n signalling pathways, denoted by X.

3.4.2 Model description

To investigate possible network topologies for each of 27 compound logic functions identified in the previous section, it is necessary to develop a model of signalling networks for regulation of a single gene. As dictated by the PDP framework, regulatory network for a single gene is described with three layers of nodes. The state of each node is determined by functions providing phenomenological description of the relationship between inputs and output of the node:

$$X_i = f(S_1, S_2), \tag{3.1}$$

$$P = g(X_1, X_2, X_3, ..., X_n), \tag{3.2}$$

where $i = 1, 2, 3, ..., n$. Fig. 3.6 shows two network structures that will be examined during the subsequent analyses. Here the state of a transcription factor X_i is determined by function f, and the promoter P by function g. Genes controlled by *homotypic* promoters (containing a single type of TFBS) are represented by g with $n = 1$. Genes with *heterotypic* promoters (containing two types of TFBS) are described by g with $n > 1$. Topology of the overall signalling network is described by a combination of any function f, and g. For simplicity, we only investigated two cases, $n = 1$ and $n = 2$.

The transcription factor X_i is considered to conduct logical operation of S_1 and S_2 to determine its state, which is assumed to be in binary form (active, inactive). Table 3.1 shows the truth table resulting from 7 possible logical functions used for describing f (excluding the function that is zero for all three stress conditions). Although the individual steps of signal transduction are not explicitly modelled by f, each logical function still confers how the two stresses interact with each other in regulating the activity of targeted TF.

The state of individual TF consequently determine the state of promoter, any change in which results in change in gene expression. Determining the logical functions to describe promoter, however, is more difficult because the states of promoter must be described with ternaries

in order to describe increase, no change, and decrease in the level of gene expression. We resolve this by introducing two types of promoters, where one is targeted by an transcriptional activator ($X_{i,A}$) and the other by an transcriptional repressor ($X_{i,I}$). Positive changes in gene expression are assumed to be facilitated by actions of transcriptional activator, while negative changes occur when transcriptional repressor is triggered by stress inputs. The distinction means that a TF can only act either as activatory or inhibitory pathway, not both.

Based on these assumptions, we defined 12 logical functions describing how states of one or multiple TFs affect the state of promoter (Table 3.2). The first two functions (g_1, g_2) represent the genes with promoters targeted by a single type of TF. The other ten functions describe the genes regulated by two pathways, which were selected to represent commonly observed types of interactions between two TFs. 'Independent' functions (g_3, g_4, g_5, g_6) depict promoters that are bound by two TFs with no interaction. 'Competitive' functions (g_7, g_8, g_9, g_{10}) describe the promoter that preferentially binds to one TF. 'Cooperative' functions (g_{11}, g_{12}) describe the promoters that require presence of both TFs to be active.

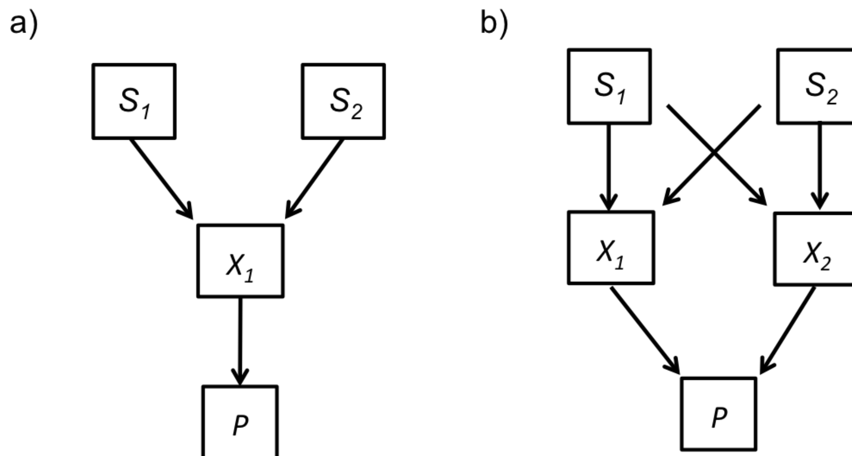


Figure 3.6: **Overview of a logical network model for description of signalling network topology**. a) Gene regulatory network with a single signalling pathway ($n = 1$), b) Gene regulatory network with two signalling pathways ($n = 2$).

Table 3.1: **Logical functions describing regulation of TF activity by multiple external inputs.**

Function	$S_1 = 1$ $S_2 = 0$	$S_1 = 0$ $S_2 = 1$	$S_1 = 1$ $S_2 = 1$
$f_1(S_1)$	1	0	1
$f_2(S_2)$	0	1	1
$f_3(S_1, S_2)$	1	1	1
$f_4(S_1, S_2)$	0	1	0
$f_5(S_1, S_2)$	1	0	0
$f_6(S_1, S_2)$	0	0	1
$f_7(S_1, S_2)$	1	1	0

Table 3.2: **Logical functions describing the states of promoters regulated by one or two TFs.**

Function	$X_1 = 1$ $X_2 = 0$	$X_1 = 0$ $X_2 = 1$	$X_1 = 1$ $X_2 = 1$	Type of TF interaction
$g_1(X_{1,A})$	1	0	1	Not applicable
$g_2(X_{1,I})$	-1	0	-1	
$g_3(X_{1,I}, X_{2,A})$	-1	1	0	Independent
$g_4(X_{1,A}, X_{2,I})$	1	-1	0	
$g_5(X_{1,A}, X_{2,A})$	-1	-1	-1	
$g_6(X_{1,I}, X_{2,I})$	1	1	1	
$g_7(X_{1,I}, X_{2,A})$	-1	1	-1	Competitive
$g_8(X_{1,I}, X_{2,A})$	-1	1	1	
$g_9(X_{1,A}, X_{2,I})$	1	-1	-1	
$g_{10}(X_{1,A}, X_{2,I})$	1	-1	1	
$g_{11}(X_{1,I}, X_{2,I})$	0	0	-1	Cooperative
$g_{12}(X_{1,A}, X_{2,A})$	0	0	1	

3.4.3 Genes regulated by homotypic promoters ($n = 1$)

We now investigate which network topologies (combination of function f and g) reproduce which compound logic function defined in the previous section. The network topologies for genes regulated by a single signalling pathway can be described by any of the first two functions of g, combined with any of seven available f, leading to 2×7 possible network topologies. Each of the resulting 14 structures could be assinged to a specific logical function.

Because g only specifies the sign of change in gene expression in this case, the type of K associated with the given network topology is mainly governed by the choice f. Some of the results are intuitive: for example, systems implementing f_1 and f_2, which describe the TFs that are activated only under one type of input, show no change in response when two inputs are combined together. Implementation of function f_3, which represents the TF that are inducible by both inputs, lead to regulatory changes of identical sign across all input conditions. Implementation of f_4 and f_5 leads to TF activities being suppressed upon combination of inputs, describing nullification of downstream gene expression. Finally, TFs represented by f_6 and f_7 are only active when both inputs are present, leading to gene regulatory outcome specific to combination of two inputs.

While the simple logical network models implementing a single TF provide simple and intuitive explanation of the possible behaviours of genes, those could not reproduce all functions.

3.4.4 Genes regulated by heterotypic promoters ($n = 2$)

Because the proposed logical network models implementing single type of TF could not capture all of the 27 compound logic functions (Fig. 3.4), we investigated the models with two types of TF. Description of the network topologies for genes regulated by two TFs requires choice of g for promoter and f for both signalling pathways, which leads to $10 \times 7 \times 7$ possible

Table 3.3: **Logical functions reproduced by models of regulatory network consisting of a single TF.**

Structure	$S_1 = 1$ $S_2 = 0$	$S_1 = 0$ $S_2 = 1$	$S_1 = 1$ $S_2 = 1$	Compound logic function
$g_1(f_1(S_1))$	1	0	1	K_{24}
$g_1(f_2(S_2))$	0	1	1	K_{18}
$g_2(f_1(S_1))$	-1	0	-1	K_4
$g_2(f_2(S_2))$	0	-1	-1	K_{10}
$g_1(f_3(S_1, S_2))$	1	1	1	K_{27}
$g_2(f_3(S_1, S_2))$	-1	-1	-1	K_1
$g_1(f_4(S_1, S_2))$	1	0	0	K_{23}
$g_1(f_5(S_1, S_2))$	0	1	0	K_{17}
$g_2(f_4(S_1, S_2))$	-1	0	0	K_5
$g_2(f_5(S_1, S_2))$	0	-1	0	K_{11}
$g_1(f_6(S_1, S_2))$	0	0	1	K_{15}
$g_1(f_7(S_1, S_2))$	1	1	0	K_{26}
$g_2(f_6(S_1, S_2))$	0	0	-1	K_{13}
$g_2(f_7(S_1, S_2))$	-1	-1	0	K_2

network topologies. Although the resulting 490 profiles are able to reproduce all 27 variants of function K, each function K could not be attributed to a single network topology.

Fig. 3.7 shows the number of different network topologies that produce the outcomes described by each function K. The result shows that each regulatory profile can be explained by approximately 12 to 18 network topologies with an exception of the K_1 and K_{27}, which are reproduced by 40 network topologies. In addition, 38 network topologies fail to induce change in expression under any of the three stress conditions (K_{14}).

Given that the interactions of two signalling pathways occurring at the promoter level leads to increase in the number of possible network topologies for each regulatory profile, determining

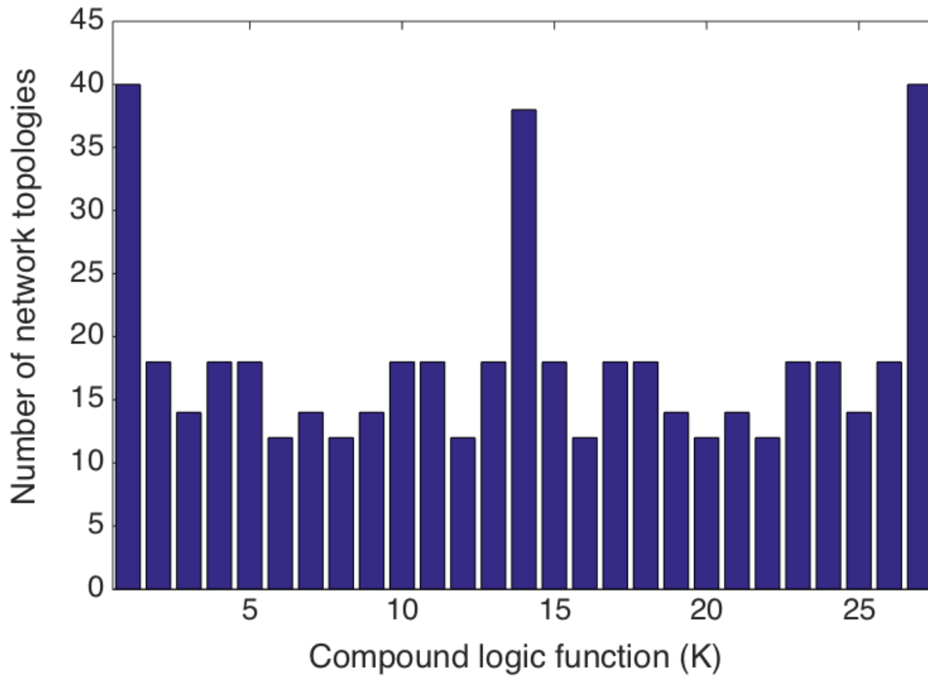


Figure 3.7: **Number of possible network topologies with two signalling pathways for each regulatory profile**.

the network topology that explains the observed regulatory profile requires integration of additional information.

3.5 Discussion

In this chapter, we have built up on the compound logic framework originally developed by Tanaka and Kimura [239] for systematic classification of possible signal integration outcomes. Ternary logic was used to describe up-regulation, down-regulation or no change in expression. This offers a unique advantage to the binary framework, which can only describe changes in gene expression in terms of “on” and “off” states. Limiting the number of possible states of genes led to identification of 27 unique behaviours via compound logic. Having clearly

defined possible gene regulatory outcomes from integration of multiple input signals, we explored whether each of the defined behaviours can be associated with a particular network topologies. This was achieved by developing simple models of signalling network inspired by the PDP network.

In order to model a gene regulatory network modulated by multiple environmental inputs as a PDP network, it is essential to describe the behaviours of genes with logical (discrete) variables. In a dynamical systems perspective, one way to interpret such discrete behaviours of genes is transition from one equilibrium point to another upon perturbation by input signals. Specifically, use of such description requires the following two conditions to be assumed:

1. The system of interest has reached an equilibrium after sufficient time has passed from initial input perturbation, and
2. It is possible to fully determine the steady-state behaviours of system with a finite (three in the case of ternary compound logic) number of fixed point attractors.

Assuming these two conditions introduces several limitations associated with the compound logic method and modelling of gene regulatory networks with PDP network models. The first condition forces us to ignore time and trajectories the system has taken to reach equilibrium, which means that the method cannot be used to describe any transient behaviours exhibited by a gene. The second condition also excludes the possibility of a gene showing other types of stable behaviours, such as convergence to limit cycles or strange attractors [235]. Although it is known that many genes behave in discrete manners through well-known motifs such as bistable switch [203] and such qualitative changes in the state of the gene are often what matters in decision-making of a cell [56, 93, 132], it is biologically unrealistic to apply the same assumptions to all genes. For instance, circadian rhythm in plants plays a critical role in regulating cellular processes independently to environmental conditions, and the state of the genes affected by the clock usually do not converge to a single equilibrium value over a long period of time [92].

Despite its limitations, however, the compound logic method can still provide new insights regarding the behaviour of genes under combinatorial input conditions provided that it is carefully used. By describing a gene's behaviour under different input conditions with a pre-defined compound logic function, it is possible to rapidly identify how multiple input signals interact with each other to regulate the gene. Several compound logic functions defined above describe complex processing of two input signals (e.g. reversal of regulatory outcome seen in K_6 and K_{22} - see Fig. 3.4), and presence of genes that can be approximated by those compound logic functions may reveal biologically important regulatory interaction between the two input signals. Furthermore, it is possible to reduce a potentially large model space into a manageable number of models through construction of PDP network, subsequently allowing a systematic investigation of the system structure responsible for the observed behaviours. Given that the PDP network models are also readily scalable into larger networks, we argue that the analytical framework proposed altogether in this chapter can be particularly useful in enquiring large gene expression datasets to generate the first insight regarding connectivities between input signals and different nodes in the upstream regulatory network.

If the proposed analytical framework using compound logic and PDP network can be used to analyse gene expression dataset and provide insights about important regulatory structures of gene regulatory network, how does it compare with conventional methods used for reconstruction of gene regulatory network structures? Numerous existing techniques already use transcriptomic microarrays data to infer topology of gene regulatory network: for instance, clustering algorithms can be used to organise gene expression data by grouping similar expression profiles together [94], based on degree of similarity between a pair of genes using Euclidean distance and Pearsons Correlation Coefficient [54]. Alternatively, Bayesian inference methods have been used for more accurate prediction of causal relationships from microarray expression profiles [70, 277]. Irrespective of the theoretical basis and assumptions specific to each method, those existing gene regulatory network inference methods are based on the premise that promoters of those similarly-behaving genes are regulated

by common transcription factors and upstream network elements, and aim to translate the information from gene clusters into a graph with regulatory interactions (edges) between genes (nodes). Indeed, those methods have led to highly successful examples of large-scale gene regulatory network inferred from transcriptomic datasets [101, 139]. However, we note that these methods cannot fully replace the analytical framework we propose in this chapter due to their own limitations: importantly, the existing network inference methods focus exclusively on gene-gene relationships, and are unable to consider complex interaction between multiple environmental input signals. The graph models resulting from those methods are therefore limited to description of overall topological structure of gene regulatory networks, and offer little mechanistic insights regarding the origin of behaviours observed amongst the genes.

In conclusion, we argue from the results presented in this chapter that comparison between gene regulatory changes to single and combined stresses may also be a viable source of information, providing a basis for further distinction between different groups of genes that are potentially regulated by common transcriptional regulatory mechanisms. We explore this idea more thoroughly in the subsequent chapters.

Chapter 4

Analysis of transcriptomic changes under combined stresses in *Arabidopsis thaliana* using compound logic

4.1 Introduction

In this chapter, we apply the analytical framework developed in Chapter 3 to a real gene expression datasets related to plant stress responses and demonstrate how the framework can be used to extract information about the regulatory network structure. Before moving onto actual application of the compound logic framework, we briefly review the status quo of the transcriptomic studies conducted in various plant species upon single and combined environmental stresses, and the previous attempts to infer structure of stress response gene network from the datasets resulting from those studies.

Transcriptomic changes under various single stress condition have been well documented across different plant species: In *Arabidopsis*, several studies have independently investigated transcriptomic responses to individual abiotic or biotic stresses [122, 129, 222]. Using conventional gene regulatory network inference methods (discussed in Section 3.5), many studies successfully identified several features regarding global organisation of gene regula-

tory system in response to various stresses [145]. For instance, association of the module identified from transcriptomic datasets with known TF-Binding Site (TFBS) motifs suggests for presence of a ubiquitous stress response module, which acts in response to a variety of stress types [155]. While the studies specifically focusing on stress-related transcriptional network are scarce, numerous studies have successfully integrated the transcriptomic datasets with the results of genomic sequence analysis (e.g. identification of TFBS and resulting TF-gene relationships) to map genome-wide transcriptional regulatory networks, which are made publicly accessible via online databases: for Arabidopsis, two most notable examples include the AraNet [138] and AthaMap [233] databases.

On the other hand, the datasets showing transcriptomic responses to combined stress in different species became only recently available [110, 182, 206, 211]. Measuring direction of change in expression profiles across the entire transcriptomes, these studies commonly demonstrated that the set of genes whose behaviours are changed by combined stresses show little overlap with those affected by single stresses. At one end, such unique expression profiles observed under combined stress indicates that regulatory cues on genes from singly applied stresses are modulated by addition of another stress. Some genes that are not induced by singly applied stresses at all were shown to be regulated under combined stress conditions. Such dissimilarity between the transcriptomic responses to single and combined stresses suggests that interaction between multiple stress types to regulate target sets of genes is not restricted to synergistic and antagonistic relationship, and there are mechanisms responsible for processing of simultaneous stress signals to produce such unique regulatory profiles. Although few, there have been attempts to make use of these transcriptomic datasets from combined stress to infer transcriptional regulatory network topology: for instance, Barah et al. [21] separately analysed parts of the global transcriptional network responsive to different stress conditions. The study found that some TFs, especially those with roles associated to circadian clock and developmental processes, act as master regulators of stress responsive genes, while other TFs mediate regulate subset of genes specific to a particular type of stress condition.

Despite these interesting developments we have seen so far, the current approaches for reconstruction of stress-responsive transcriptional regulatory network in plants share a fundamental limitation. The insights generated from those approaches are largely restricted to topological structures of the transcriptional network, which do not provide mechanistic explanation of the non-additive behaviours exhibited by the genes in response to combined stress. In this regard, we apply the compound-logic description of gene expression profiles and derivation of PDP network models from those profiles as an alternative method to overcome this challenge. Here, we chose to specifically examine the transcriptomic dataset generated by Rasmussen et al. [206], which contains information about gene expression changes upon 10 different (single and combined) stress conditions.

The chapter executes as follows: first, the genes in the transcriptomic dataset are analysed according to the compound logic functions defined in the previous chapter for detailed comparison of the responses of the genes to single and combined stress treatments. We then explore possible network topologies that may give rise to each regulatory profile based on the PDP network models. The models that are developed for each of the pre-defined regulatory profiles highlight possible origins of the regulatory profiles within the upstream signalling networks in terms of control of TF activities and promoter interaction. By analysing the genes that are regulated by HSF and WRKY pathways for heat and biotic stress signalling as the examples, we show that the model can be applied to understand the expression profiles of smaller subsets of genes controlled by common cis-elements. The result is a framework that allows integration of information from transcriptomic profiles and promoter architecture to generate new hypotheses regarding the upstream signalling network structure.

4.2 Regulation of *Arabidopsis* transcriptome by combined environmental stresses

4.2.1 Regulatory Profiles Matrix

For comprehensive visualisation of the transcriptomic responses to various stress conditions, we obtained a Regulatory Profiles Matrix (RPM) (Fig. 4.1) from the data published by Rasmussen et al. [206]. RPM summarises the qualitative changes in the expression profile of each gene in response to each stress as a series of ternary logical outcomes: up-regulation (1), no change (0) and down-regulation (-1).

Source of data The original dataset was downloaded from Gene Expression Omnibus (Accession number: GSE41935), an online public database repository containing gene expression profile data from a wide variety of organisms and experimental settings [57].

Experimental conditions The original dataset contains expression profiles measured under 12 stress treatment conditions: no stress, NaCl (N), Heat (H), High Light (HL), Cold (C), Flg22 peptides as the inducer of biotic stress response (F), and six pairs of those stresses (S+H, S+HL, H+HL, C+HL, H+F, and C+F). For detailed description of the treatment settings, see the orignal paper [206]. The dataset also contains the expression profiles data for 10 *Arabidopsis* ecotypes. For our analysis, however, we used the expression profiles

Figure 4.1: **Regulatory Profiles Matrix - an illustrative example**.

Stress	N	H	HL	C	F	N+H	N+HL	H+HL	C+HL	H+F	C+F
$Gene_1$	1	1	0	0	-1	0	0	1	0	0	0
$Gene_2$	0	0	0	1	0	0	0	0	1	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
$Gene_{25430}$	1	0	0	-1	0	0	0	0	0	0	-1

data from only *Columbia* (*Col*) ecotype. This was because the information used for model development in latter parts of the chapter is mostly based on the *Col* platform.

Identification of significantly regulated genes The expression profiles of 25430 annotated genes were analysed, which excludes the measurements for mitochondrial and chloroplastic genes in the original dataset. Student's t tests were carried out between the triplicate control samples and the triplicate samples treated with each stress ($p \leq 0.1$) for each gene. Directions of regulatory changes amongst the strongly regulated genes were determined by comparing the mean expression indices from the control samples and those from the treated samples.

4.2.2 Overview of transcriptomic responses to single and double stresses

To compare transcriptomic responses to various stress conditions, we first analysed the RPM obtained from the microarray data of Rasmussen et al. [206] based on percentage of genome regulated, and the degree of overlap between the subset of genes regulated under each stress condition.

Analysis of the first five columns of RPM allows comparison of the transcriptomic profile induced by singly applied stresses (Fig. 4.2). The result revealed that percentage of genome up-regulated slightly varies across different stresses; High Light stress and Flg22 induced positive changes in greater number of genes (4.3% and 5.2% of genome, respectively), while NaCl and Cold stresses positively affected relatively smaller number of genes (2.3% and 2.5%, respectively). Amongst the genes detected as down-regulated, a stark contrast was observed between percentage of genome affected by abiotic stresses (NaCl, Heat, High Light, and Cold), which only induces changes in only 1.8-2.2% of genome, and the biotic stress (Flg22), which induces changes in 12.3% of genome. This indicates the response to biotic stress mainly involves down-regulation of the genes that are constitutively expressed in absence of the stress.

The subsets of genes detected as regulated by each stress were further classified into those specifically responding to the corresponding stress and those with promiscuous roles in other types of stresses (Fig. 4.2). Here, another clear difference between the role of abiotic stress response genes and biotic stress response genes emerged. Approximately half of the genes regulated by each abiotic stress is regulated by one or two other types of stresses (turquoise), suggesting the responses to those stresses show some overlap. A small number of genes is found to change in response to more than three stress type (yellow), suggesting that these genes might be the ubiquitous regulators of the stress response. In contrast, the majority of genes responsive to Flg22 do not respond to any of other type of stress (dark blue), suggesting that transcriptomic profile to biotic stress is mainly specific to the stress.

Analysis of the genes regulated under pairs of stress (Fig. 4.3) showed that the transcriptomic responses to combined stress are not additive of the transcriptomic responses to each member of the pair. The percentages of genome regulated under combinations of the stress do not show significant increase with respect to the percentage of genome regulated under individually applied stress, which is expected if the responses to combined stresses are additive. In fact, combination of certain stresses even reduced the size of regulated sets of genes compared to individually applied stresses; pairing of Flg22 stress with either heat or cold stress leads to decrease in the number of down-regulated genes, from 12% to 7.9% to 9.1%, respectively (Fig. 4.3b).

An evidence for non-additive interactions between different pairs of stresses is also seen from the degree of overlap between the sets of genes regulated by combined stress and single stress. Amongst the genes regulated by each pair of stresses, only a limited number showed similar regulatory profiles to those observed under each member of the pair applied individually (dark blue, turquoise). Majority of the genes regulated by combined stresses showed regulatory profiles that do not resemble the behaviours observed under single stresses (yellow). These genes represent the genes that are specific to the combination of stresses, whose presence suggests for the uniqueness of transcriptomic responses to combination of stress.

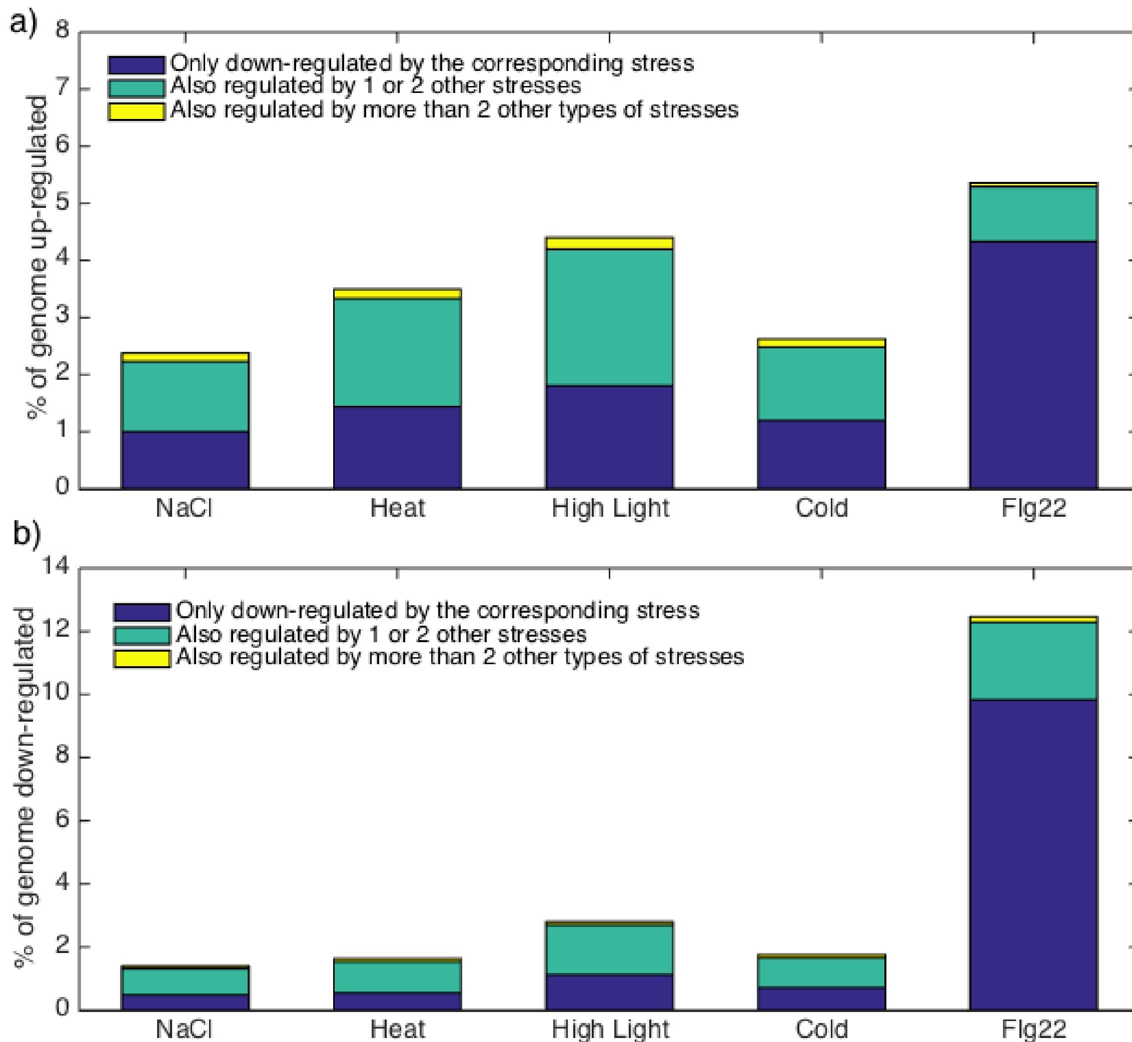


Figure 4.2: **Transcriptomic responses to singly applied stresses from *Col* ecotype**. a) Percentage of the genes up-regulated by each stress, and b) down-regulated by each stress within the whole 25430 annotated genes. The subset of genes regulated by each stress was divided into three subclasses depending on the number of stresses that the constituting genes can respond to.

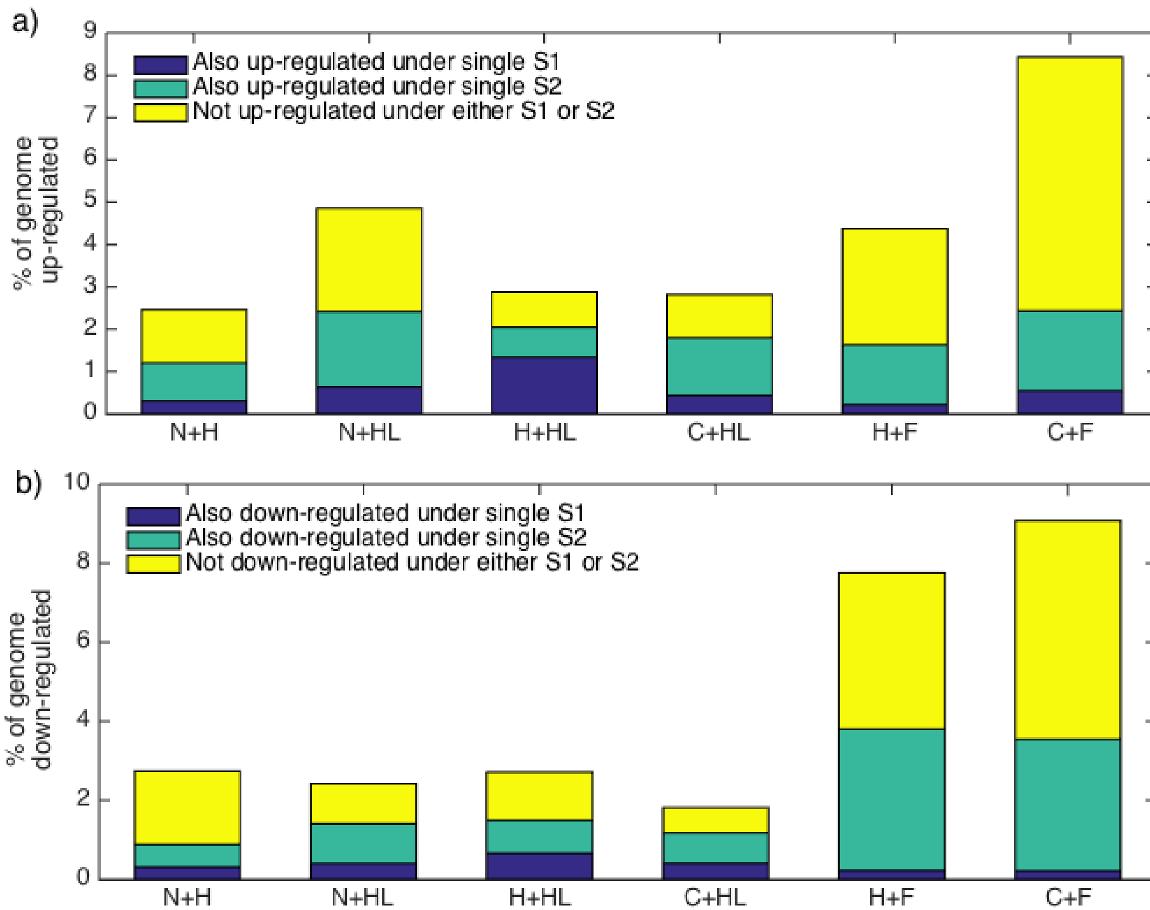


Figure 4.3: **Comparison of transcriptomic responses to combinations of two stresses against the responses to singly applied stresses**. The transcripts detected as significantly regulated by each combination of two stresses S1 + S2 were classified into three portions, including those showing identical regulatory profile to the first stress S1 (dark blue), those showing identical regulatory profile to the second stress S2 (turquoise), and those behaving differently with respect to their response to either of the stress applied individually (yellow).

4.2.3 Classification of gene regulatory profiles under single and combined stress conditions

In their original paper, Rasmussen et al. [206] proposed their own framework for classifying gene regulatory outcomes, which categorises the genes five modes of transcriptional regulation based on comparison between single stress responses and combined stress response:

- **Combinatorial** mode describes the genes showing the unique regulatory change under a pair of inputs that is not observed under either of single input.
- **Cancelled** mode describes combining two inputs nullifies the response elicited by single input.
- **Prioritised** mode describes the genes where combination of two inputs follow the response triggered by one of the singly applied input, where the two singly applied inputs induce contradictory changes.
- **Independent** mode describes the genes that respond only to one type of singly applied stress, where response to the stress combined with another leads to identical expression profile.
- **Similar** mode describes the genes that show same type of regulatory change under both singly applied stresses and their combination.

