THE UNIVERSITY OF WARWICK

# Library Declaration and Deposit Agreement

## 1. STUDENT DETAILS

*Please complete the following:*

Full name: ……

University ID number: ……

## 2. THESIS DEPOSIT

2.1 I understand that under my registration at the University, I am required to deposit my thesis with the University in BOTH hard copy and in digital format. The digital version should normally be saved as a single pdf file.

2.2 The hard copy will be housed in the University Library. The digital version will be deposited in the University's Institutional Repository (WRAP). Unless otherwise indicated (see 2.3 below) this will be made openly accessible on the Internet and will be supplied to the British Library to be made available online via its Electronic Theses Online Service (EThOS) service.
[At present, theses submitted for a Master's degree by Research (MA, MSc, LLM, MS or MMedSci) are not being deposited in WRAP and not being made available via EthOS. This may change in future.]

2.3 In exceptional circumstances, the Chair of the Board of Graduate Studies may grant permission for an embargo to be placed on public access to the hard copy thesis for a limited period. It is also possible to apply separately for an embargo on the digital version. (Further information is available in the *Guide to Examinations for Higher Degrees by Research*.)

2.4 *If you are depositing a thesis for a Master's degree by Research, please complete section (a) below. For all other research degrees, please complete both sections (a) and (b) below:*

(a) Hard Copy

I hereby deposit a hard copy of my thesis in the University Library to be made publicly available to readers (please delete as appropriate) EITHER immediately OR after an embargo period of ……..... months/years as agreed by the Chair of the Board of Graduate Studies.

I agree that my thesis may be photocopied. YES / NO (*Please delete as appropriate*)

(b) Digital Copy

I hereby deposit a digital copy of my thesis to be held in WRAP and made available via EThOS.

Please choose one of the following options:

EITHER My thesis can be made publicly available online. YES / NO (*Please delete as appropriate*)

OR My thesis can be made publicly available only after…..[date] (Please give date)
YES / NO (*Please delete as appropriate*)

OR My full thesis cannot be made publicly available online but I am submitting a separately identified additional, abridged version that can be made available online.
YES / NO (*Please delete as appropriate*)

OR My thesis cannot be made publicly available online. YES / NO (*Please delete as appropriate*)

## 3. GRANTING OF NON-EXCLUSIVE RIGHTS

Whether I deposit my Work personally or through an assistant or other agent, I agree to the following:

Rights granted to the University of Warwick and the British Library and the user of the thesis through this agreement are non-exclusive. I retain all rights in the thesis in its present version or future versions. I agree that the institutional repository administrators and the British Library or their agents may, without changing content, digitise and migrate the thesis to any medium or format for the purpose of future preservation and accessibility.

## 4. DECLARATIONS

(a) I DECLARE THAT:

- I am the author and owner of the copyright in the thesis and/or I have the authority of the authors and owners of the copyright in the thesis to make this agreement. Reproduction of any part of this thesis for teaching or in academic or other forms of publication is subject to the normal limitations on the use of copyrighted materials and to the proper and full acknowledgement of its source.

- The digital version of the thesis I am supplying is the same version as the final, hard-bound copy submitted in completion of my degree, once any minor corrections have been completed.

- I have exercised reasonable care to ensure that the thesis is original, and does not to the best of my knowledge break any UK law or other Intellectual Property Right, or contain any confidential material.

- I understand that, through the medium of the Internet, files will be available to automated agents, and may be searched and copied by, for example, text mining and plagiarism detection software.

(b) IF I HAVE AGREED (in Section 2 above) TO MAKE MY THESIS PUBLICLY AVAILABLE DIGITALLY, I ALSO DECLARE THAT:

- I grant the University of Warwick and the British Library a licence to make available on the Internet the thesis in digitised format through the Institutional Repository and through the British Library via the EThOS service.

- If my thesis does include any substantial subsidiary material owned by third-party copyright holders, I have sought and obtained permission to include it in any version of my thesis available in digital format and that this permission encompasses the rights that I have granted to the University of Warwick and to the British Library.

## 5. LEGAL INFRINGEMENTS

I understand that neither the University of Warwick nor the British Library have any obligation to take legal action on behalf of myself, or other rights holders, in the event of infringement of intellectual property rights, breach of contract or of any other right, in the thesis.

---

*Please sign this agreement and return it to the Graduate School Office when you submit your thesis.*

Student's signature: .....…… Date: .....

AUTHOR: **Johanna Rhodes**

DEGREE: **PhD**

TITLE: **Identifying gene regulatory networks common to multiple plant stress responses**

DATE OF DEPOSIT: .....

I **agree** that this thesis shall be available in accordance with the regulations governing the University of Warwicktheses.

I **agree** that the summary of this thesis may be submitted for publication.

I **agree** that the thesis may be photocopied (single copies for study purposes only).

Theses with no restriction on photocopying will also be made available to the British Library for microfilming. The British Library may supply copies to individuals or libraries. subject to a statement from them that the copy is supplied for non-publishing purposes. All copies supplied by the British Library will carry the following statement:

> "Attention is drawn to the fact that the copyright of this thesis rests with its author. This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author's written consent."

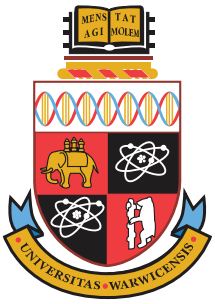
AUTHOR'S SIGNATURE: .....

---

**USER DECLARATION**

1. I undertake not to quote or make use of any information from this thesis without making acknowledgement to the author.
2. I further undertake to allow no-one else to use this thesis while it is in my care.

| DATE | SIGNATURE | ADDRESS |
|---|---|---|
| ..... | ..... | ..... |
| ..... | ..... | ..... |
| ..... | ..... | ..... |
| ..... | ..... | ..... |
| ..... | ..... | ..... |




# Identifying gene regulatory networks common to multiple plant stress responses

by

**Johanna Rhodes**

**Thesis**

Submitted to the University of Warwick

for the degree of

**Doctor of Philosophy**

*Supervisor:* Dr. Katherine Denby

Systems Biology Doctoral Training Centre

September 2012

SYSTEMS BIOLOGY DOCTORAL TRAINING CENTRE

THE UNIVERSITY OF WARWICK

# Contents

# List of Tables

# List of Figures

# Acknowledgements

I would like to acknowledge the following people, without whom this PhD would not have been possible. I am grateful to everyone mentioned here.

Firstly, I wish to acknowledge my supervisor Dr. Katherine Denby, for her enthusiastic supervision. Thank you for pushing me in the right direction! Secondly, to Dr. Sascha Ott, to whom I am eternally grateful for all the advice given regarding the theory side of my PhD. I would also like to acknowledge my advisory committee, Dr. Jens Steinbrenner and Dr. Miriam Gifford, for being incredibly supportive and always having an open door. Many thanks to Dr. Claire Hill, without whom I would still be lost in the laboratory cloning! Thank you for giving me peace of mind.

Also, many thanks to Dr. Steven Kiddle for many a conversation in the laboratory and for also giving me inspiration to start Wigwams. A big thank you to Dr. Stuart McHattie for driving Steve and myself to Scotland in the snow! Many thanks to Peijun Zhang, Dr. Oliver Windram, Dr. Priya Madhou, Alison Jackson, Emily Breeze, Dr. Chris Penfold, Dr. Dafyd Jenkins, Dr. Justyna Prusinka, Emma Cooke, Aleksey Jironken, Dr. Vicky Buchanen-Wollason, Dr. Jim Beynon and the rest of the PRESTA group. A heartfelt thank you to Laura Lewis for lots of emotional support! And thank you to my office friends, Sarah Harvey, Tony Carter, Charlotte Nellist, Femi Akande, Mato Lagator, David Schafer, Gemma Hough, Ailidh Woodcock, Oliver Smith, Heather Riley and Richard Tyson.

Finally I wish to acknowledge and thank my family, without whom I would never have made it so far. To my parents: my mother, Susan, thank you for always believing in me and sacrificing so much so I could get here; to my dad, Michael, thank you for being the voice of reason. To my grandad, Frank, thank you for your unwavering support and always having faith in me. To my Uncle Peter, thank you for pushing me to do my best. To my Uncle Harry, thank you for many

years of maths tuition! To my brother Nick, thank you for letting me be a part of your life. A big thank you to my husband Anthony: thank you for putting up with me in my most stressful hours.

Ultimately, this thesis is dedicated to three people who I owe so much, but unfortunately are no longer with me. I wish you were here. To my nan, Dorothy: thank you for telling you were proud of me, even when I felt like I was letting you down. To my Auntie Marion: together with Uncle Harry, you have had a big impact on my life, and without the two of you I would never have gone to university. Thank you for always having an open door and a shoulder to cry on. To my other dad, David: I always strive to make you proud of me and hope that you are. I miss you all every day.

This thesis was written in LaTeX $2_\varepsilon$

# Declaration

This thesis is presented in accordance with the regulations for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree except where otherwise stated. The work in this thesis has been undertaken by myself except where otherwise stated.

# Abstract

Stress responses in plants can be defined as a change that affects the homeostasis of pathways, resulting in a phenotype that may or may not be visible to the human eye, affecting the fitness of the plant. Crosstalk is believed to be the shared components of pathways of networks, and is widespread in plants, as shown by examples of crosstalk between transcriptional regulation pathways, and hormone signalling.

Crosstalk between stress responses is believed to exist, particularly crosstalk within the responses to biotic stress, and within the responses to abiotic stress. Certain hormone pathways are known to be involved in the crosstalk between the responses to both biotic and abiotic stresses, and can confer immunity or tolerance of *Arabidopsis thaliana* to these stresses. Transcriptional regulation has also been identified as an important factor in controlling tolerance and resistance to stresses.

In this thesis, networks of regulation mediating the response to multiple stresses are studied. Firstly, co-regulation was predicted for genes differentially expressed in two or more stresses by development of a novel multi-clustering approach, Wigwams Identifies Genes Working Across Multiple Stresses (Wigwams). This approach finds groups of genes whose expression is correlated within stresses, but also identifies a strong statistical link between subsets of stresses. Wigwams identifies the known co-expression of genes encoding enzymes of metabolic and flavonoid biosynthesis pathways, and predicts novels clusters of co-expressed genes. By hypothesising that by being co-expressed could also infer that the genes are co-regulated, promoter motif analysis and modelling provides information for potential upstream regulators.

The context-free regulation of groups of co-expressed genes, or potential regulons, was explored using models generated by modelling techniques, in order to generate a quantitative model of transcriptional regulation during the response to *B. cinerea*, *P. syringae* pv. tomato DC3000 and senescence. This model was subsequently validated and extended by experimental techniques, using Yeast 1-Hybrid to investigate the protein-DNA interactions, and also microarrays. Analysis of mutants and plants overexpressing a predicted regulator, *Rap2.6L*, by gene expression analysis identified a number of potential regulon members as downstream targets.

*Rap2.6L* was identified as an indirect regulator of the transcription factor members of three potential regulons co-expressed in the stresses *B. cinerea*, *P. syringae* pv. tomato DC3000 and long day senescence, allowing the confirmation of a predicted gene regulatory network operating in multiple stress responses.

# Abbreviations

| | |
|---|---|
| 3AT | 3-amino-1,2,3-triazole |
| $A_{340nm}$ | Absorbance at wavelength 340nm |
| ABA | Abscisic acid |
| ABRE | Abscisic acid responsive element |
| AGI | Arabidopsis genome identifier |
| ANOVA | Analysis of variance |
| APPLES | Analysis of Plant Promoter-Linked Elements |
| aRNA | Anti-sense ribonucleic acid |
| BATS | Bayesian Analysis of Time Series |
| *B. cinerea* | *Botrytis cinerea* |
| bHLH | Basic helix-loop-helix |
| BiNGO | Biological Network Gene Ontology |
| bp | Base pair |
| bZIP | Basic leucine zipper |
| CATMA | Complete Arabidopsis Transcriptome MicroArray |
| CBF | C-repeat/dehydration responsive element binding factor |
| cDNA | Complementary deoxyribonucleic acid |

| | |
|---|---|
| ChIP | Chromatin Immunoprecipitation |
| cm | Centimetres |
| $CO_2$ | Carbon dioxide |
| Col-0 | Columbia-0 |
| CRT | C-repeat |
| DAS | Days after sowing |
| DC3000 | *Pseudomonas syringae* pathovar *tomato* DC3000 |
| dd | Double distilled |
| DEPC | Diethylpyrocarbonate |
| DNA | Deoxyribonucleic acid |
| DNase | Deoxyribonuclease |
| DO | Drop out |
| DRE | Dehydration responsive element |
| DREB | Dehydration responsive element binding |
| *E. coli* | *Escherichia coli* |
| EDA | Edge attribute file |
| EDTA | Ethylenediaminetetraacetic acid |
| ERF | Ethylene response factor |
| EtBr | Ethidium bromide |
| EtOH | Ethanol |
| ET | Ethylene |
| FDR | False-discovery-rate |
| g | Gram |

| | |
|---|---|
| *g* | Centrifugal force (relative to gravitational pull) |
| GO | Gene ontology |
| GP2S | Gaussian process 2 sample |
| GRN | Gene regulatory network |
| GST | Gene-specific sequence tag |
| $H_2O$ | Water |
| His | Histidine |
| h.p.i. | Hours post infection |
| hrpA | *Pseudomonas syringae* Hrp pili protein A |
| HSF | Heat shock factor |
| HSP | Heat shock protein |
| JA | Jasmonic acid |
| kb | Kilobase |
| KCl | Potassium Chloride |
| L | Litre |
| LB | Lysogeny broth |
| Leu | Leucine |
| LiAc | Lithium acetate |
| $\mu$g | Microgram |
| $\mu$L | Microlitre |
| $\mu$M | Micromolar |
| M | Molar (moles per litre) |
| m | Metre |

| | |
|---|---|
| mL | Millilitre |
| mM | Millimolar |
| mg | Milligram |
| $MgCl_2$ | Magnesium Chloride |
| mm | Millimetres |
| MH-VBSSM | Metropolis-Hastings Variational Bayesian State Space Modelling |
| MAANOVA | MicroArray ANalysis Of VAriance |
| MCMC | Markov Chain Monte Carlo |
| MOPS | 3-(N-morpholino)propanesulfonic acid |
| MTC | Multiple testing correction |
| NaCl | Sodium Chloride |
| nm | Nanometres |
| NaOH | Sodium hydroxide |
| $OD_{600}$ | Optical density at 600nm |
| ORF | Open reading frame |
| PCR | Polymerase chain reaction |
| PEG | Polyethylene glycol |
| pH | $-\log_{10}[H^+]$ |
| ppm | Parts per million |
| PRESTA | Plant Responses to Environmental STress in Arabidopsis |
| PSSM | Position Specific Scoring Matrix |
| *P. syringae* | *Pseudomonas syringae* |
| qPCR | Quantitative polymerase chain reaction |

| | |
|---|---|
| RGB | Red/green/blue values |
| RNA | Ribonucleic acid |
| ROS | Reactive oxygen species |
| mRNA | Messenger ribonucleic acid |
| SA | Salicylic acid |
| SAG | Senescence associated gene |
| *S. cerevisiae* | *Saccharomyces cerevisiae* |
| SD | Synthetic defined (base media for yeast) |
| SD-LT | Synthetic defined minus leucine and tryptophan |
| SD-LTH | Synthetic defined minus leucine, tryptophan and histidine |
| SDS | Sodium dodecyl sulphate |
| SOC media | Super Optimal broth with Catabolite repression |
| SSM | State space model |
| TSS | Transcriptional start site |
| T-DNA | Transfer DNA |
| Tris | Tris(hydroxymethyl)aminomethane |
| Trp | Tryptophan |
| tRNA | Total RNA |
| UV | Ultraviolet |
| V | Volts |
| VBSSM | Variational Bayesian State Space Modelling |
| WS | Wassilewskija |
| W | Watts |

| | |
|---|---|
| Wigwams | Wigwams Identifies Genes Working Across Multiple Stresses |
| Y1H | Yeast-1-hybrid |
| Y2H | Yeast-2-hybrid |
| YPDA | Yeast peptone dextrose adenine |

# Chapter 1

# Introduction

## 1.1 The model plant organism *A. thaliana*

*Arabidopsis thaliana* is the model organism for genetic research of plants, and occasionally, other eukaryotic organisms (200). The move of researchers to study this organism gained speed when the analysis of the genomic sequence of *A. thaliana* was published in 2000 (14). The main advantages to using *A. thaliana* for genomic analysis are its short generation time, small size, small genome, and the ability to grow the plants in many different environments, such as petri dishes, greenhouses or under fluorescent lights in the laboratory (200). With the development of transgenic lines carrying T-DNA insertions from *Agrobacterium tumefaciens* (56), using mutant lines to study how the absence of a gene effects a biological system has proved to be straightforward.

Using *A. thaliana* as a model organism can allow sequence comparisons which would be beneficial for genetic analysis in commercially important crop species. However, due to the diploid nature of the *A. thaliana* genome, and given most staple crop species are monocots, mainly the 'cereal' wheat, rice and maize, comparative genomics may be complex (207). Despite this, there is a general belief that beneath the variation, there is a genetic, developmental and physiological structure that is intrinsic and conserved amongst all plants, which can be understood when studying any plant species (91).

Findings from studying processes in *A. thaliana* have profound relevance to processes such as tolerance to environmental stress and disease, providing hypotheses on the functions of homologs of

candidate genes in commercially important plants (91). Although the tomato is an established model organism for studying pathogen infection, the availability of the complete *A. thaliana* genome allowed Mysore *et al.* to investigate functional analysis of tomato genes using map-based cloning, on the assumption of chromosomal synteny between the two plant species (215). Several articles explore using *A. thaliana* for genetic analysis of stress responses as an alternative to crop species, as there is a lack of field and laboratory screening tests, and physiological and molecular markers for understanding stress responses are available for crop species (345, 111, 312). These failures in crop species make *A. thaliana* an attractive option for analysis. Much research has shown that stress responsive genes found in crop species have isologs or paralogs in *A. thaliana* (345): Wilkinson *et al.* observed that *A. thaliana* possessed an identical ethylene receptor which usually mediates fruit-ripening response in tomatoes (320). Zhang and Blumwald also noted that increased expression of the membrane $Na^+/H^+$ antiporter, *AtNHX1*, resulted in increased salt tolerance in *A. thaliana* (340). Ohta *et al.* furthered this work to show that the rice ortholog of *AtNHX1* increased the tolerance of *Oryza sativa* rice plants when exposed to salt (221).

### 1.1.1 Motivation for studying plant defence responses

Crop yields have increased throughout the 20th century, mainly due to the introduction of new farming practices and cultivars (204). The Broadbalk classical experiments at Rothamsted, which started in the late part of the 19th century, demonstrated an increase in the yield of wheat from 1940 due to the introduction of herbicides and fungicides. These practices are now common-place in farming (204), and an increase in crop yields has brought about greater food security.

However, with greater availability of food comes an increase in population: the current world population stands at approximately 7.038 billion individuals (United States Census Bureau), which is estimated to increase to 8.3 billion individuals by 2030 (United Nations). The availability of food therefore is likely to become a limiting factor, and due to the increase in demand for food, prices will rise. The ability to produce larger quantities of food, therefore, depend on a variety of crops that can produce higher yields, availability of arable land, environmental conditions and the prevalence of diseases that can affect crop yield.

Abiotic stresses, such as salt, cold and drought stress can significantly affect crop yield. Indeed, drought stress is responsible for losses of up to 60% in cereal crops worldwide (45): 15 million

$km^2$ of land surface is dedicated to growing crops (239), however, only 16% of this area has proper irrigation (264). Therefore, drought stress in the remaining 84% is inevitable. The prevalence of drought stress is likely to increase further due to climate change and competition for water supplies, making the demand for drought tolerant crop species higher.

Pathogen infection by fungi and bacteria is responsible for 16% potential yield loss in commercially important crops worldwide annually (220). Infection of plum trees by *Pseudomonas syringae* pathovar (pv.) *syringae* has been reported to cause up to 30% mortality in Germany alone (123). *P. syringae* is particularly problematic for crops, due to the large number of pathovars within the *syringae* species that are pathogenic to over 180 plant species (42). Each pathovar has a high level of variation among strains and a broad host range, with a capacity to cause significant yield loss to many crop species (148).

*Botrytis cinerea*, the necrotrophic fungal pathogen responsible for grey mould, has a host range of over 200 crop species worldwide (300, 321). The mode of infection of young tissues of crop plants, where it stays dormant initially, followed by the degradation of tissues post-harvest is catastrophic, causing massive financial losses (321). *B. cinerea* is particularly difficult to control, due to the multiple modes of attack used against the host (300). Growth of *B. cinerea* was previously controlled by fungicides: however, resistant strains of *B. cinerea* have been isolated, which is though to be due to the high genetic variability of *B. cinerea* (171, 170, 162). Therefore, a grasp of the host mechanisms employed during infection is essential.

For the reasons mentioned here, it is important to investigate the mechanisms of stress responses in plants: traditional methods of farm management, such as the use of pesticides and herbicides are expensive, and are being met with increasing resistance from target pest species. The changing climate, changing geographical location of pathogens and increased demand from a growing human population also create extra pressure to tackle abiotic stresses. Due to the sessile nature of plants, they are also often met with multiple stresses, whereas current research has focused mainly on plant responses and adaptations to single stresses. Plants are often subject to multiple stresses at once: plants under high light intensities are more prone to dehydration or extreme changes in temperature (235).