While such classification of genes according to mode of regulation offers intuitive biological insights regarding the potential effects on gene expression triggered by combined stress inputs, its description can also be ambiguous, as each mode can describe multiple regulatory profiles. To provide more precise description, we clarified the set of mathematical rules in the definitions of these regulatory modes (Table 4.1). As before, the response of a gene to a specific stress condition is denoted with function $K(S_1, S_2)$, which leads to three possible output $\{-1, 0, 1\}$ under three possible input conditions (S_1, S_2 and $S_1 + S_2$). Note that Combinatorial, Cancelled and Prioritised logic demonstrates nonlinear interaction between the two stresses, as the other three logical outcomes are concordant to the properties of

Table 4.1: **Mathematical rules used for definition of the five regulatory modes** extracted from the description of each mode by Rasmussen et al. [206].

Mode	Single stress condition	Combined stress condition
Combinatorial	$K(1,0) = K(0,1)$	$K(1,1) \neq K(1,0)$ and $K(0,1)$
Cancelled	$K(1,0) \neq K(0,1)$	$K(1,1) = 0$
Prioritised	$K(1,0) \neq K(0,1)$ $K(1,0)$ and $K(0,1) \neq 0$	$K(1,1) = K(1,0)$ or $K(0,1)$
Independent	$K(1,0) \neq K(0,1)$, $K(1,0)$ or $K(0,1) = 0$	$K(1,1) = K(1,0)$ or $K(0,1)$ $K(1,1) \neq 0$
Similar	$K(1,0) = K(0,1)$	$K(1,1) = K(1,0)$ and $K(0,1)$

linear system.

Establishing the set of mathematical rules also enabled direct comparison between logical functions identified in the previous chapter with the five modes of classification. It was found that not all of the logical functions can be assigned to the mode of regulation defined by Rasmussen et al. [206], suggesting that the five modes of regulation is insufficient to account for all of the possible signal integration outcomes. Fig. 4.4 shows assignment of each compound logic function presented in Chapter 3 into one of the five transcriptional regulatory modes proposed by Rasmussen et al. [206], excluding the function that does not show response to any of the three stress conditions ($K(1,0) = K(0,1) = K(1,1) = 0$). Notably, there were four profiles that could not be assigned to any of the original five modes based on the rules defined in Table 4.1. We therefore classified these functions into a sixth mode of regulation called Reversed, which describes the genes where the direction of change in response to one type of singly applied stress is reversed by combination of two stresses.

Based on the compound logic functions defined in the previous chapter, we subsequently conducted a more detailed analysis of the transcriptomic dataset by examining the behaviours

A: Combinatorial

$R(S_1)$	$R(S_2)$	$R(S_1+S_2)$	ID
0	0	1	A1
0	0	-1	A2
1	1	0	A3
-1	-1	0	A4
1	**1**	**-1**	**A5**
-1	**-1**	**1**	**A6**

B: Cancelled

$R(S_1)$	$R(S_2)$	$R(S_1+S_2)$	ID
1	0	0	B1
0	1	0	B2
-1	0	0	B3
0	-1	0	B4
1	-1	0	B5
-1	1	0	B6

C: Prioritised

$R(S_1)$	$R(S_2)$	$R(S_1+S_2)$	ID
1	-1	1	C1
1	-1	-1	C2
-1	1	-1	C3
-1	1	1	C4

D: Independent

$R(S_1)$	$R(S_2)$	$R(S_1+S_2)$	ID
1	0	1	D1
-1	0	-1	D2
0	1	1	D3
0	-1	-1	D4

E: Similar

$R(S_1)$	$R(S_2)$	$R(S_1+S_2)$	ID
1	1	1	E1
-1	-1	-1	E2

F: Reversed

$R(S_1)$	$R(S_2)$	$R(S_1+S_2)$	ID
1	0	-1	F1
-1	0	1	F2
0	**1**	**-1**	**F3**
0	**-1**	**1**	**F4**

Figure 4.4: **Classification of all possible regulatory profiles via compound logic**. Each of the 26 profiles, which shows nonzero response to at least one stress condition, was assigned to one of the five modes of regulation (A-F). The functions in each mode of regulation were further indexed numerically. The regulatory profiles in bold represents the behaviours that were not considered by Rasmussen et al. [206].

of individual genes. Each of 25430 genes in the RPM was assigned with one of the 26 logical function for each of the six stress combinations measured in the experiment. Note that we gave new indexing of each logical function according to the six mode of regulation (A-F, Fig. 4.4). We then analysed the distribution of six regulatory modes and their constituting regulatory profiles across the whole *Arabidopsis* transcriptome under different pairs of stresses (Fig. 4.5).

The results showed that Combinatorial and Cancelled modes occur frequently under all stress pairs, where they account for more than 75% of the total transcriptomic response together (Fig. 4.5a). Independent and Similar modes, which represent the overlap between the set of genes regulated by single and combined stresses, occupy approximately 20% of the total transcriptomic responses. Prioritised and Reversed modes are observed from only a small number of genes, affecting less than 5% of transcriptome under all stress pairs. Overall, the differences between stress pairs are not apparent from classification of six modes, with one exception: pairs of abiotic and biotic stress (H+F, C+F) do not invoke Similar mode, which contrasts the combinations of abiotic stresses (N+H, N+HL, H+HL, C+HL). This suggests that the number of genes with ubiquitous roles in abiotic and biotic stress response is negligible.

Distributions of individual logical functions led more detailed visualisation of similarities and differences between the transcriptomic responses to the stress pairs (Fig. 4.5b). From consideration of individual logical function, it can be seen that prevalence of a regulatory mode within the overall transcriptomic response is mainly due to high occurence of only a few logical functions. For example, most of the genes regulated via Combinatorial mode show the function A1 and A2 under all stress pairs, while almost no genes are described with A5 and A6. Similarly, B6 in Cancelled mode is also rarely observed across all stress pairs.

Analysing the similarity between most and least frequently observed logical functions inform us about the mechanisms that are commonly used, and those that are not. The most

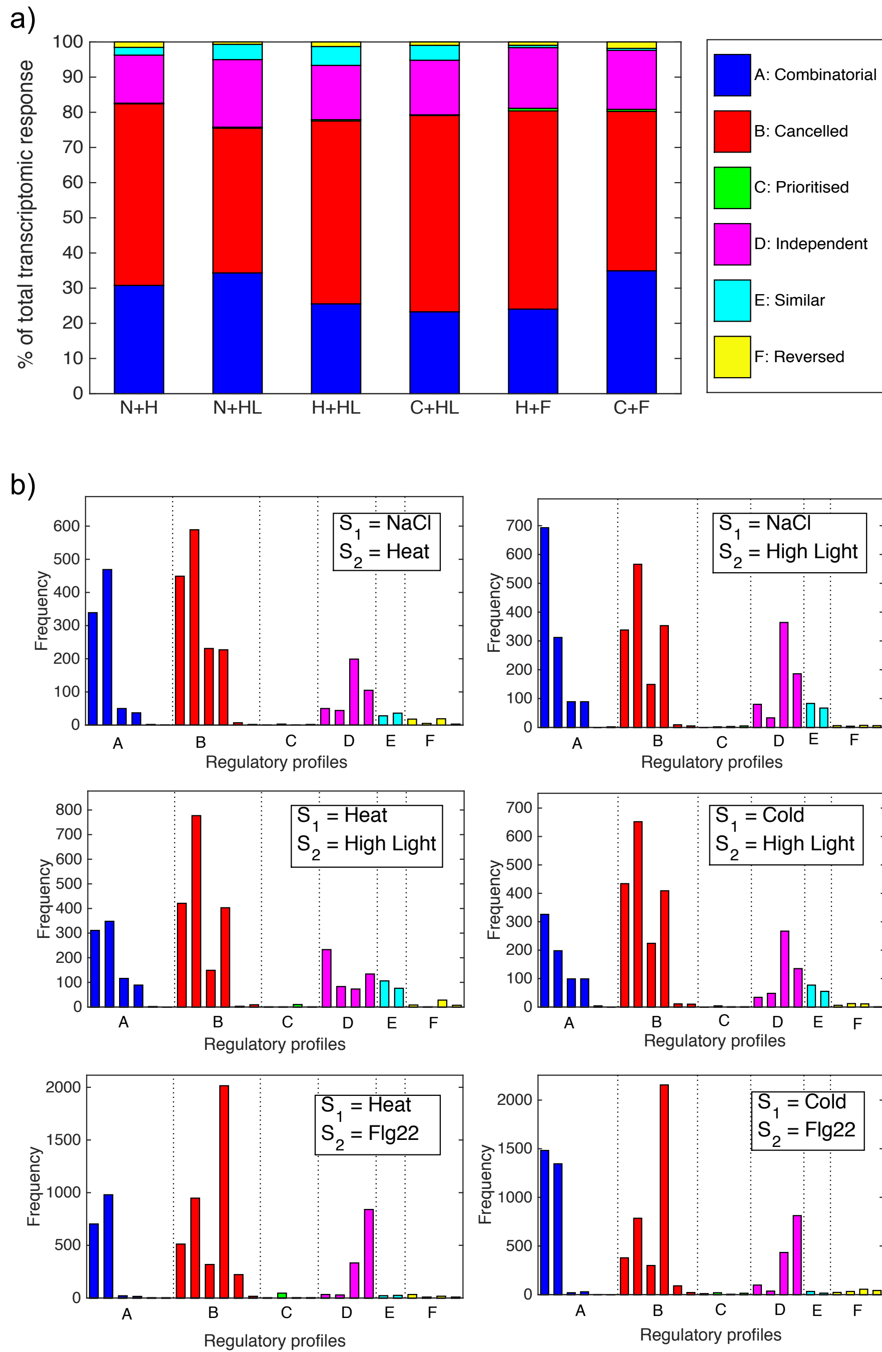


Figure 4.5: **Regulation of individual genes across transcriptome under different pairs of stresses**. a) Distribution of the six regulatory modes the total transcriptomic set induced by each pair of stresses. b) Breakdown of each mode into individual regulatory profiles (in ascending orders of ID). The height of each bar shows the number of genes showing the corresponding regulatory profile under the designated pair of stresses.

frequently observed functions, such as A1, A2, B1, B2, B4 all describe the genes that are specifically responsive to one type of stress condition, either single or combined. Presence of those genes suggests that their corresponding regulatory network act in autonomous manner, responding only to one type of stress amongst the tested combination. In contrast, the least frequently observed functions, such as A5, A6, B6 along with all of Prioritised and Reversed profiles, describe the genes showing incoherent responses to different stress conditions, suggesting that it is uncommon for genes to respond positively to one type of stress while responding negatively to another. An exception to this observation is B5 in the pairs of abiotic and biotic stresses (H+F, C+F), which describes the genes responding positively to singly applied abiotic stresses and negatively to Flg22 treatment. Enrichment of this function indicates that there may be genes with contradictory roles in coordinating the responses to abiotic and biotic stress.

4.3 Prediction of signalling network topologies from the transcriptomic dataset

4.3.1 Heat and biotic stress-inducible genes regulated by HSF- and WRKY-mediated signalling pathways

Having analysed the distribution of compound logical functions across Arabidopsis transcriptome, we investigate the topology of signalling network responsible for regulating stress response genes upon single and combined stress inputs. This is achieved by applying the PDP network models developed in section 3.4 for the set of genes displaying particular compound logic function. Given that multiple topologies can be used to describe a given compound logic function, however, additional information regarding the promoter design of the genes is required. We therefore choose a specific stress pair and examine the compound behaviours of a smaller set of genes to that are known to respond to that pair. Here, we choose the

combination of Heat + Biotic stress as a demonstrative example. The key regulators of transcriptional response to heat stress are Heat Shock Factors (HSF), which bind to Heat Shock Element (HSE) motifs in the promoters of target genes [234]. Transcriptional responses to Flg22 is triggered by activation of WRKY family of transcription factors [193], which bind to conserved W-box motifs in the promoters of target genes [47].

Although the HSF and WRKY pathways are known to transduce heat and biotic stress signals, they are not the sole mediators of transcriptional responses to the streses [185, 255]. Furthermore, presence of HSE or Wbox motifs in cis-regulatory regions does not guarantee the genes' response to the corresponding stress. Thus, we assume that HSF and WRKY pathways regulate only a partial subset of genes responding to Heat stress and/or Flg22.

Fig. 4.6 shows how we isolated the genes that are regulated by HSF and WRKY pathways by comparing the list of genes showing significant changes under singly applied Heat or Flg22 from microarray data set (p-value < 0.1) and the list of genes known to contain HSE or Wbox motifs in their promoters. The latter list, which contain 693 genes (205 genes containing HSE and 488 genes containing Wbox), was obtained from the Stress-responsive transcription factor database (STIFDB2)[i] [178]. The database hosts information about the enriched Transcription Factor Binding Site (TFBS) motifs for the stress-responsive genes. We then looked for the genes located in the intersect between the different lists of genes, identifying 3 subsets of genes.

4.3.2 Distribution of regulatory profiles amongst the genes regulated by HSF and WRKY pathways

Once the genes regulated by the HSF and WRKY pathways are identified, we classified each gene with one of the compound logic function defined in Fig. 4.4, and the resulting distribution of compound logic functions is shown in Fig. 4.7. The compound logic functions observed from less than 5% of the chosen set of genes were omitted to account for the

[i]Available online at http://caps.ncbs.res.in/stifdb2/

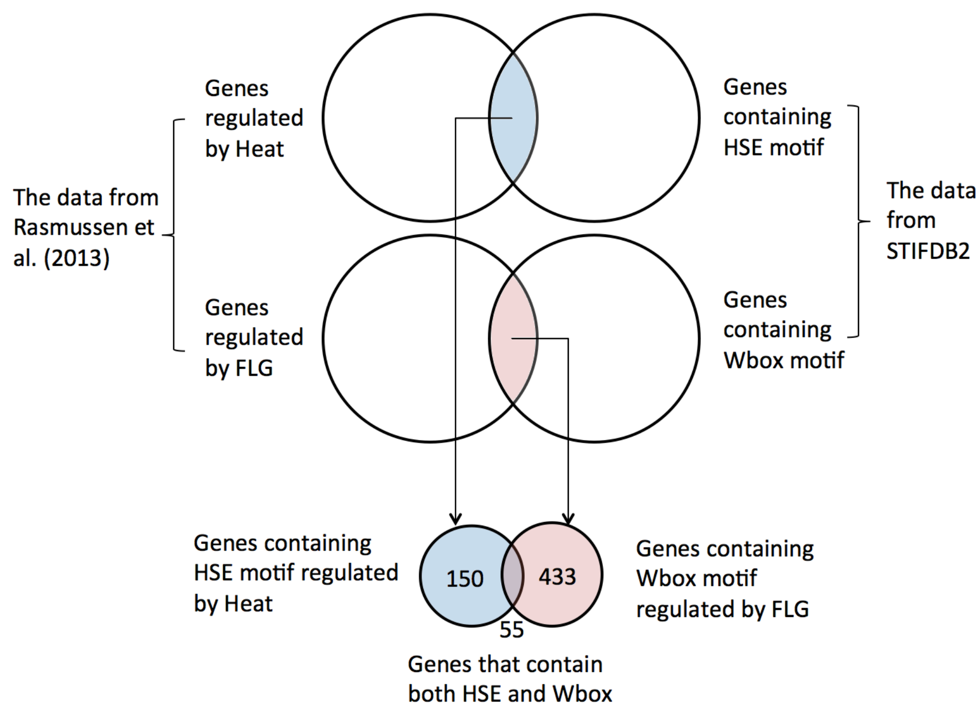


Figure 4.6: **Identification of gene subsets controlled by Heat stress and/or Flg22**. To identify the specific subsets of genes controlled by HSF- and WRKY-mediated pathways, we integrated two sources of information - the list of genes showing statistically significant change upon Heat stress or Flg22 treatment in the dataset of Rasmussen et al. [206], and the list of genes that contain HSE and/or Wbox motifs in their upstream promoter region from STIFDB2 database [178]. The resulting sets contain 693 genes in total, which are responsive to either Heat stress, Flg22 treatment, or both via the HSF- and the WRKY-mediated pathways.

possibility of erroneously identified functions from data uncertainty. Overall, the result indicates that the chosen subset of genes show wide variety of compound logic functions. The fact that the behaviours of the genes with common type of TFBS cannot be described a single compound logic function suggests for presence of multiple signalling pathways targeting the same TFBS, possibly by acting on different isoforms of HSF and WRKY transcription factors.

The majority of genes in all three subsets of genes show Cancelled mode of regulation (B1-B6), suggesting that the regulatory cues transduced via most of the HSF and WRKY pathways are nullified upon combination of the two stresses. However, presence of Independent

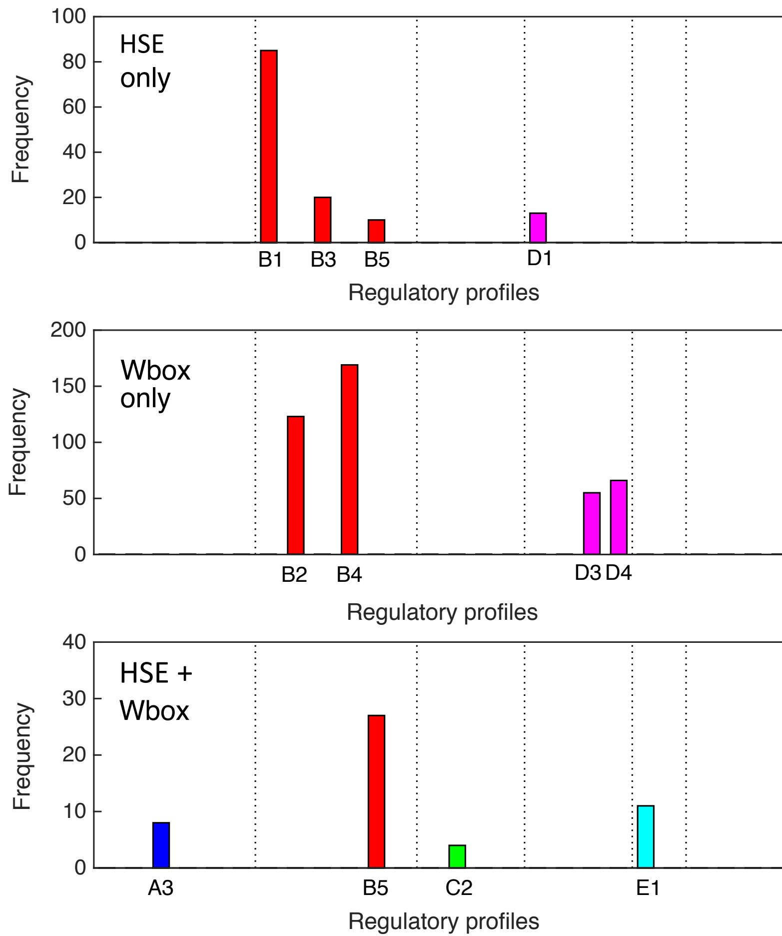


Figure 4.7: **Distribution of regulatory profiles amongst the genes containing either HSE, Wbox or both**. The regulatory profiles (in ascending order of ID) are organised into six modes of regulation: Combinatorial (A), Cancelled (B), Prioritised (C), Independent (D), Similar (E) and Reversed (F). The result shows that the genes with only HSE motifs in their upstream promoter (top panel) mainly show Cancelled modes (B1, B3, B5), indicating that the heat and Flg22 stress signals are acting antagonistically to regulate this subset of genes. On contrary, the genes controlled only by Wbox motifs (middle panel) show both Cancelled (B2, B4) and Independent modes (D3, D4), suggesting for presence of multiple WRKY pathways that respond differently to heat stress. The genes with both motifs (bottom panel) mainly show nullification of regulatory change upon stress combination (A3, B5) and similar behaviours upon all stress conditions (E1), with a small minority of them prioritisation of Flg22 response (C2). See Fig. 4.4 on page 65 for definition of each regulatory profile.

(D1-D4) and Similar modes (E1, E2) also indicate that there are numerous genes whose behaviours are unaffected by addition of the second stress. Several genes that contain both HSE and Wbox motifs show Combinatorial (A3) and Prioritised mode of regulation (C2).

A notable difference between the HSE-only and Wbox-only subsets is that most genes belonging to the former exhibit positive change upon introduction of single stress input, whereas the genes in the latter subset show both positive and negative response to single stress input. For instance, the genes up-regulated by Heat (B1) is observed more frequently amongst the genes containing only HSE motif compared to the genes down-regulated by the same stress (B3). This contrasts what is observed from the genes containing only Wbox motif, where the number of genes down-regulated by Flg22 (B4) is comparable to those up-regulated by the same stress (B2). Together, these observations suggest that the HSF pathways are mainly activatory, whereas the WRKY pathways play both activatory and inhibitory roles.

We acknowledge, however, that some features are artefacts of the method for identifying the subsets of genes regulated by HSF and WRKY pathways. Because the genes were assigned to corresponding subset based on their ability to respond to single Heat or Flg22 stress, the genes that specifically respond to Heat + Flg22 combination (A1 and A2) are not identified. The method of identifying genes also cause several regulatory profiles to occur exclusively in certain subset of genes, such as B1, B3 in the HSE-only subset and B2, B4 in the Wbox-only subset. Because the HSE+Wbox subset was obtained from finding the intersection between HSE-only and Wbox-only subset, all regulatory profiles in HSE + Wbox subset represent the genes that are inducible by both single Heat and single Flg22 stress.

4.3.3 Gene ontology analysis

Having identified the group of genes for each mode of transcriptional regulation, we examined whether the group of genes are associated with particular biological function (Table 4.2) using AmiGO analysis tool [36]. Due to small number of genes, statistically significant enrichment of gene ontology terms was seen only from the whole set (150 HSE containing genes, 433

Table 4.2: **Enrichment of gene ontology terms amongst the genes containing either HSE or Wbox motifs.**

	HSE containing genes		**Wbox containing genes**	
	Enriched biological processes	P-value	Enriched biological processes	P-value
All	response to oxygen-containing compound	1.73E-10	single-organism process	7.88E-13
	response to jasmonic acid	5.68E-04	**response to oxidative stress**	6.74E-10
	response to salt stress	9.73E-04	organonitrogen compound metabolic process	8.59E-09
	response to light stimulus	1.03E-03	sulfur compound metabolic process	1.80E-07
	response to osmotic stress	2.16E-03	**response to osmotic stress**	3.44E-06
	response to radiation	2.78E-03	response to cadmium ion	4.10E-06
	circadian rhythm	3.57E-03	pyruvate metabolic process	4.89E-06
	monocarboxylic acid biosynthetic process	4.28E-03	**response to wounding**	6.28E-06
Cancelled by another stress	**monocarboxylic acid biosynthetic process**	4.74E-03	**response to oxidative stress**	7.98E-06
	response to salt stress	5.92E-03	**response to wounding**	2.07E-05
	response to osmotic stress	1.07E-02	**response to osmotic stress**	5.63E-03

Wbox containing genes) and the genes showing cancelled mode amongst those genes. The analysis shows that combining Heat and Flg22 stress leads to attenuation of specific biological processess: for HSE containing genes, monocarboxylic acid biosynthesis, response to salt and osmotic stress are nullified, whereas responses to oxidative stress, wounding and osmotic stress are suppressed.

4.3.4 Network topologies for the genes with only HSE or Wbox motifs

Classification of the gene behaviours with compound logic functions allows description of HSF and WRKY pathways as PDP networks. Figure 4.8 shows the topologies of HSF and WRKY pathways expressed as PDP networks, derived from the genes controlled by homotypic promoters (i.e. either or HSE or Wbox only).

The resulting structures suggest that the HSF and WRKY systems consist of two subsystems.

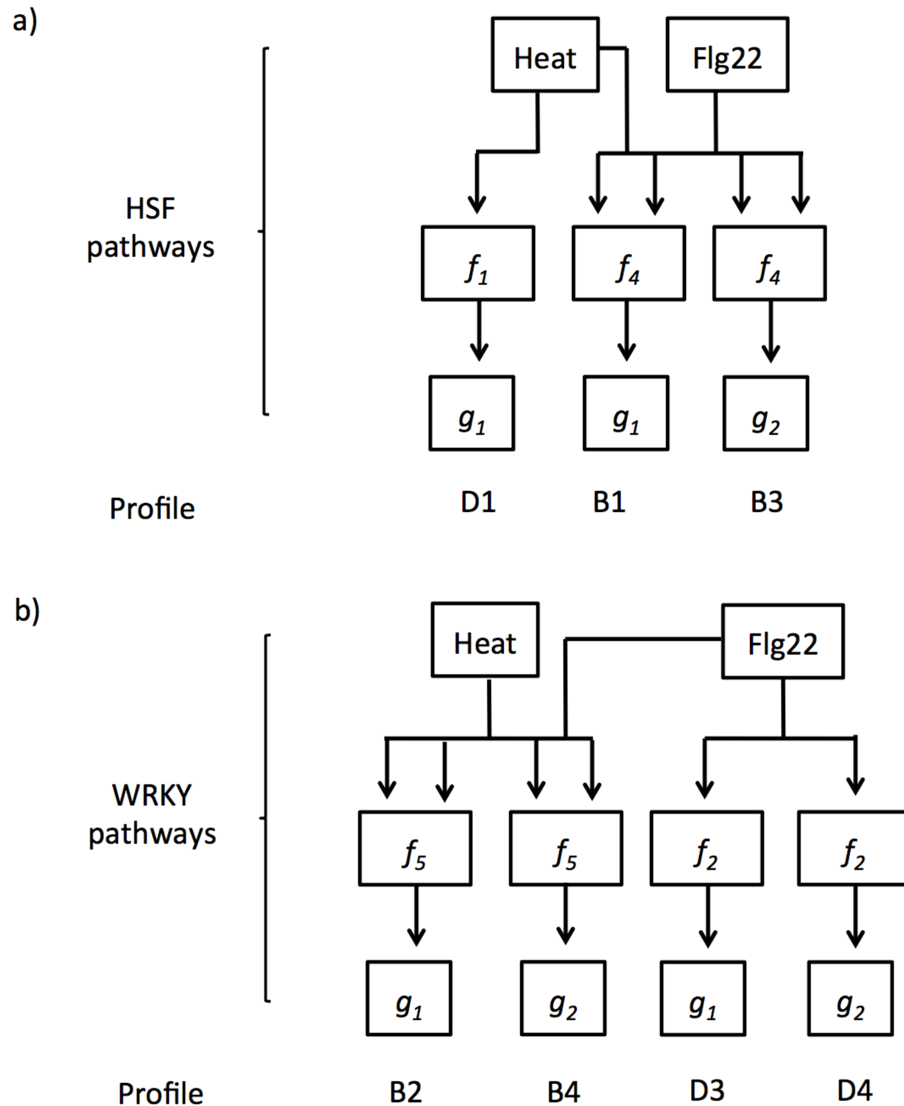


Figure 4.8: **Possible Parallel Distributive Processing (PDP) network topologies for regulatory profiles of the genes with homotypic promoter**. Each observed profile is described by a combination of functions f and g described in Chapter 3 (Table 3.1 and 3.2). a) The structure of the HSF pathways predicted from distribution of compound logic functions suggests that there are three types of HSF pathways. Two pathways are antagonistically regulated by Heat and Flg22, whereas the other pathway is independent of Flg22. Notably, this heat-specific pathway only involves transcriptional activation. b) The predicted structure of the WRKY pathways suggests for four types of WRKY pathways, with two pathways showing antagonism between Heat stress and Flg22 and the other two pathways responding only to Flg22.

One is only dependent on cognate stress input, such as Heat for the HSF pathway and Flg22 for WRKY pathway, described by f_1 and f_2. The genes regulated by these pathways are unaffected by combination of two stresses. On the other hand, there is another type of signalling pathways that are inactivated by addition of non-cognate stress input described by f_4 and f_5. Both signalling pathways dependent on one or two stresses trigger transcriptional activator and inhibitor.

Most of the logical functions observed from the chosen subset of genes could be explained with the identified network topologies with a single TFBS, with one exception. Several genes in the HSE-only subset displayed function B5, indicating that those genes are able to respond to singly applied Flg22 stress even though they do not contain W-box motif in their promoter. This suggests for the presence of another biotic-stress signalling pathway that does not involve Wbox motif.

4.3.5 Network topologies for the genes with both HSE and Wbox motifs

The genes with both HSE and Wbox motifs are regulated simultaneously by HSF and WRKY signalling pathways ($n = 2$). Because it was shown in the previous chapter that there are numerous network topologies that describe each of the 27 regulatory profiles (Fig. 3.7), it is impossible to determine which network topology from the distribution of regulatory profiles in HSE+Wbox subset alone.

From the analysis of HSE-only and Wbox-only subsets, however, it is known that the HSF pathway is described by either f_1 or f_4, while the WRKY pathway is described by either f_2 or f_5. Assuming that the same signalling pathways act on the genes with dual binding sites, it is possible to restrict the number of possible network topologies involving only those four functions. Table 4.3 shows the choice of the network topologies based on this restriction. The result indicates that profiles A3, C2, and E1 observed from the genes in HSE+Wbox

Table 4.3: **Possible Parallel Distributive Processing (PDP) network topologies for the genes containing both HSE and Wbox motifs.**

Regulatory Profile	X_1	X_2	P	Type of TF Interaction
A3	f_4	f_5	g_6	Independent
B5	f_1	f_2	g_4	Independent
	f_4	f_5	g_4	Independent
	f_4	f_5	g_7	Competitive
	f_4	f_5	g_8	Competitive
C2	f_1	f_2	g_9	Competitive
E1	f_1	f_2	g_6	Independent

subset could be described by a single network topology, except B5 which can be described by 4 network topologies.

Function A3 could be reproduced if the genes are simultaneously targeted by activatory HSF and WRKY pathways whose activities are nullified under combination of stresses (f_4 and f_5). Function E1 is also reproduced when both HSF and WRKY pathways are activatory, but acting independently to each other. On the other hand, functions B5 and C2 could be reproduced by the network topology with activatory HSF pathway and inhibitory WRKY pathways. When the two pathways are inducible only by their cognate stress inputs (f_1 and f_2), the mode of interaction between TFs determine the regulatory profile. For example, the activatory and inhibitory cues may cancel each other out (g_4) if the HSF and WRKY transcription factors independently bind onto cis-regulatory region, leading to function B5. Function C2 may be reproduced if the inhibitory WRKY signalling pathway outcompetes positive HSF pathway. An alternative explanation for function B5 is that the signalling pathways are attenuated by the combination of two stresses prior to TF-DNA binding (f_4 and f_5), resulting in no change in expression regardless of the mode of TF interaction.

4.4 Discussion

In this chapter we have explored how plants choose subset of genes to be regulated under different stress conditions by analysing the differential cDNA microarray dataset of Rasmussen et al. [206]. Our analysis of Regulatory Profiles Matrix obtained from the dataset led to the conclusion similar to the original study that subsets of genes regulated by individually applied stresses significantly differ to the subsets of genes regulated by combination of stresses. By describing the behaviours of each gene in terms of the compound logic functions developed in the previous chapter, we were able to visualise the patterns amongst transcriptomic profiles under different pairs of stress. This also allowed further investigation for the sources of the observed behaviours using the logical network models. We have classified putative targets of HSF- and WRKY-signalling pathways according to compound logic, which led to a newly proposed structure of the Heat/Flg22 signalling network integrating the HSF- and WRKY-mediated pathways (Fig. 4.9). The structure exhibits two features that are proposed for HSF- and WRKY-mediated signalling pathways: independent or competitive binding between HSF and WRKY transcription factors at the level of transcriptional regulation, and cross-inhibition of HSF- and WRKY pathways by Flg22 and Heat signals are proposed.

Although the evidence of antagonism between heat and pathogenic signals at physiological level already exists [74], the molecular mechanisms of interaction between Heat and pathogenic signals such as Flg22 are unknown. Because our model constructed as a logical network was restricted to the set of genes that are known to be regulated by two types of TFs, a direct hypothesis regarding the connection between the signalling networks upstream of the chosen TFs could be generated. A potential mechanism for crosstalk between the WRKY signalling pathway with heat signals may involve Calmodulin (CaM), which perceives heat-induced increase in cytosolic Ca^{2+} (Chapter 2), as numerous WRKY proteins were found to contain CaM-Binding Domain (CaMBD), which is a conserved structural motif found in group IId of the WRKY protein family [19, 194]. On the other hand, members

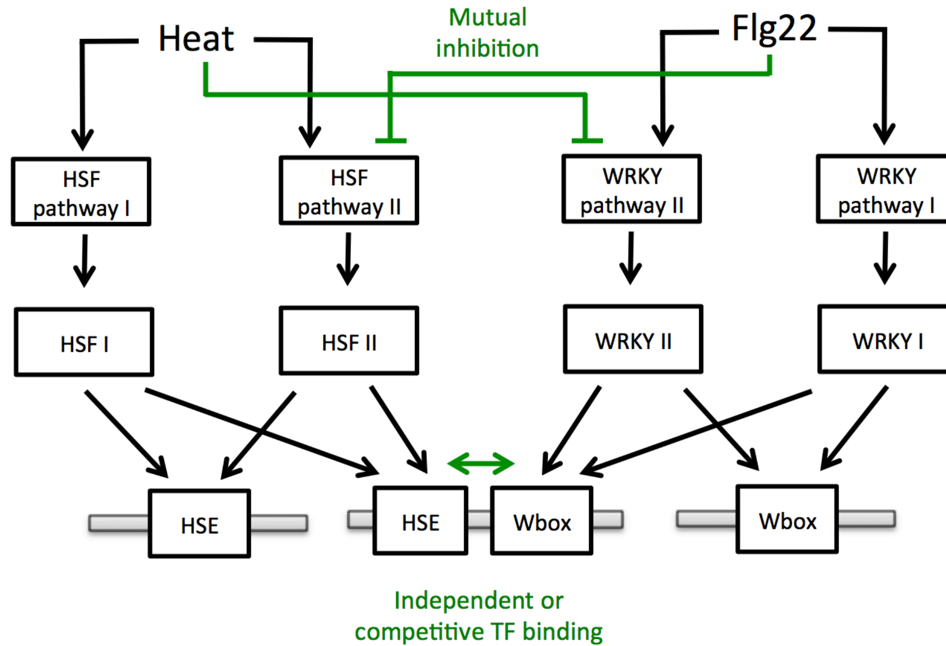


Figure 4.9: **Proposed structure of HSF- and WRKY-mediated signalling pathways**. Green arrows represent novel mechanisms that are predicted from the regulatory profiles observed amongst the chosen subset of genes.

of MAPK cascades appear to be viable candidates for mediating crosstalk between HSF signalling pathways and the Flg22 signal, based on a recent paper reporting interaction between MAPK6, the principal signalling kinase of PAMP-induced MAPK cascade, with HsfA2, a major HSF regulator of heat stress along with HsfA1 [61]. Because whether those existing crosstalk connections facilitate negative regulation as predicted in the logical network model is not known yet, more detailed biochemical characterisation of the aforementioned proteins upon single and combined Heat and Flg22 stress conditions will further clarify the nature of interaction between the two stresses.