### 1.1.2 Motivation for studying the role of gene regulatory networks in plant defence

Due to the recent advances in 'omics' technologies, researchers have realised that stress responses are not controlled solely by single genes and gene products, but by complex regulatory networks controlling gene expression (295). Integrated transcriptome, genome-wide transcription factor binding, and proteome analyses have been used to infer functional interactions between genes and proteins (211), in order to formulate gene regulatory networks underlying important biological processes, such as stress responses. Researchers have also begun to understand that regulatory networks involved in plant stress responses are complex, and require genome-scale analyses to elucidate them.

Many stresses affecting crop plants are under transcriptional control (341). Therefore, transcription factors provide ideal targets for research in *A. thaliana*, with a view to understanding the roles of transcription factors in the regulatory network controlling stress tolerance and resistance (64). Research into translating knowledge gained in *A. thaliana* into crop species has been applied to rice, maize and barley (139, 79, 269), allowing the genetic engineering of regulatory networks in crops.

Dembinsky *et al*. generated a map of global gene expression of the *A. thaliana* root using fluoresecence-activated cell sorting (FACS) of different root cell type via protoplasting cell-type specific promoter fused to GFP. Different cell types were analysed by microarray hybridisation. These experiments provided candidate genes involved in pericycle specification, which, with the identification of homologs in maize, enables the analysis of a network of genes involved in pericycle specification and lateral root initiation (79).

Sreenivasulu *et al*. used a novel way to translate research from *A. thaliana* into barley crop species: the *A. thaliana* data evaluation tools MapMan (297) and PageMan (298), which were developed to map *A. thaliana* transcriptome data into functional categories, to analyse time course data into functional gene groups and to map functional categories onto pathways, were adapted to map barley transcriptome data (269).

More recently, investigations into the architectural structure of the root of *A. thaliana* has led to advancements in increasing the lateral root branching, and therefore increase the surface area of the root system, in maize and rice crop species (267): genes in maize *Zea mays* and rice *Oryza*

*sativa* (*rocs* and *arl1/crl1* respectively) were found to be closely related to the *A. thaliana* genes *LBD16* and *LBD29*, whose functions are involved in lateral root branching (222, 263). However, recent focus has shifted to studying more complex biological organisation and processes within *A. thaliana* (165): this approach has been chosen due to the realisation that the relationship between a genotype and phenotype is not linear, and phenotypes are often dependant on environmental conditions. Therefore, results for certain genes seen in *A. thaliana* will not necessarily translate to the field because of our lack of understanding of the complexity within plants.

## 1.2 Crosstalk between multiple plant defence responses

Crosstalk, whereby components of different signalling pathways can influence or interact with each other (41), is thought to provide the plant with the necessary regulatory mechanisms to adapt to multiple changes in their environment (159). It is favourable for plants to allow crosstalk between multiple response pathways, as defence is costly and reduces plant fitness (120). Therefore, plants have evolved the ability to induce responses only in the presence of stress, and to promote crosstalk between signalling pathways (280). In this section, the aspects of crosstalk between phytohormone signalling pathways, and known gene regulatory network employed in multiple stress responses will be discussed. Firstly, however, a brief overview of the plant defence to pathogens and how there is basis for crosstalk in the biotic stress responses, will be given.

### 1.2.1 The plant defence response against pathogens

Plants, unlike humans, rely solely on their innate immune system i.e. they do not have an adaptive immune system, which generates antibodies to fight infection. Plants defend themselves against pathogen infections (such as viruses, bacteria, fungi and oomycetes) using their basal defence response (26). However, plants have evolved a sophisticated immune system comprising of a two-branched response: one branch makes use of transmembrane pattern recognition receptors (PRRs); the other branch uses protein products encoded by *R*-genes (142).

PRRs recognise microbial- or pathogen-associated molecular patterns (MAMPs and PAMPs respectively), a consequence of which is PAMP triggered immunity (PTI) in the plant. Bacterial flagellin, which is decognised by the plants as a PAMP, is capable of triggering PTI in a variety of plants (110). Previous research has shown that application of the peptide flg22, which represents

flagellin, is capable of inducing defence-related genes and trigger resistance to pathogenic bacteria, such as *P. syringae* in *A. thaliana* (347).

Pathogens have evolved ways to suppress host defences by secreting effectors which dampen PTI, causing effector triggered susceptibility (ETS) (26) which seeks to suppress PTI. Some bacterial, fungal and oomycete species are known to suppress effectors, such as *P. syringae* and *Phytophthora infestans* (115). *P. syringae* secretes effectors which aim to suppress the MAPK signalling pathway, thereby suppressing the expression of genes required for plant immunity (40).

The plant *R* genes offset this attempt of the pathogen by recognising the effector, thereby causing effector-triggered immunity (ETI) (142). This 'arms race' between plant and pathogen is known as the 'zigzag model' (142). Activation of *R*-genes by the recognition of a pathogen effector signals crosstalk between different response pathways (108), particularly the salicylic acid (SA), jasmonic acid (JA) and ethylene (ET) pathways: for example, ETI represses salicylic acid-dependant cell death in cells adjacent to sites of infection (290).

This initial defence response described here illustrates the mode of pathogen detection undergone by the plant. The subsequent activation of *R*-genes leads to the induction of different response pathways under the control of various phytohormone signalling pathways. Plants are known to co-ordinate their defence response to many different pathogens via crosstalk between these signalling pathways (164). It is this crosstalk between signalling pathways that could provide initial candidate targets for genes working across biotic stresses. Previous genomic studies have shown comparable responses of plants to pathogens and environmental stresses, therefore suggesting crosstalk is common between biotic stress and abiotic stress responses (240). For example, Reymond *et al.* carried out a set of time series microarray experiments under mechanical stress identified differential expression of *PR1*, *PR2* and *PR3* genes (242) which are known to be involved in the defence response (107).

### 1.2.2 Crosstalk of phytohormone signal transduction in the role of multiple plant stress responses

As mentioned in the previously crosstalk exists within the biotic stress response in the form of hormone signalling pathways. Plant hormones (also referred to as phytohormones) are chemicals that regulate a number of processes in plants, such as growth and development, but also play an

important role in the defence response to not only pathogen attack, but all insect herbivory, drought cold and heat stress (92). Multiple hormone pathways can interact to form regulatory networks, which ultimately influence a plants defence response to stress (77). Although many hormone pathways are capable of crosstalk, the genetic interactions between the SA, JA and ET pathways are the most studied, and hence will be the main topic in this section. Therefore, the study of crosstalk of phtyohormone signalling pathways, and the genes involved in these processes, provide a logical starting point for identifying a gene regulatory network capable of being expressed in multiple stress responses.

The JA signalling pathway is vital for the plant response to wounding (such as that caused by insects), but is also implemented in the response to biotic and abiotic stress responses, such as osmotic stress and drought (92). During the stress response to pathogens, namely necrotrophs such as *B. cinerea*, JA signalling causes the expression of two groups of genes: genes regulated by *MYC2*, a MYC transcription factor, which suppress resistance to pathogens (85); and genes not regulated by *MYC2*, which promote resistance to pathogens (107). The response to abiotic stress, however, relies heavily on the *COI1* complex, with *Skp1*, *Cul1* and *RBX1* (102, 107).

Levels of SA increase in the plant following pathogen infection. The expression of pathogenesis-related (PR) genes are dependent on SA (164), thus rendering SA as an important mediator of the plant response to pathogens. The PR gene *PR1* is an important component of the SA signalling pathway, which confers resistance to the bacterial pathogen *P. syringae*, by limiting the growth of this pathogen (107). Other genes important in the SA signalling pathway include *EDS1* (1), *PAD4* (344), *EDS5* (247), *SID2* (217) and NPR1 (109, 256).

The ET signalling pathway has roles in senescence (55), programmed cell death (193), and also in the detection and initial response to stress and pathogens (311). Ethylene is recognised by a family of receptors, which include *ETR1/ETR2* (249), *ERS1* (128) and *EIN4* (129). Members of the EIN family are positive regulators of the ethylene response, and act downstream of *CTR1*, which itself interacts with *ETR1* and *ERS1* (114). The transcription factor *EIN3* regulates *ERF1*, a transcription factor capable of binding to the GCC-boc present in ethylene-inducible defence genes (117).

#### 1.2.2.1 Interactions of the Salicylic acid, Jasmonic acid and Ethylene signalling pathways are important in the responses to both biotic and abiotic stresses

Previous research has suggested that there is extensive crosstalk between SA and JA: SA has been shown to target the JA pathway downstream of JA biosynthesis (159), as well as being antagonistc to the JA defence pathway (251). Also, the activation of *R* genes results in crosstalk in the hormone pathways activated, in order to distinguish between different pathogens, such as biotrophs or necrotrophs (108). As shown in Figure 1.1, SA signalling is effective against biotrophic pathogens, such as *P. syringae*, whereas the JA- and ET-mediated defined response is mostly effective against necrotrophic pathogens, such as *Botrytis cinerea* (285, 108).

Figure 1.1 highlights but a small portion of our current understanding of hormone signalling in *A. thaliana*. Recent investigations into the role of *WRKY33* have shown this transcription factor to have a key role in regulating multiple hormone signalling pathways (38). Previous studies shoed that *WRKY33* was required for resistance to *B. cinerea*, and therefore positively regulates JA- and ET-responsive pathways, whilst negatively regulating SA-mediated signalling (343). However, Birkenbihl *et al*. demonstrated that the early stages of JA signalling are independent of *WRKY33* regulation (38).

The regulation of signalling pathways in plants is extremely complex, and is not yet completely understood. It is thought that plants prioritise SA induction over JA, which causes the downregulation of JA-responsive genes, such as *PDF1.2* and *VSP2* (169). However, since the ET and JA signalling pathways are not antagonistic, as demonstrated by SA and JA signalling pathways, if the ET signalling pathway is prioritised, the JA response is also activated (168). ET is also a crucial component for SA suppression of JA: the presence of ET signalling can suppress JA, but only in the presence of SA (280). When ET is not present, SA can suppress JA via the gene *NPR1* (280).

Many genes are involved in the orchestration of the antagonism between SA and JA, including *WRKY70* (173) and *MYC2* (146): *WRKY70* is activated in an *NPR1*-dependent manner by SA and repressed by JA (173). In transgenic plants that were unable to accumulate SA, the expression of *WRKY70* was not induced, confirming that *WRKY70* expression requires direct regulation by *NPR1*, and indirectly, SA. However, by testing *coi1* mutant plants, which fail to accumulate JA, expression of *WRKY70* was unaffected (173), confirming that regulation of *WRKY70* is not dependant on JA.

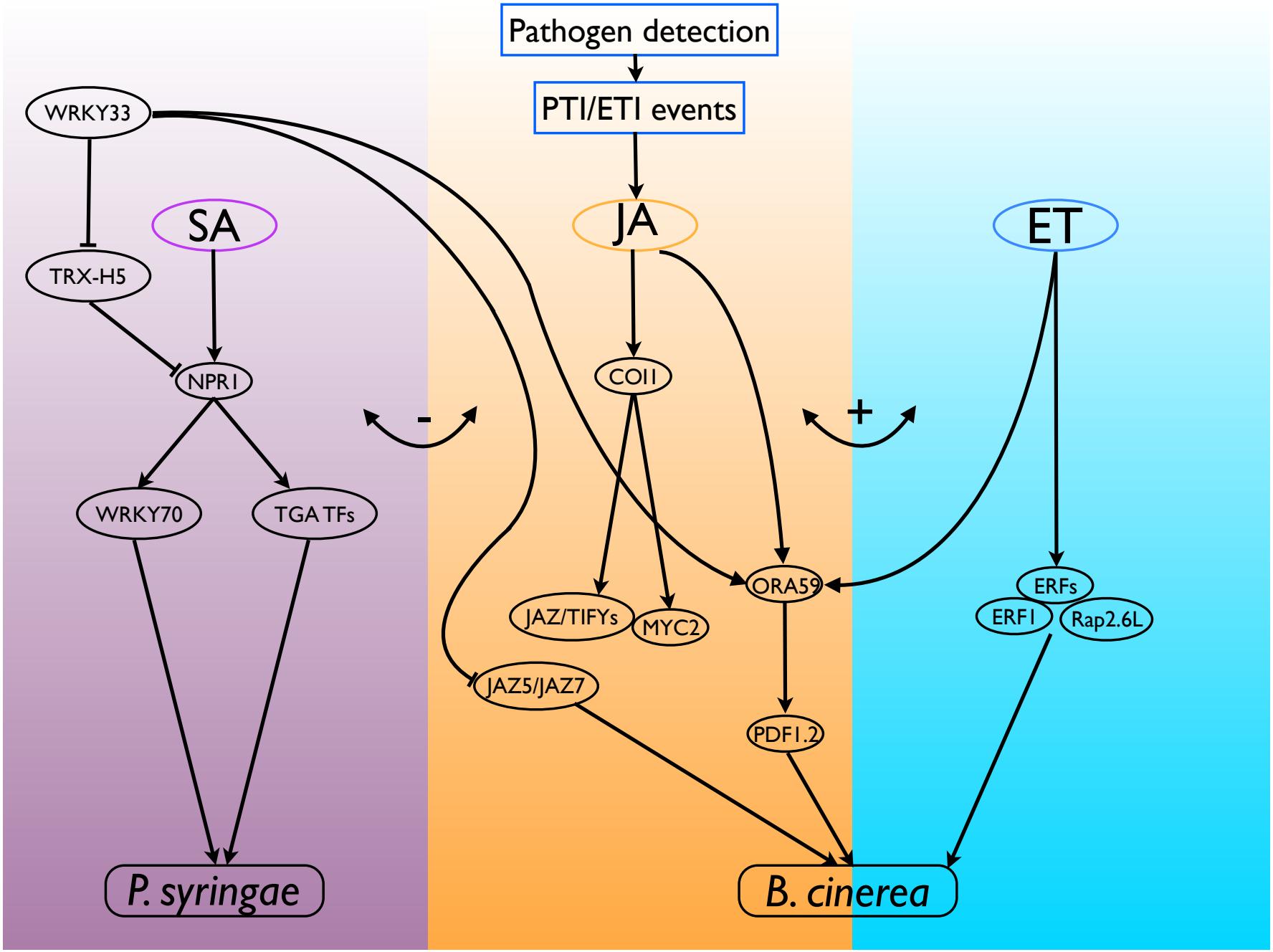



Figure 1.1: Network of hormone signalling in *A. thaliana* important in different necrotrophic (example: *B. cinerea*) and biotrophic (example: *P. syringae*) stress responses, adapted from Thomma *et al*. (285), Birkenbihl *et al*. (38) and Adie *et al*. (6). Here, the SA pathway is required for the response to *P. syringae*, whereas a JA/ethylene-dependent pathway is required for the response to *B. cinerea* (284). The SA and JA pathways are mutually antagonistic, whereas the JA and ET pathways exhibit extensive crosstalk. Transcription factors, and transcription factor families, activated in the relevant hormone pathways are shown.

CORONATINE INSENSITIVE 1 (*COI1*), a protein required for wounding-induced JA regulation of stress response genes, is the principal target of JA signalling (as shown in Figure 1.1) (331): when bound by active JA, the JA-*COI1* complex recruits members of the jasmonate ZIM-domain (JAZ) family as co-receptors (146). The JAZ proteins are transcriptional repressors of *MYC2*, a transcriptional activator, at low levels of JA (146). *MYC2* itself is a hub in the crosstalk between many phytohormone signalling responses, namely JA, ET and ABA. Its role in Abscisic acid-mediated responses, which will be discussed later, is mainly involved in the response to abiotic stress (11). Anderson *et al*. completed in-depth research into the role of *MYC2* in JA/ET-mediated signalling during biotic stress responses, revealing that Methyl-JA treatment induced *MYC2* expression, whilst ET treatment suppressed *MYC2* expression (11). Coupled with data from *myc2* mutant screen, which revealed decreased susceptibility to a necrotrophic fungal pathogens (183), it can be concluded that *MYC2* is a negative regulator of plant defence (11).

*MYC2* becomes repressed when JA-mediated defence is required to be activated: JAZ proteins interact with the co-repressors *NINJA* and *TOPLESS* (*TPL*) (15), where *NINJA* subsequently represses *MYC2* activity (229) by suppressing the activity of the RNA polymerase II-Mediator complex (280).

JAZ proteins are also a crucial component of the mediation of antagonistic effects seen between SA and JA. Two transcription factors, *EIN3* and *EIL1* are repressed by a JAZ protein/*HDA6* complex (146). *EIN3* and *EIL1* in turn repress *SALICYLIC ACID INDUCTION DEFICIENT 2* (*SID2*), which encodes an enzyme required for the biosynthesis of SA (59). At low JA levels, the *COI1-JAZ* complex inhibits *EIN3* and *EIL1*, which allows the biosynthesis of SA. However, in the presence of necrotrophic pathogens, JA levels rise, and JAZ proteins are degraded (144), allowing *EIN3* and *EIL1*to inhibit *SID2*, and thereby down regulating SA biosynthesis. *EIN3* and *EIL1* are also capable of up regulating the expression of JA and ET responsive genes, such as *ORA59* and *ERF1*. Therefore, JAZ proteins also function as regulators of the synergistic JA/ET crosstalk (246). It is likely that *EIN3* and *EIL1* are the hubs of the crosstalk between these two pathways (146).

#### 1.2.2.2 The role of Abscisic acid, and its crosstalk with JA and ET, in the stress response

Abscisic acid (ABA) is involved in the regulation of leaf senescence, abiotic stress responses (such as drought, cold, heat, high light and salt stress) (262), and also the negative, and sometimes

positive, regulation of plant defence against both biotrophs and necrotrophs (6).

ABA has previously been shown to enhance susceptibility of plants to disease, however, its interaction with the JA and ET signalling pathways have implications in the expression of genes important in plant defence and resistance to pathogen stress (6). Previous research has shown that wounding can induce several genes involved in the JA, ABA, and ET pathways (66). As mentioned previously, *MYC2* is involved in the ABA-mediated response to abiotic stresses, and is a positive regulator of ABA signalling (3), as well as regulating the interaction between JA and ET in the defence response (26).

Overall, ABA is a negative regulator of the plant defence response to pathogens. Mutations in the *ABA DEFICIENT 2* (*ABA2*) gene in *A. thaliana* rendered the plant less susceptible to the necrotrophic fungus *Fusarium oxysporum* (11) and the oomycete pathogen *Hyaloperonospora arabidopsidis* (210). The *aba2* mutants also exhibited significantly higher transcript levels of defence genes which are regulated by the JA/ET pathways (11), highlighting the crosstalk between these three hormones. Similarly, mutations in genes important for the synthesis of ABA also increased the resistance of *A. thaliana* to *Pseudomonas syringae* pv. tomato (76) and *B. cinerea* (6).

However, examples of ABA acting as a positive regulator of plant defence have also been reported, where mutants in genes affected by the ABA defence pathway were less resistant to the fungal pathogens *Pythium irregular* and *Alternaria brassicicola* (6). This suggests that although ABA is a negative regulator of defence responses to the majority of pathogens, ABA can also be a positive regulator of plant defence.

ABA induces a number of genes important in the stress response, and tolerance, of abiotic stresses. Seki *et al.* found evidence of crosstalk between ABA and JA when investigating ABA-inducible genes using microarrays: genes involved in the metabolism of JA, and genes known to be regulated by JA were found to also be inducible by ABA (255), emphasising the link between these two hormone signalling pathways.

Plants can manipulate the relationship between the ABA and JA/ET-synergistic pathways in order to reduce the overall fitness cost when dealing with either biotic or abiotic stress (11, 120).

#### 1.2.2.3 Summary of the role of phytohoromone signalling in plant stress responses

In summary, there is substantial crosstalk exhibited by the hormone signalling pathways in *A. thaliana*. Much of the crosstalk is mediated at the transcriptional level: Abe *et al*. demonstrated that *MYC2* was involved in the ABA-mediated response to abiotic stresses (3), with Bari *et al*. identifying *MYC2* as a regulator of the JA and ET interaction in the defence response to pathogens (26). Figure 1.1 illustrates that a number of genes encoding transcription factors are important for the regulation of stress responses, via the hormone signalling pathways: *WRKY33* regulates components of the SA, JA, and ET signalling pathways, whereas *ORA59*, a member of the ERF transcription factor family, is known to be regulated by JA and ET signalling pathways. In view of this, transcriptional regulation can be seen to provide a convergence point in the networks of hormone signalling in *A. thaliana*.

### 1.2.3 Gene regulatory networks involved in multiple stress responses

As mentioned previously, the main focus of research has moved away in recent years from investigating single genes in specific responses and pathways, to investigating regulatory networks involved in stress responses (295). However, by using genome-scale analysis in *A. thaliana*, the function of regulatory networks can be investigated on a whole-system scale, identifying how genes potentially work together (211).

Gene regulatory networks (GRNs) consist of a set of genes which are expressed in a specific spatial and/or temporal pattern or process, such as development or stress. In living systems, genes do not work independently: groups of genes work together, and can interact indirectly, through their protein products, and these interactions create a network of genes (281), illustrating the complex interactions of transcription factors and their target genes, for example. GRNs allow the organism to respond to altering conditions through the dynamic co-ordination of expression of gene members of the network (287, 74).

A number of GRNs have subsequently been identified as important to the response, or tolerance, of individual stresses. However, it has become increasingly evident that plants employ crosstalk between response mechanisms to adapt to different environmental changes (159), and in order to improve plant fitness (120). Therefore, identification of GRNs employed over multiple stresses is

required.

From the perception of the stress signal to the expression of genes important in the response to the stress, transcription factors and *cis*-acting regulatory elements in the promoters of stress-responsive genes have been identified as having important functions in the plant response to stress (330). Gene regulatory networks, particularly transcriptional regulatory networks, involved in the response to multiple plant stresses will be discussed here.

Genes induced during the stress conditions have been classified into two groups (255, 161). The first group contains genes which function in stress tolerance, such as chaperones and mRNA-binding proteins. The second group contains proteins involved in the regulation of gene expression involved in the stress response, such as transcription factors (330).

Plants are widely affected by the abiotic stresses drought, osmotic and cold, severely limiting plant growth and ultimately, the production of crops. For these reasons, much investigation has been carried out not only on these individual stresses, but crosstalk between them. The methods employed to realise the genes under the influence of multiple stresses, such as microarray experiments, has led us to apply these methods to other combinations of stresses (265, 189, 216). Studies investigating the crosstalk in abiotic stresses have identified transcription factors as important in stress responses (22, 37), and also as points of convergence in multiple stress responses (99).

Figure 1.2 shows the crosstalk and points of convergence (i.e. genes involved in multiple stress responses) between osmotic and cold stress responses. Here it can be seen that a number of transcription factors (*DREB1*, *CBF4* and *DREB2*) involved in the response to osmotic and cold stress all regulate *rd29A* via the *cis*-acting element DRE/CRT, found in its promoter (260). The ABRE element is also found in the promoter of *rd29A*, but this is targeted by members of the ABF and AREB transcription factor family. The binding of different transcript factors to these different *cis*-acting elements allows the distinction between different signals in response to cold and osmotic stresses (179). The subsequent induction of *rd29A* leads to the expression of genes involved in the stress response (260).

A number of transcription factor families are also highlighted as having a role in the response to either the osmotic or cold stress individually, notably *MYC*, *MYB* and *ANAC* families. Within these families, a number of transcription factors have roles within individual responses. For example,

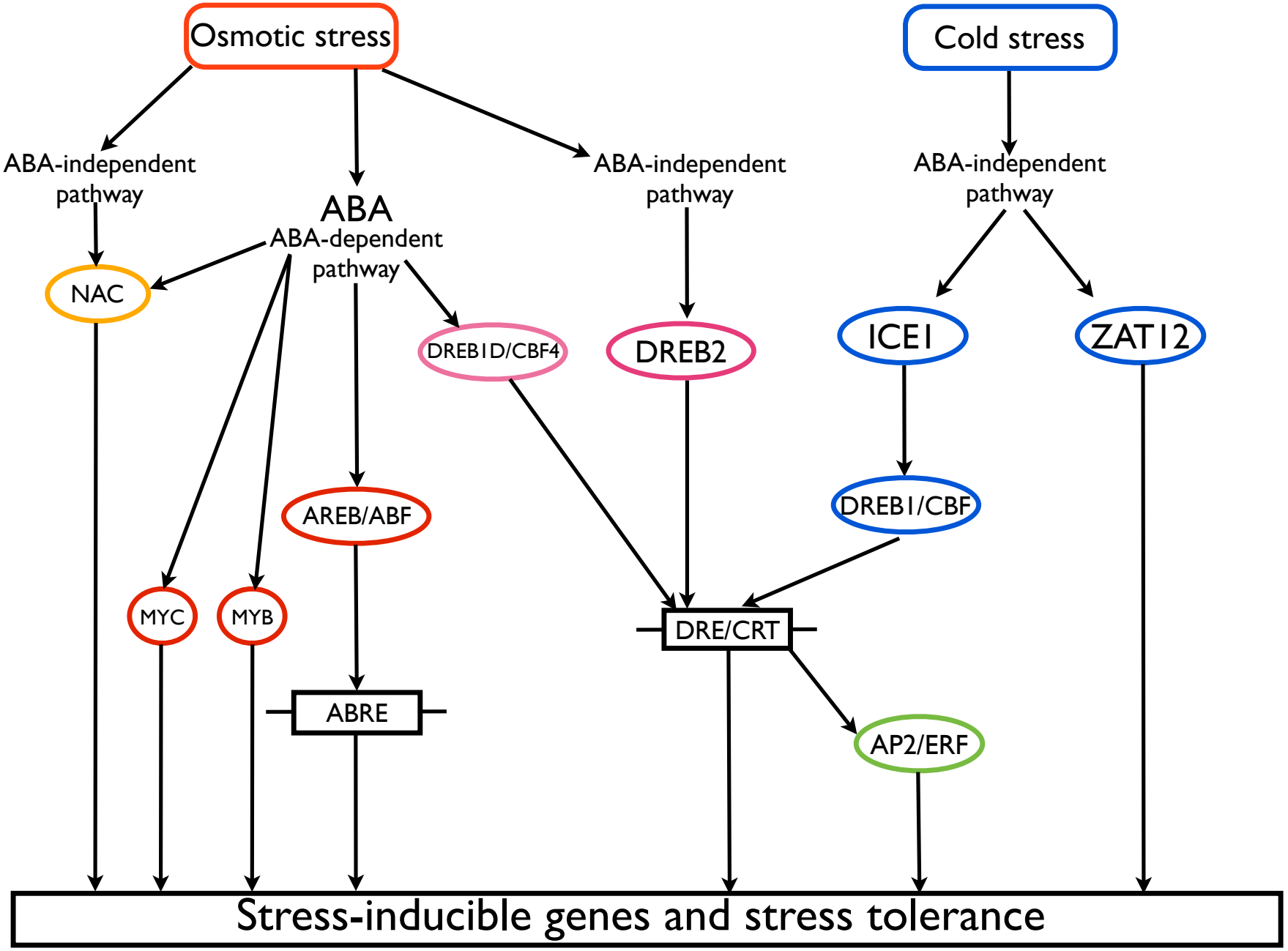



Figure 1.2: A proposed regulatory network of transcription factors and *cis*-acting regulatory elements (CAREs) involved in osmotic and cold stress, modified from Yamaguchi-Shinozaki and Shinozaki (330). Transcription factors involved in these stress responses are shown in coloured ovals. Crosstalk between the two stress responses occurs at the CARE (DRE/CRT), found in the promoter of the gene *rd29A*, where multiple DREB transcription factors bind.

*MYB2* and *MYC2* are known to be required for the regulation of *RD22* and *ADH1* during osmotic stress (2), and *ANAC072* (also known as *RD26*) is required for positively regulating the expression of the glyoxalase-encoding gene *GLY* during drought stress (98).

Elements of the ABA-independent pathway also exhibit crosstalk between the response to salt stress (261). *A. thaliana* plants over expressing genes encoding DREB transcription factors exhibited an increased tolerance to freezing (179). DREB proteins contain two subclasses, DREB1 and DREB2, which both control a large number of genes involved in stress tolerance (260). Overexpression of *DREB1A* also showed increased tolerance to drought stress (104). Microarray analyses carried out on *A. thaliana* leaf tissue over expressing *DREB1A* identified more than 40 genes as downstream targets (255). A number of genes functioned in stress tolerance, whilst some downstream genes also included transcription factors, such as *STZ* and AP2/ERF transcription factors. Expression of *STZ* is strongly induced in response to drought and cold stresses: *A. thaliana* transgenic plants over expressing *STZ* exhibited growth retardation and increased drought tolerance, suggesting *STZ* acts as a transcriptional repressor (250).

Since ABA only plays a part in the response to drought, and not cold stress (260), it can be seen that the crosstalk between the drought and cold stress responses is at the transcriptional level, through the DRE/CRT CARE and the DREB transcription factors.

The drought specific part of the GRN shown in Figure 1.2 has also been seen to exhibit crosstalk with the response to biotic stresses via the *MYB2*, *MYC2* and *ANAC072* transcription factors, which are targets of the JA signalling pathway (261). Both the *MYC2* and *MYB2* transcription factors have been found to bind to *cis*-elements in the promoter of *RD22* (3), a gene regulated by ABA, and involved in the response to cold stress, osmotic stress and salt stress. Microarray analysis carried out on *A. thaliana* plants over expressing *MYB2* and *MYC2* identified a number of ABA- and JA-inducible genes. Overexpression of these two genes also conferred an ABA-hypersensitive response and decreased drought susceptibility in plants (3). *ANAC072* is seen to be induced by both ABA and JA, whereas the downstream targets of this gene are JA-inducible. Typical ABA-inducible genes are not regulated by *ANAC072* expression (98).

The network illustrated in Figure 1.2 was generated via the accumulation of many experimental findings. However, with the generation of gene expression datasets, various modelling tools have

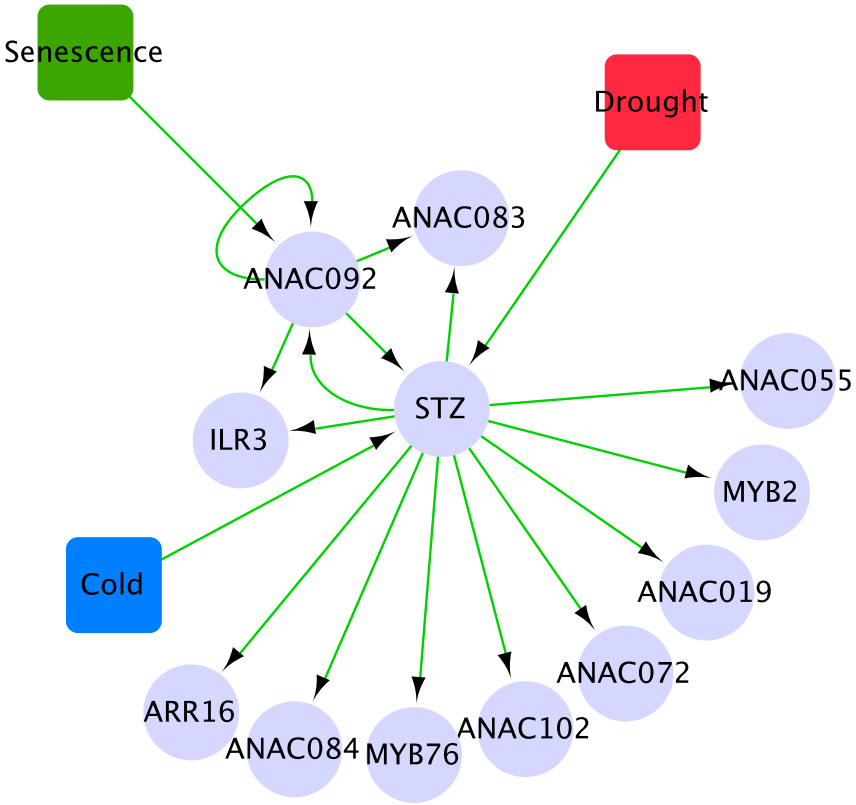



Figure 1.3: A generated network model containing genes enhanced by senescence using variational Bayesian state space modelling (VBSSM), adapted from Breeze *et al*. (43). This model was generated using time series expression data during the senescence response to elucidate a gene regulatory network involved in the response to this stress. This model predicts that *STZ* is regulated by *ANAC092*, a transcription factor known to be involved in the senescence stress response (22). *STZ*, here seen to act as a hub gene in this response, has also been identified experimentally as inducible by drought and cold stresses by Sakamoto *et al*. (250).

also been developed to mine this data for signals of regulation, in order to infer a models of gene regulation, without the laborious process of experimentation. Such methods will be described in greater detail in Section 1.3.

Breeze *et al*. (2011) (43) presented a predicted network model for a small set of genes, generated using a the modelling technique, variational Bayesian state space modelling (VBSSM) method (30) on senescence time series data. The resulting model (shown in Figure 1.3) demonstrated that *STZ* is regulated by *ANAC092* (43), drought and cold stress, and also indirectly by senescence. Senescence has previously been shown to enhance the expression of *ANAC092*, and this regulation is thought to be transcriptional (22). However, the network downstream of *ANAC092* remains unknown (22): the predicted network shown in Figure 1.3 provides hypothesised downstream targets of *ANAC092* in the senescence stress response. Indeed, a number of the genes predicted to be downstream of *ANAC092* in this network model (*ANAC083*, *ARR16* and *ANAC084*) were identified as being differentially expressed in a microarray experiment using an overexpressor line of *ANAC092* (22).

The networks mentioned here are mainly devoted to abiotic and developmental stress responses. The interaction between abiotic and biotic stress responses in plants is still a relatively new area of research (124), however, recent studies have shown that ABA does influence the temporal regulation of the pathogen defence system (289). It seems that ABA is involved in the response to many pathogen stresses, as ABA crosstalks between the stress response and developmental processes. However, the exact gene regulatory network involved in the biotic/abiotic crosstalk has yet to be deciphered.

## 1.3 Theoretical methods for discovering genes important to multiple plant stress responses and constructing gene regulatory networks

The elucidation of GRNs can be costly and time-consuming using laboratory based techniques alone. For example, the discovery of the circadian clock in *A. thaliana* took many years and many laboratory-based experiments in order to confirm its existence: genomic analysis to identify clock components was not feasible until the 1970s, and experiments were laborious (196). The first plant clock mutant was not generated until 1995, when Miller *et al*. (205) produced a *toc1* transgenic

line. With the advent of a computational model of the circadian clock network in 2006 (180), the research in this field has rapidly expanded. Consequently, there is a great need for computational approaches to predict GRNs, in order to reduce both the time and effort spent on experimental validation by predicting candidate groups of genes to analyse.

Many theoretical methods are now freely available to analyse gene expression data to infer GRNs, and to identify genes important to multiple stress responses (30, 272, 150, 95, 230). Mjolsness *et al*. conducted extensive research into which essential steps should be completed for predicting GRNs using expression data from microarrays (209). In order to identify transcriptional regulatory mechanisms, such as co-expression and co-regulation, data should be clustered by their expression into sets of co-expressed genes, and time-courses should be modelled in each condition the data is measured in (209). Integration of co-expressed genes, inferred regulation from modelling techniques, gene ontology (GO) annotations and promoter motif information provides a logically inferred network of genes with specific biological functions (302). Some clustering and biclustering approaches for inferring co-expressed genes in single and multiple stresses, and modelling approaches used for inferring GRNs will be discussed here.

### 1.3.1 Predicting gene regulatory networks using modelling techniques

Modelling algorithms aim to describe the transcriptional events occurring within a biological system in a graphical manner, using gene expression transcriptome data from microarrays. Modelling methods use data to infer the regulatory interactions (edges) between genes (nodes) (211). Although modelling techniques have been successfully applied to data from mammalian and human systems (7, 306), the situation is more complex in plants: a lack of information on transcription factors, and the promoter motifs they are capable of binding to, means this prior information cannot be incorporated into modelling techniques.

Recently, a comparison study between various theoretical approaches for GRN construction was completed (231). In this study, ordinary and stochastic differential equations (ODEs and SDEs respectively) were compared with Bayesian and dynamic Bayesian network (BNs and DBNs respectively) methods. Penfold *et al*. found that DBNs, such as VBSSM (30), and Casual Structure Identification (CSI) (Penfold, C.P., University of Warwick, *in preparation*) outperformed other GRN construction methods, such as ODEs and SDEs, when handling time-series datasets. It was

also concluded that the 'best' networks were recovered using time-series data, confirming the previous results of Bansal *et al*. (23), which suggests that time-series observations are the appropriate choice for inferring gene regulatory network structure.

DBNs have been used to model time-series gene expression data, as they present an advantage in the form of managing hidden variables (such as protein levels or mRNA decay) to infer network models of interacting components. However, there are also disadvantages to to using DBNs to model time-series expression data: the direction of regulation is difficult to determine; and the incorporation of prior information can be difficult, and would have to be included manually.

The VBSSM (30) algorithm can be applied to high resolution time series gene expression datasets. Advantages of VBSSM over other modelling methods is the ability to model combinatorial regulation, and also the ability to detect feedback loops. Figure 1.4 describes the state-space model used for time series gene expression data, where the outputs from the previous time step are used as an input for the current time step, in order to infer interactions between genes over time.

VBSSM is limited, however, by the number of genes it is capable of modelling together. Therefore, a method to select which genes should be modelled is required. Selection criteria can be based on pre-existing data, such as genes with known function in the stress response. However, due the evidence stating that stress responses are governed at the transcriptional level (341), including transcription factors would be a prudent choice. By limiting the group of genes to be modelled to transcription factors which are differentially expressed over time to a stress, previously unknown cases of transcriptional regulation may be uncovered by the resulting predicted model.

#### 1.3.1.1 Graph theory as a basis for modelling time series gene expression data

DBNs are a class of modelling approaches that have a basis in graph theory (30). Since GRNs can be described using parameters from graph theory (17), using DBNs to model gene expression data is a logical choice for a modelling approach.

In graph theory, a 'graph' is a set of nodes (which, in the context of GRNs would represent genes), which are connected by 'edges' (which, in the context of GRNs would represent regulation) (25). Graphs can be 'non-directional', where no definition is given between the nodes associated with an edge, for example, in a protein-prontein interaction network (25, 252). In directed graphs, the edge

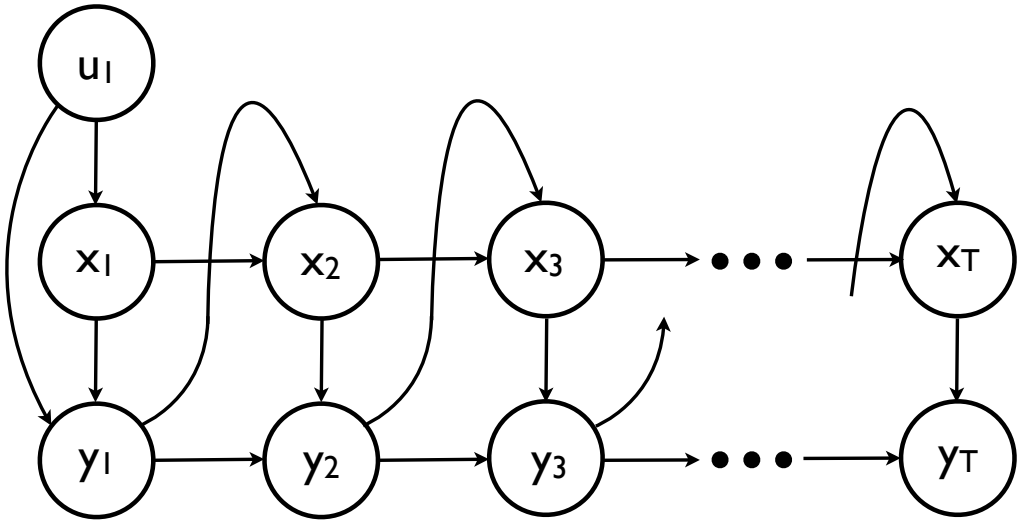



Figure 1.4: The VBSSM feedback model, adapted from Beal *et al*. (30), where outputs feed back into the inputs. At time $t$, gene expression levels are defined as $y_t$, whereas hidden variables are defined as $x_t$

between modes of the graph has a defined direction, which, in the context of GRN, will represent the direction of regulation from a transcription factor to its target gene (25).

VBSSM was developed to approximate the marginal likelihood of Bayesian dynamical systems, using variational methods, with the intention of applying it to microarray data to elucidate interactions between regulatory transcription factors and their target genes (29). Beal *et al*. incorporated directionality into VBSSM, so that the direction of regulation could be inferred. This means VBSSM differs from other modelling techniques, such as Markov random fields (MRFs) models (214, 319), which do not distinguish which direction the edge trajectory is facing. However, by incorporating directionality, VBSSM becomes a more complex and computationally intensive modelling technique to employ (214).

### 1.3.2 Inferring regulation using clustering

Cluster analysis, or clustering, is used when one wishes to assign a set of objects into groups, which are called 'clusters'. Objects in the same cluster are similar to each other, but are dissimilar to objects in other clusters. Ultimately, clustering algorithms aim to organise a set of observations, allowing the researcher to develop an understanding of the data being studied. Clustering algorithms are primarily used to organise gene expression data by grouping similar expression profiles together (277). This requires a mathematical definition of similarity in order to measure the behaviour of two genes, based on their expression profiles. Suitable measurements include the Euclidean distance and Pearson's Correlation Coefficient, which are the most widely used.

The Euclidean distance, $d$, is a measure of distance between two points, and is based on the Euclidean norm. The Euclidean distance defines the distance between pairs of expression profiles by the length of the distance between gene expression points. The Pearson's Correlation Coefficient is a similarity measure which scores pairs of expression profiles as highly correlated if they are linearly similar. Both Pearson's Correlation Coefficient and Euclidean's distance score genes that are functionally related highly (334).

Eisen *et al*. published a highly significant paper, which has since become the standard for gene expression data clustering (88), and as such, clustering has become a routine feature of analysis of gene expression data. There are many advantages to the clustering of gene expression data: the function of an unannotated gene can be predicted if the functions of the remaining genes in the cluster are known, as functionally similar genes are often co-expressed (88); by clustering the arrays themselves, rather than the genes, classes of samples can be identified; co-expressed genes are likely to share a common regulatory mechanism, allowing the identification of transcription factor binding sites (116); finally, clusters of genes can be mapped onto metabolic networks, allowing the detection of overrepresented metabolic pathways (299).