Similar analysis and modelling of interactions as was presented for Heat and Flg22 stress can be conducted for other stress pairs. Classification of genes into pre-defined regulatory profiles offers unambiguous distinction of the possible outcomes on gene expression from combining two types of stress. Overall, the variation observed in the distributions of the

regulatory profiles between for different pairs of stresses suggest that interaction between two stresses for regulation of gene expression cannot simply be summarised as synergistic or antagonistic. However, it was observed that combination of two types of stress results in extensive transcriptional reprogramming, with suppression of the genes that are induced by singly applied stresses (Cancelled), or expression of the genes that are normally inactive under single type of stress (Combinatorial). This supports the notion that the responses to the combination of stresses may indeed be unique compared to the responses to single stresses.

In the previous chapter, we discussed about the limiting assumptions used in the compound logic method and the resulting description of signalling networks with PDP network models (Section 3.5). How do these assumptions affect the validity of the prediction presented above? Because the dataset analysed here was taken only at a single time point (4 hours after introduction of stress), it is highly likely that regulatory outcomes observed from each gene may be transient. If the compound logic method is applied to a different dataset taken at a different time point, the result of analysis can change: for instance, temporal difference in gene expression profiles will affect the distributions of compound logic function shown in Figs 4.5 and 4.7. In fact, distributions of compound logic functions, and the resulting prediction about the upstream regulatory network structure, are susceptible to change if behaviours of the subjected genes vary significantly across time.

To obtain more reliable prediction about the network structures, it is therefore essential to investigate further about temporal changes in transcriptomic profiles. A suggested course of action is to make comparison across different time points and investigate whether the assignment of genes with compound logic functions changes across time. Functions that are consistently observed across different time points within the target set of genes may indicate that the behaviour of genes can be reasonably described by qualitative changes, enhancing reliability of the prediction made based on the PDP interpretation of the signalling network. If a gene cannot be described by a single function across different time points, it is possible that temporal profile may play an important role in the gene's function. Unfortu-

nately, it would not possible to extract meaningful information about the regulatory network topologies at upstream of those time-varying genes.

In conclusion, we have shown that the compound logic description can be used to analyse large-scale transcriptomic data upon single and combined stress, and make predictions about potential connectivities between the signalling pathways transducing different types of stress signals. Given that behaviours of genes cannot always be described qualitatively, however, we also learnt that more careful consideration of potential temporal behaviours is necessary. The focus of next chapter is therefore on how simultaneous application of multiple stress factors affect temporal dynamics of gene expression, a topic largely left unexplored in current literature.

Chapter 5

Experimental measurement of ***RD29A*** expression dynamics in ***Arabidopsis thaliana***

5.1 Introduction

In this chapter, we study how combinations of different stress stimuli affect the temporal dynamics of stress response gene expression. Because an experimental study of temporal gene expression dynamics at a transcriptomic scale is highly costly and time-consuming, here we focus on a single gene called *Response-to-Dehydration 29A* (*RD29A*) in *Arabidopsis thaliana*, a model stress-responsive gene encoding a 78kD hydrophilic protein [269] of unknown function [174]. *RD29A* has played an important role in abiotic stress research particularly osmotic, salinity and dehydration stresses due to its inducibility by multiple types of stresses [149, 270]. This makes the choice of *RD29A* as the model gene is particularly relevant for the study of gene regulation in response to combined stress. In addition, the mechanism of *RD29A* expression regulation is relatively well understood, with two cis-regulatory elements identified in the *RD29A* promoter that participate in regulation by stress signals. Multiple copies of ABRE and a DRE provide a basis for ABA-dependent and ABA-independent

regulation of the *RD29A* promoter, respectively [270].

Multiple simultaneous stress stimuli leads to various types of combinatorial effect in regulating *RD29A* expression. For example, Xiong et al. [264] measured *RD29A* expression under combinations of NaCl and ABA, and reported that combinations of NaCl and ABA treatment leads to synergistic activation of *RD29A* expression, where expression level of the gene in response to combined NaCl and ABA exceeded the sum of the levels induced by individual stress inputs. Complementary to this finding, it has been suggested that the two types of TF-binding motifs in *RD29A* promoter, DRE and ABRE (discussed in Chapter 2), interact with each other to induce synergistic *RD29A* expression [183]. However, the synergy between NaCl and ABA in inducing *RD29A* expression reported in these studies is deduced from single time-point measurements, which do not inform us how the combined NaCl and ABA stimuli affect dynamics of *RD29A* expression.

To address this gap, we subsequently performed real-time quantitative PCR (qRT-PCR) experiments to quantify *RD29A* transcript abundance at different durations of treatment. The results indicate that the *RD29A* transcript abundance fluctuates over time even in absence of NaCl or ABA, suggesting that *RD29A* expression may be dependent on circadian rhythm operating independently to the gene regulatory mechanisms triggered by NaCl and ABA. To discern the effect of NaCl and ABA stimuli on *RD29A* expression, we remove such stress-independent fluctuation from NaCl- and ABA-induced *RD29A* expression profiles by normalising them with the *RD29A* expression profile under unstressed control. The resulting time course profiles reveal distinct qualitative features that are associated with single and combined NaCl and ABA treatments.

5.2 Materials and Methods

5.2.1 Stress treatment and Sample preparation

Arabidopsis thaliana (ecotype Col-0) seedlings were stratified at 4 °C for 48 hours, followed by growth on agar plates containing Murashige-Skoog medium for 5-6 weeks in constant temperature (20 °C). The seedlings were entrained with 12 hour light / 12 hour dark cycle (9:00 to 21:00 hours in real time) during growth, illuminated under 60 μmol $m^{-2}s^{-1}$ white light. Seedlings were then hydroponically treated, where each treatment medium with varying concentrations of NaCl and ABA was prepared by dissolving appropriate amount of NaCl and crystallised ABA in deionised water. The treatment conditions include control (H_2O only), 150mM NaCl, 300mM NaCl, 50μM ABA, 100μM ABA, 300mM NaCl + 100μM ABA, and 150mM NaCl + 50μM ABA. Samples ($n = 3$) were collected at different durations of stress treatment (0, 0.5, 1, 2, 3 and 5 hours after initial exposure to stress). Three replicate samples were made from collection of several randomly selected seedlings into three Eppendorf tubes, with each tube approximately weighing 60mg (fresh weight) in total. The collected samples were then immediately frozen in liquid N_2 and stored at -80 °C prior to extraction of RNA. Initiation of stress treatment and sample collection occurred at the same time of day for all experiments, from 7:00 to 12:00 hours in Zeitgeber time (15:00 to 20:00 hours in real time). Zeitgeber time indicates the specific point of the light / dark cycle under which the plants were grown, with 0 ZT and 12 ZT indicating the start of light and dark period, respectively.

5.2.2 Sample processing

Tissue disruption and RNA extraction were carried out using RNEasy mini kits (Qiagen). RNA integrity was verified by using Nanodrop Spectrophotometer (ND-1000, Thermo Scientific Inc.); the samples with relatively high RNA yield (500-800 ng/l) and high DNA to RNA, RNA to salt separation ratio were selected. The resulting transcriptome samples were

Table 5.1: **List of primers**

Primer	Sequence
RD29A forward	5'-CCGGAATCTGACGGCCGTTTA-3'
RD29A reverse	5'-CCGTCGGCACATTCTGTCGAT-3'
actin-2 forward	5'-TCCTCACTTTCATCAGCCG-3'
actin-2 reverse	5'-ATTGGTTGAATACATCAGCC-3'

converted into cDNA using Quantitect Reverse Transcription kits (Qiagen). During this step, the samples were diluted accordingly to give the uniform concentration of 500 ng/nl, and were treated with DNAse to remove any trace of genomic DNA and obtain high-quality transcriptome samples. Real-time quantitative PCR experiments were carried out (Rotor-gene Q cycler, Qiagen) to measure the fold changes in RD29A expression compared to that of a control gene, *actin-2*. The two genes are specifically amplified using the pre-prepared primers (Table 5.1):

The reaction conditions are to be prepared using Rotor-gene Syber Green PCR kits. For optimal results, the reaction samples were diluted again, such that the template cDNA amount is 20ng per reaction. Prior to the qPCR experiments, each sample was divided further into these technical replicates in order to achieve high technical accuracy.

5.2.3 Quantitative RT-PCR Data analysis

Cycle-Time (CT) data was obtained from the resulting fluorescence data of quantitative PCR experiments by setting a threshold value (normalised fluorescence = 2.5 x 10^{-3} RFU). The CT values for *actin-2* transcript abundance were then subtracted from that of *RD29A* transcript abundances to obtain ΔCT. Calculating $log_2(\Delta\mathrm{CT}(S, \mathrm{ZT}))$ quantifies the relative level *RD29A* transcript abundance with respect to that of *actin-2*, given under stress condition S for the specific time of day measured in Zeitgeber time (ZT). To allow comparison

between different time course profiles, we calculate fold change m with

$$m(S, \mathrm{ZT}) = log_2(\Delta\mathrm{CT}(S, \mathrm{ZT})) - log_2(\Delta\mathrm{CT}(0, \mathrm{ZT}_0)), \tag{5.1}$$

where $log_2(\Delta\mathrm{CT}(0, \mathrm{ZT}_0))$ indicates the transcript abundance in absence of stress inputs ($S = 0$) at specific reference time (ZT = 7:00 hour). Normalising transcript abundance as such allows direct quantitative comparison between the time course profiles, as all profiles start from fold change of 1.

5.3 Results

The experimentally measured temporal profiles of *RD29A* expression in response to various treatment conditions is shown in Fig. 5.1. The chosen concentration of NaCl and ABA in the hydroponic treatment media represent the level of stress at which *RD29A* expression is strongest (300mM NaCl and 100μM ABA), and their half (150mM NaCl and 50μM ABA), based on a previous study [264]. Each data point indicates mean relative fold change, $\bar{m}(S, \mathrm{ZT})$ with $n = 3$. Error bars associated with the time points represent standard deviation calculated from the triplicate samples, where $\mathrm{SD}(S,\mathrm{ZT}) = \sqrt{\frac{1}{3}\sum_{i=1}^{3} |m_i(S, \mathrm{ZT}) - \bar{m}(S, \mathrm{ZT})|^2}$.

The time course expression profile obtained under H_2O control (i.e. $\bar{m}(0, \mathrm{ZT})$) shows that *RD29A* expression level in absence of NaCl or ABA does not remain constant over time (Fig. 5.1a). Although this fluctuation is weaker than changes induced in *RD29A* expression level by stress inputs (Fig. 5.1b-d), it is still non-trivial as the induced *RD29A* expression profiles may contain some components the intrinsic variability seen from the H_2O control. Assuming that the effect of time-variability of the control on induced expression profile is multiplicative, we remove stress-independent variability of control from each sample by calculating

$$M(S, t) = \frac{m(\mathrm{S,ZT})}{\bar{m}(0, \mathrm{ZT})}, \tag{5.2}$$

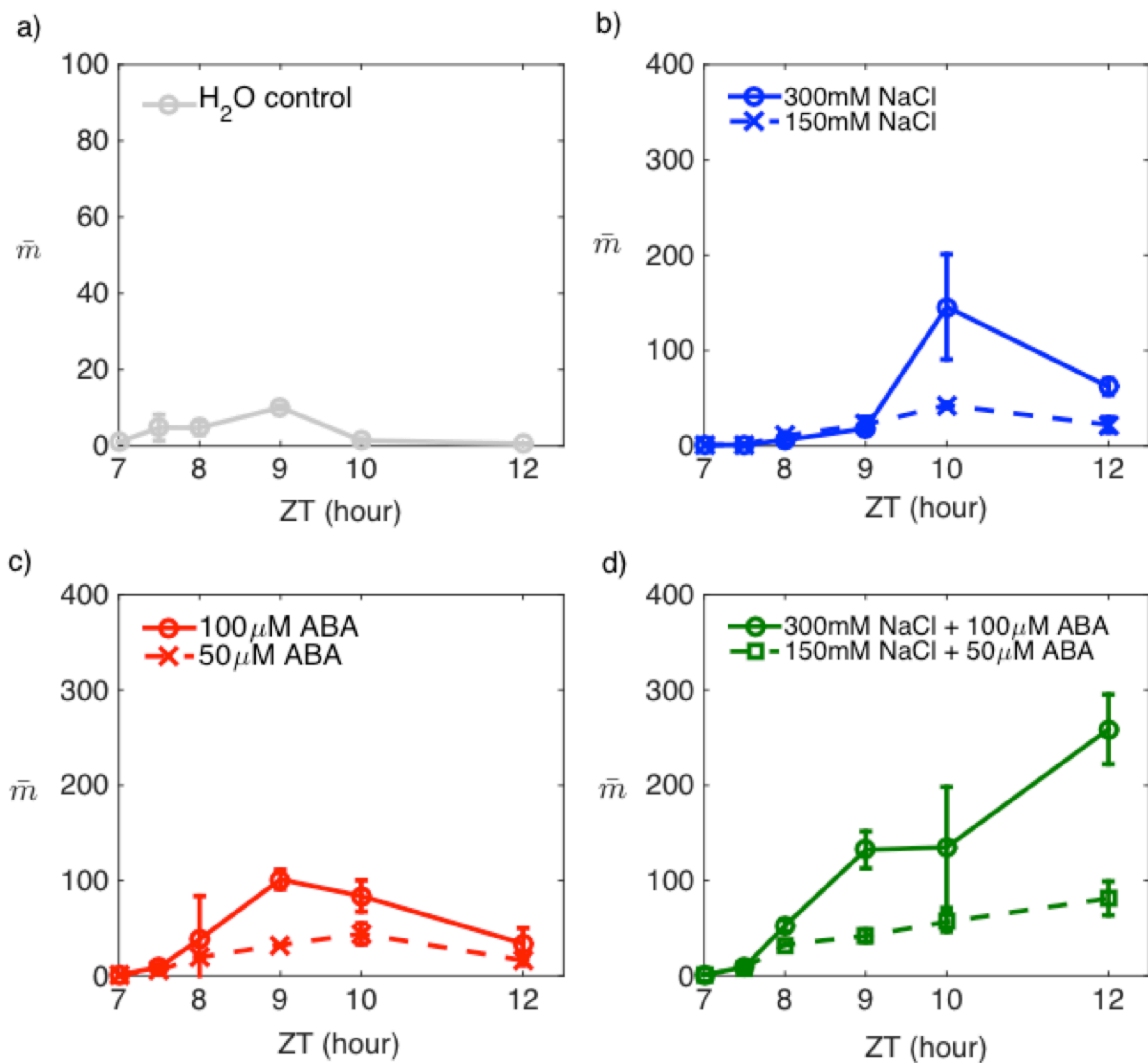


Figure 5.1: **Experimental measurements of *RD29A* expression level at different time and treatment conditions**. a) *RD29A* expression profile under negative control treatment (H_2O), b) NaCl treatments at full (300mM) and half-stregnth (150mM), c) ABA treatments at full (100μM) and half-strength (50μM), and d) combined NaCl and ABA treatments at full (300mM NaCl + 100μM ABA) and half-strength (150mM NaCl + 50μM). Horizontal and vertical axes represent time of day at which the measurements were taken (ZT) and fold change of the transcript abundance measured each time point with respect to start of treatment, respectively. Each data point represent mean fold change, $\bar{m}$, and the error bars represent $\pm$ SD.

where M denotes fold change in *RD29A* expression level that is independent of the intrinsic variability seen from the H_2O control. Since there is no longer a intrinsic variability in control dependent on time-of-day (ZT), the normalised fold changes are now measured with respect to durations of treatment, denoted by t. Furthermore, normalising each sample with mean of the H_2O control also means that the error bars are scaled by constants, such that the new measure of error SD' = $\mathrm{SD}/\bar{m}(0, \mathrm{ZT})$.

Figure 5.2 shows the time course profiles obtained under different stress conditions without the intrinsic time-variability of the control, which are equivalent to the responses triggered purely by the stress inputs. It is observed that the responses to all stress treatments consist of two distinct phases. During the early phase ($t \leq 2$), only a small increase of expression is observed with a negligible increase induced by 300mM NaCl and approximately 10-fold increase induced by 100μM ABA (Fig. 5.2a, b). Transcript abundance during the late phase ($t > 2$) is significantly greater than that in the early phase, where 300mM NaCl induces up to a 110-fold increase in transcript abundance, while 100μM ABA induces up to a 60-fold increase. Combined stimulation resulted in much larger increases in *RD29A* transcript abundance, up to 460-fold by the combined NaCl and ABA stresses at full strength, and up to 150-fold at half strength (Fig. 5.2c). Abrupt changes in transcript abundance are observed under all stress conditions between 2 and 3 hours post-stress, suggesting that the main production of *RD29A* transcripts initiates mainly after 2 hours of stress exposure.

It is possible to more clearly visualise the qualitative features unique to each stress condition by comparing the time course profiles in logarithmic scale (Fig. 5.2d). First, there is a negligible increase in the level of *RD29A* transcript during the first 2 hours of NaCl stress treatment (blue), whereas the ABA and combined NaCl and ABA treatments induce weak yet rapid induction of *RD29A* expression during the early phase. Second, the responses to single NaCl and single ABA treatments, regardless of strength of the stress inputs, appear to reach plateau levels from 3 hours of treatments. This observation is supported by the fact that there is no significant statistical difference between the groups of samples obtained at 3 and 5 hours of treatment (Table 5.2). In contrast, the responses to combined NaCl and ABA do not

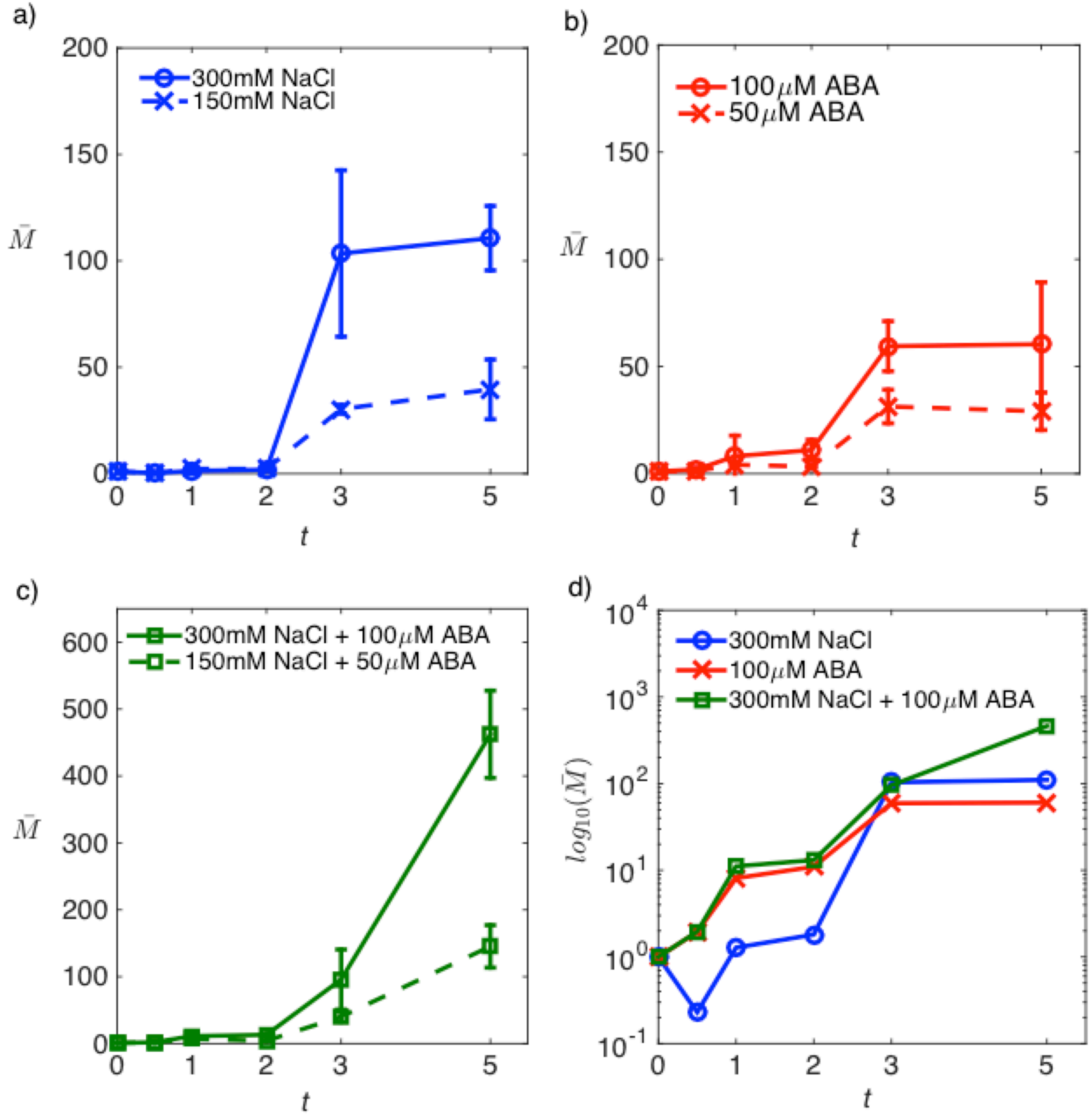


Figure 5.2: **Temporal dynamics of *RD29A* expression with intrinsic stress-independent variability removed**. a) Mean normalised fold change $\bar{M}$ upon single NaCl treatments at full- and half-strength, b) upon single ABA treatments at full- and half-strength, and c) upon combined NaCl+ABA treatment at full- and half-strength. Error bars represent ± SD' (see text). d) Comparison between the time course profiles obtained under single and combined stress conditions at full strength at logarithmic scale. Horizontal axes for all panels represent duration of treatment, t hours.

Table 5.2: Comparison of *RD29A* expression level between 3 and 5 hours of treatment (two-sample t-test)

Treatment condition	$\bar{M}$ (σ^2)		p-value
	3 hour	5 hour	
300mM NaCl	103.4 (39.1)	110.7 (15.1)	0.825
150mM NaCl	30.3 (1.92)	39.8 (14.1)	0.605
100μM ABA	59.4 (11.7)	60.3 (28.9)	0.969
50μM ABA	31.3 (7.84)	29.8 (8.73)	0.789
300mM NaCl + 100μM ABA	95.7 (44.9)	462.3 (65.2)	0.007**
150mM NaCl + 50μM ABA	40.2 (7.44)	145.3 (31.7)	0.005**

$\bar{M}$ and SD' represents normalised fold change and standard deviation, respectively.

** Significant difference at p<0.01.

converge onto single values within 5 hours of treatment, which is supported by statistically significant difference between the samples obtained at 3 and 5 hours of treatments (Table 5.2). The observed differences between the expression level at 3 and 5 hours of treatment suggest that the responses to single NaCl or single ABA treatments may be reaching equilibrium more quickly compared to the responses to combined NaCl and ABA treatments.

Remarkably, halving the concentration of the stress inputs do not affect any of the qualitative features above, with the time course profiles obtained under half-strength inputs all showing two-phase response, and the differences in the early and late phase responses between different stress conditions. This indicates that the strength of input stimuli only affects the magnitude of fold increase in expression, not its dynamics. In fact, we observe that halving ABA concentration reduces expression fold change by approximately half across all data points, raising the possibility that strengths of stress inputs are linearly related to the magnitude of fold increase induced by the stress inputs.

5.4 Discussion

In this chapter, we have investigated the temporal dynamics of *RD29A* expression by experimentally measuring its transcript abundance at different durations of NaCl and ABA treatments. The resulting data suggests that NaCl and ABA and their combination leads to temporal profiles with qualitative features that are specific to treatment conditions. A key observation made from the analysis of the data is that *RD29A* expression fluctuates even without stress (Fig. 5.1a), which led to subsequent removal of the intrinsic time-of-day variability in *RD29A* expression seen from the H_2O control from all stress-induced time course profiles (Fig. 5.2). Based on purely stress-induced changes in the time course profiles, we identify three qualitative features characteristic to NaCl and ABA regulation of *RD29A* expression:

Feature 1 Induction of *RD29A* expression by all combinations of NaCl and ABA occurs in two phases, with weak initial increase in the level of expression followed by strong increase after 2 hours of treatment.

Feature 2 Strength of the input stimuli affects the magnitude of expression, but does not change the qualitative features described above.

Feature 3 *RD29A* expression remains at a constant level after 3 hours of treatment with single NaCl and ABA, while it continues to increase under combined NaCl and ABA treatment.

The treatment with combination of NaCl and ABA eventually leads to the level of expression greater than the sum of individual responses: this agrees well with the observations previously made in the study by Xiong et al. [264], which compared *RD29A* transcript abundance induced by single NaCl and ABA with combined NaCl + ABA after 4 hours of treatment, confirming the presence of synergy between NaCl and ABA signals in activating *RD29A* expression. However, our measurement of time course profiles provides a novel insight that combination of NaCl and ABA does not only affect the magnitude of expression, but also

the dynamical profile of the response. Under single NaCl and ABA, *RD29A* appears to maintain a constant level of transcript after approximately after 3 hours of treatment, which indicates that the rates of transcript production and degradation may reach equilibrium within 5 hours (Fig. 5.2a, b). The *RD29A* expression induced by combinations of NaCl and ABA, on the other hand, do not reach a steady state level within the same time frame (Fig. 5.2c). This suggests that interaction between NaCl and ABA inputs may possibly modulate the signalling process resulting in slowing down of the approach towards equilibrium.

If the the observed time course profiles are the properties of stress-dependent part of the *RD29A* regulatory system, what might be causing the fluctuation of *RD29A* transcript level in absence of stress? One possible explanation is that such variability of *RD29A* expression dependent on time-of-day originates from the effects of intrinsic circadian oscillation. In fact, it has been already established that regulation of *RD29A* expression in response to several types of abiotic stress is *gated* to circadian clock, where environmental stimuli of identical strengths applied at different time of day leads to transcriptional response of different intensities [99]. For example, the magnitude of *RD29A* expression in response to cold depends on time of day because the activity of its cold-induced upstream regulator, DREB1, oscillates with circadian rhythm [68]. Similarly, several components of ABA signal transduction pathways were shown to interact with TOC1, a key regulator of circadian clock [142]. This suggests that the *RD29A* responses to ABA treatment is also circadian-gated [170, 224]. Although the exact mechanism of how the circadian clock influences *RD29A* expression is still unknown, these evidences in addition to our observation of stress-independent variability of the control suggest that careful consideration of potential circadian effects on regulation of *RD29A* (and any other stress-responsive genes) is an essential prerequisite for investigation of the relationship between stress inputs and temporal dynamics of induced gene expression.

Another important consideration that must proceed the investigation of stress-induced *RD29A* expression dynamics relates to reliability and reproducibility of the data. Unfortunately, the experiments conducted for this chapter were based on the small sample size ($n = 3$), which

makes the results more prone to experimental errors. For instance, presence of outliers in samples severely increased variability associated with certain data points such as at 10 ZT under single NaCl at full strength (Fig. 5.1b), 8 ZT under single ABA at full strength (Fig. 5.1c), and 10 ZT under combined NaCl and ABA at full strength (Fig. 5.1d). Given that such large error bars are only observed from the samples treated with full-strength stress, one can suspect that the errors may be of a systematic origin, with the overly strong stresses undermining the ability of the subjected plants to adequately induce gene expression. On the other hand, contamination in some samples could also have led to the outliers. Thus, it is essential that further repeats of the experiments with a large sample size are conducted to improve reliability of the data.

Although the issues regarding the experimental design and errors may undermine reliability of the data in a quantitative manner, it is possible to assess whether the qualitative features observed from the data sufficiently reproducible via comparison with the other similar data in literature. Several previous studies have also experimentally measured dynamics of *RD29A* expression in response to single NaCl and ABA treatments, using the *RD29A:LUC* luminscence reporter system [89, 252, 265]. Their results bear close resemblence to our non-normalised measurement of *RD29A* transcript abundance data (Fig. 5.1): for example, delayed induction by NaCl stress compared to ABA treatment is also observed in the data from those studies [252, 265]. The results from our own preliminary investigation conducted before the experiments conducted for this chapter (Section A.1 in Appendix) also suggest that the qualitative features unique to treatment conditions are fairly reproducible. For example, the previous results (Fig. A.1, A.2) also show that the responses to NaCl are delayed at least for one or two hours, while ABA responses seem to occur earlier compared to NaCl. However, it is unclear whether the qualitative features from the results of the similar experiments, both conducted by us and the others, originate from the regulatory system triggered by NaCl or ABA signals or intrinsic circadian oscillation because those studies did not consider the time of day at which the measurements were taken.

Chapter 6

Mathematical modelling of *RD29A* regulatory system

6.1 Introduction

The characteristic qualitative features observed from the time course *RD29A* expression profiles upon various combinations of NaCl and ABA prompt further investigation of their mechanistic origin. To address this, we construct and analyse the mathematical model of *RD29A* regulatory network. Mathematical modelling is a powerful tool for investigating the causal relationship between the understanding of the inner workings of a biological system and its dynamical behaviours [6]. The main advantage of constructing mathematical models include abstraction of complex network of the system into simpler understanding of control mechanism, observe the effects of various perturbations in the system via computational experiments, and generation of new hypotheses that can be validated via further experiment [126].

The first aim of this chapter is to describe development of the model: structures of the signalling pathways transducing NaCl and ABA signals for regulation of *RD29A* expression are understood reasonably well, with detailed information regarding the functions of some of their signalling components including transcription factors and their upstream regulators.

We integrate the existing knowledge of *RD29A* regulatory system into a unified structure, making several simplifying assumptions to address the lack of required information. The resulting model representing *RD29A* regulatory network topology is translated into a network of kinetic processes described by a system of Ordinary Differential Equations (ODEs). Due to absence of experimentally measured kinetic data, the model parameters are globally optimised by fitting the model to the experimental data.

The second aim is to explain the qualitative behaviours observed from the gene expression dynamics in the light of the current knowledge of the *RD29A* regulatory network. Here, we specifically examine whether the topology of the *RD29A* regulatory network reflected in the ODEs can reproduce the qualitative features of the experimental data after suitable parameterisation. While some of the observed features are explained by the properties of the currently known signalling network, the result indicates that the model is insufficient to simultaneously explain all features observed from the experimental data. To bridge the gap between the model outcomes and the experimental data, various modifications in the model structure are introduced, implementing hypothetical interactions between the NaCl and ABA signals that are not evidenced in literature yet. We identify several mechanisms allowing the model to reproduce all features observed from the experimental data, and evaluate feasibility of those mechanisms by a further experiment and literature analysis. The insight resulting from this analysis of the model reveals the possible mechanism of interaction between NaCl and ABA stress signals.

Contribution: development of the methodologies leading to the main results of this chapter, particularly regarding the analytical solutions to the ODE model and parameter estimation, received some contributions from Dr. Neville J. Boon.

6.2 Structure of *RD29A* regulatory system

The current understanding of the *RD29A* regulatory system is summarised in Fig. 6.1. The *RD29A* regulatory system describes induction of *RD29A* expression by NaCl and ABA stress via accumulation of transcriptionally active DREB2 and AREB proteins, which is facilitated by stress-induced post-translational modification and TF gene expression.

Plant responses to high salinity stress occur in two phases [176]: Phase I responses are mainly triggered by osmotic stress signals which typically last several days after introduction of salt stress. Phase II responses are triggered by increased intracellular ion concentration, which arise from failures of Phase I response and occur in the timescale of weeks . Because our experiments in Chapter 5 were conducted in a short timespan (0 to 5 hours of stress treatment), only the signalling mechanisms for Phase I responses are relevant for our model. Several transmembrane sensors in the plasma membrane are responsible for initial detection of osmotic stress and an increase in the extracellular Na^{+} content [30]. These sensors subsequently trigger increase in intracellular Ca^{2+} as the secondary messenger [44], which is picked up by downstream signalling pathways.

Activation of the ABA-dependent pathways including the AREB-mediated pathway occurs in response to increase in internal ABA concentration. NaCl stress, along with water deficit and other stress types leading to osmotic imbalance, increases endogenous ABA concentration via ABA biosynthetic pathways (Reviewed in Section 2.3). On the other hand, exogenous ABA entering from cell exterior can also trigger the same ABA-induced responses: there is at least one designated transmembrane ABA importer, namely ABCG40 [114]. Since its loss-of-function mutants do not display severe phenotypic defects seen from ABA-deficient mutants [190], however, other import mechanisms including passive diffusion across pH gradient also play significant roles in ABA transport across plasma membrane [223].