Transcriptional regulation can be inferred from gene expression: promoters of co-expressed genes may share regulatory elements, suggesting a common regulator; also, expression profiles of target genes may be correlated to the expression profile of the regulator, but there may be a time-shifted delay between the expression of the regulator and its target (237). Also, if a transcription factor negatively regulates its targets, the expression profiles may be inversely correlated (237). The Tem-

poral Clustering by Affinity Propagation (TCAP) algorithm (151) infers transcriptional regulation of genes using time series expression data, by analysing complex temporal events which may not be identified using standard clustering techniques. This technique will be discussed in more detail in the subsequent sections.

#### 1.3.2.1 Hierarchical clustering

This method of clustering analysis seeks to generate a hierarchy of clusters, by one of two approaches: 'bottom-up' (agglomerative), where each observation, or gene expression profile, starts in its own cluster, followed by pairs of clusters merging; a 'top-down' approach (divisive), where all gene expression profiles start in one cluster, and clusters splinter off from this cluster as the algorithm moves down the hierarchy. Both approaches generally result in the production of a dendrogram to demonstrate the clusters.

Hierarchical clustering of gene expression data demonstrates a lack of robustness when analysing data that contains a level of noise, however (187). Hierarchical clustering methods are also unable to re-evaluate the results it generates, making clusters difficult to interpret when a large amount of gene expression data is involved (187).

The Self-Organising Tree Algorithm (SOTA) is an unsupervised hierarchical clustering algorithm based on the self-organising map (SOM) (158) and growing cell structures (96), which was developed in order to achieve robustness when analysing noisy gene expression data (86). The SOTA method was also at an advantage to other hierarchical clustering methods due to its divisive nature: by clustering 'top-to-bottom', the highest levels of clusters were resolved before the details of the clusters at the lowest levels of the hierarchy. SOTA also allowed the user to stop the growing of the 'tree' at a desired hierarchical level (86). Luo *et al*. applied SOTA to gene expression data of 3000 yeast cell cycle genes, which were originally described in (68): SOTA identified 25 clusters which were functionally enriched when mapped to the functional category of the MIPS database (203), and contained between 88 and 276 genes (186). The clusters were functionally enriched for categories such as 'cell cycle', 'DNA processing', and 'Organisation of the cytoplasm', suggesting SOTA is capable of detecting 'real' functionally similar clusters (186).

#### 1.3.2.2 Time-series data clustering

Time-series experiments have become a increasingly popular method to analyse the temporal process of gene expression (24). Time-series experiments differ to static gene expression experiments, where a single snapshot of the gene expression in different samples are measured. Static data is also only measured from a sample population, and are assumed to be independently and identically distributed. Time-series data, on the other hand, generates gene expression data where there is a strong similarity between consecutive data points, which can be observed as a function of the time separation between data points (otherwise known as autocorrelation) (24). Therefore, autocorrelation can be defined as the correlation between the values of gene expression at different time points.

Time-series expression datasets provide challenges also: the larger the number of genes, the greater the computational effort to discover co-expressed genes is needed; there is also an issue of how great the temporal resolution should be. Although the microarray experiments needed to generate the time-series gene expression data are expensive, as there are multiple arrays needed for each time point measured (89), the temporal resolution gained has high value, as gene expression is itself a temporal process (24).

Eisen *et al*. performed clustering of time-series gene expression data using both a similarity measure (a variation of the Pearson's Correlation Coefficient), and hierarchical clustering (88). Multiple time-series data from the budding yeast *Saccharomyces cerevisiae*, including sporulation (70), and the diauxic shift (82), were analysed. Using both clustering methods, Eisen *et al*. found that genes represented by two or more probes on the microarray, or genes with high sequence similarity, were clustered next to each other, concluding that the location on an array does not affect the gene expression profile observed. Also, when cluster were examined more closely, it could be seen that the gene members shared a common role or cellular process. In the diauxic shift dataset (82), 126 genes clustered together were seen to be strongly down regulated. 112 of the gene members of the cluster were genes encoding ribosomal proteins, as well as other genes encoding proteins involved in translation, such as initiation and elongation factors, and tRNA synthetases (88).

Time series expression data can also be used to infer transcriptional regulation: by identifying similarly expression genes, which are co-expressed, the regulator may too be co-expressed to its

target genes, but with a time delay in expression (237). Standard static and time-series clustering algorithms have the advantage of identifying cases of simultaneous co-regulation, but are not ideal at identifying genes which regulate each other. The task of identifying the regulating transcription factor is also a challenge. TCAP (151) is an approach for finding gene clusters, or 'modules', which incorporate temporal features, such as time lags and inversions, by exploiting Affinity Propagation (AP; (94)), and the Qian similarity measure (237). Whilst this method is capable of detecting transient co-expression, it is not as sensitive as biclustering algorithms to expression events that occur in a short window of time, due to the approximate nature of the Qian similarity measure.

However, by applying TCAP to a time-series dataset of gene expression in *A. thaliana* leaves during infection by *B. cinerea* (Windram *et al.* 2012, *in press*), 338 clusters were identified. A cluster, where the genes were seen to have a circadian rhythm, contained two genes encoding the circadian clock components *LHY* and *GI*: *GI* was observed to have a delayed and inverted expression profile to *LHY*. Another member of this cluster, *At1g56300*, also demonstrated a similar expression profile to that of *GI*. *At1g56300* is a member of the Rapid Wounding Response (RWR) genes, which were identified by (309) to be regulated by the circadian clock.

### 1.3.3 Identifying genes important in multiple stress responses using biclustering

Biclustering algorithms, sometimes referred to as 'co-clustering' or 'two-mode clustering', mine time series gene expression data by clustering both the rows (genes) and columns (conditions) of a data matrix simultaneously. The biclustering of gene expression data is a relatively new method of analysing time series datasets in multiple conditions, first introduced by Cheng and Church (65). Previously, the term 'biclustering' was used by Mirkin, but not in the context of gene expression data (172).

A biclustering algorithm will generate 'biclusters', which are subsets of rows with similar behaviour, such as similar expression profiles, across subsets of columns (or vice versa). Subsequently, biclusters can possibly form overlapping groups of genes. Genes belonging to the same bicluster are hypothesised to be regulated in the same manner (co-regulated), usually by a transcription factor, and subsequently these biclusters can be used to form transcriptional networks (211). Biclustering methods overcome the disadvantages of standard clustering methods to yield groups of genes that are co-expressed under subsets of conditions. However, biclustering algorithms do

produce limitations of their own: methods can generate overlap between clusters; and not all of the algorithms are deterministic, due to a lack of a 'gold-standard', and the unsupervised approach many of the algorithms employ (190).

Biclustering differs from standard clustering algorithms, which partition the genes or conditions into groups that are mutually exclusive from each other, based on a high similarity score of genes or conditions in an expression matrix. The aim of biclustering is to identify the role of a gene or condition in multiple pathways (65). Given the importance of identifying subsets of genes that are functionally related by the exploratory analysis of time-series data in recent years, this challenge has since been extended to include identifying functionally related genes that are co-expressed together in multiple datasets. Hence, biclustering algorithms have become more sought after.

The leading biclustering algorithm is the Extended Dimension Iterative Signature Algorithm (EDISA) (272), an extension of the ISA algorithm (132, 131). EDISA generates initial biclusters, or 'modules', which are refined by removing genes and conditions until the module fits a 'definition'. EDISA covers three such module definitions: single response, coherent and independent response.

Single response modules aim to identify genes associated with one condition in order to discover specific mechanisms employed in the response to that condition (illustrated in Figure 1.5, bottom row). Coherent modules identify genes that are co-expressed in multiple conditions (illustrated in Figure 1.5, top row). The modules identified in the single response and the coherent response could potential be co-regulated by a common transcription factor. The independent response identifies modules where the regulatory mechanism is hypothesised to be different for each condition the module is co-expressed in (illustrated in Figure 1.5, middle row).

EDISA was used in an initial analysis of the PRESTA time series datasets (see Section 4.1.3). However, due to the limitations of EDISA in the number of genes it is capable of analysing, we were restricted to analysing a set of genes which were differentially expressed in response to *B. cinerea* infection, short day and long day senescence. These genes (2774) were found to be differentially expressed in all three datasets. Non-differentially expressed genes were not included in this analysis, as if they are not differentially expressed in a particular stress they are not likely to form a part of the response to that stress.

Eight modules were identified by EDISA as being significantly co-expressed in a combination of

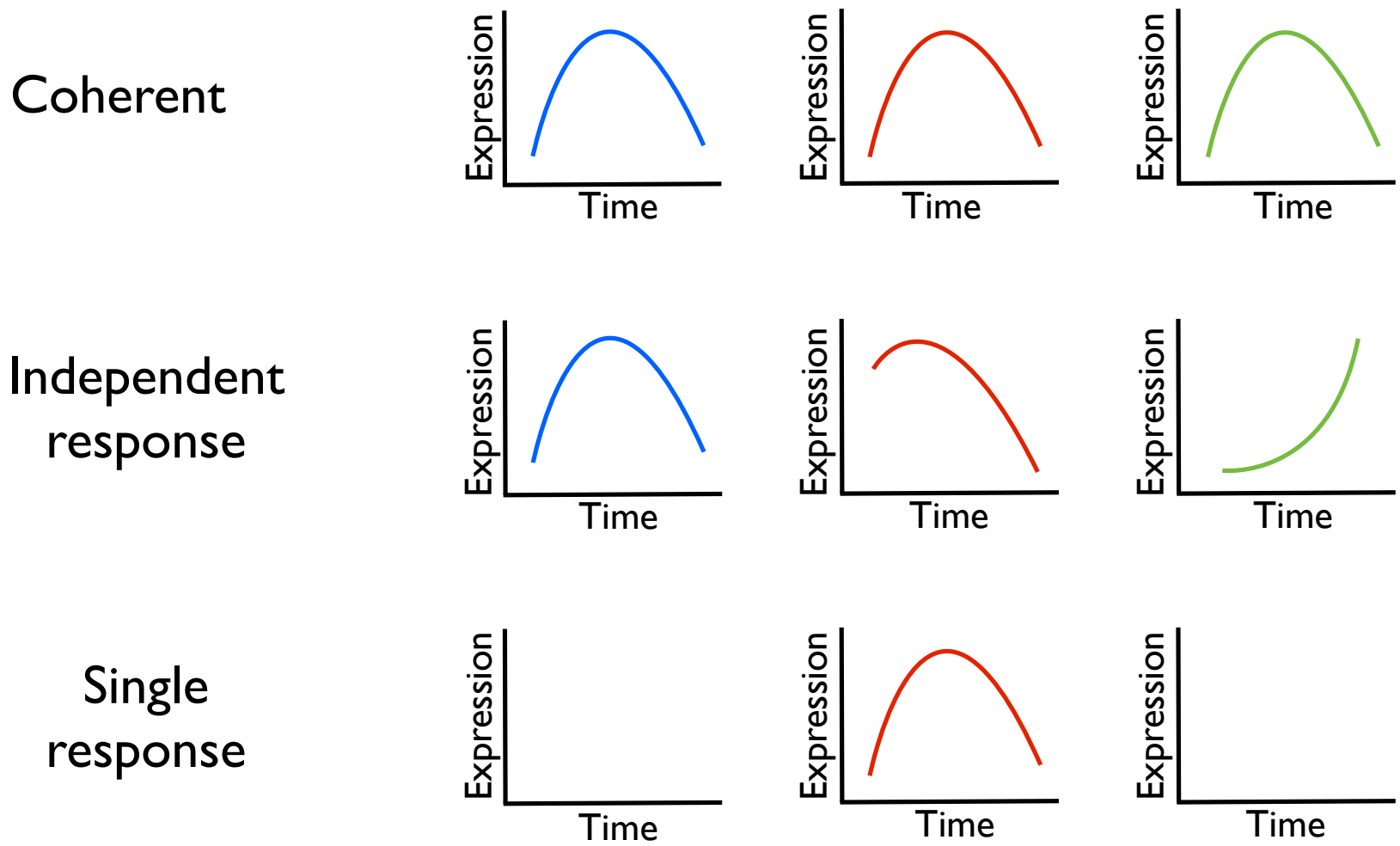





Figure 1.5: EDISA module definitions: coherent, independent response and single response. Adapted from Supper *et al*. (272). Coherent modules identify co-expressed genes in multiple conditions. The single response defines modules as gene associated with just one condition. The independent response identifies modules where the regulatory mechanism is hypothesised to be different for each condition the module is co-expressed in.

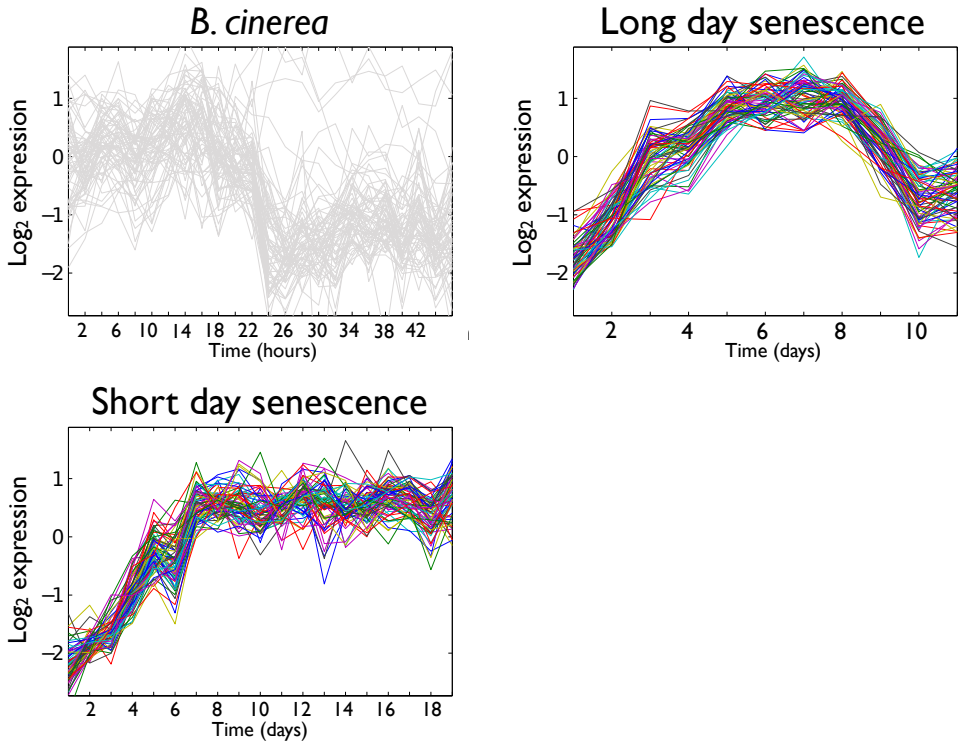



Figure 1.6: EDISA module of a set of 68 genes significantly co-expressed in long and short day senescence. Here, expression profiles of the 68 genes are shown in each of the stresses: *B. cinerea*, short day and long day senescence. EDISA has identified these genes are significantly co-expressed in short and long day senescence only (indicated by the coloured expression profiles). These genes are not significantly co-expressed in *B. cinerea*, as indicated by the grey expression profiles.

these three stresses. An example of the output from EDISA, and one of the modules, is illustrated in Figure 1.6. EDISA identified 68 genes as significantly co-expressed in short day and long day senescence (as indicated by coloured expression profiles in Figure 1.6), but not in response to *B. cinerea* stress (as indicated by grey expression profiles in Figure 1.6). However, the expression profiles of these genes in response to *B. cinerea* infection in Figure 1.6 suggest that there is an interesting biological event taking place: the expression profiles of the majority of the genes are being down regulated in response to *B. cinerea*. This raises the question: why aren't these genes co-expressed in all three stresses?

Wigwams (see Chapter 3) analysed the same 68 genes in an attempt to establish whether this EDISA module was a false negative, and should have been identified as co-expressed in all three stresses. Wigwams identified these genes as significantly co-expressed in all three stresses, as

shown in Table 1.1, which shows the *p*-values associated with each pairwise comparison of stresses.

Table 1.1: *p*-values associated with each pairwise comparison of stresses, calculated by Wigwams using the hypergeometric test

| Pairwise stress combination | *p*-value |
|---|---|
| *B. cinerea* and long day senescence | $8.13e^{-03}$ |
| *B. cinerea* and short day senescence | $2.51e^{-03}$ |
| Long day senescence and short day senescence | $3.98e^{-57}$ |

The disadvantages of EDISA include a lack of *p*-values assigned to the resulting modules, to state the significance of the level of co-expression observed. In addition, although EDISA is also capable of searching for three categories of modules, the user has to state which type of definition they wish to use for the analysis of the time series expression data. Therefore, EDISA does not provide evidence of all three definitions in one execution of the algorithm. EDISA is also limited to the number of genes the number of conditions and time points it capable of mining for biclusters. These drawbacks present an opportunity to develop a biclustering algorithm that does not possess the limitations on the gene expression data it is capable of mining for significantly co-expressed genes across multiple time-series datasets, which also presents evidence for coherent and independent responses.

### 1.3.4 Comparison of the theoretical methods available for the construction of GRNs

A number of theoretical methods, which can be used to infer and predict GRNs from gene expression data, have been described in this chapter. To summarise, these findings are presented in Table 1.2 below, which describes the limitations and advantages of modelling techniques (VBSSM (30)), clustering techniques (TCAP (151)) and biclustering approaches (EDISA (272) and BIGA (273)).

VBSSM (30), the dynamic Bayesian modelling approach to time-series expression data, is extremely advantageous in that it can identify combinatorial regulation and feedback loops. However, this approach is limited to modelling approximately 70 genes over approximately 20 timepoints (with fewer timepoints, more genes can be modelled). This obviously raises the issue of selection of genes to initially model, which is a major drawback. Whilst VBSSM has the advantage over other DBNs in that it can infer the direction of regulation (30, 29, 231), it cannot, however, distinguish between direct and indirect regulation. Figure 1.7 illustrates how the transcription factor encoded by gene *A* indirectly regulates gene *C* via gene *B*. VBSSM, however, may not interpret

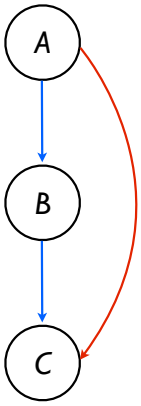



Figure 1.7: VBSSM cannot distinguish between direct and indirect regulation. Gene *A* encodes a transcription factor which regulates the expression of gene *B*. Gene *C* is regulated by the transcription factor encoded by gene *B*. Therefore, the gene *A* product indirectly regulates gene *C*, as shown by the blue arrows. However, VBSSM may not distinguish gene *B* as an intermediate, and would instead provide an output where the transcription factor encoded by gene *A* directly regulates gene *C*, as shown by the red arrow. Therefore, when observing output from VBSSM, it is important to consider that inferred regulation can be direct or indirect.

this indirect regulation, and would instead provide an output where the transcription factor encoded by gene *A* is seen to directly regulate gene *C*. Therefore, when observing output from VBSSM, it is important to consider that inferred regulation can be direct or indirect. It is important to also consider that although VBSSM infers the direction of regulation, this is still a predicted inference, and would need experimentation to be confirmed.

TCAP (151) is a temporal clustering approach for time series gene expression data, which can take time delays and inversions into account, and therefore infer regulatory relationships. This method is computationally inexpensive, and was shown to correctly cluster together members of gene regulatory network (151). TCAP also provides candidate genes which suggest hypothesises to analyse experimentally. However, TCAP can only analyse one condition dataset at a time, so therefore cannot identify regulatory relationships across multiple conditions.

Standard hierarchical clustering approaches usually achieve a lack of robustness when analysing noisy gene expression data (187). SOTA, however, was developed to overcome this issue, and gave robust clusters when analysing noisy gene expression data (86). Despite this, SOTA still experiences the same limitations as other hierarchical clustering methods, namely that any results generated are difficult to interpret, due to being unable to re-evaluate results (187).

Both EDISA (272) and BIGA (273) are biclustering approaches for gene expression data. EDISA outperformed its predecessor, ISA (132, 131), when it was developed in 2007, and was seen to be the leading biclustering algorithm at that time, as this approach allowed for a comprehensive view of gene expression responses in different conditions by identifying three biologically relevant module types (coherent, independent and single response modules) (272). However, EDISA is based on calculating the similarity between two genes, using Pearson correlation coefficient $r$, to identify co-expressed genes, and does not take into account time-delayed or inverted expression profiles, which would infer regulation. BIGA (273) aimed to overcome the reliance on similarity measures in biclustering methods by employing an iterative genetic algorithm to search for biclusters. However, BIGA is computationally intensive, and does not utilise time-series data, ruling out its use on the PRESTA datasets.

From Table 1.2 it can be seen that each approach to generating a GRN has its limitations. Therefore, it seems logical to perform analyses using more than one method to uncover biological relationships (273).

### 1.3.5 Integration of gene expression data and promoter motif data

The previous sections have described some current methods available for generating hypothesised regulatory links between genes, and subsequently form a transcriptional network, from transcriptome data, as well as methods for identifying co-expressed genes in single and multiple stresses. However, networks generated by modelling techniques do not discriminate between direct and indirect regulation, nor the direction of regulation. Although experimental techniques such as ChIP-Seq would enable this distinction to be made, these techniques are expensive and time-consuming (see Section 5.1 for a more detailed comparison of ChIP-based methods, and their drawbacks). Promoter motif prediction methods would identify transcription factor binding motifs located in the promoter sequences of genes, and assist the understanding of transcription regulation (211). By integrating data from transcriptional network modelling and promoter motif analysis, direct and indirect targets of transcription factors can be clarified.

However, promoter motif analysis does not provide information on the expression changes that occur when a transcription factor is bound to the promoter of a target gene: a large amount of genes with a regulatory transcription factor bound respond transcriptionally by altering the level

Table 1.2: Comparing various theoretical approaches available for elucidating GRNs using gene expression data, highlighting input data needed; whether the approaches can utilise single or many datasets; and limitations and advantages of each method

| | Type of method | Input data | Single or multi-dataset | Limitations | Advantages |
|---|---|---|---|---|---|
| VBSSM (30) | Dynamic Bayesian network modelling using hidden states | Time-series gene expression data | Single | Limited to modelling 70 genes over 20 time points, computationally intensive | Identifies combinatorial regulation and feedback loops |
| TCAP (151) | Time-series clustering | Time-series gene expression data | Single | Not as sensitive as biclustering methods to events occurring in short periods of time | Infers regulation, little user input, computationally inexpensive |
| EDISA (272) | Biclustering | Time-series gene expression data | Multi | Limited to 2500 genes over 3 datasets (total of 54 time points), no *p*-values associated with output, false negatives | Computationally inexpensive |
| BIGA (273) | Biclustering based on genetic algorithm (GA) | Gene expression data | Multi | High gene overlap (gene found in many biclusters), cannot utilise time-series data, computationally intensive | Large gene coverage |
| SOTA (86) | Hierarchical clustering | Gene expression data | Single | Cannot re-evaluated generated results | Achieves robustness with noisy data, 'top-to-bottom' approach |

of expression (211): it is thought that less than 10% of directly bound genes exhibit a significantly altered expression level (166). This suggests that binding of transcription factors to their respective targets may be silent, which leads to another avenue of investigation to discover the function of these silent regulatory events.

## 1.4 Aims

This thesis will present the method development of a tool to analyse multiple large-scale time series datasets in order to identify genes important to stress responses in *A. thaliana*. The aims of each chapter are as follows:

**Chapter 3 - Method development of Wigwams (Wigwams Identifies Genes Working Across Multiple Stresses)** To develop a new method which is capable of discovering groups of co-expressed genes in subsets of multiple time series gene expression datasets, which are hypothesised to also be co-regulated. These groups of co-expressed genes are predicted to have a shared regulatory mechanism, which is activated in multiple stresses. This method will be validated by identifying known examples of co-expressed and co-regulated genes, as well as being applied to a simulated dataset of randomly generated data.

**Chapter 4 - Applying Wigwams to multiple time course gene expression datasets** To apply the Wigwams to high resolution time series microarray datasets generated by the PRESTA group to investigate the effect of stress on gene expression changes over time. Wigwams will identify potential regulons found to be significantly co-expressed in subsets of these datasets. By using bioinformatical analyses, such as motif analysis and gene ontology (GO) term analysis, it can be inferred whether these potential regulons have a common regulator or function, respectively, and are therefore likely to be truly co-expressed and co-regulated.

**Chapter 5 - Biological validation of a theoretically predicted gene network** To integrate modelling techniques with potential regulons identified as significantly co-expressed over time, using the PRESTA time series datasets, in order to find potential transcription factor regulators of predicted co-expressed and co-regulated genes. Experimental validation of these predictions will be established using high throughput matrix Yeast 1-Hybrid (Y1H) and microarray analysis. These results will subsequently be used to produce a common gene regulatory network which is activated

in multiple plant stress responses.

# Chapter 2

# Experimental methods for the elucidation of gene regulatory networks

## 2.1 Growth conditions of *A. thaliana* plants

*A. thaliana* plants were grown in a climate controlled environment with the following conditions: constant temperature of 20 °C; fluorescent tungsten irradiation of ~110 $\mu$mols photons·$m^{-2}$·$s^{-1}$; 8:16 hour light: dark cycle; and a relative humidity of 70%. *A. thaliana* seed was stratified for three days in 500 - 1000 $\mu$L of sterile 0.1% (w/v) agar at 4 °C. Stratified seed was sown on soil suitable for *A. thaliana*, mixed to a ratio of six parts Scotts Levingtons F2s compost, one part silica sand and one part fine grade vermiculite (Horticultural Services, Wellesbourne Campus, University of Warwick).

## 2.2 Phenotyping screens of *A. thaliana* mutant lines

### 2.2.1 Phenotyping screens of *A. thaliana* using the necrotrophic fungal pathogen *B. cinerea*

#### 2.2.1.1 *B. cinerea* growth and isolation

*B. cinerea* strain pepper (80) was subcultured biweekly onto sterile tinned apricot halves (Tesco's own) in deep petri dishes. Subcultures were incubated in constant darkness at 25 °C. *B. cinerea*

spores were isolated two weeks post subculture for an infection. Spores were harvested from subcultures in 3 mL of sterile dd$H_2O$, and were separated from fungal hyphae by pipetting the water containing the spores into a sterile syringe, which contained glass wool. The spores were then filtered through the glass wool. The number of spores per mL was calculated by pipetting 17 $\mu$L onto a haemocytometer slide, and counted using a light microscope. Spore suspensions were prepared by diluting *B. cinerea* spores to a concentration of 1 x $10^5$ spores/mL, using sterile half strength grape juice (Tescos own).

#### 2.2.1.2 *B. cinerea* infection of *A. thaliana* leaves

After 4 weeks (28 days) of growth, one leaf per plant was detached and placed on 800 mL of 0.8% agar that had set in the base of a propagator tray. The leaves were inoculated with one 10 $\mu$L droplet of either a mock or *B. cinerea* spore suspension. Mock solution consisted of sterile half strength grape juice. *B. cinerea* spore solution similarly consisted of half strength grape juice, but also contained *B. cinerea* spores pepper strain. Leaves were covered with propagator lids and placed in a Sanyo SGC970 growth cabinet, set to a constant temperature of 20 °C, fluorescent tungsten irradiation of ~120 $\mu$mols photons·$m^{-2}$·$s^{-1}$ (16 hour light-dark cycle with the light period starting at 2am and ending at 6pm), 350 ppm $CO_2$ and 70% relative humidity. Digital photographs of each tray were taken at 24, 48, and 72 hpi (hours post infection).

#### 2.2.1.3 Analysis of *B. cinerea* infection

Photographs taken at each timepoint were analysed using image analysis software ImageJ (4) in order to determine the lesion area of each leaf. The average lesion area of 20 leaves from individual plants of T-DNA knockout or overexpressor lines were compared to Col-0 wild type (or other suitable background to mutation) using a two-tailed t-test, assuming equal variance. Phenotypes were noted that had a *p*-value < 0.05 at 72 hpi.

### 2.2.2 Dark induced senescence phenotyping screen of *A. thaliana* plants

#### 2.2.2.1 Senescence screen

Three replicate petri dishes per line (labelled A, B and C) were prepared, with filter paper in the lid, wet with 3 mL sterile dd$H_2O$. Nine representative rosettes per line were cut at the base, along the

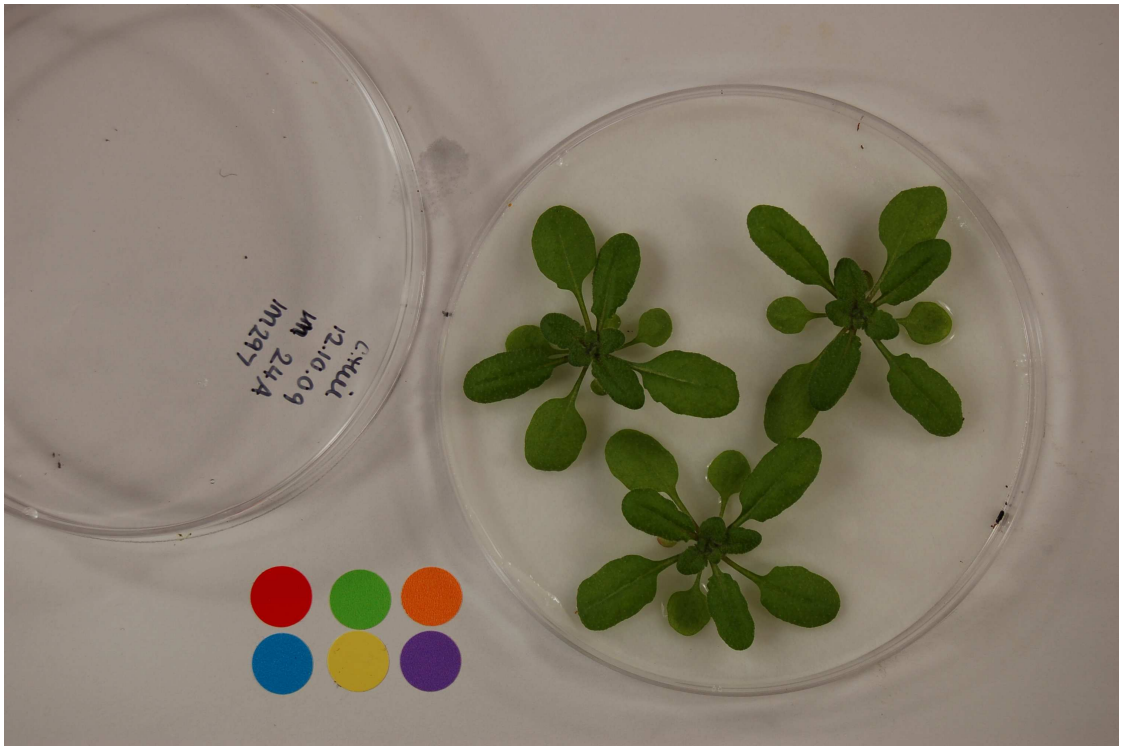


Figure 2.1: Image of the arrangement used in senescence screens: three *A. thaliana* rosettes are arranged on a petri dish and stored in complete darkness. In order to measure the progression of senescence over time, petri dishes were placed on a piece of white paper and photographed in the same position each day.

level of the soil, using scissors and arranged three per plate. Plates were stored in complete darkness, in a plastic container with a lid, in the 20 °C room to maintain constant temperature. A pure white piece of paper was prepared with a selection of coloured spots, measuring approximately 1 cm in diameter per spot. A circle was drawn onto the paper to ensure subsequent dishes can be placed in the same position. This arrangement is illustrated in Figure 2.1. Digital photographs of each plate were taken every day until the all the rosettes were completely yellow in a photo studio using an incandescent light rig.

#### 2.2.2.2 Analysis of senescence screen

Image analysis software ImageJ (4) was used to analyse the photographs and generate red/green ratios for each line compared to wild type. For each photo, the RGB values for a circle in a white area of the paper was measured. The RGB values of leaf 5 on each rosette were also measured. RGB values were obtained using the 'colour histogram' option within ImageJ.

The RGB values were transformed using an R (279) script written by Stuart McHattie (University of Warwick, unpublished). This script normalised the data using the RGB values for the white

area of the paper, and transformed over the whole image. Subsequently, the red/green ratios for each leaf were generated using a second R (279) script written by Stuart McHattie (University of Warwick, unpublished), which provide a ratio indicating the degree of senescence. The ratio between the average red and the average green measurements of a leaf allows a direct comparison of senescence between two rosettes in the experiment. A ratio of one is approximately 50% of the way through to full senescence.

The ratios are used to plot curves of the progress of senescence against time. Where a difference between the curves of mutant and wild type lines can be seen by eye, this indicates an altered phenotypical response compared to the control.

## 2.3 Cloning

### 2.3.1 Promoter fragment cloning using Gateway technology

Oligonucleotides were designed to amplify overlapping promoter fragments of approximately 400 bp, to cover 1000 bp upstream of the transcriptional start site (TSS), with 100 bp overlap between fragments. *att*B sites needed to be incorporated into these oligonucleotides (see Appendix A), and are shown by lower case letters in the oligonucleotides (see Appendix A, Table A.1).

Promoter regions were amplified from genomic DNA (Col-0) using the aforementioned oligonucleotides and KOD polymerase master mix (Roche, Welwyn) according to the manufacturer's instructions using a two-step PCR approach (PCR conditions can be found in Tables 2.1 and 2.2). Amplification of template (genomic) DNA was completed using specific primers (shown in Table A.1, Appendix A). Step 2 of the PCR involved amplifying the product of the first step using universal *att*B adapter primers (shown in Table A.1, Appendix A). 5 $\mu$L of PCR product was loaded onto 2% agarose gels to confirm the size of the product. Once the correct product size was confirmed, the PCR products were purified with the QIAquick PCR purification kit (Qiagen, West Sussex) according to the manufacturer's instructions, to remove any remaining *att*B primers or *att*B primer-dimers.

Once the promoter fragments were purified, entry clones were created using the recombination reaction method. The BP recombination method, illustrated in Figure 2.2, allowed the transfer of the promoter fragment in the *att*B PCR product into an *att*P containing vector. In this case, the

Table 2.1: PCR conditions for multistep Gateway cloning using KOD master mix: Step 1

| | | |
|---|---|---|
| 95 °C | 2 minutes | |
| 95 °C | 15 seconds | |
| 55 °C | 15 seconds | 11 cycles |
| 68 °C | 2 minutes | |

Table 2.2: PCR conditions for multistep Gateway cloning using KOD master mix: Step 2

| | | |
|---|---|---|
| 95 °C | 2 minutes | |
| 95 °C | 15 seconds | |
| 45 °C | 15 seconds | 5 cycles |
| 68 °C | 2 minutes | |
| 95 °C | 15 seconds | |
| 55 °C | 15 seconds | 35 cycles |
| 68 °C | 2 minutes | |
| 68 °C | 5 minutes | |
| 4 °C | 15 minutes | |

*pDONRZeo* entry vector (Invitrogen, Paisley) was used. The recombination reaction, took place at 25 °C overnight, and contained 1 $\mu$L of BP Clonase enzyme mix (Invitrogen, Paisley), 1 $\mu$L of *pDONRZeo* entry clone (concentration 150 ng/$\mu$L), 1 $\mu$L *att*B PCR product (final concentration 150 ng/$\mu$L) and 2 $\mu$L of sterile dd$H_2O$.

The BP reaction mixture was then transformed into gold standard DH5$\alpha$ competent *E. coli* cells (Bioline, London). Competent cells were defrosted on ice for 10 minutes. 1 $\mu$l of BP reaction was added to 10 $\mu$l of competent cells, and incubated on ice for 30 minutes. The cells were then heat shocked at 42 °C for 30 seconds, followed by incubation on ice for 2 minutes. 250 $\mu$l of SOC media (see Table 2.3) was added to the cells, which were then incubated, with shaking, at 37 °C for an hour. These cells, which contain the *pDONRZeo* entry vector (Invitrogen, Paisley) containing the promoter fragment of interest, were plated onto selective media containing Zeocin at a concentration of 30 ng/$\mu$L, in order to select for cells containing the entry clone, and incubated overnight at 37 °C.

The plasmids of successfully transformed cells are then extracted from overnight cultures grown in liquid selective media using the QIAprep® spin miniprep kit (Qiagen, West Sussex) and sequenced (see Table A.3 for sequencing oligonucleotides) by GATC Biotech (Konstanz, Germany).

Once the entry clone had been successfully generated and confirmed as being correct using sequencing, the LR recombination reaction, illustrated in Figure 2.3, was performed to transfer the

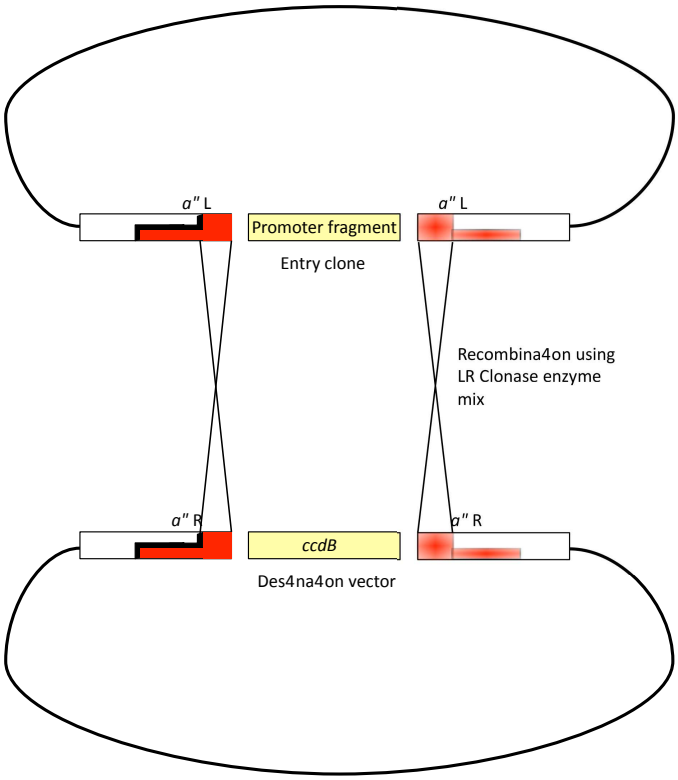



Figure 2.2: The BP reaction, catalysed by BP Clonase enzyme mix, allows the recombination of an *att*B PCR product with an *att*P donor vector to create an *att*L entry clone. Oligonucleotides designed to amplify promoter fragments have the *att*B sites incorporated (see Appendix A, Table A.1).

Table 2.3: SOC media

| Reagents (Sigma-Aldrich, Gillingham) |
|---|
| 2% (w/v) bacto-tryptone (20 g) |
| 0.5% (w/v) bacto-yeast extract (5 g) 8.56 mM NaCl (0.5 g) |
| 2.5 mM KCl (0.186 g) |
| 10 mM $MgCl_2$ (0.952 g) |
| 20 mM glucose (3.603 g) |
| dd$H_2O$ to 1000 mL |

promoter fragment into an *att*R-containing destination vector. The *pHis2Leu2* vector was converted into a Gateway compatible destination vector using the Gateway Vector Conversion System (Invitrogen, Paisley), and will now be referred to as *pHis2Leu2GW* (conversion by Claire Hill, University of Warwick). The recombination reaction took place, according to manufacturer's instruction, at 25 °C overnight, and contained 1 $\mu$L of LR Clonase enzyme mix (Invitrogen, Paisley), 1 $\mu$L of *pDONRZeo* entry clone (concentration 150 ng/$\mu$L), 1 $\mu$L destination vector *pHis2Leu2GW* (final concentration 150 ng/$\mu$L) and 2 $\mu$L of sterile dd$H_2O$.

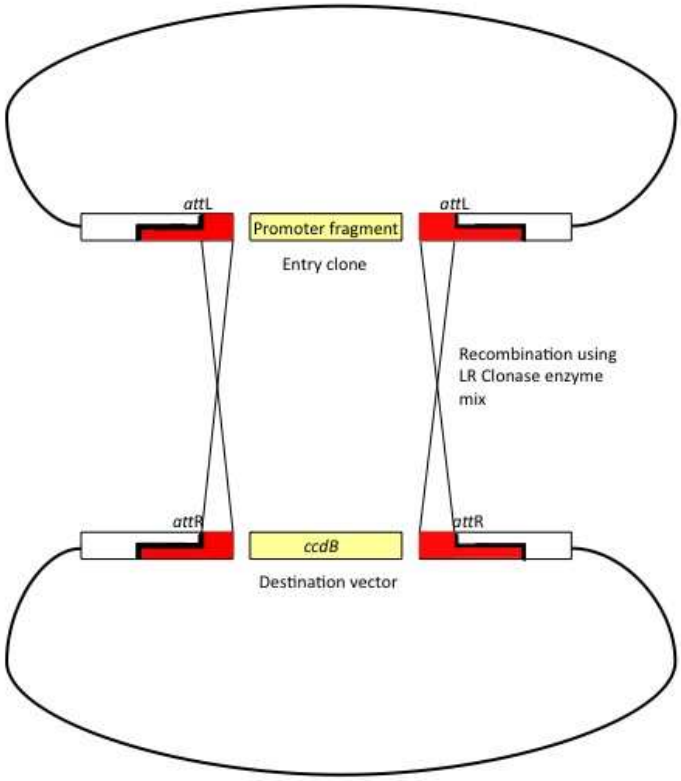



Figure 2.3: The LR reaction, catalysed by LR Clonase enzyme mix, allows the recombination of an *att*L entry clone containing the promoter fragment of interest, with an *att*R substitute (destination vector) to create an *att*B expression clone.

The LR reaction mixture was then transformed into gold standard DH5$\alpha$ competent *E. coli* cells (Bioline, London) using the same method as described for transforming the BP reaction. Cells containing the destination vector *pHisLeu2GW*, which contains the promoter fragment of interest, are plated onto selective media containing Kanamycin at a concentration of 50 ng/$\mu$L, in order to select for cells containing the destination clone, and incubated overnight at 37 °C. Plasmids from successfully transformed cells were extracted and sequenced as described above.

## 2.4 Yeast-1-Hybrid (Y1H) (307)

Differential gene expression drives development, and is mainly controlled by transcriptional regulation, via transcription factors binding to *cis*-regulatory elements in promoters of genes (182). By using microarray technology to capture the expression profiles of genes over time, and then using clustering analyses to group genes with similar profiles together we can identify genes involved in common biological processes (88). It can be hypothesised that if genes have a shared expression

profile then they also may share elements in their promoters, such as transcription factor binding motifs (254). Therefore, it is likely that these genes would be regulated by common upstream transcription factors. This is termed as 'co-regulation', and the clustered genes are referred to as 'co-expressed' genes (254). Techniques have been developed (ChIP, DNase I foot-printing) to establish whether certain genes form a co-ordinated response. Y1H allows the investigation of which transcription factors bind to a particular section of a promoter (182). Additionally, Y1H is more applicable for discovering regulatory elements within the promoter regions of genes that may have been predicted bioinformatically using motif analysis, for example (182). Y1H is primarily used to investigate protein-DNA interactions, rather than protein-protein interactions, which are found using Y2H (308).