The intracellular stress signals post-translational modification of their corresponding TF proteins via enzymatic cascades. For instance, AREB requires phosphorylation prior to binding its target motif, ABRE [77, 250]. As described in Chapter 2, the pathway leading

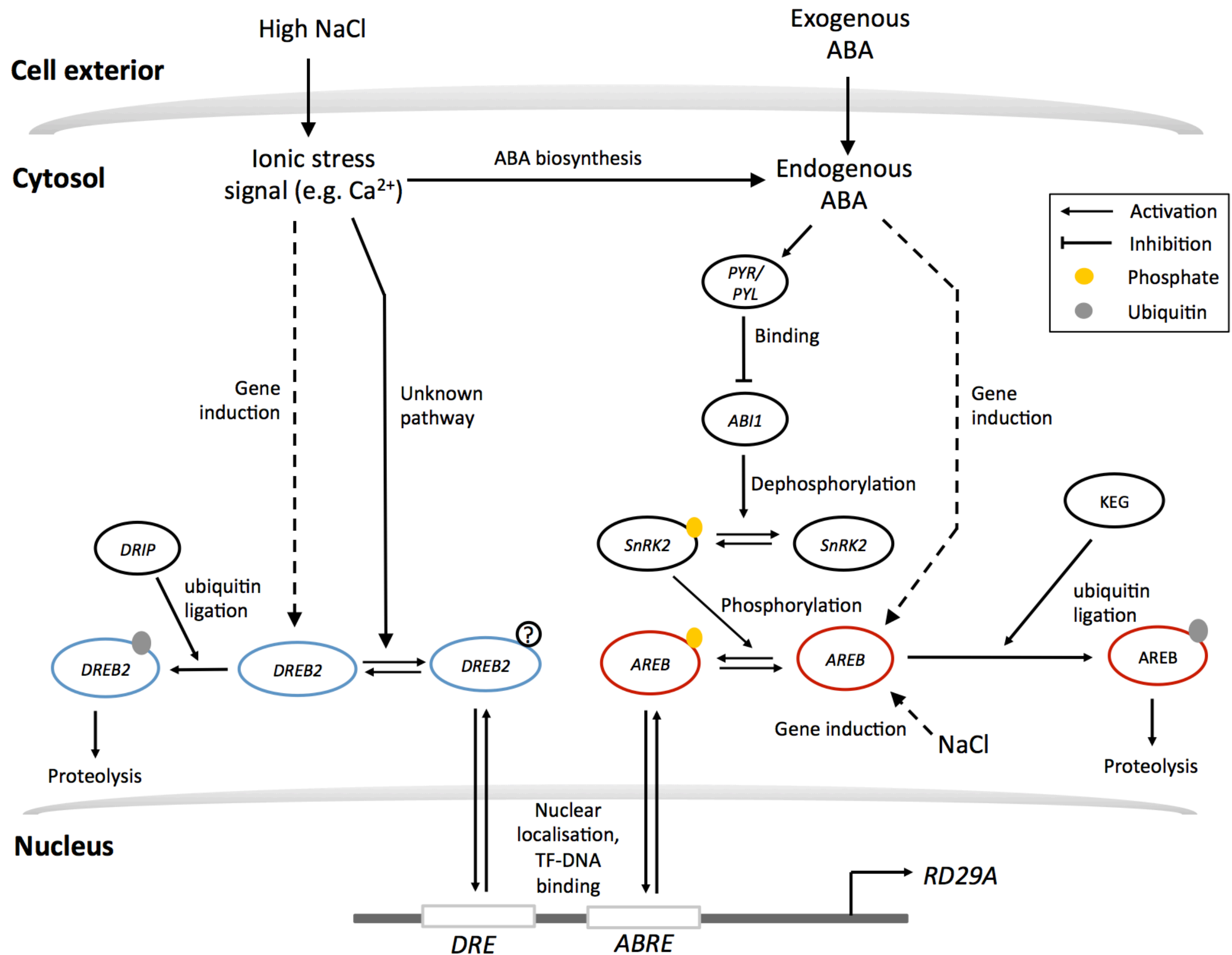


Figure 6.1: **Overview of the *RD29A* regulatory system involving the DREB2 and AREB pathways**. *RD29A* expression can be induced by both high salinity stress and exogenous ABA applied from the cell exterior, which trigger production of secondary intracellular stress signals. For example, high external NaCl concentration causes ionic and osmotic imbalance, which leads to production of *de novo* ABA biosynthesis for activation of the ABA-dependent pathways and other stress signals picked up by the ABA-independent signalling pathways. Exogenous ABA can also be imported to directly activate the ABA-dependent pathways. Notably, there are key similarities in the structures of ABA-independent and ABA-dependent pathways, which govern the dynamics DREB2 and AREB proteins: 1) both TF proteins are activated via post-translational modification (phosphorylation cascade for AREB, unknown mechanism for DREB2), 2) both TF proteins can be produced further by the stress signals via induction of their genes, and 3) both TF proteins are subject to ubiquitin-mediated proteolysis.

to post-translational AREB activation is well understood; upon binding of ABA, PYR/PYL receptor sequestrates the activity of Protein Phosphatase 2 C (PP2C) such as ABI1 [195], which prevents auto-phosphorylation of SNF-related Kinase 2 (SnRK2) in absence of ABA. Accumulation of phosphorylated SnRK2 subsequently leads to AREB phosphorylation [75]. DREB2 is also considered to require post-translational activation prior to binding of DRE motifs, as its transcriptional activity of DREB2 is not proportional to its abundance [172, 214]. The exact mechanism of DREB2 post-translational activation is currently unknown, with no direct experimental studies available yet.

There are also evidences suggesting that NaCl and ABA cross-regulate the DREB2 and the AREB signalling pathway, respectively. NaCl and ABA stresses also induce expression of TF genes. DREB2A and DREB2B genes are induced by NaCl and osmotic stresses but not significantly by ABA [149]. All AREB genes are inducible by presence of exogenous ABA as well as NaCl [75, 250]. Although whether the protein levels of DREB2 and AREB TFs increase upon presence of the stress inputs is yet to be confirmed, it is reasonable to assume TF protein population increases with increased transcript levels of the TF genes. Combined with the stress cues leading to post-translational activation of the TF proteins, such increase in the inactive TF protein population results in stronger induction of *RD29A* expression.

Lastly, TF protein concentration is negatively regulated via ubiquitin-dependent proteolysis. This is mediated by RING domain E3 ubiquitin ligases such as DREB2-Interacting Proteins (DRIP) for DREB2 [205] and KEEP ON GOING (KEG) for AREB [42]. Those E3 ubiquitin ligases are responsible for keeping the level of stress-inducible TFs low in absence of stresses to avoid unwanted expression of stress response genes and waste of cellular resources in maintaining the TF population. Whether activities of these E3 ubiquitin ligases are modulated by stress inputs is still being debated: the DRIP and KEG genes both do not exhibit strong changes upon introdution of stress inputs [122, 129], and there is currently no evidence suggesting that their activities are modulated at protein levels.

6.3 Model development

Based on the structure of the *RD29A* regulatory system, we developed a mathematical model of the DREB2 and AREB pathways. This section explains how the model captures relative increase in *RD29A* transcript abundance in response to NaCl and ABA stress inputs, based on the structure of the upstream regulatory pathway. In particular, several simplifying assumptions have been adopted during model development are stated in each section.

6.3.1 Stress input dynamics

Since the experiments are measuring the molecular changes occuring inside cells, there must first be a description of intracellular stress signal dynamics based on stress inputs introduced to cell exteriors. The experiments in Chapter 5 introduced NaCl and ABA stress input to *Arabidopsis* seedlings by transferring them to hydroponic media of given concentration of stress agent from normal unstressed condition. To reflect this abrupt change in stress environment, the dynamics of intracellular salt stress signal $S_1(t)$ can be described by a simple step increase:

$$S_1(t) = \begin{cases} \frac{[\mathrm{NaCl}]_{\mathrm{ext}}}{[\mathrm{NaCl}]_{\mathrm{max}}} & \text{if } t > 0, \\ 0 & \text{if } t \leq 0. \end{cases} \tag{6.1}$$

Here, $[\mathrm{NaCl}]_{\mathrm{ext}}$ and $[\mathrm{NaCl}]_{\mathrm{max}}$ represent the actual and maximum (300mM) strength of the external NaCl stress input, respectively. The duration of stress treatment, t, is in the unit of hours, where $t = 0$ represents the time at which the treatments commenced. Note that by describing salt stress input dynamics with Eqn. 6.1, we are assuming that

1. Intracellular NaCl stress signal $S_1(t)$ is linearly proportional to $[\mathrm{NaCl}]_{\mathrm{ext}}$.

In other words, the intracellular salt stress signal $S_1(t)$, ranging from 0 to 1, is equivalent to salt input strength scaled by a constant. We set $[\mathrm{NaCl}]_{\mathrm{max}} = 300\mathrm{mM}$ for it typically is

considered as the saturating level of stress (i.e. cells are unable to function normally when exposed to the stress beyond this concentration) [264].

On the other hand, the dynamics of intracellular ABA signal (endogenous ABA) is described as follows:

$$S_2(t) = \begin{cases} \frac{f_{\mathrm{ABA}}(S_1,t)+[\mathrm{ABA}]_{\mathrm{ext}}}{\max f_{\mathrm{ABA}}+[\mathrm{ABA}]_{\mathrm{max}}} & \text{if } t > 0, \\ 0 & \text{if } t \leq 0. \end{cases} \tag{6.2}$$

The intracellular ABA signal $S_2(t)$, also ranging from 0 to 1, is described as above because the amount of endogenous ABA can increase via two routes (Fig. 6.1): ABA is synthesised directly in presence of the salt stress by the function f_{ABA} with a fixed maximum value, or is imported from an exogenous pool located in cell exterior, the size of which is determined by $[\mathrm{ABA}]_{\mathrm{ext}}$. We set $[\mathrm{ABA}]_{\mathrm{max}} = 100\mu\mathrm{M}$ [264]. Note that describing $S_2(t)$ with Eqn. 6.2 required the following assumptions:

2. The intracellular ABA signal $S_2(t)$ is linearly dependent to $[\mathrm{ABA}]_{\mathrm{ext}}$.

As with the case of salt stress input, the second assumption simplifies the model by setting the endogenous ABA concentration equal to $[\mathrm{ABA}]_{\mathrm{ext}}$ scaled by a constant. In fact, Eqn. 6.2 can be simplified even further to

$$S_2(t) = \begin{cases} \frac{[\mathrm{ABA}]_{\mathrm{ext}}}{[\mathrm{ABA}]_{\mathrm{max}}} & \text{if } t > 0, \\ 0 & \text{if } t \leq 0, \end{cases} \tag{6.3}$$

based on assumption that

3. The amount of ABA internally produced from *de novo* production is negligible compared to the amount imported from exterior, such that $\max f_{ABA} \ll [\mathrm{ABA}]_{\mathrm{max}}$.

This third assumption is biological reasonable because there are evidences for such significant difference between the levels of endogenous ABA produced via *de novo* synthesis and im-

ported from cell exterior. For instance, the data from Ren et al. [207] suggests that $S_2 \approx 1.8 \times 10^{-9}$g mol^{-1} fresh weight found when severely dehydrated (another stress similar to NaCl able to induce *de novo* ABA biosynthesis), while the data from Windsor et al. [263] suggests that $S_2 \approx 1.6 \times 10^{-7}$g mol^{-1} fresh weight found intracellularly when $[\text{ABA}]_{\text{ext}} = 100\mu$M. For the subsequent parts of the chapter, we will use Eqn. 6.3 to describe the dynamics of the intracellular ABA signal.

6.3.2 Transcription factor dynamics

The intracellular stress signals, S_1 and S_2, affect the dynamics of DREB2 and AREB proteins, the main transcription factors for *RD29A* regulation. The known structure of the DREB2 and AREB pathways (Fig. 6.1) show that the stress inputs NaCl and ABA control the dynamics of transcription factor proteins via two separate mechanisms: (1) production of the TF proteins in their inactive form (TF_i) via induction of the TF genes, and (2) conversion of the inactive TF proteins to the post-translationally activated form (TF_i^*). The dynamics of $TF_1(t)$, $TF_1^*(t)$, $TF_2(t)$ and $TF_2^*(t)$ is described by a set of differential equations

$$\dot{TF_i} = r_i + r_i^\tau S_i(t-\tau) + C_i + d_{-i} TF_i^*(t) - [d_{ib} + \alpha_i S_i(t) + u_i + \delta_i] TF_i(t), \quad (6.4)$$

$$\dot{TF_i^*} = [d_{ib} + \alpha_i S_i(t)] TF_i(t) - (d_{-i} + \delta_i) TF_i^*(t), \quad (6.5)$$

for $i = 1, 2$, where the parameters represent the rates of biochemical processes such as production (r_i, r_i^τ), degradation (u_i, δ_i) and post-translational modification of TF proteins (α_i, d_{ib}, d_{-i}). Detailed biological meanings of the parameters, with description of whether they can be fixed from literature information, is shown in Table 6.1 . Note that not all of the known intermediate signalling components shown in Fig. 6.1, such as SnRK2 and DRIP, are described as intrinsic state variables. Instead, the the model consolidates the effect of intermediate signalling components into kinetic processes that directly relate TF proteins against stress inputs. To model the overall relationship between the strength of stress inputs

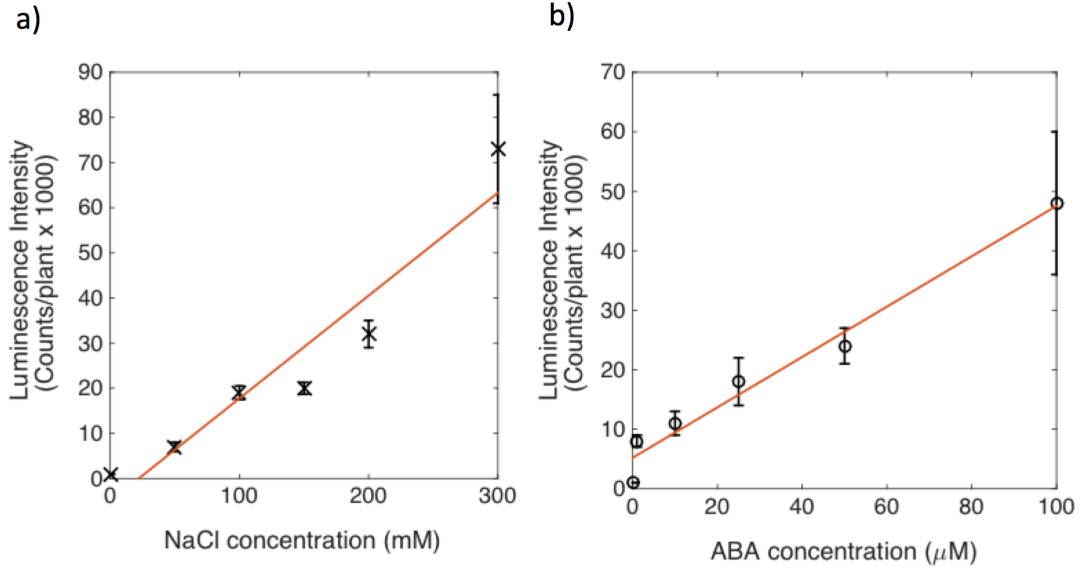


Figure 6.2: ***RD29A* expression level upon varying NaCl and ABA concentrations (data taken from Xiong et al., 1999 [264])**. Points indicate the data from the original authors, and lines represent linear regression fit the the data points. The dose-response characteristics for a) NaCl-induced and b) ABA-induced *RD29A* expression profiles are well approximated by linear models.

and the amount of transcriptionally active TF_i^*, we have adopted a simplest description, where

4. the effects of each of S_i on both TF_i activation (α_i) and production (r_c^τ) are linear.

We argue that the fourth assumption is biological reasonable based on the existing evidence for approximate linear dependence of *RD29A* expression on varying concentrations of NaCl and ABA (Fig. 6.2).

The parameter τ represents the time delay for the input S_i to affect accumulation of inactive TF_i via induction of its genes. Time delay is implemented here because unlike post-translational activation of TF_i that is triggered by enzymatic cascade, the pathways leading to *de novo* production of functional TF proteins are likely to involve many hidden intermediate steps including production of mRNA transcript, post-transcriptional processing, polypeptide synthesis and post-translational processing for the TF genes.

Table 6.1: **Description of model parameters.** All parameters, except τ that is measured in hours, have the unit of hours^{-1}. See Table A.1 in A.3 in Appendix for the values.

Pathway	**Name**	**Biological process**	**Method of Determination**
TF_1	r_1	Basal TF_1 production rate	Fixed [149, 214]
	δ_1	Natural decay rate	Fixed [202]
	d_{1b}	Basal TF_1 activation rate	Fixed [214]
	r_1^τ	S_1-induced TF_1 production rate	Parameter optimisation
	α_1	S_1-induced TF_1 activation rate	Parameter optimisation
	d_{-1}	Basal TF_1^* deactivation rate	Parameter optimisation
	u_1	TF_1 ubiquitination rate	Parameter optimisation
TF_2	δ_2	Natural decay rate	Fixed [202]
	d_{2b}	Basal TF_1 activation rate	Parameter optimisation
	r_2	Basal TF_1 production rate	Fixed by analytical derivation (Eqn. A.6, page 134)
	r_2^τ	S_2-induced TF_2 production rate	Parameter optimisation
	α_2	S_2-induced TF_2 activation rate	Parameter optimisation
	d_{-2}	Basal TF_2^* deactivation rate	Parameter optimisation
	u_2	TF_2 ubiquitination rate	Parameter optimisation
	r_c^τ	S_1-induced TF_2 production rate	Parameter optimisation
Both	τ	Time delay before TF_i production	Fixed (our experimental data)

The term C_i in Eqn. 6.4 represents the effect of adjacent stress on production of TF_i. We assume $C_1 = 0$ because the DREB2 expression is independent to ABA [149]. We assume $C_2 = r_c^\tau S_1(t-\tau)$ with r_c^τ representing the rate of TF_2 production induced by S_1, since AREB expression is known to be triggered by NaCl [75, 250].

6.3.3 mRNA synthesis

Lastly, we describe production of *RD29A* mRNA transcript, the final outcome of the *RD29A* regulatory system, from the actions of active DREB2 (TF_1^*) and AREB (TF_2^*). Although transcriptional regulation in eukaryotes (including plants) is achieved through complex series of molecular processes such as recruitment of RNA polymerase and chromatin restructing [135], it is common to quantitatively describe the phenomenon via thermodynamic models just based on fractional occupancy of DNA binding sites by TFs [17].

A thermodynamical model of transcription assumes that the rate of mRNA production is proportional to the number of sites on DNA occupied by their corresponding TF proteins at equilibrium. There are two TFs binding to separate sites in the case of the *RD29A* promoter, (Fig. 6.1). First, we assume the followings:

5. There is no non-specific binding, such that DREB2 cannot bind to ABRE and vice versa. This means that occupancy of each cis-elements is function of only the corresponding TF such that $\theta_{\text{DRE}} = f_{\text{DRE}}([\text{DREB2}])$ and $\theta_{\text{ABRE}} = f_{\text{ABRE}}([\text{AREB}])$, with binding f_{DRE} and f_{ABRE} representing the binding equations.
6. DREB2 and AREB do not interact cooperatively between each other, both before and after binding onto the cis-regulatory elements. Thus $k_c[\text{DREB2}][\text{AREB}] = 0$, where k_c is the association constant for cooperative DREB2-AREB binding.

Based on these assumptions, we can specify f_{DRE} and f_{ABRE} such that

$$\theta_{\mathrm{DRE}} = \frac{[\mathrm{DRE_{bound}}]}{[\mathrm{DRE_{total}}]} = \frac{K_{A,\mathrm{DREB2^*}}[\mathrm{DREB2^*}]}{1 + K_{A,\mathrm{DREB2^*}}[\mathrm{DREB2^*}]}, \tag{6.6}$$

$$\theta_{\mathrm{ABRE}} = \frac{[\mathrm{ABRE_{bound}}]}{[\mathrm{ABRE_{total}}]} = \frac{K_{A,\mathrm{AREB^*}}[\mathrm{AREB^*}]}{1 + K_{A,\mathrm{AREB^*}}[\mathrm{AREB^*}]}, \tag{6.7}$$

where $[\mathrm{DREB2^*}](t)$ and $[\mathrm{AREB^*}](t)$ represent the free concentration of transcriptionally active (i.e. post-translationally modified) DREB2 and AREB transcription factors and $K_{A,\mathrm{DREB2^*}}$ and $K_{A,\mathrm{AREB^*}}$ denote equilibrium association constants for binding of those TF onto the targeted sites. More information about derivation of Eqn. 6.6 and 6.7 can be found in [24]).

From Eqns. 6.6 and 6.7, we subsequently define the probability of *RD29A* promoter being active as a whole (P_{active}) with

$$P_{\mathrm{active}} = \frac{n_{\mathrm{DRE}}}{n_{\mathrm{DRE}} + n_{\mathrm{ABRE}}}\theta_{\mathrm{DRE}} + \frac{n_{\mathrm{ABRE}}}{n_{\mathrm{DRE}} + n_{\mathrm{DRE}}}\theta_{\mathrm{ABRE}}, \tag{6.8}$$

where n_{DRE} and n_{ABRE} represent the copy numbers of DRE and ABRE motifs present in the *RD29A* promoter, respectively. The copy number of each motif determines how much contribution each motif makes towards total mRNA production. For example, if the copy number of both motifs is one, each motif is responsible for half of the overall *RD29A* promoter activity.

Eqn. 6.8 shows that P_{active} has nonlinear relationship with $[\mathrm{DREB2^*}]$ and $[\mathrm{AREB^*}]$. For further simplification, the following assumption was employed:

7. The amounts of TFs are non-saturating, such that $[\mathrm{DREB2^*}]$ and $[\mathrm{AREB^*}] \ll 1$.

At these non-saturating concentrations of TF proteins, we can then assume linear relationship between the promoter activity and the TF amounts because $\theta_{\mathrm{DRE}} \approx [\mathrm{DREB2^*}]$ and $\theta_{\mathrm{ABRE}} \approx$

[AREB*]. We subsequently define

$$m = [\mathrm{DREB2}^*] + k[\mathrm{AREB}^*], \tag{6.9}$$

where $k = n_{\mathrm{ABRE}}/n_{\mathrm{DRE}}$. Note that m is a unitless quantity which changes with the amount of active DREB2 and AREB proteins present, thus providing an arbitrary representation of promoter activity. Such linear assumption appears to be a reasonable simplification, as *RD29A* expression outputs linearly increase with increasing concentrations of NaCl and ABA stress inputs ([264], Fig. 6.2).

The last simplification we make for the purpose of modelling the *RD29A* regulatory system involves how temporal dynamics of the transcriptional system relates to the dynamics of TF proteins, previously described in Section 6.3.2:

8. Molecular processes leading to production of mRNA, such as RNAP recruitment and polymerisation of nucleotides, occur at much faster rates compared to those that govern the promoter activity, such that $\dot{[\mathrm{mRNA}]} = \epsilon \dot{m}$, with $\epsilon \to 0$.

This assumption essentially introduces timescale separation between the mRNA production dynamics and TF regulation dynamics, such that the dynamics of mRNA transcript is solely governed by the those processes leading to accumulation of active TF proteins and increase in the promoter activity.

The experiments in Chapter 5 measured relative increase in the amount of mRNA transcript across differing durations of stress treatments (t) with respect to the basal level under unstressed condition ($t = 0$). Thus, the final model output $M(t)$ is also defined as the relative increase of transcript abundance induced by stresses with respect to its basal level in absence of stress:

$$M(t) = \frac{m(t)}{m(0)} = \frac{[\mathrm{DREB2}^*](t) + k[\mathrm{AREB}^*](t)}{[\mathrm{DREB2}^*](0) + k[\mathrm{AREB}^*](0)} = TF_1^*(t) + TF_2^*(t), \tag{6.10}$$

where $TF_1^*(t)$ and $TF_2^*(t)$ are the dimensionless state variables representing quantities of active DREB2 and AREB. The variables $TF_1^*(t)$ and $TF_2^*(t)$ represent the dynamics of all DREB2 and AREB protein isoforms, and are the primary factors that govern the dynamics of the system as a whole.

6.3.4 Summary

A graphical representation of the mathematical model to be analysed in the subsequent sections is shown in Figure 6.3.

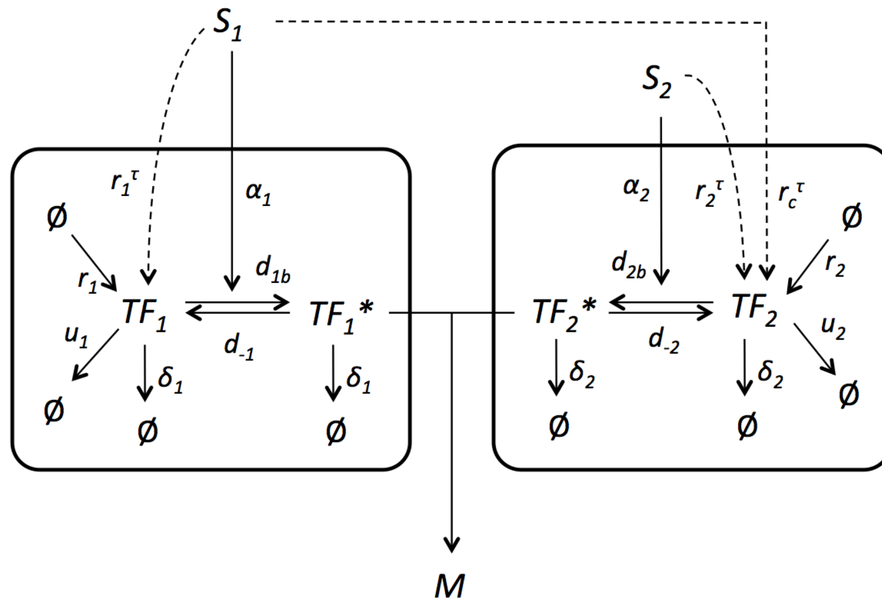


Figure 6.3: **The proposed mathematical model of *RD29A* regulatory system**, representing the structure of the DREB2 (TF_1) and AREB (TF_1) pathways. Asterisks (*) denote post-translationally modified form of the protein. Solid arrows represent kinetic processes leading to change in the state variable, while dashed arrows the kinetic processes associated with time delay. Arrows originating from null sets (∅) denote de novo production of proteins, while arrows pointing towards null sets represent degradation of proteins. Model parameters for kinetic rates of the signalling processes are shown next to the corresponding arrows.

6.4 Results

6.4.1 Comparison of experimental data and model outcomes

Temporal profiles for RD29A expression calculated from the model were obtained from 6.10. The analytical solutions to Eqns. 6.4, 6.5 can be found in section A.2 in Appendix. As was shown already in Table 6.1, we fixed 6 parameters based on either the existing evidence in the literature, or the assumptions derived from our experimental data. The remaining parameters were estimated by fitting the model to the experimental data using Monte Carlo Simulated Annealing (MCSA) algorithm. Further details about the methods of parameter estimation as well as the nominal parameter set used for subsequent results are described in section A.3 in Appendix. The results (Fig. 6.4) showed that the model could reproduce the features observed from the responses to single NaCl and ABA treatments well, but not the synergistic effect observed from the responses to combined NaCl and ABA treatments. Here, we subsequently analyse how the model reproduces each of the three features identified in the previous Chapter (section 5.4).

Feature 1: Induction of *RD29A* expression by all combinations of NaCl and ABA occurs in two phases, with weak initial increase in the level of expression followed by strong increase after 2 hours of treatment.

The model reproduces the two-phase induction of *RD29A* expression under all treatment conditions by implementing two types of signalling process with one representing the fast signalling processes via enzymatic cascade, and the other representing the delayed signalling processes via induction of genes. Given that the experiments for all treatment conditions shows abrupt changes in the behaviours of the system somewhere between 2 to 3 hours after treatment, we assumed that the length of delay associated with the latter processes (τ) is equal to 2.5 hours. Thus, the increase in *RD29A* expression level after 2 hours observed from all treatment conditions can be attributed to increased amounts of TF_i, facilitated by

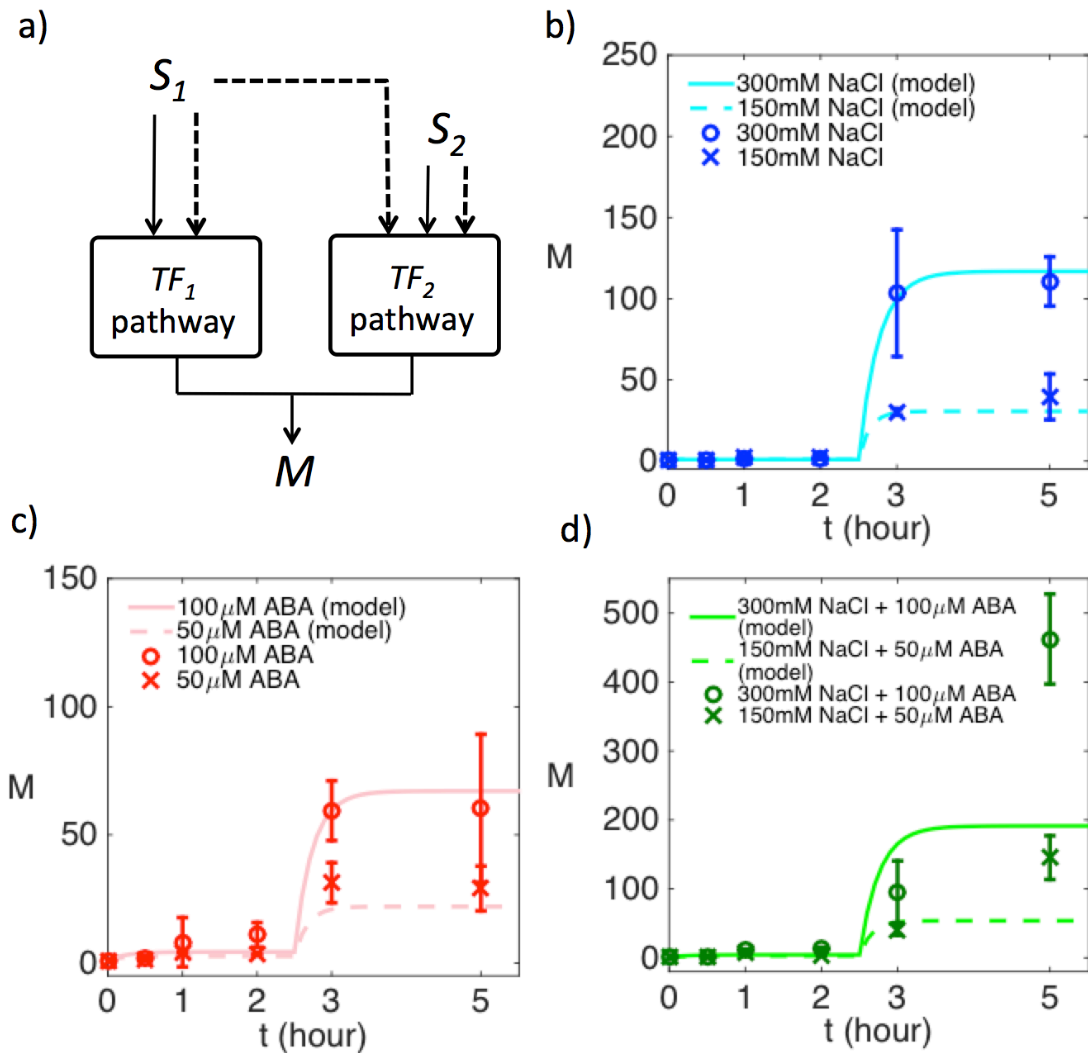


Figure 6.4: **Comparison of the experimental data and the outcome of the *RD29A* regulatory system model**. a) A simplified overview of the system structure describing the regulatory cues from each stimulus. Solid and dashed arrows represent fast and delayed regulatory cues, respectively. The model outcomes (solid and dashed lines) calculated based on the identified parameter set (Table A.1 in Appendix) are compared with the corresponding experimental data, obtained under b) single NaCl treatments, c) single ABA treatments, and d) combined NaCl + ABA treatments.

induction of the corresponding TF genes.

Other parameters also play important role in shaping the two-phase characteristic of *RD29A* expression. Weak expression during the early phase of the response ($0h \leq t \leq 2h$) compared to the late phase ($t > 2h$) indicate that there is limited amount of TF proteins available immediately after treatments are introduced. In the model, scarcity of TF proteins arises from negligible or weak basal expression of TF genes (r_i), combined with strong constitutive degradation of TF proteins maintained by ubiquitin-mediated proteolysis (u_i). While the enzymatic cascades are immediately switched on after introduction of stress, small amount of inactive TF proteins restricts the amount of post-translationally active TF proteins being produced, thereby limiting the level *RD29A* expression inducible by the input signals.

Feature 2: Strength of the input stimuli affects the magnitude of expression, but does not change the qualitative features described above.

The model reproduces Feature 2 by assuming that $S_i(t)$ exerts linear control over the signalling processes considered in the model (See assumptions 4, 7). Although the experimental measurements at only two levels of input strength (full- and half-strength) are insufficient to determine whether the effects of NaCl or ABA on the signalling processes in the *RD29A* regulatory system are actually linear, assuming this simplifies the model without compromising its ability to reproduce magnitude of fold increases with respect to strength of the input stimuli.

Feature 3: *RD29A* expression remains at a constant level after 3 hours of treatment with single NaCl and ABA, while it continues to increase under combined NaCl and ABA treatment.

The model reproduces the constant level of expression after 3 hours of single NaCl and ABA treatment in terms of the steady-state response of the system. However, the continued accumulation of *RD29A* transcript from 3 hours of combined stress response, is not reproduced

by the model. The model can only describe the dynamics of combinatorially induced *RD29A* expression profile as the sum of the dynamics of singly induced profiles (Fig. 6.4d, lines) because there are no nonlinear interactions between the two stresses, which can act as the potential source of synergistic effect. Given that model formulation based on the literature fails to capture the greater than additive interactions between the NaCl and ABA stimuli, the model must be modified in order to capture the greater than additive expression upon combination of two stresses.

6.4.2 Cross-input modulation of intracellular signalling processes

The comparison between the experimental data and the model outcomes showed that the model based on the current understanding of the *RD29A* regulatory system is unable to capture the synergistic effect under combined NaCl and ABA treatments. This indicates the model has failed to implement a unknown mechanism mediating interaction between the NaCl and ABA signals. In this section, we systematically explore possible mechanisms that can act as the potential source for the observed synergy between NaCl and ABA in regulating *RD29A* expression.

We hypothesised that the synergistic effect observed from the combined NaCl and ABA stress responses originates from cross-input modulation, where S_1 and S_2 cross-modulate TF_2 and TF_1, respectively. More specifically, we define a cross-input modulation as a regulatory cue produced by the non-cognate stimulus (S'), leading to enhancement (E) or inhibition (I) of the kinetic rates associated with the targeted signalling process (Fig. 6.5b). The effect of cross-input modulation is implemented in the model by replacing the targeted parameter, p_j in the model equations (6.4 and 6.5) with

$$E(p_j) = p_j(1 + c_j^E S'(t-\tau)), \quad (6.11)$$

$$I(p_j) = \frac{p_j}{(1 + c_j^I S'(t-\tau))}, \quad (6.12)$$

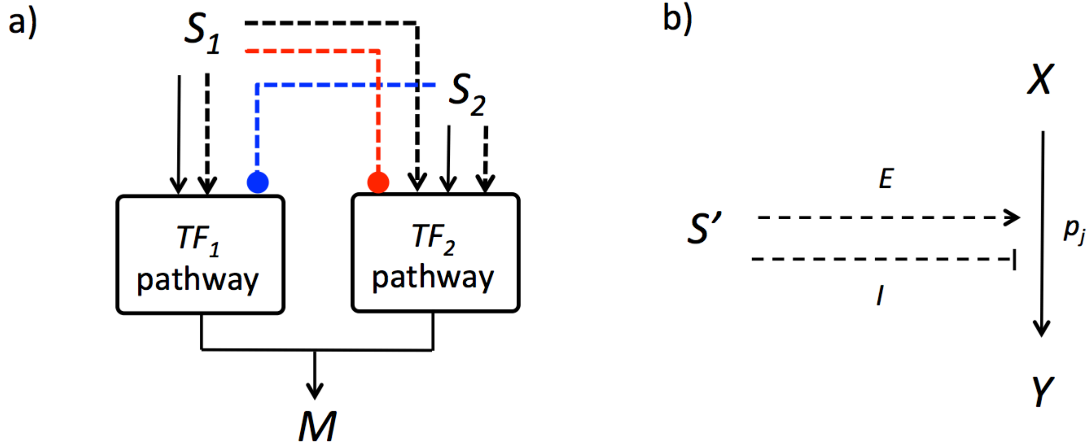


Figure 6.5: **Definition of cross-input modulation**. a) Synergistic interaction between the two input signals can be achieved by modulation of the signalling processes in the adjacent pathway by the non-cognate input signals. The blunt arrows coloured red and blue represent the regulatory cues of unknown effect emerging from S_1 and S_2, respectively. Dashed lines indicate that the regulatory cues involve induction of gene. b) A cross-input modulation occurs where a non-cognate input signal (S') targets a signalling process of the adjacent pathway (p_j). Regulatory outcome of the crosstalk interaction can be either by enhancement (E) or inhibition (I) of p_j.

where the variable S' denotes the non-cognate input, which can be either S_1 or S_2 depending on which pathway p_j belongs to. Note that S' is delayed by τ, because the synergistic effect in the experimental data appears most pronounced during the late phase of stress response. This is equivalent to assuming that cross-input modulation affects the expression of the genes responsible for the target signalling process. The parameters c_j^E and c_j^I represent the strength of enhancement and inhibition of p_j, respectively. The condition $c_j^E = c_j^I = 0$ corresponds to the case where there is no cross-input modulation, which is equivalent to the model presented in Figure 6.4.

Amongst the 16 parameters in the model, there are 9 signalling processes in the *RD29A* regulatory system ($r_2^\tau, \alpha_2, d_{-2}, u_2, r_1^\tau, \alpha_1, d_{-1}, u_1, r_c^\tau$) that can be targeted by cross-input modulation. This is because the other 7 processes ($r_1, r_2, d_{1b}, d_{2b}, \delta_1$, and δ_2) are stress-independent by definition and cannot form cross-input modulation (revisit Table 6.1 for further explanation of their stress-independent nature).

Implementing enhancement or attenuation crosstalk interactions for each of these 9 processes led to 18 modified system structures (Fig. 6.6a). The parameters for each system structure were identified *ab initio*, using the routine described in Appendix (Section A.3). We found 5 system structures that qualitatively reproduce the third qualitative feature (synergistic effect), as well as the other two features (Fig. 6.6a, red and blue). The ability of each system structure in reproducing the observed synergistic effect was assessed based on how well the model fits the combined stress response data after parameter optimisation.

The 5 system structures, $E(\alpha_1), I(d_{-1}), I(u_1), I(d_{-2})$ and $I(u_2)$, qualitatively reproduce all three features observed from the experimental data (Fig. 6.6b, d, f). These system structures show a common topological feature, where the non-cognate input signals enhances the post-translational production of TF_i^*. In system structures $I(d_{-1})$ and $I(d_{-2})$, non-cognate input signals enhances accumulation of TF_i^* by attenuating the rate of its post-translational deactivation. In $I(u_1)$ and $I(u_2)$, the non-cognate input signals indirectly increases TF_i^* by attenuating degradation of TF_i, which results in increased net forward conversion rate into TF_i^*. $E(\alpha_1)$ is also a valid crosstalk interaction because it enhances the post-translational processing of TF_1 into TF_1^*. Thus, the results suggest that enhanced rate of TF_i^* accumulation by non-cognate input signals may account for the synergistic effect observed from the experimental data.

The remaining 13 system structures (Fig. 6.6a, grey) cannot reproduce the synergistic effect qualitatively, no matter how the parameters of the original *RD29A* regulatory system models and cross-input modulation are chosen (Fig. 6.6c, e, g). These fail to reproduce the synergistic effect because they do not lead to selective enhancement of either pathway. For example, cross-input modulation in some system structures such as $E(d_{-1})$ or $I(\alpha_2)$ decreases the rate of TF_i^* accumulation, leading to attenuation of the targeted pathway instead of enhancement. Modulating production of TF proteins via gene induction $(r_1^\tau, r_2^\tau, r_c^\tau)$ also does not lead to effective enhancement of the selected pathway because it increases the population of TF_i rather than increasing its stability, only influencing the magnitude of *RD29A* expression at steady state.

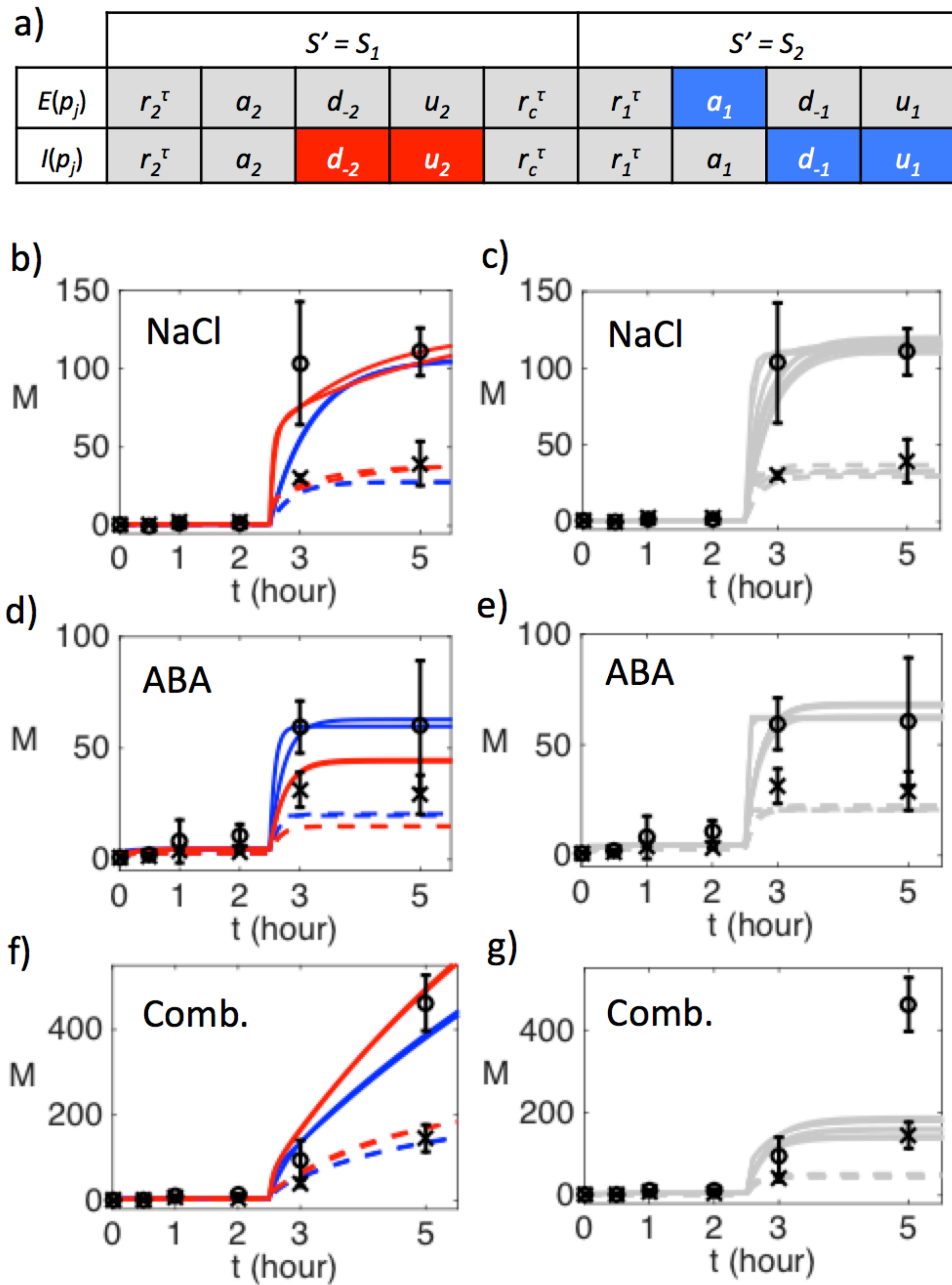


Figure 6.6: **Comparison of the system structures implementing different mechanisms of cross-input modulation**. a) An outline of all 18 possible system structures, organised by regulatory outcome (E or I) and the non-cognate stress input (S_1 or S_2). The 5 system structures that reproduce the observed synergistic effect are highlighted in colour (red = S_1 modulates TF_2 pathway, blue = S_2 modulates TF_1 pathway). b)-g) Comparison of the experimental data (black) with the best-fitting model outcomes obtained from the 5 selected system structures (2 red and 3 blue lines) and the other structures (13 grey lines), superimposed onto single plots according to treatment conditions.

6.4.3 Evidence of individual crosstalk interactions in microarray datasets

To further investigate the feasibility of the crosstalk interactions predicted from 5 identified system structures $I(d_{-1}), I(u_1), E(\alpha_1), I(u_2)$ and $I(d_{-2})$, we subsequently examined expression profiles of the genes that are potentially responsible for mediating the five cross-input-modulated processes $(d_{-1}, u_1, \alpha_1, u_2, d_{-2})$ from two publicly available transcriptome-wide cDNA microarray datasets produced by Kreps et al. [129] and Kilian et al. [122] (Table 6.2).

Suppression of DREB2 degradation in $I(u_1)$ is not supported by the expression profiles observed from both datasets because expression of DRIP, an E3 ubiquitin ligase responsible for targeted proteolysis of DREB2, appears independent to various abiotic stresses including ABA [122]. The expression profile of KEG obtained from one dataset [122] shows independence to NaCl stress, which contradicts attenuation of the AREB pathway claimed by $I(u_2)$. Both datasets contradict $I(d_{-2})$ by showing expression of AHG3, a gene encoding ABI-clade Phosphatase [154], is up-regulated in presence of NaCl stress. Note that this observation does not prove that crosstalk interaction of opposite regulatory outcome i.e. $E(d_{-2})$ exist, because ABA is known to strongly inhibit the protein activity of AHG3 [8].

Consequently, two system structures, $E(\alpha_1)$, (ABA enhances DREB2 post-translational activation α_1) and $I(d_{-1})$ (ABA attenuates post-translational deactivation of active DREB2 d_{-1}), remain as viable system structures. The information regarding those system structures could not be extracted from the microarray datasets because the identities of the genes responsible for DREB2 post-translational modification are yet unknown. This result suggests that ABA-induced enhancement of post-translational activation of DREB2 via $E(\alpha_1)$ or $I(d_{-1})$ is responsible for the observed synergistic effect.

Table 6.2: Comparison of the system structures with cDNA microarray datasets

Viable system structures reproducing the synergistic effect			**Evidence of crosstalk interactions within the existing experimental dataset**			
Type	Name	Proposed mechanism	Candidate gene (locus)	Molecular function	Expression profiles from cDNA microarray datasets	
					Kreps *et al.*, 2002	Killian *et al.*, 2007
Enhancement of DREB2 outputs by ABA	$I(u_1)$	ABA attenuates DREB2 ubiquitination (u_1)	*DRIP1* (At1g06770) *DRIP2* (At2g30580)	E3 ubiquitin ligase (Qin et al., 2008)	Data not available	Expression independent to abiotic stress (NaCl, drought, osmotic stresses)
	$E(\alpha_1)$	ABA enhances DREB2 post-translational activation (α_1)	Unknown	N/A	N/A	N/A
	$I(d_{-1})$	ABA attenuates post-translational deactivation of active DREB2 (d_{-1})	Unknown			
Enhancement of AREB outputs by NaCl	$I(u_2)$	NaCl attenuates AREB ubiquitination (u_2)	*KEG* (At5g13530)	E3 ubiquitin ligase (Chen et al., 2013)	Data not available	Expression independent to abiotic stress (NaCl, drought, osmotic stresses)
	$I(d_{-2})$	NaCl attenuates phospho-AREB dephosphorylation (d_{-2})	*AHG3* (At3g11410)	Protein phosphatase 2C (Lynch et al., 2012)	Expression up-regulated by NaCl	Expression up-regulated by NaCl

6.4.4 Reduction of synergistic effect upon reduced input strength

In parallel to the finding from the analysis of cDNA microarray datasets, we further validate the ability of each identified system structures in predicting the qualitative synergistic effects in combined stress conditions. Because the identified system structures implements cross-input modulation in different directions ($S_1 \rightarrow TF_2$ pathway or $S_2 \rightarrow TF_1$ pathway), the model predicts that halving the dose of either one in combined stress input would result in asymmetric reduction in the synergistic effect. Reduction in the strength of input that triggers the cross-input modulation results in greater reduction of the synergistic effect: for instance, the cross-input modulation for structures $E(\alpha_1), I(d_{-1})$ and $I(u_1)$ originate from S_1 signal, and reduction in value of S_1 leads to reduced synergistic effect. Subsequently, we conducted additional measurements of *RD29A* expression profile using the identical experimental settings (Chapter 5), except with combinations of NaCl and ABA with different concentration (300mM NaCl + 50μM ABA, 150mM NaCl + 100μM ABA). The new data is shown in Fig. 6.7.

Comparison between the prediction made by two selected structures, $I(d_{-1})$ and $I(u_2)$, with the newly obtained experimental data is shown in Figure 6.8. Note that the model outcomes shown in Fig. 6.8a and 6.8b are obtained using the same parameter set used to fit each structure to the previous experimental data (Fig. 6.6). The model outcomes for both system structures do not provide quantitative description of the new experimental data, but still qualitatively captures the difference in the degree of reduction in the synergistic effect, which can be approximated by gradient of increase in expression between 3 and 5 hour time frame.

Inability of the model predictions to quantitatively reproduce the new sets of experimental data means that it is not possible to compare those two with conventional means, such as quantification of model-data differences via the residual sum of squares. Since we only consider whether the model can reproduce the third qualitative features observed from the experimental data, we introduce our own metric to better analyse the difference between the

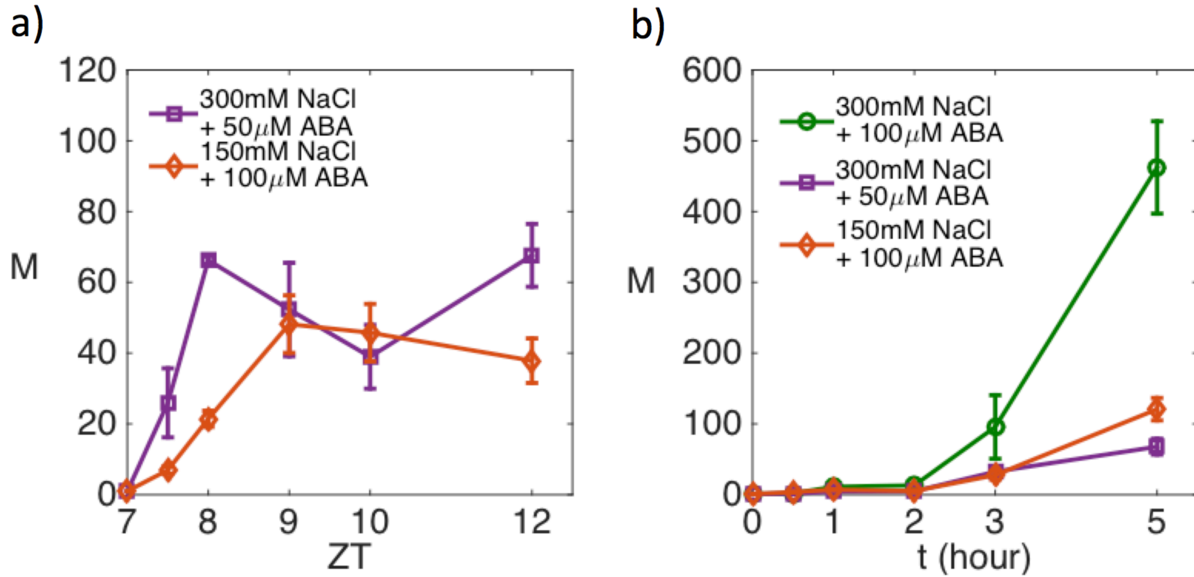


Figure 6.7: **Effect of combining NaCl and ABA at different strengths**. a) Measurement of *RD29A* transcript abundance induced by full-strength NaCl and half-strength ABA treatment and vice versa. The effect of circadian oscillation has not been removed. b) Comparison of the *RD29A* expression profiles induced by the combination of full-strength NaCl and ABA treatment from the previous experiment against the profiles with the strength of one input halved. Effect of circadian oscillation has been removed by normalising with the H_2O profile obtained from the previous experiment. Note that decrease in either NaCl or ABA concentration leads to significant reduction in the total level of expression.

two treatment conditions. We calculated a ratio, R, which compares the degree of reduction in gradient of *RD29A* expression fold increase at 3 to 5 hours post-stress triggered by halving of either NaCl or ABA stress (Fig. 6.8c). If halving of NaCl leads to greater reduction in the synergistic effect compared to halving of ABA, R will be greater than 1. If halving of ABA leads to greater reduction in the synergistic effect, R will be less than 1. Calculating R for all five system structures indicate that the predictions from $E(\alpha_1)$, $I(d_{-1})$ and $I(u_1)$ match that of the experimental data, while the predictions from $I(d_{-2})$ and $I(u_2)$ do not (Fig. 6.8d). This suggests that enhanced production of post-translationally active DREB2 by ABA is responsible for the synergistic effect unique to the profiles from combined NaCl and ABA treatments.

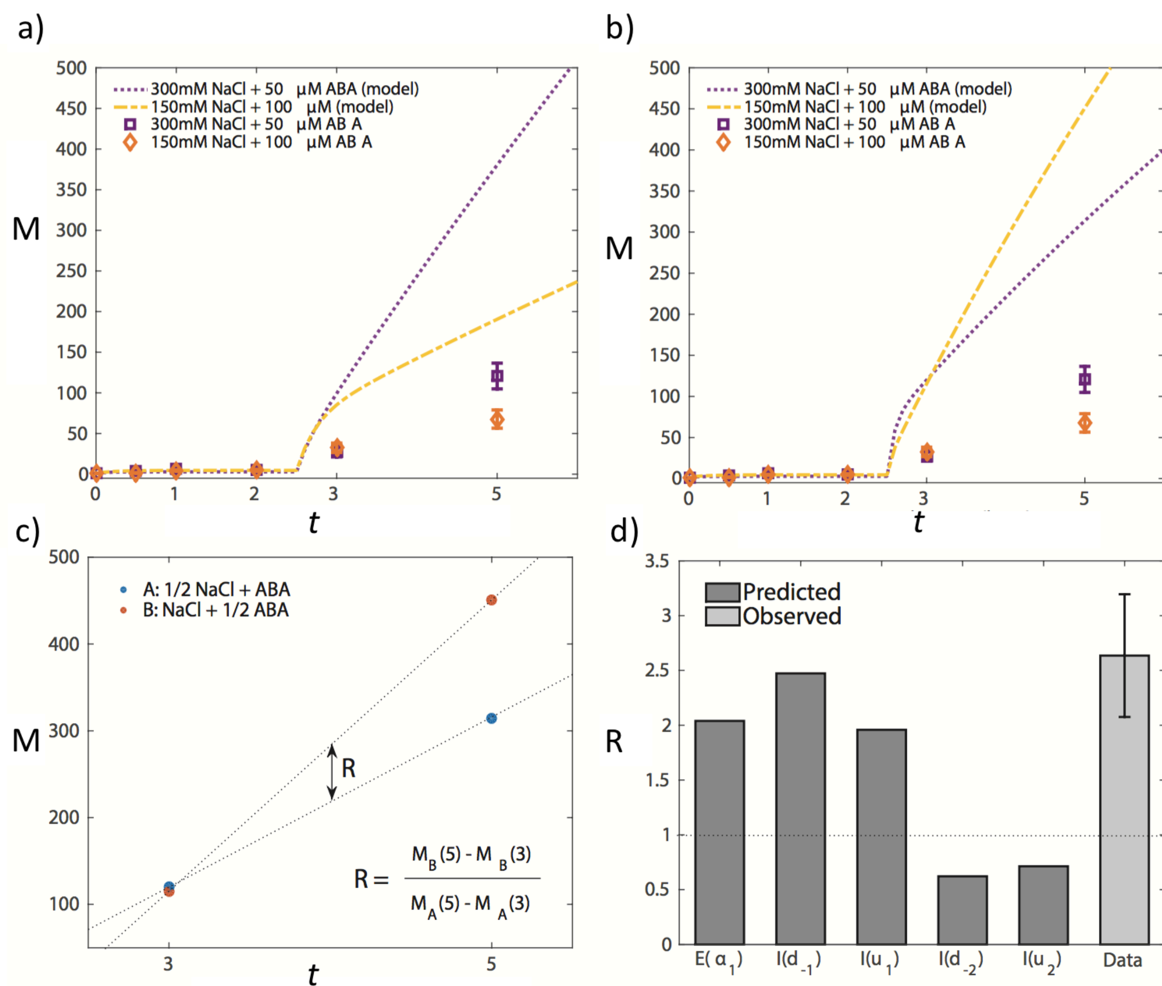


Figure 6.8: **Reduction of synergistic effect upon halving the dose of one stress in a combined stress input.** a) Comparison between the predicted model outcome from the structure $I(d_{-1})$ in response to reduction of either NaCl or ABA concentration, with the corresponding experimental data. b) Similar comparison made for structure $I(u_2)$. c) Quantification of the synergistic effect arising from combined stress by the measuring the slope of increase in *RD29A* transcript abundance, M, between 3 and 5 hours of treatment. The effect of halving either stress is compared by calculating the ratio, R. d) Comparison of the ratio R calculated from experimental data with the predicted ratios from each of the five identified system structures.

6.5 Discussion

In this chapter, we investigated the signalling mechanisms that regulate the temporal dynamics of *RD29A* expression in response to NaCl, ABA, and their combinations. We explained *RD29A* expression dynamics using a simple mathematical model developed from the existing knowledge of the DREB2 and AREB pathway structures. The current model could reproduce the features observed in the dynamics of *RD29A* expression under individual NaCl stress and ABA, but was insufficient to reproduce the synergistic effect observed under the combinations of NaCl stress and ABA. Via systematic investigation of plausible system structures, we identified five mechanisms of cross-input modulation that may be responsible for the observed synergistic effect. Analysis of the existing microarray datasets combined with a further hypothesis-driven experiment identified cross-input modulation mediating ABA-dependent DREB2 post-translational activation as the potential sources of the observed synergistic effect.

Our mathematical model aimed to extract the core regulatory structures of the DREB2 and the AREB pathways in order to understand the complex mechanisms that regulate expression of stress response genes. While addressing this aim, several assumptions were adopted to obtain a simplified description of numerous biological processes (Section 6.3). To ensure the biological conclusion derived from analysis of this simplified model is valid, it is necessary to examine how each of these assumptions affects the results obtained from the model. Table 6.3 provides the recap of the 8 assumptions taken during model development.

Assumptions 1, 2 and 4 are responsible for an overall linear dependence of the amount of *RD29A* transcript on external NaCl stress and ABA exhibited by the model. Adopting these assumptions may be a strong simplification of the real biological processes, as the processes affected by these assumptions are likely to be nonlinear. For example, the magnitude intracellular salt stress signal (e.g. Ca^{2+}) may linearly increase with the concentration of external NaCl only up to a certain point, beyond which there is no further increase. Such saturating behaviours can also applied to other signalling processes such as activation and

Table 6.3: **List of assumptions used in the model of the *RD29A* regulatory network.**

	Assumption	Model Interpretation	Eqn.
1	Intracellular NaCl stress signal is linearly proportional to external NaCl concentration.	$S_1 \propto [\text{NaCl}]_{\text{ext}}$.	6.1
2	Endogenous ABA concentration is linearly proportional to exogenous ABA concentration.	$S_2 \propto [\text{ABA}]_{\text{ext}}$.	6.2
3	The amount of ABA internally produced from *de novo* production is negligible compared to the amount imported from exterior.	$S_2 = f_{\text{ABA}} + [\text{ABA}]_{\text{ext}}$, $\max f_{\text{ABA}} \ll [\text{ABA}]_{\text{max}}$.	6.3
4	The effects of intracellular stress signals on TF activation and production are linear.	$\dot{TF_i} \propto S_i$,	6.4
		$\dot{TF_i^*} \propto S_i$.	6.5
5	There is no non-specific binding, such that DREB2 cannot bind to ABRE and vice versa.	$\theta_{\text{DRE}} = f_{\text{DRE}}([\text{DREB2}])$	6.6
		$\theta_{\text{ABRE}} = f_{\text{ABRE}}([\text{AREB}])$	6.7
6	DREB2 and AREB do not interact cooperatively between each other.	$k_a[\text{DREB2}][\text{AREB}] = 0$.	6.8
7	The amounts of TFs are non-saturating.	$[\text{DREB2}] \ll 1$, $[\text{AREB}] \ll 1$.	6.9
8	Molecular processes leading to production of mRNA occur at much faster rates compared to those that govern the promoter activity.	$[\dot{\text{mRNA}}] = \epsilon \dot{m}$, $\epsilon \to 0$.	6.10

production of TF proteins. Since the previous data on *RD29A* response to NaCl and ABA doses shows linearly relationship within the range of stress inputs considered in our experiments (Fig. 6.2), however, assuming linearity appears to be an adequate simplification of the processes involved for the purpose of analysing the data presented in Chapter 5. Implementing more complex nonlinear relationships to describe the effect of external stress on the signalling processes may enhance explanatory power of the model over broader range of the stress input strengths, but at the expense of significantly increasing the number of model parameters.

Another strong simplification of a known biological process relates to exclusion of NaCl-

induced ABA biosynthesis (Assumption 3). We have chosen to ignore *de novo* ABA biosynthesis from NaCl stress inputs. Adopting this assumption changes the way the results are interpreted, such that the temporal dynamics of *RD29A* expression induced by singly applied NaCl stress is solely controlled by the DREB2 pathway. Such interpretation is incorrect, as it is well established that singly applied NaCl stress can also activate the AREB pathway and the stress responsive genes controlled by ABRE [149, 250]. However, we argue that expanding the model to include additional processes involved in ABA biosynthesis does not alter the conclusion regarding the difference between singly-induced and combinatorially-induced response (Feature 3, regarding the synergistic effect unique to the combined NaCl and ABA treatments). Because the amount of endogenous ABA is likely to be much smaller than the amount of imported ABA in combined stress settings [207, 263], it is reasonable to assume that activation of the AREB pathway under these treatment conditions is mainly conducted by ABA imported from outside. Based on this interpretation, the reason why the synergistic effect is not observed under singly applied NaCl stresses even though ABA biosynthesis occurs is because the amount of ABA produced *de novo* is too small to trigger additional ABA-dependent post-translational activation of DREB2, the main finding of this chapter. Even if the amount of endogenous ABA produced *de novo* is comparable to the amount imported externally such that the response triggered by single NaCl stress is a product of both DREB2 and AREB pathways, our experimental data showing that the synergistic effect is specific to combined NaCl and ABA treatments suggests it is still the addition of externally supplemented ABA that cause this feature.

The last four assumptions (5-8) simplify the processes regarding transcriptional regulation. One amongst those assumptions represents a particularly strong simplification: the assumption that there is no interaction between DREB2 and AREB is likely to be incorrect, as an *in vitro* study comparing *RD29A* promoter activities upon recombinant DREB2 and AREB showed that activation by combined DREB2 and AREB results in greater-than-additive transcript production [183]. Again, we argue that relaxing this assumption does not affect the final conclusion of this chapter, as the effects of cooperativity between DREB2 and AREB

proteins will only occur in the magnitude of *RD29A* induction, not its temporal dynamics. Specifically, greater-than-additive transcript production from simultaneous presence of active DREB2 and AREB will change the dose response to combined NaCl and ABA into a nonlinear curve, which does not concern how *RD29A* transcript level changes over time. One important factor that can influence the temporal response to stress inputs is validity of the last assumption (8), which leads us to interpret the results that the observed temporal dynamics of *RD29A* expression arise solely from the TF dynamics, governed by the signalling network structure. If this assumption is wrong, it is possible that the dynamics of gene expression may change with the mRNA production and degradation rates, which are intrinsic properties of the transcriptional regulatory system. Their effects on the dynamics of *RD29A* expression, however, will be indifferent between all stress conditions, suggesting that the dynamics of transcriptional regulation is not responsible for the unique differences in dynamical features of *RD29A* expression profiles induced by single and combined NaCl and ABA.

Having established that the simplified model is able to provide an adequate description of the biological processes involved in *RD29A* regulation, it is interesting that such simplified model has a certain degree of predictive power over the qualitative behaviour of the system upon changes in combined stress inputs (Fig. 6.8). The apparent inter-dependence of the NaCl and ABA signals in modulating both DREB2 and AREB signalling pathways then raises a further question: how does the abiotic stress system distinguish and selectively express ABRE- or DRE-controlled genes in response to single NaCl or ABA if the DREB2 and the AREB pathways are cross-modulated by both NaCl and ABA? Our model shows that such stress-specific response emerges from the activation mechanism of DREB2 and AREB, which resembles a logical AND operator. Because activation of DREB2 and AREB requires simultaneous increase in post-translational modification and TF population through their gene expression, activation of either one mechanism is insufficient to induce their transcriptional activity. Given that the suggested cross-input modulation affects only the post-translational modulation, the abiotic stress response system can avoid any unwanted outcomes of crosstalk

interactions and produce outputs specific to DREB2 and AREB pathways under single stress conditions.

Cross-regulation of DREB2 and AREB signalling pathways by NaCl and ABA has wider implication for understanding abiotic stress response as a whole. Although both DRE and ABRE regulons contain stress response genes, ABRE regulons also control biological processes other than stress response such as long-term developmental response [179]. Thus, we propose that selective enhancement of DRE regulon upon combined stress conditions ensures prioritisation of the DRE-driven gene regulation to the stress without committing to long-term effects induced by ABA. More detailed ontology analysis of genes constituting the DRE and ABRE regulons, coupled with measurements of their expression profiles under combined NaCl and ABA stress settings, will help in understanding the physiological role of the crosstalk interaction.

Considering 61% of *Arabidopsis* stress-induced genes display various types of compound effect upon exposure to multiple abiotic and/or biotic stresses [206], studies of crosstalk interactions between stress signalling pathways attracts increasing attention. To this end, there have been studies on the crosstalk mechanisms in combinations of various stresses, such as drought with heat [212], drought with high light [58, 84], heat with high light [34], and abiotic stresses with biotic stress [13, 282]. Our approach provides a platform to integrate the existing knowledge of individual stress signalling pathways generated from the conventional genetic and biochemical framework, and systematically highlight the gap in understanding through comparison with experimentally observed gene expression profiles under different stress combinations.

Chapter 7

Concluding Remark

Compound gene regulation refers to the control of gene expression outcomes by combinations of multiple stresses, and is a fundamental aspect of plant adaptation to hostile environments. From the observations made in the previous studies as well as ours, it has been shown that regulation of gene expression upon a combination of stresses does not take place as a sum of responses to individual types of stress. While significant knowledge has accumulated regarding plant stress response gene regulation, simple assembly of the intracellular signalling pathways individually characterised for single types of stress is insufficient to explain the non-additive outcomes observed under combined stress conditions. This suggests that the current understanding of molecular stress response system is unable to address signal integration mechanisms, which ultimately allows plants to search for optimal responses to given environments. In this thesis, we have investigated how the experimental observations of the gene behaviours under single and combined stress conditions can be used to inform about the structure of plant stress signalling network.