In the last decade, Y1H has been made compatible with Gateway technology (81), which reduces the need for a transcription factor library biased towards highly expressed genes, seen in libraries constructed using cDNA from total RNA. Cloned libraries using the Gateway technology can be pooled to make the process high-throughput.

### 2.4.1 Transformation of yeast strain Y187 with *pHisLeu2GW* plasmid

Yeast *Saccharomyces cerevisiae* strain Y187 was grown overnight in 10 mL of yeast peptone dextrose adenine (YPDA) liquid media at 30 °C at 400 *g*. 1 mL of this culture (sufficient for ten transformations) was centrifuged at 400 *g* for 5 minutes. Cells were resuspended in 1 mL of 0.1 mM LiAc, centrifuged a second time, and resuspended again as before. Cells were then incubated in a 30 °C water bath for one hour.

3 $\mu$L of the *pHis2Leu2GW* plasmid containing a promoter fragment (concentration between 500ng and 1000ng) was added to 4 $\mu$L of boiled single-stranded carrier DNA (Clontech, Saint-Germain-en-Laye, France), and mixed with 290 $\mu$L of 50% (v\v) PEG 3350. This mixture was heated to 30 °C. 100 $\mu$L of cell suspension was added to the PEG mix and incubated in a 30 °C water bath for 50 minutes. Cells were subsequently heat shocked for 15 minutes at 42 °C, followed by centrifugation at 1000 *g* for five minutes. The resulting supernatant was removed, and pelleted cells were re-suspended in sterile water and plated onto SD-Leu (minimal SD and amino acid DO supplements from Clontech, Saint-Germain-en-laye) agar plates. These plates were incubated at 30 °C until colonies appeared.

Table 2.4: YPDA liquid media

| Reagents |
|---|
| 20 g glucose |
| 20 g peptone |
| 10 g yeast extract |
| 100 mg adenine |
| up to 1 L dd$H_2O$ |

### 2.4.2 Transcription factor library

The transcription factor library was generated by Dr. Claire Hill and Alexandra Tabrett (both University of Warwick), as described in Windram (2010) (323). The libraries contain 1037 transcription factor clones (808 full length ORFs and 229 partial ORFs) with N-terminal GAL4 activation domain fusions in the *pDEST22* vector (Invitrogen, Paisley), which were provided by Franziska Turck (University of Cologne). Using Gateway technology (Invitrogen), an additional 332 full length ORFs in the *pDEST22* vector were generated by Claire Hill and Alexandra Tabrett (both University of Warwick). The vectors were transformed into the yeast *S. cerevisiae* strain AH109. The libraries are arranged so 24 clones are pooled into each well in a 96-well plate. There are two alternative arrangements, to account for spatial bias, giving two 96-well plates in total.

### 2.4.3 Transcription factor clone transformation

The *S. cerevisiae* $\alpha$ strain was kindly provided by Claire Hill (University of Warwick). For individual one-on-one Y1H screens (as outlined in Sections 2.4.6 and 2.4.7), the *pDEST22* (Invitrogen, Paisley) plasmid, kindly provided by Alison Jackson (University of Warwick), containing a *Rap2.6L* cDNA clone, was transformed into an $\alpha$ strain of yeast, AH109. Transformations took place as described in Section 2.4.1, with the exception to the yeast strain used being AH109 as opposed to Y187 strain and cultures were grown on SD-Trp (minimal SD and amino acid DO supplements from Clontech, Saint-Germain-en-Laye, France).

A *pDEST22::GFP* plasmid, kindly provided by Steve Kiddle (University of Warwick), was used as a control and was transformed into AH109 as mentioned above.

### 2.4.4 Transcription factor library subculture

For each 96-well glycerol stock library, 500 $\mu$L of SD-Trp (minimal SD and amino acid DO supplements from Clontech, Saint-Germain-en-Laye, France) was added to each well in a 2.2 mL deep 96-well plate (see Table 2.5). The 96-deep well replicator (V and P Scientific Inc., San Diego) was sterilised using 70% ethanol and used to subculture the library onto the plate containing SD-Trp (minimal SD and amino acid DO supplements from Clontech, Saint-Germain-en-Laye, France). These plates were sealed with a gas permeable cover, and incubated in a 30 °C incubator for 96 hours on a vigorous shaker to promote yeast growth.

Table 2.5: SD-Trp liquid media

| Reagents |
|---|
| 26.7 g minimal SD base |
| 0.74 g -Trp DO supplement |
| up to 1 L dd$H_2O$ |

### 2.4.5 Matrix high-throughput Y1H by mating and auxotrophic selection

Cultures of yeast strain Y187 that had been transformed with the *pHis2Leu2GW* plasmid containing promoter fragments were grown up in 10 mL of SD-Leu (minimal SD and amino acid DO supplements from Clontech, Saint-Germain-en-Laye, France) overnight (see Table 2.6) at 30 °C at 400 $g$. 3 $\mu$L of overnight culture was spotted onto a YPDA master plate in a 96-well plate matrix layout. Each spot contained 3 $\mu$L of yeast strain Y187 that had been transformed with the *pHis2Leu2GW* plasmid containing promoter fragments in SD-Leu. Once the spots were dry, 3 $\mu$L of each well of the transcription factor library was spotted on top of the Y187 overnight culture spot, to allow mating to occur. Plates were incubated overnight at 30 °C.

Table 2.6: SD-Leu liquid media

| Reagents |
|---|
| 26.7 g minimal SD base |
| 0.69 g -Trp DO supplement |
| up to 1 L dd$H_2O$ |

YPDA master plates (see Table 2.7) were replica plated using velvets the following day onto the following plates: SD-LT (Table 2.8), SD-LTH (Table 2.9), and SD-LTH with various concentrations of 3AT (predominantly, 25 mM, 50 mM, and 100 mM concentrations were used). These

Table 2.7: YPDA plates. Poured in a lamina flow cabin and allowed to dry for 1 hour

| Reagents |
|---|
| 20 g glucose |
| 20 g peptone |
| 10 g yeast extract |
| 100 mg adenine |
| 18 g agar |
| up to 1 L dd$H_2O$ |

Table 2.8: SD-LT plates. Poured in a lamina flow cabin and allowed to dry for 1 hour

| Reagents |
|---|
| 26.7 g minimal SD base |
| 0.64 g -Leu/-Trp DO supplement |
| 18 g agar |
| up to 1 L dd$H_2O$ |

plates were incubated overnight at 30 °C. The following day, each plate was replica cleaned using three velvets per plate. The plates were then incubated at 30 °C for 96 hours.

Photos of each plate were taken using a G:BOX imager (SynGene, Cambridge) after 96 hours incubation, using upper white light, in order to visualise any growing colonies. Successfully growing colonies were patched onto new plates of the same selection as the original growth had been observed, in order to confirm colony growth.

Colonies growing on SD-LTH and SD-LTH 3AT agar plates were picked into 10 $\mu$L of 20 mM NaOH. Colony PCR was performed using Taq polymerase (see Table A.4 for primers; Taq from Invitrogen, Paisley), according to the manufacturer's instructions. PCR products were cleaned up using a MultiScreen HTS PCR 96-well plate (Millipore, Watford) according to the manufacturer's instructions. Cleaned PCR products were sequenced by GATC Biotech (Konstanz, Germany) in order to identify the interacting transcription factors using the forward oligonucleotide in Table A.4.

### 2.4.6 Individual one-on-one Y1H by mating and auxotrophic selection

Individual transcription factor-promoter interactions can be tested one-on-one. Overnight cultures of *S. cerevisiae* strain Y187, which have already been transformed with a promoter fragment containing the *pHis2Leu2GW* plasmid, were grown in 10 mL of SD-L. These cultures were incubated

Table 2.9: SD-LTH plates. Poured in a lamina flow cabin and allowed to dry for 1 hour

| Reagents |
|---|
| 26.7 g minimal SD base |
| 0.62 g -Leu/-Trp/-His DO supplement |
| 18 g agar |
| up to 1 L dd$H_2O$ |

at 30 °C on a vigorous shaker. As a control, *S. cerevisiae* strain AH109 was transformed with *pDEST22::GFP*. Overnight cultures of AH109 were made using 10 mL of SD-Trp liquid media (see Table 2.5), and incubated at 30 °C at 400 g. 3 $\mu$L of Y187 culture was spotted onto YPDA plates (see Table 2.7) in a 96-well grid format. Once dry, 3 $\mu$L of AH109 culture was spotted on top of the Y187 spot, and allowed to dry. Plates were incubated overnight at 30 °C for mating to occur.

The following day, YPDA plates were replica plated using velvets onto the following plates: SD-LT (Table 2.8), SD-LTH (Table 2.9), and SD-LTH with various concentrations of 3AT (predominantly, 25 mM, 50 mM, and 100 mM concentrations were used). These plates were incubated overnight at 30 °C. The following day, each plate was replica cleaned using three velvets per plate. The plates were incubated at 30 °C for 96 hours. Photos of each plate were taken using a G:BOX after 96 hours incubation, using upper white light, in order to visualise any growing colonies.

### 2.4.7 Individual Y1H by co-transformation and auxotrophic selection

Co-transformation is another method, like the method outlined in Section 2.4.6, that allows transcription factor-promoter interactions to be tested one-on-one (35). A yeast strain Y187, which has already been transformed with *pHis2Leu2GW* containing a promoter fragment, was transformed again with a *pDEST22* plasmid, which contains a transcription factor clone. As a control, a Y187 strain can also be co-transformed with a *pDEST22::GFP* plasmid to act as a control. This transformation was carried out as described in Section 2.4.1. Once yeast were transformed, cultures were grown in SD-LT liquid media (see Table 2.10).

An overnight culture of the co-transformed strain of Y187 was incubated at 30 °C on a vigorous shaker in SD-LT (see Table 2.10). The concentration of cells in the culture was determined using optical density and an $OD_{600}$ table. Cultures were adjusted to give a final concentration of $10^8$

cells per mL. Serial dilutions of each culture were made using 96-well plates, by taking 20 $\mu$L of culture and adding 180 $\mu$L of $H_2O$ and mixing. This step was repeated until a concentration of $10^4$ cells per mL was obtained. 3 $\mu$L of each dilution was spotted on the following plates: SD-Leu, SD-Trp, SD-LT, SD-LTH, and SD-LTH with various concentrations of 3AT (predominantly, 25 mM, 50 mM, and 100 mM concentrations were used). These plates were incubated at 30 °C for 2 to 3 days, or until colonies appeared. Photos were taken of the plates using the upper white light in a G:BOX.

Table 2.10: SD-LT liquid media

| Reagents |
|---|
| 26.7 g minimal SD base |
| 0.64 g -Leu/-Trp DO supplement |
| up to 1 L dd$H_2O$ |

## 2.5 Analysis of gene expression changes using microarrays

Microarray analysis using CATMA slides was carried out on plants overexpressing *Rap2.6L*, as well as the background control, *A. thaliana* ecotype WS (both kindly provided by Dr. Nataraj Kav, Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Canada). This line is originally described in (163).

### 2.5.1 RNA extraction

*A. thaliana* leaf samples were snap frozen in liquid nitrogen and stored at -80 °C. Leaf samples were ground for 1 minute in 2 mL tubes (Eppendorf) using a Dremel drill. tRNA was extracted from ground leaf tissue using TRIzol® (Invitrogen, Paisley) according to the manufacturer's instructions. Total RNA (tRNA) was purified using the RNeasy® Mini Kit (Qiagen, West Sussex), following the manufacturer's instructions, and eluted in 50 $\mu$L of DEPC-treated nuclease free water. The concentrations of each tRNA sample were measured using the nano drop. Approximately 1.5 $\mu$L of each tRNA sample was run on an Agilent 2100 bioanalyser using RNA 6000 Nano Chip kit (Agilent) to assess the integrity of the extracted RNA. Total RNA was amplified using the MessageAmp$^{TM}$aRNA Amplification Kit (Ambion Biosystems), according to the manufacturer's instructions.

### 2.5.2 RNA sample labelling and CATMA array hybridisation

Table 2.11: Pre-hybridisation buffer

| Reagents |
|---|
| 10 mg/mL bovine serum albumin (BSA) |
| 5x SSC |
| 0.1% (w/v) SDS |

Three amplified RNA (aRNA) samples per line were pooled. Approximately 5 $\mu$g of pooled aRNA was reverse transcribed, using 0.5 $\mu$L of random nonamers (2 $\mu$ g/$\mu$L) (Invitrogen), 0.5 $\mu$L of RNaseOUT$^{TM}$Inhibitor (Invitrogen), and nuclease free water added to give a total of 10.5 $\mu$L per sample. This solution was incubated at 70 °C for 10 minutes. 4 $\mu$L of 5x SuperScript$^{TM}$II First Strand Buffer, 2 $\mu$L 0.1 M of dithiothreitol (DTT), 1 $\mu$L of dNTP mix (10 mM dATP, 10 mM dGTP, 10 mM dTTP, 2 mM dCTP), 1 $\mu$L of SuperScript$^{TM}$II Reverse Transcriptase (all Invitrogen), and 1.5 $\mu$L of 25 nmol Cy3- or Cy5-dCTP (GE Healthcare) was added to the denatured aRNA samples. Labelling reactions were incubates for 2 hours and 30 minutes at 42 °C. Labelled samples were purified using the QIAquick® PCR purification kit (Qiagen), according to the manufacturer's instructions, and eluted in 30 $\mu$L of Elution buffer.

Table 2.12: Wash solution 1

| Reagents |
|---|
| 2x SSC |
| 0.1% (w/v) SDS |

40 *p*mol of purified labelled cDNA were mixed i.e. Cy3-*35S:Rap2.6L* with Cy5-WS, including a dye swap, and concentrated using freeze drying. The samples were re-suspended in 50 $\mu$L of pre-hybridisation buffer (see Table 2.11 for reagents). This mix was incubated for 5 minutes at 95 °C, before being applied to CATMA version four slides (8). Technical replicates were used to control for variation in dye, printing and spatial variations. Arrays were covered with a cover slip (Sigma Aldrich, Gillingham) and placed in a high humidity environment and incubated at 42 °C for 16 hours in a hybridisation oven.

Arrays were washed in three wash solutions: wash solution 1 (see Table 2.12), preheated to 42 °C for 5 minutes on a shaker; wash solution 2 (see Table 2.13) for 10 minutes on a shaker; and wash solution 3 (see Table 2.14) four times for 1 minute each on a shaker. Finally, arrays were briefly

Table 2.13: Wash solution 2

| Reagents |
|---|
| 0.1x SSC |
| 0.1% (w/v) SDS |

washed in isopropanol, and spun dry for 1 minutes at 2000 g.

Table 2.14: Wash solution 3

| Reagents |
|---|
| 0.1x SSC |

### 2.5.3 Array scanning

Arrays were scanned on a 428 Affymetrix scanner at wavelengths of 532 nm and 635 nm for Cy3 and Cy5 respectively. The 'gain' setting (laser power) was optimised to give the most favourable spot saturation and minimal background fluorescence. Both scans for Cy3 and Cy5 were combined and processed using ImaGene version 8.0 (BioDiscovery) in order to extract the raw intensity data values for each spot on each array. Six CATMA v4 slides were processed, as three samples were hybridised to six arrays in a pairwise manner.

### 2.5.4 Data processing of CATMA arrays

Expression values for individual spots were processed using LimmaGUI (318) in R (279). Data was normalised using print tip Loess within-array normalisation, which eliminates spatial effects that can occur during hybridisation. Between array quantile normalisation was also used to ensure the data was more evenly distributed. By fitting a linear model to the data, using the least square method, differentially expression genes in the *35S:Rap2.6L* and WS comparison were identified. A Benjamini and Hochberg FDR (34), to eliminate false positives, and a maximum $p$-value of 0.05 was applied to generate a list of differentially expressed genes.

## 2.6 Extraction of genomic DNA from *A. thaliana* plants

Genomic DNA was extracted from two *A. thaliana* cotyledons (see Section 2.1 for plant growth conditions) using the Extract-N-Amp$^{TM}$Tissue PCR Kit (Sigma Aldrich, Gillingham), according to the manufacturer's instructions.

## 2.7 Discovering differentially expressed genes in high-resolution time series datasets

In order to understand specific gene responses to stress, genome-wide perception of gene expression is required (227). Microarrays provide a means of investigating regulation and interactions of genes. An important output from microarray experiments is the identification of genes whose expression levels are differential in response to a condition of interest (87). By measuring gene expression in time series expression, a temporal process of expression change is measured, which is preferential to static, single time point experiments, which do not provide information for changes in differential gene expression over time (24).

The PRESTA group sought to investigate changes in *A. thaliana* gene expression in response to pathogen stress (by the necrotrophic fungal pathogen *Botrytis cinerea* and the hemibiotrophic bacterial pathogen *Pseudomonas syringae* pathovar tomato), senescence, drought and elevated light conditions (hereby referred to as 'high light' stress). The differential gene expression changes were analysed using custom CATMA (Complete Arabidopsis Transcriptome MicroArray) arrays. These custom-built arrays were designed to include high quality gene-specific sequence tags (GSTs) covering the majority of the *A. thaliana* genome. Each stress response was investigated as a temporal response, where samples were taken at equal time intervals to analyse the changes in gene expression. Thus, differential gene expression over time in response to each individual stress could be deciphered. The time points for each stress investigated are listed in Table 2.15. The time points of relevance (i.e. those which are deemed most important to observing changes in gene expression in response to a particular stress) are also listed in Table 2.15.

### 2.7.1 Experimental plans for capturing differential gene expression over time in response to stress

The microarray hybridisation experiments were designed using two-channel arrays to compare the samples (stress-treated and mock-treated) collected at each time point, and also included biological replicates for each combination of treatment and time point. The experimental designs for the *B. cinerea* and long day senescence time series are described in the Supplementary Information in Windram *et al.* (322) and Breeze *et al.* (43) respectively. The designs for the experiments

Table 2.15: Information of time points for each of the stress conditions investigated by the PRESTA group, including number of time points, scale (i.e. days or hours) of temporal resolution, the interval at which time points were taken, and number of replicates for each experiment. The timepoints of relevance are also listed for each stress condition, deciphered from personal research.

| | *B. cinerea* | *P. syringae* | Drought | High light | Long day senescence | Short day senescence |
|---|---|---|---|---|---|---|
| Number of time points | 24 | 13 | 13 | 12 | 11 | 19 |
| Scale | hours | hours | days | hours | days | days |
| Interval | 2 hrs | 1 hr | 1 day | 30 mins | 1 day | 1 day |
| Timepoints of relevance | 24-48 hrs | 1-5 days | 10 days + | 1-5 hours | 5 days + | 13 days + |
| Mock samples? | Yes | Yes | Yes | Yes | No | No |
| Number of replicates treated | 4 | 4 | 4 | 4 | 4 | 4 |
| Number of replicates mock | 4 | 4 | 4 | 4 | N/A | N/A |

investigating the responses to *P. syringae*, drought, high light and short day senescence follow similar designs to *B. cinerea* and long day senescence, whereby a complex loop design allowed the comparison between time points with biological replicates for the stress-treated samples and mock-treated samples separately. A second design allowed the comparison between stress-treated and mock-treated samples, with further comparisons between neighbouring time points. This design guaranteed that combinations of stress-treated, mock-treated and time points were replicated equally over the whole experiment, and allowed the prominent effects of stress-treatment and time point to be reliably seen.

### 2.7.2 7036 genes are differentially expressed between high light stress and mock-treated *A. thaliana* plants over time

From the MAANOVA (326) high light data output every 100th gene, based on GP2S (270). ranking, was plotted and false positives marked. A rough cut-off was decided at which false positives seemed to increase. The 1000 genes either side of this cut-off were plotted individually, and false positives marked. From the false positives marked, a threshold for the GP2S ranking was decided, which was 7154 genes. This number was rounded up to 7200 genes. This list, containing genes considered to be differentially expressed was then compared to the list of genes found to be significantly differentially expressed over time by the F-Test (326). 76 genes were considered to be differentially expressed in the F-Test list, but not in the GP2S list, and were checked by eye to decide whether they should be included in the final list of differentially expressed genes in re-

sponse to high light stress. These 76 genes were deemed to be differentially expressed, and so were included in the list. 240 probes did not hybridise to an open reading frame when the CATMA probes were mapped to *A. thaliana* gene loci. These probes were removed, as were probes found to duplicate a gene matched by another differentially expressed probe. This yielded a list of 7036 genes differentially expressed in high light stress over time.

### 2.7.3 1761 genes are differentially expressed between drought stress and mock-treated *A. thaliana* plants over time

The same process was used as described in Section 2.7.2 to identify genes differentially expressed in drought stress over time. From the false positives marked, a threshold for the GP2S ranking was decided, which was 1600 genes. 975 genes were considered to be differentially expressed in the F-Test list of genes identified as differentially expression over time, but not in the GP2S list. 814 probes did not hybridise to an open reading frame when the CATMA probes were mapped to *A. thaliana* gene loci. These were removed, as were probes which were found to duplicate a gene matched by another differentially expressed probe. This yielded a list of 1761 genes differentially expressed in drought stress over time.

### 2.7.4 Using Gene Ontology (GO) term analysis to identify overrepresented biological functions in gene clusters

BiNGO (191) is a Java-based tool used to determine which GO terms are statistically overrepresented in a set of genes. It is implemented as a plugin for Cytoscape (72).