In Chapter 1, we argued that a framework for systematic classification of possible effects of multiple stress combinations is essential. The first part of the thesis (Chapter 3, 4) addresses this challenge by proposing compound logic as a possible framework for classifying different behaviours induced by stress combinations. If changes in gene expression can be reduced to discrete variables based on steady-state assumption, compound logic can be used to limit

the outcomes of integrating multiple stress signals into a finite number of gene regulatory profiles based on qualitative comparison of expression changes under single and combined stress. Systematic investigation of the mechanisms for each profile was made possible by logical network models of intracellular signalling networks inspired from Parallel Distributive Processing (PDP) networks, with propagation of stress signals described as logical operation. The logical network models could describe the non-additive interaction between multiple stress inputs in regulating gene expression outputs based on integration of the stress signals occuring at different stages of signal transduction. The compound logic framework was then applied to *Arabidopsis* transcriptomic datasets obtained under single and combined stress conditions [206]. Analysing gene expression profiles from the transcriptomic dataset revealed various modes of interaction between pairs of stresses, which indicating greater diversity in gene regulatory outcomes beyond conventional positive and negative relationship between two stresses. Based on the assignment of logical network model to particular regulatory profile from Chapter 3, cluster of genes identified from analysis of the transcriptomic dataset was used as the potential source of information regarding the signal integration mechanisms. This was demonstrated with the specific examples (HSF and WRKY pathways for heat and biotic signal transduction, respectively).

Interaction between multiple stresses also lead to complex gene regulatory outcomes with quantitative changes that cannot be described as discrete variables. The second part of the thesis (Chapter 5, 6) investigated this in more detail, we experimentally measured the temporal expression profiles of *RD29A* induced by various combinations of NaCl and ABA stimuli. The results presented in Chapter 5 showed that the synergistic effect of combined NaCl and ABA in activating *RD29A* expression is time-dependent. This suggests that the outcomes of integrating the two stress signals cannot simply be described with logical variables, which prevents the use of the compound logic method developed in the early part of the thesis to investigate the regulatory mechanisms for the unique dynamical behaviours of *RD29A*. A mathematical model of the *RD29A* regulatory network consisting of the DREB2 and AREB signalling pathways was constructed to further interrogate the experimental data, searching

for the potential signal integration mechanism responsible for the observed synergistic effect in the form of missing edges in the given regulatory network. By systematically implementing additional edges in the model, we identified the system structures that allow the model to reproduce the synergistic effect observed from the experimental data. We also showed that the list of candidate structures can be reduced even more by checking the biological plausibility of each interaction from other sources in literature, and the experiments designed based on validating model predictions from different candidate structures. We ultimately identify two possible edges in the *RD29A* regulatory network, which offer clear direction for further experimental investigation.

With an increasing number of studies examining the effect of combined stresses in plant gene regulation [201], the needs for mechanistic models to explain the gene expression profiles observed from the resulting datasets are also becoming greater [156]. The current understanding of the plant stress signalling network is mostly based on the information obtained from single stress experiments, providing limited explanations for compound gene regulation. In addition, the current knowledge of the plant stress signalling network is largely incomplete, lacking many interactions with important functions in regulation of gene expression under combined stress inputs. By using theoretical models of signalling network, however, we showed that it is possible to make systematic inquiries on where the functionally important interactions may be located. This was achieved by first constructing a model of signalling network based on incomplete information, and subsequently comparing different the model structures with regards to their ability to describe gene regulatory outcomes from both single and combined stress inputs.

The key factor that enables such comparison is the level of abstraction at which the model is constructed. Figure 7.1 illustrates our approach of using simplified models to extract information about system structure from the data. The logical network models in Chapter 4 consolidate the intracellular signalling network into three layers, which are minimalistic representations of the network with limited descriptive capabilities. The *RD29A* regulatory network model in Chapter 6 have been developed in a similar manner by omitting some of

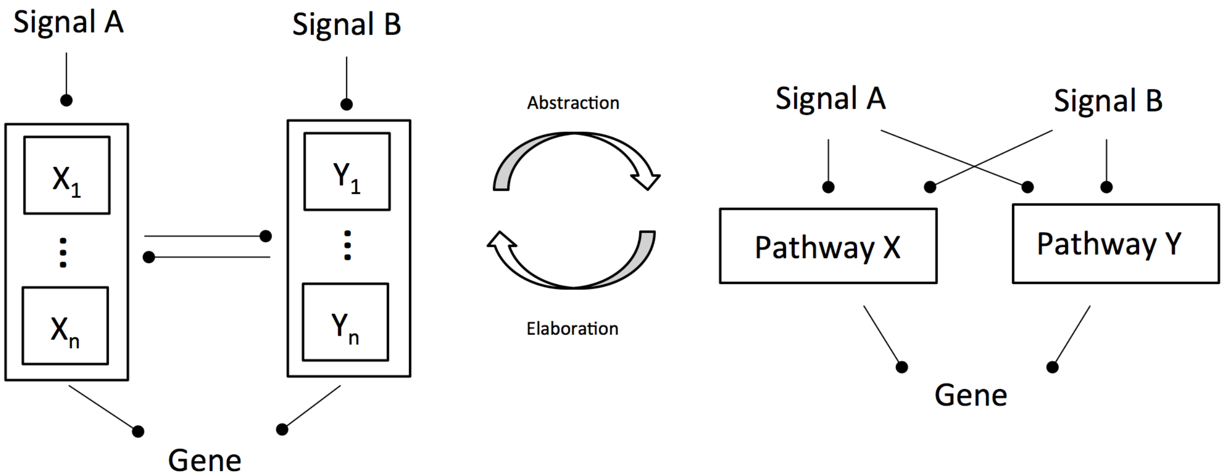


Figure 7.1: **A conceptual framework for elucidation of signal integration mechanisms from analysis of signalling network models at various levels of abstraction**. Two alternative representations of a signal-integrating network are shown, with blunt arrows depicting direction of regulation that can be either positive or negative. A detailed model of signal-integrating network can be abstracted into a simpler model, which allows identification of functionally important interaction via systematic investigation of model structure. The identified interaction validated from additional experiments can be used to elaborate the existing model of signalling network.

the known details (such as the phosphorylation cascade leading up to AREB activation), provided that such simplifications do not significantly alter the biological nature of those processes. Our reason for adopting such principle of minimalism is twofold. Because the purposes of the models are to investigate interactions between different signalling pathways with varying degree of knowledge, standardisation in the level of abstraction was necessary. The models with high levels of abstraction also provide several advantages: for example, the PDP models may act as the first starting point for mining unknown signalling pathways, as it can reveal the variety in behaviours under combined stresses amongst the genes controlled by a common type of transcriptional regulators. Simplified description of the *RD29A* regulatory network has reduced the number of plausible interactions to be tested, reducing computational burden.

It is noted that translating the insights gained from the simpler models to guide further experimental investigation requires more careful consideration. From the model of *RD29A*

regulatory system constructed in greater detail, we were able to make a clear suggestion for the new experiments that will validate the interactions required for the model to explain the observed synergistic effect. The PDP model of signalling network, however, only offers coarse picture of the network architecture. The proposed network facilitating integration of heat and biotic stress signals (Fig. 4.9) does not suggest much beyond that there may be negative crosstalk interaction between HSF and WRKY signalling pathways to achieve attenuation of the genes that are normally expressed under single stress condition upon combined stress inputs. Considering that there are numerous isoforms of HSF and WRKY proteins [59, 185], shortlisting the pathway for further experimental analysis is not possible with the information currently at hand. Further functional characterisation of the signalling pathways with more details about which signalling components are involved will shed more light on the mechanisms responsible for compound regulation of gene expression.

There are several suggestions for immediate future development on the work presented in this thesis:

- Alternative classification methods for the outcomes of compound gene regulation

  The compound logic framework rely on description of gene expression changes in binary [239] or ternary form (Chapter 3, 4). The limited number of possible classified outcomes may not adequately cover the variety of outcomes produced by interaction between multiple stress inputs. A possible study that may address this shortcoming will involve time-series microarray measurements of transcriptomic expression profiles upon combined stress inputs, and enriching common qualitative patterns amongst the expression profiles via clustering methods.

- Integration of compound gene regulation data into existing transcriptional network inference algorithms

  A further work is needed for developing methods to implement compound gene regulation data with the existing algorithms for transcriptional network inference. Numerous tools for inferring transcriptional network from time-series microarray data are already available

[20]. Further studies to scale up comparative analysis of transcriptomic profiles measured under single and combined stress conditions and to exploit the resulting information in inferring gene interaction may significantly enhance the speed and quality of the predicted transcriptional network.

- Further experiments to validate the proposed structure of *RD29A* regulatory network

  The identified interactions in the model may be tested further, time course measurement of *RD29A* expression in mutants constitutively expression DREB2 proteins such as 35S:DREB2A [214], based on *abi* (ABA-insensitive) as background is proposed. If DREB2 post-translational activity is enhanced in presence of ABA-dependent signalling mechanism, then deleting the activity of the ABA-dependent signalling mechanism is supposed to remove the synergy in combined NaCl + ABA treatment. The reason for using 35S:DREB2A transgene is to compensate for lower DREB2 expression in *abi* lines [124]. Another possible confirmatory experiment is to obtain time-course measurements using the genes that only contain DRE motif in their promoters. If DREB2 activity is enhanced by ABA, then the genes containing only DRE will exhibit synergy. Conversely, the genes only containing ABRE motif (e.g. *RD29B*) will not show synergy.

Beyond these specific suggestions, exciting oppurtunities for further studies are present in the elucidating signal integration mechanisms for regulation of various molecular processes in responses to various stress combinations [238]. Given that environmental stresses are the primary reason for reduction in crop yield worldwide, development of new crop species that are resilient to hostile environmental conditions is crucial in securing future food supply [254]. Understanding how plants perceive and respond to combination of stresses is the major step towards uncovering the mechanisms engineering tolerance in complex environments. The interdisciplinary approach proposed in this thesis may bring new perspectives to the future studies on combined stress response in plants by helping them to systematically identify the missing links in the current understanding of plant stress signalling network.

Appendix A

Supplementary Data and Information for Chapter 5 & 6

A.1 Additional *RD29A* expression dynamics dataset (preliminary experiments)

This section provides the results for stress-induced *RD29A* expression dynamics from the old experiments conducted prior to commencement of this PhD (2010, 2011). The data from these experiments are shown here to strengthen the arguments about the qualitative features of *RD29A* expression dynamics presented in Chapter 5. For example, the old datasets also show delayed response under NaCl in comparison to immediate increase in expression under ABA, and non-additive nature of the response to combined NaCl and ABA. The methods used for the old experiments are identical to what was presented in Chapter 5, unless stated otherwise below. They also add extra information regarding how the system behaves in a longer timespan. However, these datasets were not used for the analysis in Chapter 6. The 2010 dataset (Fig. A1) lacks measurement of *RD29A* expression dynamics in unstressed condition. Measurements for each stress condition in the both 2010 and 2011 (Fig. A1 and A2) dataset were also taken at different time of day, which prevented accurate measurement of time-variability of control.

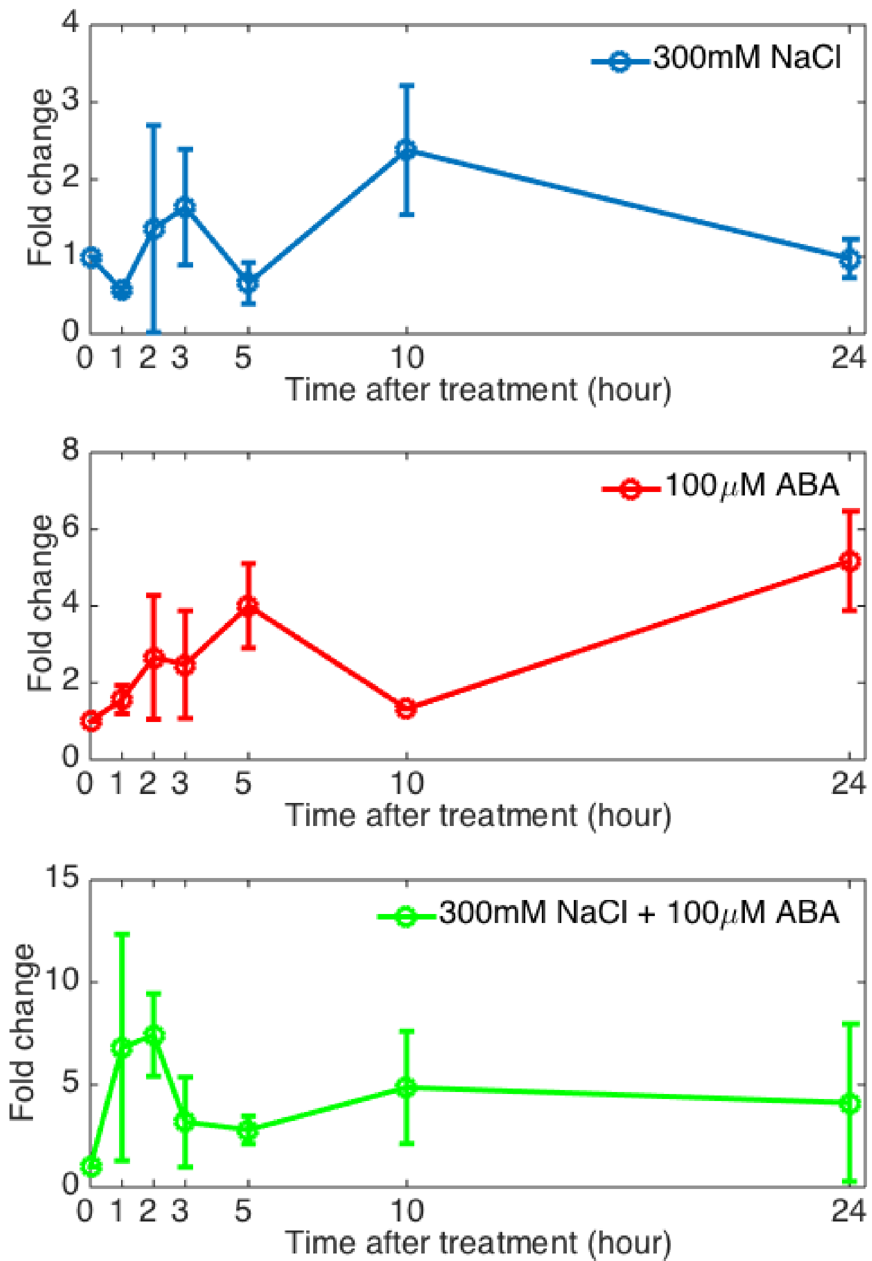


Figure A.1: **Experimental measurements of *RD29A* expression level at different time and treatment conditions (2010)**. Same methods described in Section 5.2 were used, except for the qPCR control gene where *actin-1* was used instead. All data points represent mean fold change in *RD29A* transcript level from triplicate measurements with error bars representing ± SD.

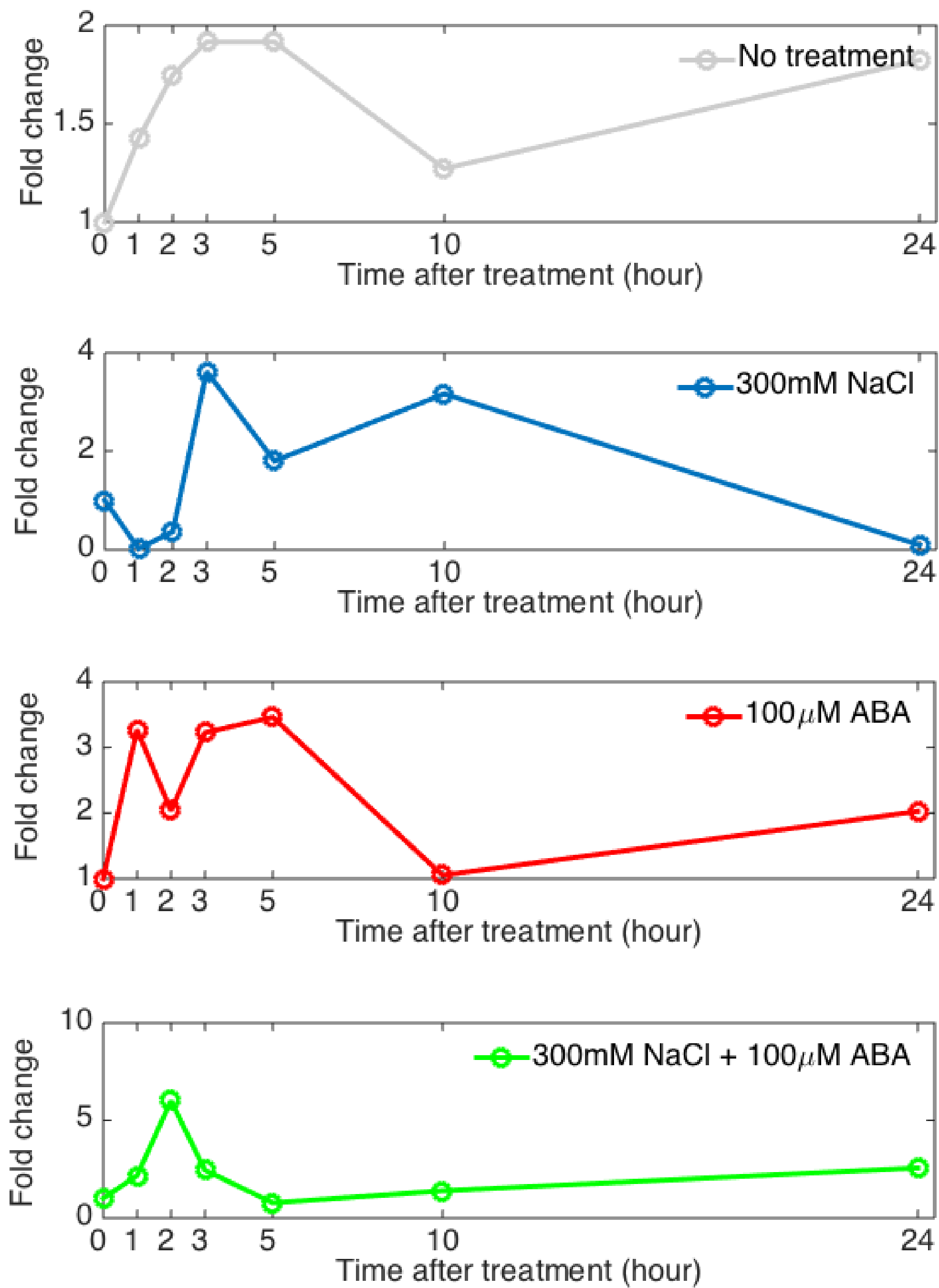


Figure A.2: **Experimental measurements of *RD29A* expression level at different time and treatment conditions (2011).** Same methods described in Section 5.2 were used, except for the qPCR control gene where *actin-1* was used instead. All data points represent fold change in *RD29A* transcript level from a single replicate.

A.2 Analytical solutions for the model of the *RD29A* regulatory system

In this section, analytical solutions for the model of the *RD29A* regulatory system presented in Chapter 6 (Eqns. 6.4 and 6.5) is described. The solutions for two time domains ($t \leq 0$ and $t > 0$) are presented separately.

A.2.1 Before treatment ($t \leq 0$)

Before stress input ($S_1 = S_2 = 0, t \leq 0$), the system is assumed to be at steady state ($\dot{TF_1} = \dot{TF_2} = \dot{TF_1^*} = \dot{TF_2^*} = 0$). Some parameters are fixed based upon the literature (as discussed in the main text) and the steady-state solutions before exposure to stress ($t \leq 0$) can be determined as follows:

$$TF_1 = 0, \tag{A.1}$$
$$TF_2 = \frac{\delta_2 + d_{-2}}{d_{2b}} \tag{A.2}$$
$$TF_1^* = 0, \tag{A.3}$$
$$TF_2^* = 1. \tag{A.4}$$

It is possible to reduce the dimension of parameter space based on the steady-state assumption. With Eqn. A.4, the solution to the ODE describing $\dot{TF_2^*}$ at ($t \leq 0$) becomes

$$\frac{d_{2b} r_2}{\delta_2^2 + \delta_2(d_{-2} + d_{2b}) + u_2 d_{-2}} = 1. \tag{A.5}$$

Rearranging above enables determination of a parameter in terms of other parameters, thus fixing the value of the chosen parameter. Here, we replace r_2 with R_2, which is a function

described in terms of δ_2, d_{2b}, d_{-2} and u_2:

$$R_2 = \frac{1}{d_{2b}}[\delta_2^2 + \delta_2(d_{-2} + d_{2b}) + u_2 d_{-2}]. \tag{A.6}$$

A.2.2 After treatment ($t > 0$)

After exposure to stress ($t > 0$) we solve our model equations by first rewriting Eqns. 6.4 and 6.5 (page 100) in the form

$$\begin{aligned} \dot{TF_i} &= \Theta_i - (\Delta_i + A_i)TF_i + d_{-i}TF_i^*, & \text{(A.7)} \\ \dot{TF_i^*} &= A_i TF_i - (\delta_i + d_{-i})TF_i^*, & \text{(A.8)} \end{aligned}$$

with $i = 1, 2$ representing the DREB2 and AREB pathway, respectively, and

$$\begin{aligned} \Theta_1 &= r_1^\tau S_1(t - \tau) + C_1, & \text{(A.9)} \\ \Theta_2 &= R_2 + r_2^\tau S_2(t - \tau) + C_2, & \text{(A.10)} \\ \Delta_1 &= u_1 + \delta_1, & \text{(A.11)} \\ \Delta_2 &= u_2 + \delta_2, & \text{(A.12)} \\ A_1 &= \alpha_1 S_1(t), & \text{(A.13)} \\ A_2 &= d_{2b} + \alpha_2 S_2(t). & \text{(A.14)} \end{aligned}$$

The functions shown above represent the gross rates of particular biological processes, such as production (Θ_i), degradation (Δ_i) and post-translational activation(A_i). Because Θ_i and A_i contain the time-varying input parameters $S_i(t)$ (Eqns. 6.1 and 6.2, page 98), the domain $t > 0$ can be separated further into two time windows:

$$T(t) = \begin{cases} t & \text{if } 0 < t \leq \tau, \\ t - \tau & \text{if } t > \tau. \end{cases} \tag{A.15}$$

The solutions for Eqn. A.7 and A.8 can then be written as:

$$TF_i(t) = K_{2,i}(t)e^{\lambda_{i+}T(t)} + K_{1,i}(t)e^{\lambda_{i-}T(t)} + \frac{\Theta_i(\delta_i + d_{-i})}{(A_i + \Delta_i)\delta_i + d_{-i}\Delta_i}, \tag{A.16}$$

$$TF_i^*(t) = \frac{1}{d_{-i}}\Bigg[K_{2,i}(t)(\lambda_{i+} + \Delta_i + A_i)e^{\lambda_{i-}T(t)} + K_{1,i}(t)(\lambda_{i-} + \Delta_i + A_i)e^{\lambda_{i-}T(t)} -\Theta_i + \frac{(A_i + \Delta_i)\Theta_i(\delta_i + d_{-i})}{(A_i + \Delta_i)\delta_i + d_{-i}\Delta_i}\Bigg], \tag{A.17}$$

where the eigenvalues of the system are

$$\lambda_{i\pm} = \frac{1}{2}\Bigg[- (A_i + \Delta_i + d_{-i} + \delta_i) \pm\sqrt{(A_i + \Delta_i + d_{-i} + \delta_i)^2 - 4(A_i\delta_i + d_{-i}\Delta_i + \delta_i\Delta_i)}\Bigg] \tag{A.18}$$

and

$$K_{1,i}(t) = \frac{1}{\lambda_{i-} - \lambda_{i+}}\Bigg(d_{-i}TF_{i,ss}^* + \Theta_i - (\lambda_{i+} + \Delta_i + A_i)TF_{i,ss} +\lambda_{i+}\frac{\Theta_i(\delta_i + d_{-i})}{(A_i + \Delta_i)\delta_i + d_{-i}\Delta_i}\Bigg), \tag{A.19}$$

$$K_{2,i}(t) = TF_{i,ss} - K_{1,i} - \frac{\Theta_i(\delta_i + d_{-i})}{(A_i + \Delta_i)\delta_i + d_{-i}\Delta_i}, \tag{A.20}$$

with

$$TF_{i,ss}^* = \begin{cases} TF_i^*(0) & \text{if } 0 < t \leq \tau, \\ TF_i^*(\tau) & \text{if } t > \tau, \end{cases} \tag{A.21}$$

$$TF_{i,ss} = \begin{cases} TF_i(0) & \text{if } 0 < t \leq \tau, \\ TF_i(\tau) & \text{if } t > \tau. \end{cases} \tag{A.22}$$

Cross-input modulation is implemented in the solution by replacing the chosen parameter p_j by $E(p_j)$ or $I(p_j)$ (Eqn. 6.11, 6.12), which subsequently changes the definition of either $\Theta_1, \Theta_2, \Delta_1, \Delta_2, A_1$, or A_2 (Eqn. A.9-A.14).

A.3 Parameter identification

This section explains the method used to fit the model presented in Chapter 6 to the data shown in Chapter 5.

A.3.1 Monte Carlo Simulated Annealing

Monte Carlo Simulated Annealing (MCSA) is an algorithm used to approximate the global optimum in a given parametric function $f(p)$ in a large search space. $f(p)$ is the objective function to be minimised, quantifying error between model simulation and experimental data. MCSA is an adaptation of the Metropolis-Hastings algorithm in a sense that a parameter set p' sampled from a given distribution is accepted under a certain probability, defined by acceptance function α,

$$\alpha(p, p', T) = \begin{cases} 1 & : f(p') < f(p) \\ \frac{1}{1+\exp\left(\frac{f(p')-f(p)}{\max(T)}\right)} & : f(p') \geq f(p) \end{cases} \tag{A.23}$$

In other words, a new parameter set p' is always accepted if the cost function is reduced by the transition $p \rightarrow p'$. Even when the cost function of the p' is higher, p' can still be accepted in order to avoid being trapped in a local minimum.

The main difference of SA to other algorithms is involvement of time-varying parameter, T, which gets decreases progressively with each algorithmic iteration. From Eqn. A.23, it can be seen that as T decreases (hence the term "annealing"), α of accepting bad parameter sets

(i.e. those giving $f(p') \geq f(p)$) also decreases. The idea of MCSA is that as T tends to zero, the remaining $f(p)$ is the estimated as the minimum.

However, $f(p')$ obtained from lowest T does not necessarily represent the global minimum. A great advantage of MCSA is that multiple minima can easily be identified via so-called "reannealing". Once a minimum in $f(p)$ is identified, T is increased again with a different starting point. Repeating annealing and reannealing steps allows approximation of the minima, and after sufficient iterations, the lowest minimum can be selected to estimate the global minimum.

A.3.2 Objective function

Whilst the values of several parameters are fixed from literature or analytical derivation (Eqn. A.6), the values for the unknown parameters were determined by fitting the model to the experimental data obtained from all experiment using a Monte Carlo Simulated Annealing (MCSA) algorithm. The algorithm seeks a parameter vector, $\mathbf{p}$, which minimises the objective function evaluating the fit between the model simulation and experimental data. The vector $\mathbf{p}$ consists of 10 parameters for the original model. The objective function $f(\mathbf{p})$, for the vector $\mathbf{p}$, is defined as

$$f(\mathbf{p}) = \sum_{S} \sum_{t} \left(\frac{D_S(t) - M_{S,\mathbf{p}}(t)}{\sigma_A D_S(t)} + \frac{\Delta D_C - \Delta M_{C,\mathbf{p}}}{\sigma_B \Delta D_C} \right)^2 . \tag{A.24}$$

The first term quantifies goodness of fit of the simulated *RD29A* expression profile, $M_{S,\mathbf{p}}(t)$ to experimental mean of *RD29A* fold-change expression, $D_S(t)$, measured at time t (0, 0.5, 1, 2, 3 or 5 hours) at treatment condition, $S = (S_1, 0), (0, S_2)$ or (S_1, S_2), with S_1 and S_2 represent NaCl and ABA input strength at arbitrary scale (0, 0.5, or 1). The error σ_A associated with each data point $D_S(t)$ is assumed to be uniform ($\sigma_A = 0.02$), hence serving as weighting coefficient of the first term in the objective function. The second term evaluates the ability of the model with $\mathbf{p}$ to reproduce the observed synergistic effect, where ΔD_C and

$\Delta M_{C,\mathbf{p}}$ respectively denote the observed and simulated slope of fold increase between 3 and 5 hours of combined stress treatment only. The error σ_B associated with the observed slope is fixed ($\sigma_B = 0.1$). MCSA Optimisation of the objective function f for each system structure identifies $\mathbf{p}'$, which approximates the vector of parameters at the global optimum of the objective function f. The parameter sets $\mathbf{p}'$ identified for each systems structure is shown in Table A.1.

A.3.3 Selection of system structure

The objective function f does not adequately represent the ability of the system structure to reproduce the synergistic effect, because the global minimum $f(\mathbf{p}')$ is designed to represent the best quantitative description of the observed *RD29A* expression profiles under all stress conditions a system structure can provide, not specifically for the combined stress. We therefore introduce a selection function, g, which evaluates the goodness of fit between the observed *RD29A* expression profile and the optimised model only under combined stress:

$$g(\mathbf{p}') = \sum_t \left(\frac{D_C(t) - M_{C,\mathbf{p}'}(t)}{\sigma} \right)^2. \tag{A.25}$$

where $D_C(t)$ represents the experimentally observed fold change at time t under combined stress treatment, and $M_{C,\mathbf{p}'}(t)$ the simulated fold change using the parameter set $\mathbf{p}'$ identified from optimising the function f. Note that g is normalised by a uniform error (=0.02). The system structures producing lowest g when implemented in the model were chosen as the viable structures (Fig. 6.6).

Table A.1: Parameter values used for the solutions presented in Chapter 6 figures. Some parameter values are fixed ($r_1 = d_{1b} = 0, \delta_1 = \delta_2 = 0.02, \tau = 2.5$)

Model type	r_1^τ	α_1	d_{-1}	u_1	d_{2b}	r_2^τ	α_2	d_{-2}	u_2	r_c^τ	c^E	c^I	Corresponding Figures
No crosstalk	843.9	762.4	24.34	334.2	411.8	311.3	1346	95.29	88.97	549			6.4
$E(\alpha_1)$	147.7	6303	35.82	245.3	361.2	807.9	1254	288.4	81.69	76.38	11.81		6.6
$E(\alpha_2)$	265	5662	95.87	144.3	353.1	278.2	1289	83.93	92.89	53.21	0.1243		
$E(d_{-1})$	1716	4781	202.2	429.7	408.5	301	1523	106.4	85.73	319.5	24.6		
$E(d_{-2})$	380.6	1755	21.05	278.4	234.4	319.8	880.8	122.4	45.6	1443	10.48		
$E(r_c^\tau)$	322.8	1178	23.56	138.3	250.4	8184	844.9	821.2	187.5	36.03	2.23		
$E(r_1^\tau)$	270.3	4780	94.55	120.5	359	2150	1171	508.2	111.2	158.4	0.021		
$E(r_2^\tau)$	340.6	1491	11.92	370.9	146.7	3874	501.25	193.4	227.36	201.5	0.047		
$E(u_1)$	479.6	6957	102.1	330.6	467.9	298.7	1717	65.55	158.2	336.7	39.65		
$E(u_2)$	375.9	1262	30.03	142	283.5	307.3	1052	64.27	100.6	1539	11.51		
$I(u_1)$	153.3	4462	135	48.61	781.4	269.9	2923	236.7	72.37	35.35		12.37	
$I(d_{-2})$	2333	1119	62.47	744.8	449.2	173.4	1687	50.76	185.2	40.37		23.07	
$I(\alpha_1)$	1294	1397	35.86	526.2	473.2	301.8	1769	77.96	135	330.5		40.76	
$I(\alpha_2)$	251.4	5534	70.08	221.6	530.1	298.7	1943	119.4	97.9	595.7		17.37	
$I(r_c^\tau)$	183.1	7366	59.57	277.9	441.6	298.7	1622	119.9	80.3	693.5		16.79	
$I(r_1^\tau)$	502.7	3754	117.4	160.5	475	302	1768	78.66	135	304.7		13.74	
$I(r_2^\tau)$	301.2	5684	83.2	185.5	646.1	298.2	2344	130.5	106.8	195.8		10.3	
$I(d_{-1})$	270.7	617.5	3.233	492.7	848.6	267.7	3120	171.1	107.9	42.89		136.4	6.6, 6.8
$I(u_2)$	1033	1205	67.49	286.5	317.9	222.9	1153	552.6	14.71	38.47		27.24	

Bibliography

[1] Abe, H., Urao, T., Ito, T., Seki, M. and Shinozaki, K. [2003], 'Transcriptional Activators in Abscisic Acid Signaling', *The Plant Cell* **15**(1), 63–78.

[2] Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D. and Shinozaki, K. [1997], 'Role of arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression.', *The Plant Cell* **9**(10), 1859–1868.

[3] Addicott, F. T., Lyon, J. L., Ohkuma, K., Thiessen, W. E., Carns, H. R., Smith, O. E., Cornforth, J. W., Milborrow, B. V., Ryback, G. and Wareing, P. F. [1968], 'Abscisic acid: a new name for abscisin II (dormin)', *Science* **159**(3822), 1493.

[4] Agarwal, P. K. and Jha, B. [2010], 'Transcription factors in plants and ABA dependent and independent abiotic stress signalling', *Biologia Plantarum* **54**(2), 201–212.