To analyse the genes for significantly overrepresented GO terms, the AGI identifiers of these genes were copied and pasted into the BiNGO GUI (version 2.44), a plugin in Cytoscape version 2.8.1. These genes were given a name, in the 'Cluster name' input box. Overrepresentation was tested, and 'no visualisation' was selected. Overrepresentation was tested using a hypergeometric test. *p*-values were corrected using Benjamini and Hochberg FDR. A maximum *p*-value of 0.05 was applied to generate a list of GO terms that were deemed significantly overrepresented.

Wigwams was completed only on genes that are differentially expressed in two or conditions. Therefore, the universe used in the GO term analysis could not be the standard universe of the whole *A. thaliana* genome. A custom annotation file, containing only the GO term identifiers

of the genes differentially expressed in two or more conditions, was used as the reference set. This reference set was chosen as we are only interested in the overrepresentation of functional categories in the clusters with respect to genes that are differentially expressed in two or more conditions, which is the same universe used initially when running Wigwams on differentially expressed genes.

In instances where the genes being tested for significantly overrepresented GO terms were not produced using Wigwams, the whole genome annotation was used as the reference set.

The organism 'Arabidopsis thaliana' was chosen in all instances, and all GO terms were used for the ontology files to search for overrepresented GO terms.

### 2.7.5 Motif analysis of promoter regions

#### 2.7.5.1 Hypergeometric test for overrepresentation of promoter motifs

The method outlined in this section was originally described in Baxter *et al*. (28).

For a given promoter, the 500 bp sequence upstream of the transcriptional start site (TSS) is examined, and a matrix similarity score (147) is generated for both strands, at each position, for all PSSMs, as originally described in (43). A third order Markov model is trained on the whole *A. thaliana* genome, which generates a random sequence of 100 million bases in length. Subsequently, *p*-values generated for each promoter are computed from a score distribution, which is obtained by applying the PSSM to the aforementioned randomly generated sequence.

A binomial test is performed for the occurrence of the top $k$ non-overlapping sites with observed $n$ values within a sequences of length 500 bp upstream of the TSS. The parameter $k$ is optimised within the range 1 to 5 for a minimum binomial *p*-value, which allows the detection of motifs without a fixed threshold per motif.

Using a threshold of $p < 0.05$, the presence or absence of a PSSM is scored for each promoter, based on the binomial probability. The frequency of each PSSM in promoters of genes in each cluster is compared with the frequency of occurrence of each PSSM in all promoters in the entire genome. Motif enrichment is calculated using the hypergeometric distribution (phyper function in the R stats package (279)). Hypergeometric *p*-values are corrected for the number of clusters

tested using Bonferroni correction. Corrected $p$-values $< 0.05$ are considered significant, indicating that associated PSSM is statistically overrepresented within the promoters of that particular cluster. Sequence logos are generated using code modified from Lenhard and Wasserman (167). All sequence analysis is performed within the APPLES software framework (28).

**Using the hypergeometric motif analysis tool** The hypergeometric motif analysis method mentioned in the previous section was subsequently developed as a web tool, and was the primary means of analysing promoter sequences for statistically significant overrepresentation of motifs in this thesis. In order for any user to analyse the promoters of genes found to be co-expressed for statistical overrepresentation of known plant promoter motifs, a tab separated file containing the cluster ID, followed by the AGI identifier of a gene on each line was uploaded. A maximum promoter length of 500 bp was selected, along with *A. thaliana* as the species in which to test. The motif clustering threshold was set to 10, which reduces the redundancy among known promoter motifs. A maximum of five motif occurrences in a single promoter was selected.

For the hypergeometric testing, all genes were considered for the universe size (i.e. the whole *A. thaliana* genome). Finally, as we were interested in statistically significant overrepresentation of promoter motifs, the option for large overlap testing was selected. Significantly large overlaps signify that a given motif is significantly overrepresented in a promoter, which could not have occurred by chance.

# Chapter 3

# Genome-wide inference of shared regulatory mechanisms from multiple gene expression time-series

DNA microarrays are a high-throughput technology which allows the concurrent monitoring of gene expression levels for thousands of genes. The transcriptome of an organism can be analysed under various conditions, providing evidence to reveal the genes involved in regulating the response to specific stresses. By clustering gene expression data generated by these microarrays, one can elucidate patterns within the dataset, with a view to understanding the regulation of genes with similar expression (138).

Clustering methods partition data into groups based on a certain attribute. In terms of clustering gene expression data, this attribute is usually similar expression profiles, as shown in Figure 3.1 c), where two genes are shown to have highly correlated expression profiles (32, 88, 299). This level of correlation is determined by a similarity measure, such as Euclidean distance. Therefore, members of a group will be more similar to each other than to members of another group. Multi-clustering methods, which allow clustering over multiple gene expression datasets, also partition the data based on a defined attribute, again, usually similar expression profiles within the dataset. Multi-clustering algorithms mine and partition subgroups of genes and conditions, where the genes share a common attribute, such as being similarly expressed, for every condition (190, 49).

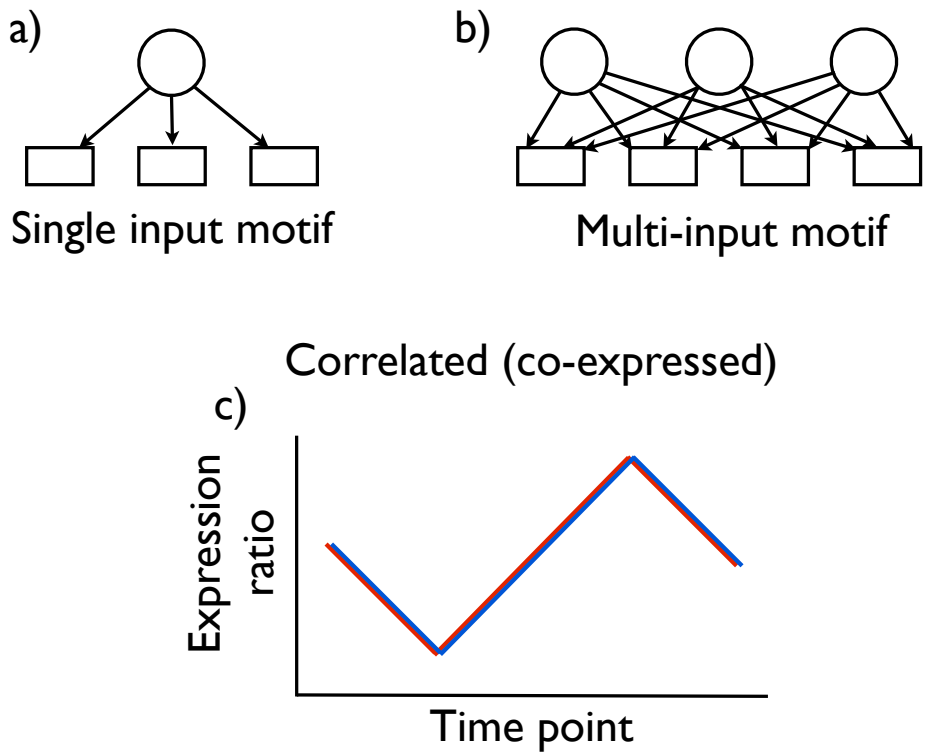



Figure 3.1: Common modes of transcriptional regulation represented by motifs, modified from Yu *et al*. (339). **a)** Single input motif, where the three genes (represented as squares) are co-regulated by a single transcription factor (represented as a circle). **b)** Multi-input motif, where the four genes (represented as squares) are co-regulated by a common set of three transcription factors (represented as circles). **c)** The blue and red gene expression profiles exhibit similarity, suggesting co-expression.

However, co-expression does not guarantee co-regulation of genes: a co-regulated set of genes is controlled by the same regulatory factor. Figures 3.1 a) and b) are examples of transcriptional co-regulation, whereby a set of genes have their expression controlled by a single transcription factor, or multiple common transcription factors, respectively (339).

## 3.1 Mechanisms of regulation of gene expression

The regulation of how information, stored as genetic code, is turned into gene products, is termed regulation of gene expression. Transcriptional regulation is a very important component of gene regulation at which expression will occur due to the involvement of a number of processes (20), such as defence responses and development (266). As the name suggests, the regulation of transcription controls when transcription takes place, and the amount of mRNA generated. Transcription factors are proteins capable of binding to the promoter regions of genes, and affect the rate of

transcription of this gene by promoting or obstructing the action of RNA polymerase. Specific factors such as co-repressors or co-activators can alter the affinity which a transcription factor binds to a promoter, and therefore influence whether transcription takes place, and ultimately, the amount of mRNA produced. Figure 3.1 shows basic schematics of transcriptional regulation. However, although transcriptional regulation in *A. thaliana* is complex (243), this type of regulation does not account for all gene regulation. Transcriptional regulation of gene expression is also limited by post-transcriptional regulation and mRNA degradation (233).

Gene expression can be regulated in other forms, such as post-transcriptionally, or by chromatin remodelling. However, we are interested in transcriptional regulation, which has been shown to control a number of processes, such as plant development and stress responses in plants (43). Transcriptional regulation is important in the immune response in humans, where gene expression needs to be regulated in order to control the response to inflammatory disease (39, 177, 103).

Transcriptional regulation of gene expression influences and controls many processes, such as the cell cycle, maintaining physiological processes, as well as response to environmental changes and pathogen attack (243). The interaction of transcription factors on their target genes, via binding at sites in gene promoters (194), is extremely complex and poorly understood (258).

Breeze *et al*. (43) discovered that groups of transcription factors were active at different stages of development and senescence, suggesting that the process of senescence was dependent on an underlying network. By analysing the promoters of co-expressed genes, potential regulators could be identified to suggest co-regulation was an important part of the stress response (43). The discovery of the importance of transcriptional regulation in the senescence response is not unique: networks of transcriptional regulation are also important in the plant response to pathogens (5). Therefore, a gene regulatory network is defined as a set of genes, where the directionality of regulation within the network can represent transcriptional regulation. An example of a gene regulatory network exhibiting transcriptional regulation is the circadian clock model in plants, where the original; and most basic version of this is shown in Figure 3.2.

A collection of co-regulated genes are termed a 'regulon' (277). Therefore, regulons can be under the control of the same transcription factor in multiple conditions. This differs from co-expressed genes, which merely have correlated expression. Co-expression, therefore, does not infer a shared

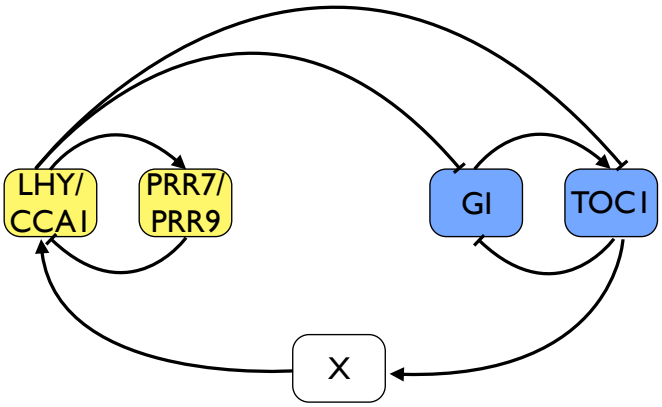



Figure 3.2: The original and standard double loop model of circadian regulation in plants, modified from Locke *et al*. (180). This example of a gene regulatory network involving transcriptional regulation of gene expression, where arrows depict the directionality of the regulation.

regulatory mechanism, and a shared regulatory mechanism does not infer co-expression. A new tool has been developed to detect evidence of these regulons, when applied to multiple time series of gene expression. Unlike existing multi-clustering algorithms, this tool is capable of detecting regulons across subsets of conditions: current multi-clustering tools generally require a feature to be present in all conditions before it can be considered a significant results.

### 3.1.1 Transcriptional regulation is important in stress responses

Transcriptional regulation is at the basis of many stress responses in *A. thaliana*, suggesting that a core network of genes is involved in the response to multiple stresses (152). By identifying gene members of a regulon, a possible method of co-regulation can be hypothesised. If a common regulator, or regulators, can be identified, a regulatory network that is active in the response to multiple conditions can be determined around this group of co-expressed genes. A network with homologous genes to this common regulatory network can be identified in commercially important

crop species. By altering the regulatory gene, or genes, which control the expression of many downstream targets, this network can be targeted in order to generate crops genetically engineered to withstand the effects of multiple stresses simultaneously. Recent studies have shown that sets of co-expressed *WRKY* genes in *A. thaliana* are also co-expressed in rice (*Oryza sativa*), and are likely to be part of the same signal transduction pathway (36). However, other studies have shown how different plant systems can yield different results: the *A. thaliana* genes *CDPK1* and *CDPK2* are induced in salt and drought stress signalling (257, 296), whilst a gene member of the same family in *O. sativa*, *OsCDPK7*, is induced by cold or salt stress (248). Furthermore, although identifying candidate genes in *A. thaliana* does yield genes to study in crop species, extensive analysis in various stress conditions in the crop plant would need to be carried out before a crop that was stress resistant or tolerant could be used in the agricultural industry.

## 3.2 Wigwams: identifying genes working across multiple stresses

Functionally similar genes are usually co-expressed (138), and more likely to be bound by a common regulatory transcription factor (9). Therefore, identifying genes that are co-expressed across multiple stress responses, and share a similar function, is an important task in the discovery of a core regulatory network in *A. thaliana*. Multi-clustering is capable of discovering co-expressed genes across multiple gene expression time series (190); however, current multi-clustering algorithms can be limited by the large size of gene expression datasets (127). Also, the nature by which standard methods cluster usually involves the partitioning of data across all conditions (138). Therefore, the output from such methods requires genes to be significantly co-expressed in all conditions. The main challenge, however, with existing multi-clustering methods is that co-expression across multiple conditions does not infer a common regulatory mechanism: the genes may be co-expressed in several conditions, but the regulatory mechanisms by which they are controlled may be independent (52). Therefore, the genes are co-expressed, not co-regulated. In light of these issues, a new tool, Wigwams (Wigwams Identifies Genes Working Across Multiple Stresses), was developed to discover potential regulons which may have a common regulatory mechanism in multiple stresses. Wigwams takes into consideration multiple gene expression time series datasets, and searches for regulons working over subsets of these conditions, in order to provide evidence for a possible shared regulatory mechanism for these sets of co-expressed genes.

### 3.2.1 Wigwams method: detecting co-expressed genes

The aim of Wigwams is to detect co-expressed genes, or 'potential regulons', that are working over subsets of stress responses using gene expression time series data. These regulons may have a shared regulatory mechanism. Wigwams is implemented in MATLAB®, and utilises statistical packages in R (279) by exploiting the Java foundations in which MATLAB® is written. The Wigwams source code is available to download at `http://www2.warwick.ac.uk/fac/sci/systemsbiology/staff/ott/tools_and_software/wigwams`.

#### 3.2.1.1 Wigwams screening: mining gene expression data for potential regulons

This section describes the initial screening method, which mines gene expression time course data for pairwise potential regulons. Multiple time course expression datasets are supplied to Wigwams, along with a list of genes, whose expression values are included in these datasets, to test for evidence of shared regulatory mechanisms. Along with this list of genes and these data, a list of values stating in which datasets these genes are differentially expressed is also required. If a gene in this list is differentially expressed in two or more stresses, this gene is now referred to as the 'seed gene', and is considered for further analysis. If a gene is differentially expressed in one, or no stresses, then it is not considered for further analysis, as we are only interested in detecting co-expressed genes in multiple datasets, and any potential shared regulatory mechanism that may exist.

The process of selecting genes that are similarly expressed to the seed gene is described in Figure 3.3. In Figure 3.3 the seed gene, which for purposes of this example is At2g02740 (AtWHY3), is differentially expressed in two or more stresses. The expression profiles of this gene are shown for *B. cinerea* (where red is the infected profile and green is the mock infected profile), short day and long day senescence are shown in Figure 3.4. Wigwams performs a correlation test between the whole expression profile of AtWHY3 and the expression profiles of genes in the dataset per stress. Pearson's product moment correlation is used for this analysis; however, other methods of correlation testing can be substituted.

These correlation values are ranked, with the most similar gene per stress ranked the highest in this list of correlated genes, as shown in Figure 3.4. Once this is completed for each stress, Wigwams

considers the overlap in gene membership of these correlated genes in a pairwise manner, as shown in Figures 3.5 and 3.6. The user specifies a number, or set of numbers, which states that only this number of genes in the correlated gene list will be considered for this overlap testing.

To test whether these pairwise overlaps are significantly large and could not have occurred by chance, the hypergeometric test is used. Figure 3.7 shows the *p*-values calculated using AtWHY3 as the seed gene, which was used as an example of how lists of correlated genes were generated in Figures 3.3 to 3.6. The equation for this calculation for the overlap between the datasets *B. cinerea* and long day senescence is shown in Equation 3.1. To calculate the *p*-value of pairwise overlaps using the hypergeometric test, a number of variables are required: the size of the overlap ($k$); the size of both lists of correlated genes that have generated this overlap ($n$); and the universe size, which is the number of genes in the dataset being tested ($m$).

$$\begin{aligned} & P(k, n, m\text{-}n, n) \\ = {} & P(11, 250, 30336 - 250, 250) \\ = {} & 2.45e^{-08} \end{aligned} \tag{3.1}$$

**Notation of inputs required for Wigwams screening**

1. For each seed gene $g \in G$ and each time course $C$ a boolean value indicating whether gene $g$ is differentially expressed in condition $C$.

2. A set of time course expression measurements for all genes in set $G$ in conditions $C$

3. A range of numbers to be used for the gene list sizes $S$

The Wigwams screening method, as illustrated in Figure 3.9, can use lists of differentially expressed genes, in order to only consider genes whose expression is differentially expressed for detection of a potential regulon. If $g$ is differentially expressed in two or more conditions, then $g$ is considered for further testing in the Wigwams screening. However, if $g$ is differentially expressed in one, or no conditions, then $g$ is not considered any further.

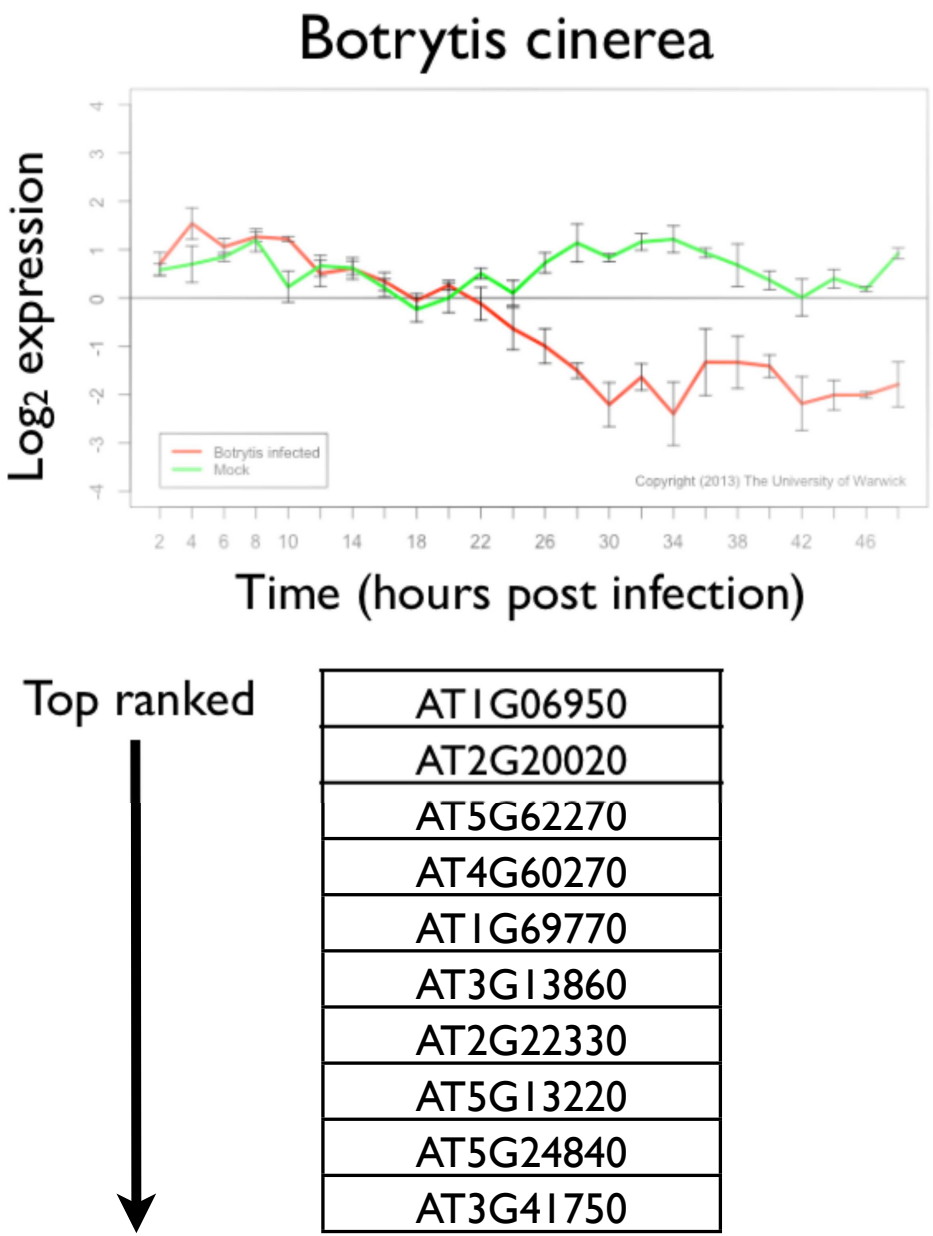



Figure 3.3: The Wigwams screening method: selecting genes with most similar expression profiles in one stress. The seed gene shown here, At2g02740 (AtWHY3), is differentially expressed in two or more stresses, and as such, is considered for further testing. Wigwams performs a correlation test between the whole expression profile of the seed gene AtWHY3 and the expression profiles of genes in the dataset. These correlation values are ranked, with the gene with the most similar expression profile ranked highest.