[5] Alba, B. M. and Gross, C. A. [2004], 'Regulation of the Escherichia coliσE-dependent envelope stress response', *Molecular microbiology* **52**(3), 613–619.

[6] Alon, U. [2006], *An Introduction to Systems Biology: Design Principles of Biological Circuits*, CRC Press.

[7] Anderson, J. P., Badruzsaufari, E., Schenk, P. M., Manners, J. M., Desmond, O. J., Ehlert, C., Maclean, D. J., Ebert, P. R. and Kazan, K. [2004], 'Antagonistic interaction between abscisic acid and jasmonate-ethylene signaling pathways modulates defense gene expression and disease resistance in Arabidopsis', *The Plant Cell* **16**(12), 3460–3479.

[8] Antoni, R., Gonzalez-Guzman, M., Rodriguez, L., Rodrigues, A., Pizzio, G. a. and Rodriguez, P. L. [2012], 'Selective Inhibition of Clade A Phosphatases Type 2C by PYR/PYL/RCAR Abscisic Acid Receptors', *Plant Physiology* **158**(2), 970–980.

[9] Arabidopsis Genome Initiative [2000], 'Analysis of the genome sequence of the flowering plant Arabidopsis thaliana.', *Nature* **408**(6814), 796–815.

[10] Asai, T., Tena, G., Plotnikova, J., Willmann, M. R., Chiu, W.-L., Gomez-Gomez, L., Boller, T., Ausubel, F. M. and Sheen, J. [2002], 'MAP kinase signalling cascade in Arabidopsis innate immunity.', *Nature* **415**(6875), 977–983.

[11] Atchley, W. R., Terhalle, W. and Dress, A. [1999], 'Positional dependence, cliques, and predictive motifs in the bHLH protein domain', *Journal of Molecular Evolution* **48**(5), 501–516.

[12] Atkinson, N. J., Jain, R. and Urwin, P. E. [2015], The Response of Plants to Simultaneous Biotic and Abiotic stress, *in* 'Combined Stresses in Plants: Physiological, Molecular and Biochemical Aspects', Springer, pp. 181–201.

[13] Atkinson, N. J., Lilley, C. J. and Urwin, P. E. [2013], 'Identification of genes involved in the response of Arabidopsis thaliana to simultaneous biotic and abiotic stresses', *Plant Physiology* **162**(4), 2028–2041.

[14] Atkinson, N. J. and Urwin, P. E. [2012], 'The interaction of plant biotic and abiotic stresses: from genes to the field', *Journal of experimental botany* **63**(10), 3523–3543.

[15] Audran, C., Borel, C., Frey, A., Sotta, B., Meyer, C., Simonneau, T. and Marion-Poll, A. [1998], 'Expression studies of the zeaxanthin epoxidase gene in Nicotiana plumbaginifolia', *Plant Physiology* **118**(3), 1021–1028.

[16] Avison, M. [2005], *Measuring Gene Expression*, Garland Science.

[17] Ay, A. and Arnosti, D. N. [2011], 'Mathematical modelling of gene expression: a guide

for the perplexed biologist', *Critical Reviews in Biochemistry and Molecular Biology* **46**(2), 1–23.

[18] Bailey, P. C., Dicks, J., Wang, T. L. and Martin, C. [2008], 'IT3F: a web-based tool for functional analysis of transcription factors in plants', *Phytochemistry* **69**(13), 2417–2425.

[19] Bakshi, M. and Oelmüller, R. [2014], 'WRKY transcription factors: Jack of many trades in plants.', *Plant Signaling & Behavior* **9**(1), 1–18.

[20] Bansal, M., Belcastro, V., Ambesi-Impiombato, A. and Di Bernardo, D. [2007], 'How to infer gene networks from expression profiles', *Molecular Systems Biology* **3**(1), 78.

[21] Barah, P., B N, M. N., Jayavelu, N. D., Sowdhamini, R., Shameer, K. and Bones, A. M. [2015], 'Transcriptional regulatory networks in Arabidopsis thaliana during single and combined stresses.', *Nucleic Acids Research* **44**(7), 3147–3164.

[22] Bari, R. and Jones, J. D. G. [2009], 'Role of plant hormones in plant defence responses', *Plant Molecular Biology* **69**(4), 473–488.

[23] Beguerisse-Diaz, M., Hernández-Gómez, M. C., Lizzul, A. M., Barahona, M. and Desikan, R. [2012], 'Compound stress response in stomatal closure: a mathematical model of ABA and ethylene interaction in guard cells.', *BMC Systems Biology* **6**, 146.

[24] Bhaskaran, S., Nair, U. P. and Nair, A. S. [2015], Hill Equation in Modeling Transcriptional Regulation, *in* 'Systems and Synthetic Biology', Springer, pp. 77–92.

[25] Bijlsma, J. J. and Groisman, E. a. [2003], 'Making informed decisions: regulatory interactions between two-component systems', *Trends in Microbiology* **11**(8), 359–366.

[26] Biswas, D. K. and Jiang, G. M. [2011], 'Differential drought-induced modulation of ozone tolerance in winter wheat species', *Journal of Experimental Botany* p. err104.

[27] Boatwright, J. L. and Pajerowska-Mukhtar, K. [2013], 'Salicylic acid: an old hormone up to new tricks', *Molecular Plant Pathology* **14**(6), 623–634.

[28] Bohnert, H., Nelson, D. and Jensen, R. [1995], 'Adaptations to Environmental Stresses.', *The Plant Cell* **7**(7), 1099–1111.

[29] Boter, M., Ruíz-Rivero, O., Abdeen, A. and Prat, S. [2004], 'Conserved MYC transcription factors play a key role in jasmonate signaling both in tomato and Arabidopsis', *Genes & development* **18**(13), 1577–1591.

[30] Boudsocq, M. and Laurière, C. [2005], 'Osmotic signaling in plants: multiple pathways mediated by emerging kinase families.', *Plant Physiology* **138**(3), 1185–1194.

[31] Bradbury, J. F. [1986], *Guide to Plant Pathogenic Bacteria.*, CAB international.

[32] Bray, E. A. [1997], 'Plant responses to water deficit', *Trends in Plant Science* **2**(2), 48–54.

[33] Bruce, T. J. and Pickett, J. a. [2007], 'Plant defence signalling induced by biotic attacks', *Current Opinion in Plant Biology* **10**(4), 387–392.

[34] Burgos, A., Szymanski, J., Seiwert, B., Degenkolbe, T., Hannah, M. a., Giavalisco, P. and Willmitzer, L. [2011], 'Analysis of short-term changes in the Arabidopsis thaliana glycerolipidome in response to temperature and light', *The Plant Journal* **66**(4), 656–668.

[35] Cao, H., Bowling, S. A., Gordon, A. S. and Dong, X. [1994], 'Characterization of an Arabidopsis mutant that is nonresponsive to inducers of systemic acquired resistance.', *The Plant Cell* **6**(11), 1583–1592.

[36] Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., Lewis, S., Lomax, J., Mungall, C., Hitz, B., Balakrishnan, R., Dolan, M., Wood, V., Hong, E. and Gaudet, P. [2009], 'AmiGO: Online access to ontology and annotation data', *Bioinformatics* **25**(2), 288–289.

[37] Carrera, J., Rodrigo, G., Jaramillo, A. and Elena, S. F. [2009], 'Reverse-engineering the

Arabidopsis thaliana transcriptional network under changing environmental conditions', *Genome Biology* **10**(9), 15 str.

[38] Cases, I., de Lorenzo, V. and Ouzounis, C. A. [2003], 'Transcription regulation and environmental adaptation in bacteria', *Trends in Microbiology* **11**(6), 248–252.

[39] Chang, C., Kwok, S. F., Bleecker, A. B. and Meyerowitz, E. M. [1993], 'Arabidopsis ethylene-response gene ETR1: similarity of product to two-component regulators', *Science* **262**(5133), 539–544.

[40] Chang S. S., Park S. K., Kim B. C., Kang B. J., Kim D. U. and Nam H. G. [1994], 'Stable genetic transformation of Arabidopsis thaliana by Agrobacterium inoculation in planta', *The Plant Journal* **5**(4), 551–558.

[41] Chaves, M. M., Maroco, J. P. and Pereira, J. S. [2003], 'Understanding plant responses to drought from genes to the whole plant', *Functional plant biology* **30**(3), 239–264.

[42] Chen, Y. T., Liu, H., Stone, S. and Callis, J. [2013], 'ABA and the ubiquitin E3 ligase KEEP ON GOING affect proteolysis of the Arabidopsis thaliana transcription factors ABF1 and ABF3', *The Plant Journal* **75**, 965–976.

[43] Chen, Z., Zheng, Z., Huang, J., Lai, Z. and Fan, B. [2009], 'Biosynthesis of salicylic acid in plants.', *Plant Signaling & Behavior* **4**(6), 493–496.

[44] Chinnusamy, V., Schumaker, K. and Zhu, J. K. [2004], 'Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants', *Journal of Experimental Botany* **55**(395), 225–236.

[45] Chinnusamy, V., Zhu, J. and Zhu, J.-K. [2007], 'Cold stress regulation of gene expression in plants.', *Trends in plant science* **12**(10), 444–451.

[46] Chugani, S. and Greenberg, E. P. [2010], 'LuxR homolog-independent gene regulation by acyl-homoserine lactones in Pseudomonas aeruginosa', *Proceedings of the National Academy of Sciences* **107**(23), 10673–10678.

[47] Ciolkowski, I., Wanke, D., Birkenbihl, R. P. and Somssich, I. E. [2008], 'Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function', *Plant Molecular Biology* **68**(1-2), 81–92.

[48] Clough, S. and Bent, A. [1998], 'Floral dip: a simplified method for Agro- bacterium-mediated transformation of Arabidopsis thaliana.', *The Plant Journal* **16**(6), 735–743.

[49] Collinge, D. B. and Slusarenko, A. J. [1987], 'Plant gene expression in response to pathogens', *Plant Molecular Biology* **9**(4), 389–410.

[50] Cornforth, D. M., Popat, R., McNally, L., Gurney, J., Scott-Phillips, T. C., Ivens, A., Diggle, S. P. and Brown, S. P. [2014], 'Combinatorial quorum sensing allows bacteria to resolve their social and physical environment', *Proceedings of the National Academy of Sciences* **111**(11), 4280–4284.

[51] Davuluri, R. V., Sun, H., Palaniswamy, S. K., Matthews, N., Molina, C., Kurtz, M. and Grotewold, E. [2003], 'AGRIS : Arabidopsis Gene Regulatory Information Server , an transcription factors', *BMC Bioinformatics* **4**(25), 1–11.

[52] de Torres Zabala, M., Bennett, M. H., Truman, W. H. and Grant, M. R. [2009], 'Antagonism between salicylic and abscisic acid reflects early host-pathogen conflict and moulds plant defence responses', *The Plant Journal* **59**(3), 375–386.

[53] De Vos, M., Van Oosten, V. R., Van Poecke, R. M. P., Van Pelt, J. A., Pozo, M. J., Mueller, M. J., Buchala, A. J., Métraux, J.-p., Van Loon, L. C., Dicke, M. and Pieterse, C. M. J. [2005], 'Signal signature and transcriptome changes of Arabidopsis during pathogen and insect attack.', *Molecular Plant-Microbe Interactions* **18**(9), 923–937.

[54] D'haeseleer, P., Liang, S. and Somogyi, R. [2000], 'Genetic network inference: from co-expression clustering to reverse engineering', *Bioinformatics* **16**(8), 707–726.

[55] Dong, C.-h., Agarwal, M., Zhang, Y., Xie, Q. and Zhu, J.-k. [2006], 'The negative regulator of plant cold responses, HOS1, is a RING E3 ligase that mediates the ubiq-

uitination and degradation of ICE1.', *Proceedings of the National Academy of Sciences* **103**(21), 8281–8286.

[56] Dubnau, D. and Losick, R. [2006], 'Bistability in bacteria', *Molecular Microbiology* **61**(3), 564–572.

[57] Edgar, R., Domrachev, M. and Lash, A. E. [2002], 'Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.', *Nucleic acids research* **30**(1), 207–210.

[58] Estavillo, G. M., Crisp, P. a., Pornsiriwong, W., Wirtz, M., Collinge, D., Carrie, C., Giraud, E., Whelan, J., David, P., Javot, H., Brearley, C., Hell, R., Marin, E. and Pogson, B. J. [2011], 'Evidence for a SAL1-PAP Chloroplast Retrograde Pathway That Functions in Drought and High Light Signaling in Arabidopsis', *The Plant Cell* **23**(11), 3992–4012.

[59] Eulgem, T., Rushton, P. J., Robatzek, S. and Somssich, I. E. [2000], 'The WRKY superfamily of plant transcription factors', *Trends in Plant Science* **5**(5), 199–206.

[60] Eulgem, T. and Somssich, I. E. [2007], 'Networks of WRKY transcription factors in defense signaling', *Current Opinion in Plant Biology* **10**(4), 366–371.

[61] Evrard, A., Kumar, M., Lecourieux, D., Lucks, J., von Koskull-Döring, P. and Hirt, H. [2013], 'Regulation of the heat stress response in Arabidopsis by MPK6-targeted phosphorylation of the heat stress factor HsfA2.', *PeerJ* **1**, e59.

[62] Finka, A., Cuendet, A. F. H., Maathuis, F. J. M., Saidi, Y. and Goloubinoff, P. [2012], 'Plasma membrane cyclic nucleotide gated calcium channels control land plant thermal sensing and acquired thermotolerance', *The Plant Cell* **24**(8), 3333–3348.

[63] Finkelstein, R. [2013], 'Abscisic Acid synthesis and response.', *The Arabidopsis Book* **11**, e0166.

[64] Finkelstein, R. R., Gampala, S. S. L. and Rock, C. D. [2002], 'Abscisic acid signaling in seeds and seedlings', *The Plant Cell* **14**(suppl 1), S15–S45.

[65] Fischer, R. A. and Maurer, R. [1978], 'Drought resistance in spring wheat cultivars. I. Grain yield responses', *Crop and Pasture Science* **29**(5), 897–912.

[66] Flowers, T. J. and Yeo, A. R. [1995], 'Breeding for salinity resistance in crop plants: where next?', *Functional Plant Biology* **22**(6), 875–884.

[67] Fobert, P. R. and Després, C. [2005], 'Redox control of systemic acquired resistance', *Current Opinion in Plant Biology* **8**(4), 378–382.

[68] Fowler, S. G., Cook, D. and Thomashow, M. F. [2005], 'Low temperature induction of Arabidopsis CBF1, 2, and 3 is gated by the circadian clock.', *Plant Physiology* **137**(3), 961–968.

[69] Fraire-Velazquez, S., Rodriguez-Guerra, R. and Sanchez-Calderon, L. [2011], Abiotic and Biotic Stress Response Crosstalk in Plants, *in* 'Abiotic stress response in plants - Physiological, biochemical and genetic perspectives', InTech, pp. 3–26.

[70] Friedman, N., Linial, M., Nachman, I. and Pe'er, D. [2000], 'Using Bayesian Networks to Analyze Expression Data.', *Journal of Computational Biology* **7**, 601–620.

[71] Fujii, H., Chinnusamy, V., Rodrigues, A., Rubio, S., Antoni, R., Park, S.-Y., Cutler, S. R., Sheen, J., Rodriguez, P. L. and Zhu, J.-K. [2009], 'In vitro reconstitution of an abscisic acid signalling pathway', *Nature* **462**(7273), 660–664.

[72] Fujimoto, S. Y., Ohta, M., Usui, A., Shinshi, H. and Ohme-Takagi, M. [2000], 'Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression.', *The Plant Cell* **12**(3), 393–404.

[73] Fujita, M., Fujita, Y., Maruyama, K., Seki, M., Hiratsu, K., Ohme-Takagi, M., Tran, L. S. P., Yamaguchi-Shinozaki, K. and Shinozaki, K. [2004], 'A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway', *The Plant Journal* **39**(6), 863–876.

[74] Fujita, M., Fujita, Y., Noutoshi, Y., Takahashi, F., Narusaka, Y., Yamaguchi-Shinozaki,

K. and Shinozaki, K. [2006], ‘Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks’, *Current Opinion in Plant Biology* **9**(4), 436–442.

[75] Fujita, Y., Fujita, M., Satoh, R., Maruyama, K., Parvez, M. M., Seki, M., Hiratsu, K., Ohme-Takagi, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. [2005], ‘AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in Arabidopsis.’, *The Plant Cell* **17**, 3470–3488.

[76] Fujita, Y., Yoshida, T. and Yamaguchi-Shinozaki, K. [2013], ‘Pivotal role of the AREB/ABF-SnRK2 pathway in ABRE-mediated transcription in response to osmotic stress in plants.’, *Physiologia plantarum* **147**(1), 15–27.

[77] Furihata, T., Maruyama, K., Fujita, Y., Umezawa, T., Yoshida, R., Shinozaki, K. and Yamaguchi-Shinozaki, K. [2006], ‘Abscisic acid-dependent multisite phosphorylation regulates the activity of a transcription activator AREB1.’, *Proceedings of the National Academy of Sciences* **103**(6), 1988–93.

[78] Gamble, R. L., Coonfield, M. L. and Schaller, G. E. [1998], ‘Histidine kinase activity of the ETR1 ethylene receptor from Arabidopsis’, *Proceedings of the National Academy of Sciences* **95**(13), 7825–7829.

[79] Gao, a. G., Hakimi, S. M., Mittanck, C. a., Wu, Y., Woerner, B. M., Stark, D. M., Shah, D. M., Liang, J. and Rommens, C. M. [2000], ‘Fungal pathogen protection in potato by expression of a plant defensin peptide.’, *Nature Biotechnology* **18**(12), 1307–1310.

[80] Garcia-Hernandez, M., Berardini, T. Z., Chen, G., Crist, D., Doyle, A., Huala, E., Knee, E., Lambrecht, M., Miller, N., Mueller, L. a., Mundodi, S., Reiser, L., Rhee, S. Y., Scholl, R., Tacklind, J., Weems, D. C., Wu, Y., Xu, I., Yoo, D., Yoon, J. and Zhang, P. [2002], ‘TAIR: A resource for integrated Arabidopsis data’, *Functional and Integrative Genomics* **2**(6), 239–253.

[81] Garg, A. K., Kim, J.-K., Owens, T. G., Ranwala, A. P., Choi, Y. D., Kochian, L. V.

and Wu, R. J. [2002], 'Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses.', *Proceedings of the National Academy of Sciences* **99**(25), 15898–15903.

[82] Gassert, F., Landis, M., Luck, M., Reig, P. and Shiao, T. [2013], 'Aqueduct global maps 2.0', *Water Resources Institute* pp. 1–20.

[83] Georghiou, G. P. [2012], *Pest resistance to pesticides*, Springer Science & Business Media.

[84] Giraud, E., Ho, L. H. M., Clifton, R., Carroll, A., Estavillo, G., Tan, Y.-F., Howell, K. a., Ivanova, A., Pogson, B. J., Millar, a. H. and Whelan, J. [2008], 'The absence of ALTERNATIVE OXIDASE1a in Arabidopsis results in acute sensitivity to combined light and drought stress.', *Plant Physiology* **147**(2), 595–610.

[85] Glazebrook, J. [2005], 'Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens', *Annual Review of Phytopathology* **43**(3), 205–227.

[86] Glazebrook, J., Rogers, E. E. and Ausubel, F. M. [1996], 'Isolation of Arabidopsis mutants with enhanced disease susceptibility by direct screening', *Genetics* **143**(2), 973–982.

[87] Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M. and Toulmin, C. [2010], 'Food security: the challenge of feeding 9 billion people', *Science* **327**(5967), 812–818.

[88] Gómez-Gómez, L. and Boller, T. [2000], 'FLS2: an LRR receptor–like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis', *Molecular Cell* **5**(6), 1003–1011.

[89] Gong, Z., Lee, H., Xiong, L., Jagendorf, A., Stevenson, B. and Zhu, J.-k. [2002], 'RNA helicase-like protein as an early regulator of transcription factors for plant chilling and freezing tolerance.', *Proceedings of the National Academy of Sciences* **99**(17), 11507–11512.

[90] Guo, A., He, K., Liu, D., Bai, S., Gu, X., Wei, L. and Luo, J. [2005], 'DATF: A database of Arabidopsis transcription factors', *Bioinformatics* **21**(10), 2568–2569.

[91] Hahn, A., Bublak, D., Schleiff, E. and Scharf, K.-D. [2011], 'Crosstalk between Hsp90 and Hsp70 chaperones and heat stress transcription factors in tomato.', *The Plant Cell* **23**(2), 741–755.

[92] Harmer, S. L., Hogenesch, J. B., Straume, M., Chang, H. S., Han, B., Zhu, T., Wang, X., Kreps, J. a. and Kay, S. a. [2000], 'Orchestrated transcription of key pathways in Arabidopsis by the circadian clock.', *Science* **290**(5499), 2110–2113.

[93] Hasty, J., McMillen, D., Isaacs, F. and Collins, J. J. [2001], 'Computational studies of gene regulatory networks: in numero molecular biology', *Nature Review Genetics* **2**(4), 268–279.

[94] Hecker, M., Lambeck, S., Toepfer, S., Van Someren, E. and Guthke, R. [2009], 'Gene regulatory network inference: data integration in dynamic modelsa review', *Biosystems* **96**(1), 86–103.

[95] Hirano, S. S. and Upper, C. D. [1990], 'Population biology and epidemiology of Pseudomonas syringae', *Annual Review of Phytopathology* **28**(1), 155–177.

[96] Hirayama, T. and Shinozaki, K. [2010], 'Research on plant abiotic stress responses in the post-genome era: Past, present and future', *The Plant Journal* **61**(6), 1041–1052.

[97] Hirt, H. [2010], *Plant Stress Biology: From Genomics to Systems Biology*, John Wiley & Sons.

[98] Hong, J. C. [2015], General Aspects of Plant Transcription Factor Families, *in* D. H. Gonzalez, ed., 'Plant Transcription Factors', Academic Press, pp. 35–56.

[99] Hotta, C. T., Gardner, M. J., Hubbard, K. E., Baek, S. J., Dalchau, N., Suhita, D., Dodd, A. N. and Webb, A. a. R. [2007], 'Modulation of environmental responses of plants by circadian clocks', *Plant, Cell & Environment* **30**(3), 333–349.

[100] Huala, E., Dickerman, a. W., Garcia-Hernandez, M., Weems, D., Reiser, L., LaFond, F., Hanley, D., Kiphart, D., Zhuang, M., Huang, W., Mueller, L. a., Bhattacharyya, D., Bhaya, D., Sobral, B. W., Beavis, W., Meinke, D. W., Town, C. D., Somerville, C. and Rhee, S. Y. [2001], 'The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant.', *Nucleic Acids Research* **29**(1), 102–105.

[101] Hwang, S., Rhee, S. Y., Marcotte, E. M. and Lee, I. [2011], 'Systematic prediction of gene function in Arabidopsis thaliana using a probabilistic functional gene network.', *Nature Protocols* **6**(9), 1429–1442.

[102] Ichimura, K., Mizoguchi, T., Yoshida, R., Yuasa, T. and Shinozaki, K. [2000], 'Various abiotic stresses rapidly activate Arabidopsis MAP kinases ATMPK4 and ATMPK6', *Plant Journal* **24**(5), 655–665.

[103] Iida, K., Seki, M., Sakurai, T., Satou, M., Akiyama, K., Toyoda, T., Konagaya, A. and Shinozaki, K. [2005], 'RARTF: Database and tools for complete sets of Arabidopsis transcription factors', *DNA Research* **12**(4), 247–256.

[104] Ingram, P. J., Stumpf, M. P. H. and Stark, J. [2006], 'Network motifs: structure does not determine function.', *BMC Genomics* **7**(5), 108.

[105] Ishitani, M., Xiong, L., Stevenson, B. and Zhu, J. K. [1997], 'Genetic analysis of osmotic and cold stress signal transduction in Arabidopsis: interactions and convergence of abscisic acid-dependent and abscisic acid-independent pathways.', *The Plant Cell* **9**(11), 1935–1949.

[106] Jacob, F. and Monod, J. [1961], 'Genetic regulatory mechanisms in the synthesis of proteins', *Journal of Molecular Biology* **3**(3), 318–356.

[107] Jaglo-Ottosen, K. R. [1998], 'Arabidopsis CBF1 Overexpression Induces COR Genes and Enhances Freezing Tolerance', *Science* **280**(5360), 104–106.

[108] Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J.,

Kroj, T. and Parcy, F. [2002], 'bZIP transcription factors in Arabidopsis', *Trends in Plant Science* **7**(3), 106–111.

[109] Jin, J., He, K., Tang, X., Li, Z., Lv, L., Zhao, Y., Luo, J. and Gao, G. [2015], 'An Arabidopsis Transcriptional Regulatory Map Reveals Distinct Functional and Evolutionary Features of Novel Transcription Factors', *Molecular Biology and Evolution* **32**(7), 1767–1773.

[110] Johnson, S. M., Lim, F.-L., Finkler, A., Fromm, H., Slabas, A. R. and Knight, M. R. [2014], 'Transcriptomic analysis of Sorghum bicolor responding to combined heat and drought stress.', *BMC Genomics* **15**, 456.

[111] Ju, C., Yoon, G. M., Shemansky, J. M., Lin, D. Y., Ying, Z. I., Chang, J., Garrett, W. M., Kessenbrock, M., Groth, G., Tucker, M. L., Cooper, B., Kieber, J. J. and Chang, C. [2012], 'CTR1 phosphorylates the central regulator EIN2 to control ethylene hormone signaling from the ER membrane to the nucleus in Arabidopsis.', *Proceedings of the National Academy of Sciences* **109**(47), 19486–19491.

[112] Jupin, I. and Chua, N. H. [1996], 'Activation of the CaMV as-1 cis-element by salicylic acid: differential DNA-binding of a factor related to TGA1a.', *The EMBO journal* **15**(20), 5679–89.

[113] Kanaoka, M. M., Pillitteri, L. J., Fujii, H., Yoshida, Y., Bogenschutz, N. L., Takabayashi, J., Zhu, J.-K. and Torii, K. U. [2008], 'SCREAM/ICE1 and SCREAM2 Specify Three Cell-State Transitional Steps Leading to Arabidopsis Stomatal Differentiation', *The Plant Cell* **20**(7), 1775–1785.

[114] Kang, J., Hwang, J.-U., Lee, M., Kim, Y.-Y., Assmann, S. M., Martinoia, E. and Lee, Y. [2010], 'PDR-type ABC transporter mediates cellular uptake of the phytohormone abscisic acid', *Proceedings of the National Academy of Sciences* **107**(5), 2355–2360.

[115] Kaplan, B., Davydov, O., Knight, H., Galon, Y., Knight, M. R., Fluhr, R. and Fromm, H. [2006], 'Rapid transcriptome changes induced by cytosolic Ca2+ transients reveal

ABRE-related sequences as Ca2+ responsive cis-elements in Arabidopsis', *The Plant Cell* **18**(10), 2733–2748.

[116] Karamanos, A. J. and Papatheohari, A. Y. [1999], 'Assessment of drought resistance of crop genotypes by means of the water potential index', *Crop Science* **39**(6), 1792–1797.

[117] Kasuga, M., Liu, Q., Miura, S., Yamaguchi-Shinozaki, K. and Shinozaki, K. [1999], 'Improving plant drought, salt and freezing tolerance by gene transfer of a single stress-inducible transcription factor.', *Nature Biotechnology* **236**(3), 287–291.

[118] Katagiri, F., Seipel, K. and Chua, N. H. [1992], 'Identification of a novel dimer stabilization region in a plant bZIP transcription activator.', *Molecular & Cellular Biology* **12**(11), 4809–4816.

[119] Kazan, K. and Manners, J. M. [2012], 'JAZ repressors and the orchestration of phytohormone crosstalk', *Trends in Plant Science* **17**(1), 22–31.

[120] Kesarwani, M., Yoo, J. and Dong, X. [2007], 'Genetic interactions of TGA transcription factors in the regulation of pathogenesis-related genes and disease resistance in Arabidopsis.', *Plant Physiology* **144**(1), 336–346.

[121] Kieber, J. J., Rothenberg, M., Roman, G., Feldmann, K. A. and Ecker, J. R. [1993], 'CTR1, a negative regulator of the ethylene response pathway in Arabidopsis, encodes a member of the raf family of protein kinases', *Cell* **72**(3), 427–441.

[122] Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., D'Angelo, C., Bornberg-Bauer, E., Kudla, J. and Harter, K. [2007], 'The AtGenExpress global stress expression data set: Protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses', *The Plant Journal* **50**(2), 347–363.

[123] Kim, J.-R., Shin, D., Jung, S. H., Heslop-Harrison, P. and Cho, K.-H. [2010], 'A design principle underlying the synchronization of oscillations in cellular systems.', *Journal of Cell Science* **123**(Pt 4), 537–43.

[124] Kim, J.-S., Mizoi, J., Yoshida, T., Fujita, Y., Nakajima, J., Ohori, T., Todaka, D., Nakashima, K., Hirayama, T., Shinozaki, K. and Yamaguchi-Shinozaki, K. [2011], 'An ABRE promoter sequence is involved in osmotic stress-responsive expression of the DREB2A gene, which encodes a transcription factor regulating drought-inducible genes in Arabidopsis.', *Plant & Cell Physiology* **52**(12), 2136–46.

[125] Kitano, H. [2002], 'Systems biology: a brief overview.', *Science* **295**(5560), 1662–1664.

[126] Klipp, E., Herwig, R., Kowald, A., Wierling, C. and Lehrach, H. [2008], *Systems biology in practice: concepts, implementation and application*, John Wiley & Sons.

[127] Knight, H. and Knight, M. R. [2001], 'Abiotic stress signalling pathways: specificity and cross-talk.', *Trends in Plant Science* **6**(6), 262–267.

[128] Koornneef, A. and Pieterse, C. M. J. [2008], 'Crosstalk in defense signaling', *Plant physiology* **146**(3), 839–844.

[129] Kreps, J. a., Wu, Y., Chang, H.-S., Zhu, T., Wang, X. and Harper, J. F. [2002], 'Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress.', *Plant Physiology* **130**(4), 2129–2141.

[130] Kunkel, B. N. and Brooks, D. M. [2002], 'Crosstalk between signaling pathways in pathogen defense', *Current Opinion in Plant Biology* **5**(4), 325–331.

[131] Kurup, S., Jones, H. D. and Holdsworth, M. J. [2000], 'Interactions of the developmental regulator ABI3 with proteins identified from developing Arabidopsis seeds', *The Plant Journal* **21**(2), 143–155.

[132] Lai, K., Robertson, M. J. and Schaffer, D. V. [2004], 'The sonic hedgehog signaling system as a bistable genetic switch.', *Biophysical journal* **86**(5), 2748–2757.

[133] Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D. L., Garcia-Hernandez, M., Karthikeyan, A. S., Lee, C. H., Nelson, W. D., Ploetz, L., Singh, S., Wensel, A. and Huala, E. [2012], 'The Arabidopsis

Information Resource (TAIR): Improved gene annotation and new tools', *Nucleic Acids Research* **40**(D1), D1202–D1210.

[134] Langridge, P. and Fleury, D. [2011], 'Making the most of 'omics' for crop breeding', *Trends in Biotechnology* **29**(1), 33–40.

[135] Latchman, D. [2002], *Gene Regulation: A Eukaryotic Perspective*, 4th edn, Nelson Thornes.

[136] Lawrence, J. G. [2002], 'Shared strategies in gene organization among prokaryotes and eukaryotes', *Cell* **110**(4), 407–413.

[137] Lawton, K., Weymann, K., Friedrich, L., Vernooij, B., Uknes, S. and Ryals, J. [1995], 'Systemic acquired resistance in Arabidopsis requires salicylic acid but not ethylene', *Molecular Plant Microbe Interactions* **8**(6), 863–870.

[138] Lee, I., Ambaru, B., Thakkar, P., Marcotte, E. M. and Rhee, S. Y. [2010], 'Rational association of genes with traits using a genome-scale gene network for Arabidopsis thaliana.', *Nature Biotechnology* **28**(2), 149–156.

[139] Lee, I., Date, S. V., Adai, A. T. and Marcotte, E. M. [2004], 'A probabilistic functional network of yeast genes.', *Science* **306**, 1555–1558.

[140] Lee, I. T., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., Zeitlinger, J., Jennings, E. G., Murray, H. L., Gordon, D. B., Ren, B., Wyrick, J. J., Tagne, J.-b. and Young, R. a. [2014], 'Transcriptional Regulatory Networks in Saccharomyces cerevisiae', *Science* **799**(2002), 799–804.

[141] Lee, K. P., Piskurewicz, U., Tureckova, V., Strnad, M. and Lopez-Molina, L. [2010], 'A seed coat bedding assay shows that RGL2-dependent release of abscisic acid by the endosperm controls embryo growth in Arabidopsis dormant seeds', *Proceedings of the National Academy of Sciences* **107**(44), 19108–19113.

[142] Legnaioli, T., Cuevas, J. and Mas, P. [2009], 'TOC1 functions as a molecular switch

connecting the circadian clock with plant responses to drought.', *The EMBO journal* **28**(23), 3745–3757.

[143] Leon-Reyes, A., Du, Y., Koornneef, A., Proietti, S., Körbes, A. P., Memelink, J., Pieterse, C. M. J. and Ritsema, T. [2010], 'Ethylene signaling renders the jasmonate response of Arabidopsis insensitive to future suppression by salicylic acid', *Molecular Plant Microbe Interactions* **23**(2), 187–197.

[144] Levi, M., Brusa, P., Chiatante, D. and Sparvoli, E. [1993], 'Cell cycle reactivation in cultured pea embryo axes. Effect of abscisic acid', *In Vitro Cellular & Developmental Biology* **29**(2), 47–50.

[145] Li, Y., Pearl, S. A. and Jackson, S. A. [2015], 'Gene Networks in Plant Biology: Approaches in Reconstruction and Analysis', *Trends in Plant Science* **20**(10), 664–675.

[146] Lima, S., Guo, M. S., Chaba, R., Gross, C. A. and Sauer, R. T. [2013], 'Dual molecular signals mediate the bacterial response to outer-membrane stress', *Science* **340**(6134), 837–842.

[147] Liu, H. T., Gao, F., Li, G. L., Han, J. L., Liu, D. L., Sun, D. Y. and Zhou, R. G. [2008], 'The calmodulin-binding protein kinase 3 is part of heat-shock signal transduction in Arabidopsis thaliana', *The Plant Journal* **55**(5), 760–773.

[148] Liu, J., Mehdi, S., Topping, J., Tarkowski, P. and Lindsey, K. [2010], 'Modelling and experimental analysis of hormonal crosstalk in Arabidopsis', *Molecular Systems Biology* **6**(1), 373.

[149] Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K. and Shinozaki, K. [1998], 'Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in Arabidopsis.', *The Plant Cell* **10**(8), 1391–406.

[150] Lockhart, D. J. and Winzeler, E. A. [2000], 'Genomics, gene expression and DNA arrays', *Nature* **405**(6788), 827–836.

[151] Lorenzo, O., Chico, J. M., Sánchez-Serrano, J. J. and Solano, R. [2004], 'JASMONATE-INSENSITIVE1 encodes a MYC transcription factor essential to discriminate between different jasmonate-regulated defense responses in Arabidopsis', *The Plant Cell* **16**(7), 1938–1950.

[152] Lu, C. and Fedoroff, N. [2000], 'A mutation in the Arabidopsis HYL1 gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin', *The Plant Cell* **12**(12), 2351–2365.

[153] Luscombe, N. M., Austin, S. E., Berman, H. M. and Thornton, J. M. [2000], 'An overview of the structures of protein-DNA complexes.', *Genome biology* **1**(1), 1–37.

[154] Lynch, T., Erickson, B. J. and Finkelstein, R. R. [2012], 'Direct interactions of ABA-insensitive(ABI)-clade protein phosphatase(PP)2Cs with calcium-dependent protein kinases and ABA response element-binding bZIPs may contribute to turning off ABA response', *Plant Molecular Biology* **80**, 647–658.

[155] Ma, S. and Bohnert, H. J. [2007], 'Integration of Arabidopsis thaliana stress-related transcript profiles, promoter structures, and cell-specific expression.', *Genome Biology* **8**(4), R49.

[156] Mahalingam, R. [2015], Consideration of Combined Stress: A Crucial Paradigm for Improving Multiple Stress Tolerance in Plants, *in* 'Combined Stresses in Plants: Physiological, Molecular and Biochemical Aspects', Springer, pp. 1–25.

[157] Maleck, K., Levine, A., Eulgem, T., Morgan, A., Schmid, J., Lawton, K. A., Dangl, J. L. and Dietrich, R. A. [2000], 'The transcriptome of Arabidopsis thaliana during systemic acquired resistance.', *Nature Genetics* **26**(4), 403–410.

[158] Mangan, S. and Alon, U. [2003], 'Structure and function of the feed-forward loop network motif', *Proceedings of the National Academy of Sciences* **100**(21), 1–6.

[159] Marcotte, W. R., Russell, S. H. and Quatrano, R. S. [1989], 'Abscisic acid-responsive sequences from the em gene of wheat.', *The Plant Cell* **1**(10), 969–976.

[160] Mcclelland, J. L. [1986], A General framework for Parallel Distributed Processing, *in* 'Parallel distributed processing: explorations in the microstructure of cognition', MIT Press, pp. 45–76.

[161] McKersie, B. D., Bowley, S. R. and Jones, K. S. [1999], 'Winter Survival of Transgenic Alfalfa Overexpressing Superoxide Dismutase1', *Plant Physiology* **119**(3), 839–848.

[162] Mehta, P., Goyal, S., Long, T., Bassler, B. L. and Wingreen, N. S. [2009], 'Information processing and signal integration in bacterial quorum sensing', *Molecular Systems Biology* **5**(325), 1–11.

[163] Meier, M., Lucchetta, E. M. and Ismagilov, R. F. [2010], 'Chemical stimulation of the Arabidopsis thaliana root using multi-laminar flow on a microfluidic chip', *Lab on a chip* **10**(16), 2147–2153.

[164] Meinke, D. W. [1998], 'Arabidopsis thaliana: A Model Plant for Genome Analysis', *Science* **282**(5389), 662–682.

[165] Merlot, S., Gosti, F., Guerrier, D., Vavasseur, A. and Giraudat, J. [2001], 'The ABI1 and ABI2 protein phosphatases 2C act in a negative feedback regulatory loop of the abscisic acid signalling pathway.', *The Plant Journal* **25**(3), 295–303.

[166] Mitsuda, N. and Ohme-Takagi, M. [2009], 'Functional analysis of transcription factors in arabidopsis', *Plant & Cell Physiology* **50**(7), 1232–1248.

[167] Mittler, R. [2006], 'Abiotic stress, the field environment and stress combination', *Trends in Plant Science* **11**(1), 15–19.

[168] Mittler, R., Finka, A. and Goloubinoff, P. [2012], 'How do plants feel the heat?', *Trends in Biochemical Sciences* **37**(3), 118–125.

[169] Miura, K., Jin, J. B., Lee, J., Yoo, C. Y., Stirm, V., Miura, T., Ashworth, E. N.,

Bressan, R. a., Yun, D.-J. and Hasegawa, P. M. [2007], 'SIZ1-Mediated Sumoylation of ICE1 Controls CBF3/DREB1A Expression and Freezing Tolerance in Arabidopsis', *The Plant Cell* **19**(4), 1403–1414.

[170] Mockler, T. C., Michael, T. P. and Priest, H. D. [2007], 'The Diurnal Project : Diurnal and Circadian Expression Profiling , Model-based Pattern Matching , and Promoter Analysis The Diurnal Project : Diurnal and Circadian Expression Profiling , Model-based Pattern Matching , and Promoter Analysis', *Cold Spring Harbor Symposia on Quantitative Biology* **LXXII**, 353–363.

[171] Mohamed, R., Meilan, R., Ostry, M. E., Michler, C. H. and Strauss, S. H. [2001], 'Bacterio-opsin gene overepression fails to elevate fungal disease resistance in transgenic poplar', *Canadian Journal of Forest Research* **31**(2), 268–275.

[172] Morimoto, K., Mizoi, J., Qin, F., Kim, J. S., Sato, H., Osakabe, Y., Shinozaki, K. and Yamaguchi-Shinozaki, K. [2013], 'Stabilization of Arabidopsis DREB2A is required but not sufficient for the induction of target genes under conditions of stress', *PLoS ONE* **8**(12).

[173] Mou, Z., Fan, W. and Dong, X. [2003], 'Inducers of plant systemic acquired resistance Regulate NPR1 function through redox changes', *Cell* **113**(7), 935–944.

[174] Msanne, J., Lin, J., Stone, J. M. and Awada, T. [2011], 'Characterization of abiotic stress-responsive Arabidopsis thaliana RD29A and RD29B genes and evaluation of transgenes', *Planta* **234**(1), 97–107.

[175] Mundy, J. and Chua, N.-H. [1988], 'Abscisic acid and water-stress induce the expression of a novel rice gene.', *The EMBO Journal* **7**(8), 2279.

[176] Munns, R. and Tester, M. [2008], 'Mechanisms of salinity tolerance.', *Annual Review of Plant Biology* **59**, 651–681.

[177] Mur, L. A. J., Kenton, P., Atzorn, R., Miersch, O. and Wasternack, C. [2006], 'The Outcomes of Concentration-Specific Interactions between Salicylate and Jasmonate Sig-

naling Include Synergy , Antagonism , and Oxidative Stress Leading to Cell Death', **140**(1), 249–262.

[178] Naika, M., Shameer, K., Mathew, O. K., Gowda, R. and Sowdhamini, R. [2013], 'STIFDB2: An updated version of plant stress-responsive transcription factor database with additional stress signals, stress-responsive transcription factor binding sites and stress-responsive genes in arabidopsis and rice', *Plant & Cell Physiology* **54**(2), 1–15.

[179] Nakashima, K., Fujita, Y., Kanamori, N., Katagiri, T., Umezawa, T., Kidokoro, S., Maruyama, K., Yoshida, T., Ishiyama, K., Kobayashi, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. [2009], 'Three Arabidopsis SnRK2 protein kinases, SRK2D/SnRK2.2, SRK2E/SnRK2.6/OST1 and SRK2I/SnRK2.3, involved in ABA signaling are essential for the control of seed development and dormancy.', *Plant & Cell Physiology* **50**(7), 1345–63.

[180] Nakashima, K., Ito, Y. and Yamaguchi-Shinozaki, K. [2009], 'Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses', *Plant Physiology* **149**(1), 88–95.

[181] Nakashima, K. and Yamaguchi-Shinozaki, K. [2006], 'Regulons involved in osmotic stress-responsive and cold stress-responsive gene expression in plants', *Physiologia Plantarum* **126**(1), 62–71.

[182] Narsai, R., Wang, C., Chen, J., Wu, J., Shou, H. and Whelan, J. [2013], 'Antagonistic, overlapping and distinct responses to biotic stress in rice (Oryza sativa) and interactions with abiotic stress', *BMC genomics* **14**(1), 93.

[183] Narusaka, Y., Nakashima, K., Shinwari, Z. K., Sakuma, Y., Furihata, T., Abe, H., Narusaka, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. [2003], 'Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of Arabidopsis rd29A gene in response to dehydration and high-salinity stresses.', *The Plant Journal* **34**(2), 137–48.

[184] Norman-Setterblad, C., Vidal, S. and Palva, E. T. [2000], 'Interacting signal pathways control defense gene expression in Arabidopsis in response to cell wall-degrading enzymes from Erwinia carotovora', *Molecular Plant-Microbe Interactions* **13**(4), 430–438.

[185] Nover, L. [1991], *Heat shock response*, CRC Press.

[186] Nover, L., Scharf, K.-D., Gagliardi, D., Vergne, P., Czarnecka-Verner, E. and Gurley, W. B. [1996], 'The Hsf world: classification and properties of plant heat stress transcription factors', *Cell Stress & Chaperones* **1**(4), 215.

[187] Oerke, E.-C. [2006], 'Crop losses to pests', *The Journal of Agricultural Science* **144**(1), 31–43.

[188] Ohme-Takagi, M. and Shinshi, H. [1995], 'Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element.', *The Plant Cell* **7**(2), 173–182.

[189] Okamoto, M., Kuwahara, A., Seo, M., Kushiro, T., Asami, T., Hirai, N., Kamiya, Y., Koshiba, T. and Nambara, E. [2006], 'CYP707A1 and CYP707A2, which encode abscisic acid 8'-hydroxylases, are indispensable for proper control of seed dormancy and germination in Arabidopsis', *Plant Physiology* **141**, 97–107.

[190] Osakabe, Y., Yamaguchi-Shinozaki, K., Shinozaki, K. and Tran, L.-S. P. [2013], 'Sensing the environment: key roles of membrane-localized kinases in plant perception and response to abiotic stress', *Journal of Experimental Botany* **64**(2), 445–458.

[191] Pääkkönen, E., Vahala, J., Pohjola, M., Holopainen, T. and Kärenlampi, L. [1998], 'Physiological, stomatal and ultrastructural ozone responses in birch (Betula pendula Roth.) are modified by water stress', *Plant, Cell & Environment* **21**(7), 671–684.

[192] Palaniswamy, S. K., James, S., Sun, H., Lamb, R. S., Davuluri, R. V. and Grotewold, E. [2006], 'AGRIS and AtRegNet. a platform to link cis-regulatory elements and transcription factors into regulatory networks.', *Plant Physiology* **140**(3), 818–829.

[193] Pandey, S. P. and Somssich, I. E. [2009], 'The role of WRKY transcription factors in plant immunity', *Plant Physiology* **150**(4), 1648–1655.

[194] Park, C. Y., Lee, J. H., Yoo, J. H., Moon, B. C., Choi, M. S., Kang, Y. H., Lee, S. M., Kim, H. S., Kang, K. Y., Chung, W. S. and Others [2005], 'WRKY group IId transcription factors interact with calmodulin', *FEBS letters* **579**(6), 1545–1550.

[195] Park, S.-Y., Fung, P., Nishimura, N., Jensen, D. R., Fujii, H., Zhao, Y., Lumba, S., Santiago, J., Rodrigues, A., Chow, T.-F. F., Alfred, S. E., Bonetta, D., Finkelstein, R., Provart, N. J., Desveaux, D., Rodriguez, P. L., McCourt, P., Zhu, J.-K., Schroeder, J. I., Volkman, B. F. and Cutler, S. R. [2009], 'Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins.', *Science* **324**(5930), 1068–1071.

[196] Pérez-Rodríguez, P., Riano-Pachon, D. M., Corrêa, L. G. G., Rensing, S. A., Kersten, B. and Mueller-Roeber, B. [2009], 'PlnTFDB: updated content and new features of the plant transcription factor database', *Nucleic Acids Research* p. gkp805.

[197] Pieterse, C. M. J., der Does, D., Zamioudis, C., Leon-Reyes, A. and Van Wees, S. C. M. [2012], 'Hormonal modulation of plant immunity', *Annual Review of Cell and Developmental Biology* **28**, 489–521.

[198] Pieterse, C. M. J., Leon-Reyes, A., der Ent, S. and Van Wees, S. C. M. [2009], 'Networking by small-molecule hormones in plant immunity', *Nature Chemical Biology* **5**(5), 308–316.

[199] Pieterse, C. M. J. and Van Loon, L. C. [1999], 'Salicylic acid-independent plant defence pathways', *Trends in Plant Science* **4**(2), 52–58.

[200] Piquerez, S. J. M., Harvey, S. E., Beynon, J. L. and Ntoukakis, V. [2014], 'Improving crop disease resistance: lessons from research on Arabidopsis and tomato', *Frontiers in Plant Science* **5**, 1–13.

[201] Prasch, C. M. and Sonnewald, U. [2014], 'Signaling events in plants: Stress factors in combination change the picture', *Environmental and Experimental Botany* **114**, 4–14.

[202] Pratt, J. M., Petty, J., Riba-Garcia, I., Robertson, D. H. L., Gaskell, S. J., Oliver, S. G. and Beynon, R. J. [2002], 'Dynamics of protein turnover, a missing dimension in proteomics.', *Molecular & Cellular Proteomics* **1**, 579–591.

[203] Prill, R. J., Iglesias, P. A. and Levchenko, A. [2005], 'Dynamic properties of network motifs contribute to biological network organization', *PLoS Biology* **3**(11), 1881–1892.

[204] Puranik, S., Sahu, P. P., Srivastava, P. S. and Prasad, M. [2012], 'NAC proteins: Regulation and role in stress tolerance', *Trends in Plant Science* **17**(6), 369–381.

[205] Qin, F., Sakuma, Y., Tran, L.-S. P., Maruyama, K., Kidokoro, S., Fujita, Y., Fujita, M., Umezawa, T., Sawano, Y., Miyazono, K.-I., Tanokura, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. [2008], 'Arabidopsis DREB2A-interacting proteins function as RING E3 ligases and negatively regulate plant drought stress-responsive gene expression.', *The Plant Cell* **20**(6), 1693–1707.

[206] Rasmussen, S., Barah, P., Suarez-Rodriguez, M. C., Bressendorff, S., Friis, P., Costantino, P., Bones, A. M., Nielsen, H. B. and Mundy, J. [2013], 'Transcriptome responses to combinations of stresses in Arabidopsis.', *Plant Physiology* **161**(4), 1783–1794.

[207] Ren, H., Gao, Z., Chen, L., Wei, K., Liu, J., Fan, Y., Davies, W. J., Jia, W. and Zhang, J. [2007], 'Dynamic analysis of ABA accumulation in relation to the rate of ABA catabolism in maize tissues under water deficit', *Journal of Experimental Botany* **58**(2), 211–219.

[208] Rhodes, D. and Nadolska-Orczyk, A. [2002], Plant Stress Physiology, *in* 'Encyclopedia of Life Sciences', John Wiley & Sons, pp. 1–7.

[209] Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O. J., Samaha, R. R., Creelman, R., Pilgrim, M., Broun,

P., Zhang, J. Z., Ghandehari, D., Sherman, B. K. and Yu, G. [2000], 'Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes.', *Science* **290**(5499), 2105–2110.

[210] Riechmann, J. L. and Meyerowitz, E. M. [1998], 'The AP2/EREBP family of plant transcription factors', *Biological Chemistry* **379**(6), 633–646.

[211] Rizhsky, L., Liang, H. and Mittler, R. [2002], 'The combined effect of drought stress and heat shock on gene expression in tobacco.', *Plant Physiology* **130**(3), 1143–1151.

[212] Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., Davletova, S. and Mittler, R. [2004], 'When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress.', *Plant Physiology* **134**(4), 1683–1696.

[213] Rosenzweig, C., Iglesias, A., Yang, X. B., Epstein, P. R. and Chivian, E. [2001], 'Climate change and extreme weather events; implications for food production, plant diseases, and pests', *Global Change & Human health* **2**(2), 90–104.

[214] Sakuma, Y., Maruyama, K., Osakabe, Y., Qin, F., Seki, M. and Shinozaki, K. [2006], 'Functional Analysis of an Arabidopsis Transcription Factor , DREB2A , Involved in Drought-Responsive Gene Expression', *The Plant Cell* **18**(5), 1292–1309.

[215] Sakuma, Y., Maruyama, K., Qin, F., Osakabe, Y., Shinozaki, K. and Yamaguchi-Shinozaki, K. [2006], 'Dual function of an Arabidopsis transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression.', *Proceedings of the National Academy of Sciences* **103**(49), 18822–18827.

[216] Sandermann, H. [2006], 'Plant biotechnology: ecological case studies on herbicide resistance', *Trends in Plant Science* **11**(7), 324–328.

[217] Schenk, P. M., Kazan, K., Wilson, I., Anderson, J. P., Richmond, T., Somerville, S. C. and Manners, J. M. [2000], 'Coordinated plant defense responses in Arabidopsis revealed by microarray analysis', *Proceedings of the National Academy of Sciences* **97**(21), 11655–11660.

[218] Schindler, U., Beckmann, H. and Cashmore, a. R. [1992], 'TGA1 and G-box binding factors: two distinct classes of Arabidopsis leucine zipper proteins compete for the G-box-like element TGACGTGG.', *The Plant Cell* **4**(10), 1309–1319.

[219] Schnable, P. S., Page, S. E. E. L., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T. A., Minx, P., Reily, A. D., Courtney, L., Kruchowski, S. S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S. M., Belter, E., Du, F., Kim, K., Abbott, R. M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S. M., Gillam, B., Chen, W., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S. and Kumari, S. [2012], 'The B73 Maize Genome: Complexity, Diversity, and Dynamics', *Science* **326**(5956), 1112–1115.

[220] Schöffl, F., Prandl, R. and Reindl, A. [1998], 'Update on Signal Transduction Regulation of the Heat-Shock Response', *Plant Physiology* **117**(4), 1135–1141.

[221] Schulze, E.-D., Beck, E. and Müller-Hohenstein, K. [2005], *Plant Ecology*, Springer.

[222] Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Taji, T., Yamaguchi-Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K. [2002], 'Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray', *The Plant Journal* **31**(3), 279–292.

[223] Seo, M. and Koshiba, T. [2011], 'Transport of ABA from the site of biosynthesis to

the site of action', *Journal of Plant Research* **124**(4), 501–507.

[224] Seung, D., Risopatron, J. P. M., Jones, B. J. and Marc, J. [2012], 'Circadian clock-dependent gating in ABA signalling networks', *Protoplasma* **249**(3), 445–457.

[225] Shen-Orr, S. S., Milo, R., Mangan, S. and Alon, U. [2002], 'Network motifs in the transcriptional regulation network of Escherichia coli.', *Nature Genetics* **31**(1), 64–8.

[226] Shen, Q. and Ho, T. H. [1995], 'Functional dissection of an abscisic acid (ABA)-inducible gene reveals two independent ABA-responsive complexes each containing a G-box and a novel cis-acting element.', *The Plant Cell* **7**(3), 295–307.

[227] Shinozaki, K. and Yamaguchi-Shinozaki, K. [1997], 'Gene expression and signal transduction in water-stress response', *Plant Physiology* **115**(2), 327.

[228] Singh, K. B., Foley, R. C. and Oñate-Sánchez, L. [2002], 'Transcription factors in plant defense and stress responses', *Current Opinion in Plant Biology* **5**(5), 430–436.

[229] Skriver, K. and Mundy, J. [1990], 'Gene expression in response to abscisic acid and osmotic stress.', *The Plant Cell* **2**(6), 503.

[230] Snedden, W. A. and Fromm, H. [1998], 'Calmodulin, calmodulin-related proteins and plant responses to the environment', *Trends in Plant Science* **3**(8), 299–304.

[231] Somerville, C. and Koornneef, M. [2002], 'A fortunate choice: the history of Arabidopsis as a model plant', *Nature Reviews Genetics* **3**(11), 883–889.

[232] Song, S., Huang, H., Gao, H., Wang, J., Wu, D., Liu, X., Yang, S., Zhai, Q., Li, C., Qi, T. and Others [2014], 'Interaction between MYC2 and ETHYLENE INSENSITIVE3 modulates antagonism between jasmonate and ethylene signaling in Arabidopsis', *The Plant Cell* **26**(1), 263–279.

[233] Steffens, N. O., Galuschka, C., Schindler, M., Bülow, L. and Hehl, R. [2004], 'AthaMap: an online resource for in silico transcription factor binding sites in the Arabidopsis thaliana genome.', *Nucleic Acids Research* **32**(Database issue), D368–D372.

[234] Storozhenko, S., De Pauw, P., Van Montagu, M., Inzé, D. and Kushnir, S. [1998], 'The heat-shock element is a functional component of the Arabidopsis APX1 gene promoter.', *Plant Physiology* **118**(3), 1005–1014.

[235] Strogatz, S. H. [2001], 'Exploring complex networks', *Nature* **410**(6825), 268–276.

[236] Suarez-Rodriguez, M. C., Adams-Phillips, L., Liu, Y., Wang, H., Su, S.-H., Jester, P. J., Zhang, S., Bent, A. F. and Krysan, P. J. [2007], 'MEKK1 is required for flg22-induced MPK4 activation in Arabidopsis plants', *Plant Physiology* **143**(2), 661–669.

[237] Suhayda, C. G., Giannini, J. L., Briskin, D. P. and Shannon, M. C. [1990], 'Electrostatic changes in Lycopersicon esculentum root plasma membrane resulting from salt stress', *Plant Physiology* **93**(2), 471–478.

[238] Suzuki, N., Rivero, R. M., Shulaev, V., Blumwald, E. and Mittler, R. [2014], 'Abiotic and biotic stress combinations', *New Phytologist* **203**(1), 32–43.

[239] Tanaka, R. J. and Kimura, H. [2008], 'Mathematical classification of regulatory logics for compound environmental changes', *Journal of Theoretical Biology* **251**(2), 363–379.

[240] Thaler, J. S., Humphrey, P. T. and Whiteman, N. K. [2012], 'Evolution of jasmonate and salicylate signal crosstalk', *Trends in Plant Science* **17**(5), 260–270.

[241] Thines, B., Katsir, L., Melotto, M., Niu, Y., Mandaokar, A., Liu, G., Nomura, K., He, S. Y., Howe, G. a. and Browse, J. [2007], 'JAZ repressor proteins are targets of the SCF(COI1) complex during jasmonate signalling', *Nature* **448**(7154), 661–665.

[242] Thomma, B. P. H. J., Eggermont, K., Penninckx, I. A. M. A., Mauch-Mani, B., Vogelsang, R., Cammue, B. P. A. and Broekaert, W. F. [1998], 'Separate jasmonate-dependent and salicylate-dependent defense-response pathways in Arabidopsis are essential for resistance to distinct microbial pathogens', *Proceedings of the National Academy of Sciences* **95**(25), 15107–15111.

[243] Thompson, A. J., Jackson, A. C., Parker, R. A., Morpeth, D. R., Burbidge, A. and

Taylor, I. B. [2000], ‘Abscisic acid biosynthesis in tomato: regulation of zeaxanthin epoxidase and 9-cis-epoxycarotenoid dioxygenase mRNAs by light/dark cycles, water stress and abscisic acid’, *Plant molecular biology* **42**(6), 833–845.

[244] Toledo-Ortiz, G., Huq, E. and Quail, P. H. [2003], ‘The Arabidopsis Basic / Helix-Loop-Helix Transcription Factor Family’, *The Plant Cell* **15**(8), 1749–1770.

[245] Tong, Z., Hong, B., Yang, Y., Li, Q., Ma, N., Ma, C. and Gao, J. [2009], ‘Overexpression of two chrysanthemum DgDREB1 group genes causing delayed flowering or dwarfism in Arabidopsis’, *Plant Molecular Biology* **71**(1), 115–129.

[246] Tran, L.-S. P., Nakashima, K., Sakuma, Y., Simpson, S. D., Fujita, Y., Maruyama, K., Fujita, M., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. [2004], ‘Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter.’, *The Plant Cell* **16**(9), 2481–2498.

[247] Turner, J. G., Ellis, C. and Devoto, A. [2002], ‘The jasmonate signal pathway.’, *The Plant Cell* **14**(Suppl 1), S153–S164.

[248] Tuteja, N. [2007], ‘Abscisic acid and abiotic stress signaling’, *Plant Signaling & Behavior* **2**(3), 135–138.

[249] United Nations Department of Economic and Social Affairs [2015], World Population Prospects : The 2015 Revision.

[250] Uno, Y., Furihata, T., Abe, H., Yoshida, R., Shinozaki, K. and Yamaguchi-Shinozaki, K. [2000], ‘Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions.’, *Proceedings of the National Academy of Sciences* **97**(21), 11632–12637.

[251] Varshney, R. K., Hoisington, D. a. and Tyagi, A. K. [2006], ‘Advances in cereal genomics and applications in crop breeding’, *Trends in Biotechnology* **24**(11), 490–499.

[252] Verslues, P. E., Kim, Y.-S. and Zhu, J.-K. [2007], 'Altered ABA, proline and hydrogen peroxide in an Arabidopsis glutamate:glyoxylate aminotransferase mutant.', *Plant Molecular Biology* **64**(1), 205–217.

[253] Vile, D., Pervent, M., Belluau, M., Vasseur, F., Bresson, J., Muller, B., Granier, C. and Simonneau, T. [2012], 'Arabidopsis growth under prolonged high temperature and water deficit: independent or interactive effects?', *Plant, Cell & Environment* **35**(4), 702–718.

[254] Vinocur, B. and Altman, A. [2005], 'Recent advances in engineering plant tolerance to abiotic stress: Achievements and limitations', *Current Opinion in Biotechnology* **16**(2), 123–132.

[255] von Koskull-Döring, P., Scharf, K.-D. and Nover, L. [2007], 'The diversity of plant heat stress transcription factors.', *Trends in Plant Science* **12**(10), 452–457.

[256] Wang, K. L., Li, H. and Ecker, J. R. [2002], 'Ethylene Biosynthesis and Signaling Networks', *The Plant Cell* **14**(Suppl 1), 131–152.

[257] Wang, Q. Y. and Nick, P. [2001], 'Cold acclimation can induce microtubular cold stability in a manner distinct from abscisic acid', *Plant & Cell Physiology* **42**(9), 999–1005.

[258] Wang, W., Vinocur, B. and Altman, A. [2003], 'Plant responses to drought, salinity and extreme temperatures: Towards genetic engineering for stress tolerance', *Planta* **218**(1), 1–14.

[259] Wang, W., Vinocur, B., Shoseyov, O. and 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.

[260] Wasilewska, A., Vlad, F., Sirichandra, C., Redko, Y., Jammes, F., Valon, C., dit Frey, N. F. and Leung, J. [2008], 'An update on abscisic acid signaling in plants and more…', *Molecular plant* **1**(2), 198–217.

[261] Wasternack, C. and Hause, B. [2013], 'Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany', *Annals of Botany* **111**(6), 1021–1058.

[262] Waters, C. M. and Bassler, B. L. [2005], 'QUORUM SENSING: Cell-to-Cell Communication in Bacteria', *Annual Review of Cell and Developmental Biology* **21**(1), 319–346.

[263] Windsor, M. L., Milborrow, B. V. and McFarlane, I. J. [1992], 'The Uptake of (+)-S- and (-)-R-Abscisic Acid by Suspension Culture Cells of Hopbush (Dodonaea viscosa).', *Plant Physiology* **100**, 54–62.

[264] Xiong, L., Ishitani, M. and Zhu, J. K. [1999], 'Interaction of osmotic stress, temperature, and abscisic acid in the regulation of gene expression in Arabidopsis.', *Plant Physiology* **119**(1), 205–212.

[265] Xiong, L., Lee, B., Ishitani, M., Lee, H., Zhang, C. and Zhu, J. K. [2001], 'FIERY1 encoding an inositol polyphosphate 1-phosphatase is a negative regulator of abscisic acid and stress signaling in Arabidopsis', *Genes & Development* **15**(15), 1971–1984.

[266] Xiong, L., Lee, H., Ishitani, M. and Zhu, J.-K. [2002], 'Regulation of Osmotic Stress-responsive Gene Expression by the LOS6/ABA1 Locus in Arabidopsis', *Journal of Biological Chemistry* **277**(10), 8588–8596.

[267] Xiong, L., Schumaker, K. and Zhu, J. [2002], 'Cell signaling during cold, drought, and salt stress', *The Plant Cell* **14**(Suppl 1), S165–S183.

[268] Xiong, L. and Zhu, J.-k. [2003], 'Regulation of Abscisic Acid Biosynthesis', *Plant Physiology* **133**(9), 29–36.

[269] Yamaguchi-Shinozaki, K. and Shinozaki, K. [1993], 'Characterization of the expression of a desiccation-responsive rd29 gene of Arabidopsis thaliana and analysis of its promoter in transgenic plants.', *Molecular & General Genetics* **236**, 331–340.

[270] Yamaguchi-Shinozaki, K. and Shinozaki, K. [1994], 'A novel cis-acting element in an

Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress.', *The Plant Cell* **6**(2), 251–264.

[271] Yamaguchi-Shinozaki, K. and Shinozaki, K. [2005], 'Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters.', *Trends in plant science* **10**(2), 88–94.

[272] Yamaguchi-Shinozaki, K. and Shinozaki, K. [2006], 'Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses.', *Annual Review of Plant Biology* **57**, 781–803.

[273] Yasuda, M., Ishikawa, A., Jikumaru, Y., Seki, M., Umezawa, T., Asami, T., Maruyama-Nakashita, A., Kudo, T., Shinozaki, K., Yoshida, S. and Others [2008], 'Antagonistic interaction between systemic acquired resistance and the abscisic acid–mediated abiotic stress response in Arabidopsis', *The Plant Cell* **20**(6), 1678–1692.

[274] Yoshida, R., Hobo, T., Ichimura, K., Mizoguchi, T., Takahashi, F., Aronso, J., Ecker, J. R. and Shinozaki, K. [2002], 'ABA-activated SnRK2 protein kinase is required for dehydration stress signaling in Arabidopsis', *Plant and Cell Physiology* **43**(12), 1473–1483.

[275] Yoshida, T., Fujita, Y., Maruyama, K., Mogami, J., Todaka, D., Shinozaki, K. and Yamaguchi-Shinozaki, K. [2014], 'Four Arabidopsis AREB/ABF transcription factors function predominantly in gene expression downstream of SnRK2 kinases in abscisic acid signalling in response to osmotic stress.', *Plant, Cell & Environment* **38**(1), 35–49.

[276] Yu, D., Chen, C. and Chen, Z. [2001], 'Evidence for an important role of WRKY DNA binding proteins in the regulation of NPR1 gene expression.', *The Plant Cell* **13**(7), 1527–1540.

[277] Yu, J., Smith, V. A., Wang, P. P., Hartemink, A. J. and Jarvis, E. D. [2004], 'Advances to Bayesian network inference for generating causal networks from observational biological data', *Bioinformatics* **20**(18), 3594–3603.

[278] Zarka, D. G., Vogel, J. T., Cook, D. and Thomashow, M. F. [2003], 'Cold Induction of Arabidopsis', *Plant Physiology* **133**(2), 910–918.

[279] Zhang, W., Zhou, R.-G., Gao, Y.-J., Zheng, S.-Z., Xu, P., Zhang, S.-Q. and Sun, D.-Y. [2009], 'Molecular and genetic evidence for the key role of AtCaM3 in heat-shock signal transduction in Arabidopsis.', *Plant Physiology* **149**(4), 1773–1784.

[280] Zhu, J.-K. [2000], 'Genetic analysis of plant salt tolerance using Arabidopsis', *Plant Physiology* **124**(3), 941–948.

[281] Zhu, S.-Y., Yu, X.-C., Wang, X.-J., Zhao, R., Li, Y., Fan, R.-C., Shang, Y., Du, S.-Y., Wang, X.-F., Wu, F.-Q. and Others [2007], 'Two calcium-dependent protein kinases, CPK4 and CPK11, regulate abscisic acid signal transduction in Arabidopsis', *The Plant Cell* **19**(10), 3019–3036.

[282] Zhu, Y., Qian, W. and Hua, J. [2010], 'Temperature modulates plant defense responses through NB-LRR proteins.', *PLoS pathogens* **6**(4), e1000844.

[283] Zhu, Z. [2014], 'Molecular basis for jasmonate and ethylene signal interactions in Arabidopsis', *Journal of Experimental Botany* **65**(20), 5743–5748.