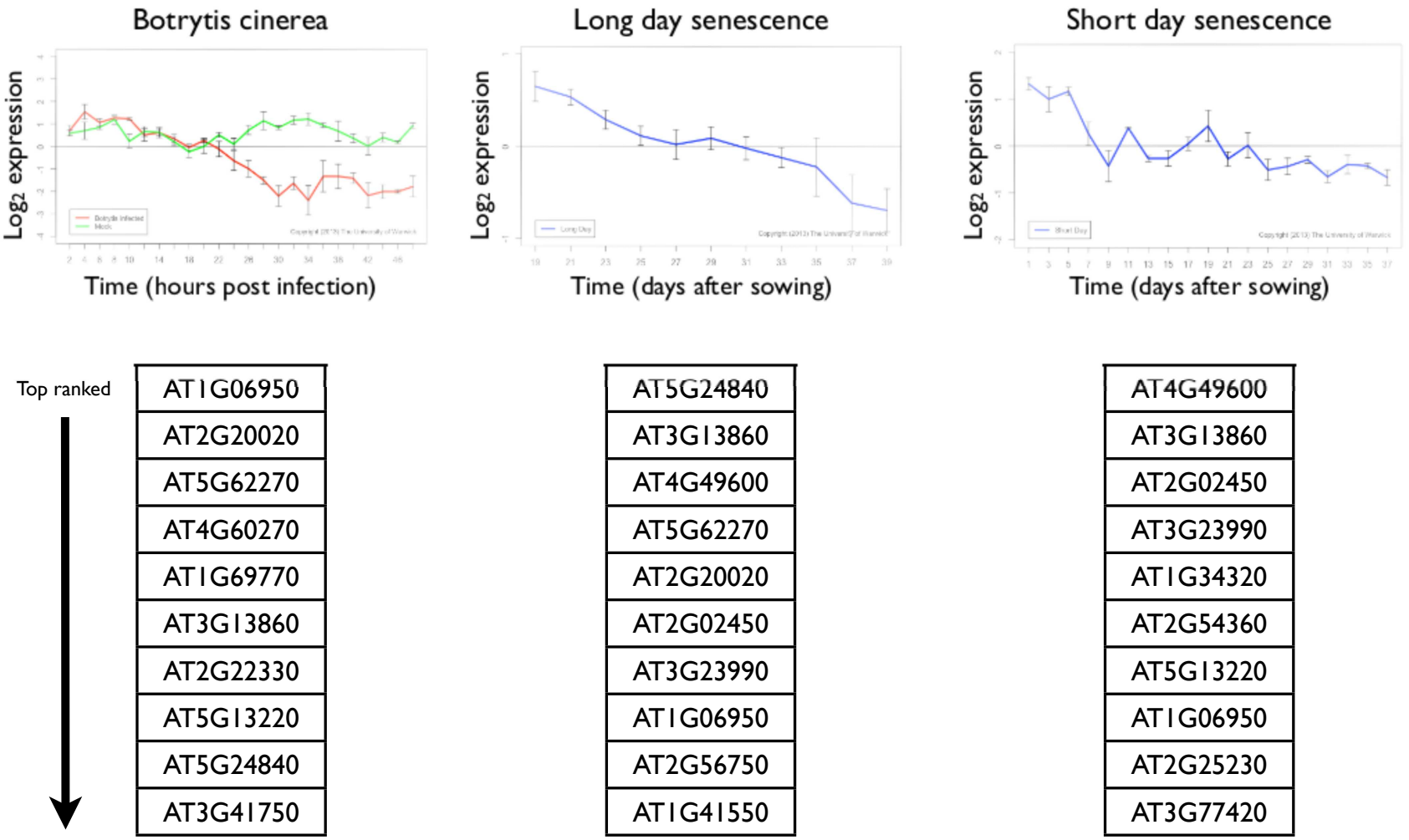



Figure 3.4: The Wigwams screening method: selecting genes with most similar expression profiles in all stresses. The seed gene is also differentially expressed in long day and short day senescence stresses. Wigwams also performs a correlation test between the whole expression profile of the seed gene AtWHY3 and the expression profiles of genes in the each dataset, before identifying genes with the most similar expression profile to the seed gene by ranking the correlation values (as described in Figure 3.3).

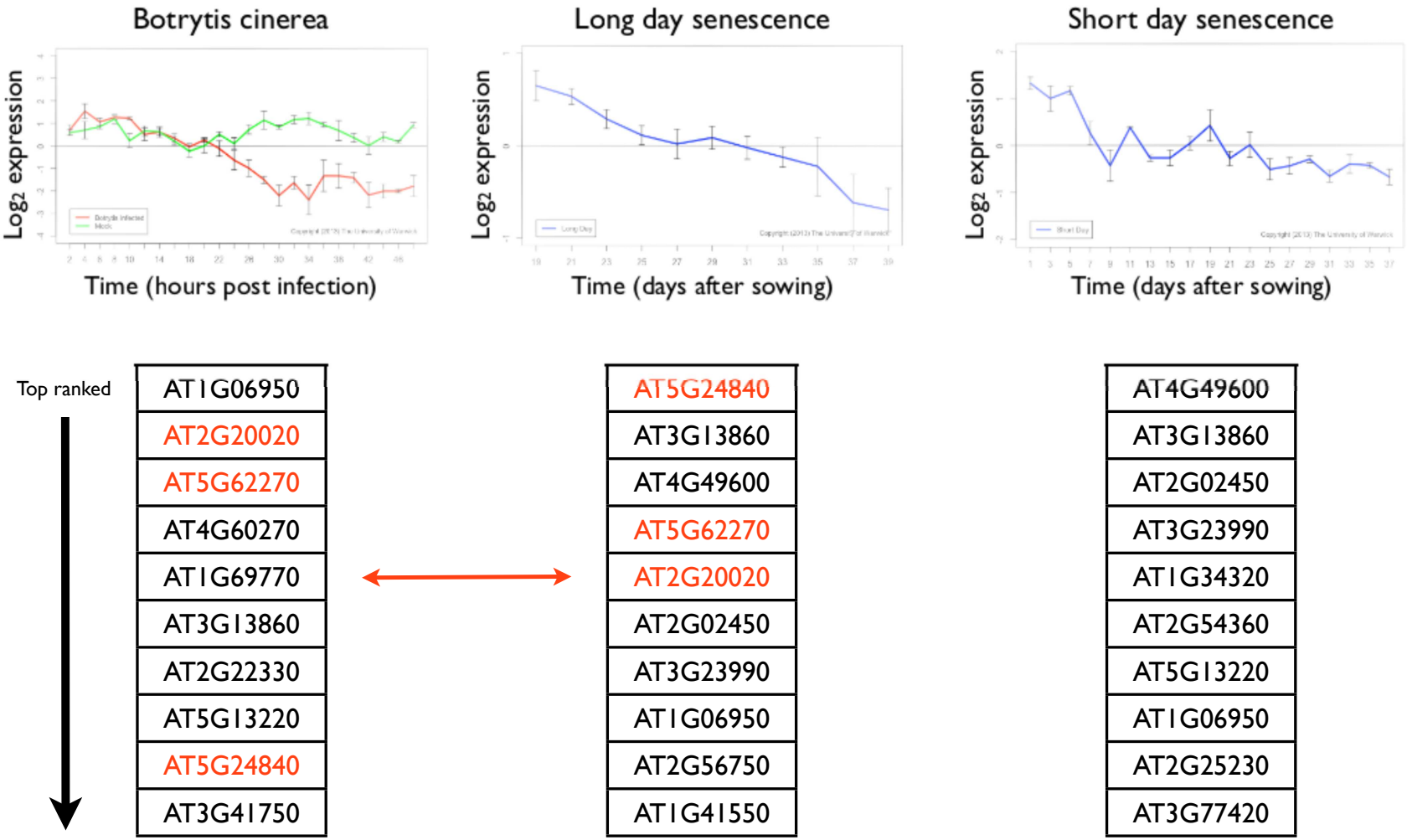



Figure 3.5: The Wigwams screening method: considering the overlap in gene membership in a pairwise manner. The genes highlighted in red are common to both the top ranked list of *B. cinerea* and long day senescence. A hypergeometric test is performed to calculate the *p*-value associated with this overlap in gene membership.

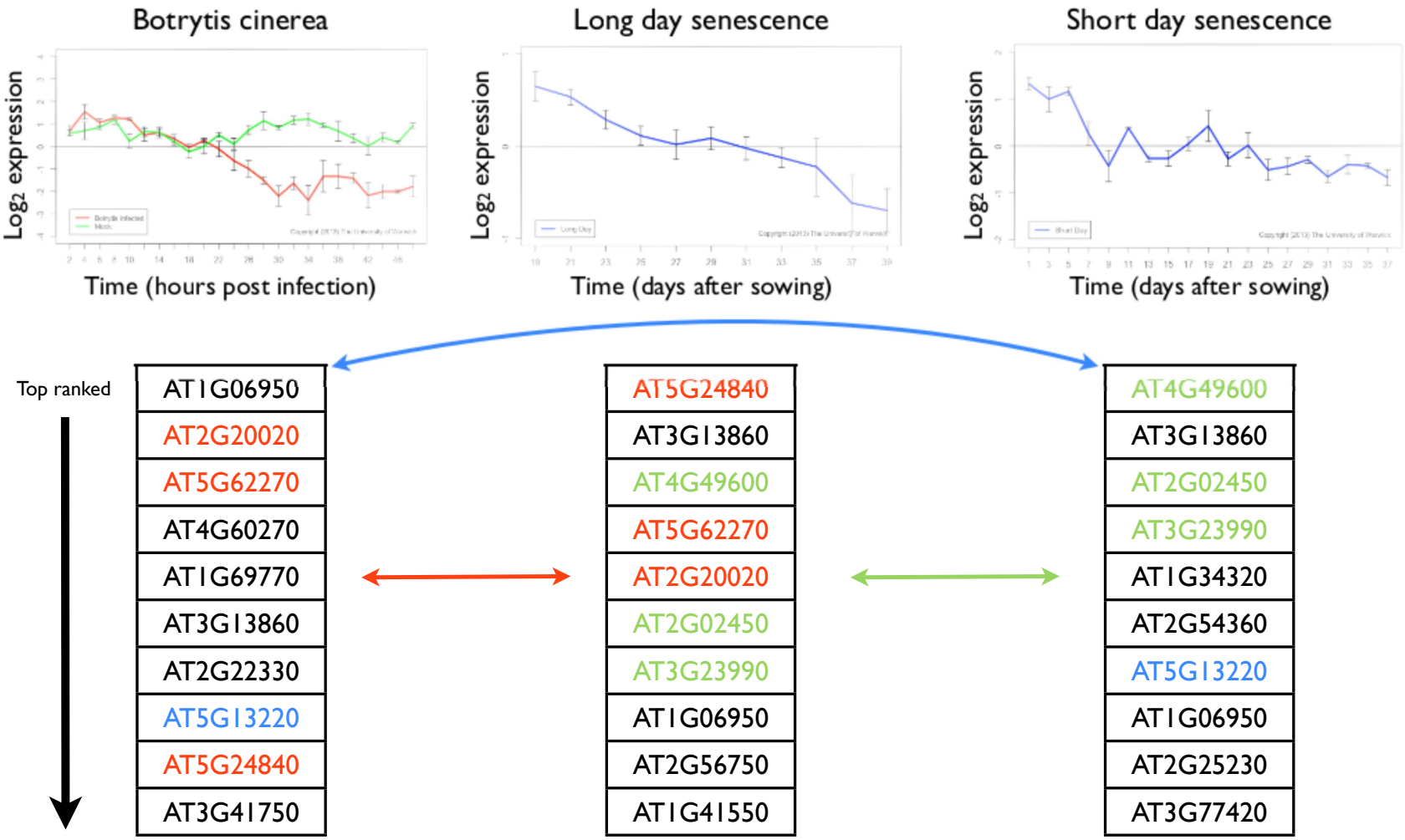



Figure 3.6: The Wigwams screening method: gene overlap comparison for each pairwise combination. Wigwams compares the overlap in gene membership for each pairwise combination of stresses. Genes highlighted in green are common to both the top ranked list of short day and long day senescence. Genes highlighted in blue are common to both the top ranked list of *B. cinerea* and short day senescence. A hypergeometric test is performed to calculate the *p*-value associated with this overlap in gene membership.

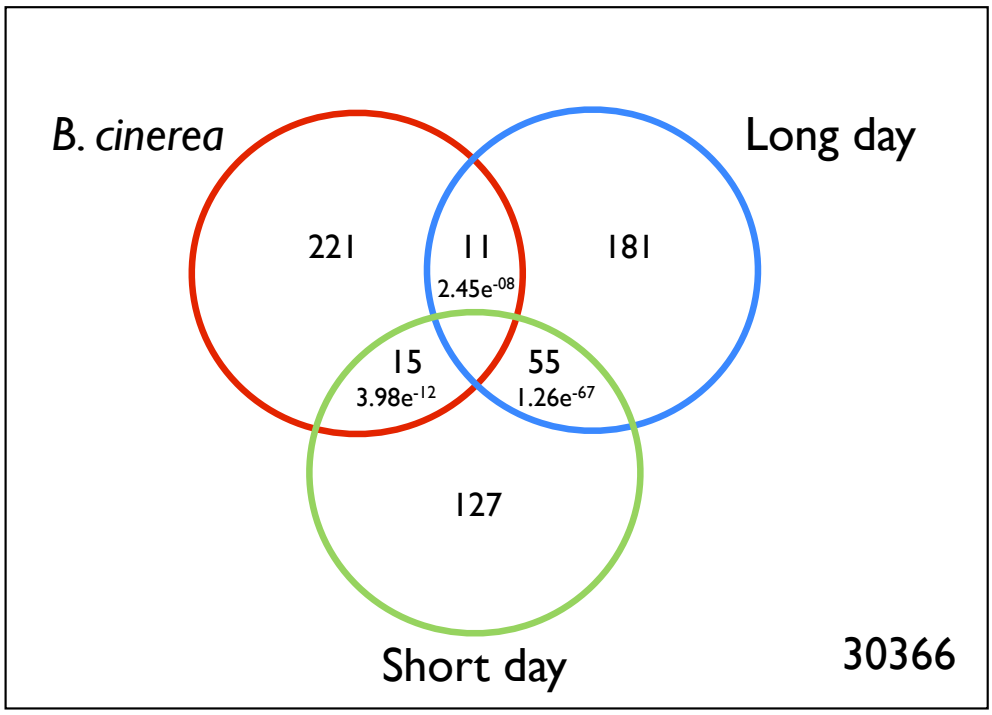



Figure 3.7: Calculating significantly large pairwise overlaps. Using seed gene AtWHY3, which was used previously in Figures 3.3 to 3.6, this illustrates the *p*-values calculated using the hypergeometric test. Only the pairwise overlaps between lists of correlated genes are considered here.

**Non-differentially expressed genes** Sets of the most similarly expressed genes are generated by picking the top $S$ correlated genes in each condition to the seed gene. Non-differentially expressed genes can still contribute to a set of correlated genes. Some genes are not differentially expressed in a condition when their expression is compared to the expression profile during the mock treatment, however, this expression is still changing over time (perhaps circadianally regulated, for instance). The contribution of non-differentially expressed genes is minimised by randomly permuting the expression values of non-differentially expressed genes. If a gene is not differentially expressed in a stress condition dataset, all expression values for this gene in this dataset are randomly permuted with the expression values of another non-differentially expressed gene in the same dataset, as shown in an example in Figure 3.8. Therefore, if a non-differentially expressed gene does become a member of a list of the most correlated genes to a test gene, it does not contribute significantly to the result calculated using the hypergeometric test, due to the permuted expression profile values.

a)

| | 2 hrs | 4 hrs | 6 hrs | 8 hrs | 10 hrs | 12 hrs | 14 hrs | 16 hrs | 18 hrs | 20 hrs | 22 hrs |
|---|---|---|---|---|---|---|---|---|---|---|---|
| AT1G01453 | -1.92 | -2.13 | -2.65 | -2.09 | -1.56 | -1.82 | -1.74 | -1.52 | -2.22 | -1.75 | -1.58 |
| AT1G01490 | -1.24 | -0.59 | -1.01 | -0.83 | -1.41 | 0.55 | 0.65 | 1.64 | 1.44 | 1.44 | 0.63 |
| AT1G01500 | 0.403 | -1.03 | -0.49 | -0.85 | -0.57 | -1.25 | -1.75 | -1.57 | -0.95 | -1.54 | 0.41 |
| AT1G01520 | 1.57 | 1.63 | 1.89 | 1.70 | 1.84 | 1.76 | 1.69 | 1.44 | 1.59 | 1.51 | 0.98 |
| AT1G01540 | 4.17 | 3.93 | 3.91 | 4.13 | 4.12 | 4.40 | 4.34 | 4.60 | 4.23 | 3.70 | 2.75 |

b)

| | 2 hrs | 4 hrs | 6 hrs | 8 hrs | 10 hrs | 12 hrs | 14 hrs | 16 hrs | 18 hrs | 20 hrs | 22 hrs |
|---|---|---|---|---|---|---|---|---|---|---|---|
| AT1G01453 | -1.92 | -2.13 | -2.65 | -2.09 | -1.56 | -1.82 | -1.74 | -1.52 | -2.22 | -1.75 | -1.58 |
| AT1G01490 | 0.403 | -1.03 | -0.49 | -0.85 | -0.57 | -1.25 | -1.75 | -1.57 | -0.95 | -1.54 | 0.41 |
| AT1G01500 | -1.24 | -0.59 | -1.01 | -0.83 | -1.41 | 0.55 | 0.65 | 1.64 | 1.44 | 1.44 | 0.63 |
| AT1G01520 | 1.57 | 1.63 | 1.89 | 1.70 | 1.84 | 1.76 | 1.69 | 1.44 | 1.59 | 1.51 | 0.98 |
| AT1G01540 | 4.17 | 3.93 | 3.91 | 4.13 | 4.12 | 4.40 | 4.34 | 4.60 | 4.23 | 3.70 | 2.75 |



Figure 3.8: Describing the permutation of expression values for non-differentially expressed genes. The expression values for eleven time points of five genes from the *B. cinerea* dataset are shown. Genes highlighted in red indicate they are not differentially expressed in response to *B. cinerea* infection. a) The expression values of the two genes *At1g01490* and *At1g01500*, which are not differentially expressed in response to *B. cinerea* infection, will be swapped with each other. b) The swapped expression values of the two non-differentially expressed genes in the *B. cinerea* dataset.

**Generation of sets of most correlated genes** In order to generate the lists of the most correlated genes, Pearson's product moment correlation coefficient is used to find the genes whose expression profiles are most similar to seed gene $g$ in each stress. Calculation of the Pearson product moment correlation coefficient $r$ is performed for the seed gene $g$ against all other genes in $G$ per time course $C_i$, based on the value of gene expression at each time point. The genes are then ranked based on the value of $r$. A list $L$ of the genes which have the most similar expression to the expression profile of seed gene $g$ is compiled for each dataset in turn. The size of the list $L$ is dependent on the user-defined parameter $S$. Therefore, list $L_1$ will contain the top $S$ most similarly expressed genes to the seed gene in dataset $C_1$. This will also be calculated analogously for list $L_2$ in dataset $C_2$.

The correlation function 'corr.m' is included in the MATLAB® package. However, Dafyd Jenkins wrote a new function, entitled 'fastcorr.m', which reduces the computational time taken to calculate the correlation test of seed gene $g$ against all other genes in $G$ per time course $C_i$.

**Finding statistically large overlaps of genes working across pairs of conditions** Two lists $L_1$ and $L_2$ are compared in a pairwise manner to determine the overlap in gene membership, as shown in Figure 3.10. A $p$-value is attributed to the size of this overlap using the 'phyper' function in R (279), which calculates the hypergeometric distribution to test for significantly large overlaps. This function requires the size of the overlap, the list size $S$ and the universe size $G$ as inputs. The log $p$-value is returned to maintain the higher order of magnitudes. The overall minimal $p$-value for $g$ over all list sizes in $S$ is determined and used as the $p$-value associated with $g$. The non-significant genes are removed, along with the lists $L$ of most correlated genes, if the minimal $p$-value does not meet the significance criteria defined by the user. For example, the $p$-value for the overlap in gene membership of the lists of genes with similar expression to AtWHY3 (as shown in Figure 3.7) in *B. cinerea* and long day senescence was calculated to be $2.45e^{-08}$. If this $p$-value was higher than a cutoff defined by the user, the seed gene AtWHY3, along with the set of most correlated genes to this seed gene, would be removed, and not used in any further analysis.

#### 3.2.1.2 Summary of Wigwams screening

Wigwams will iteratively test each gene, in a given list of genes, that is differentially expressed in two or more conditions of multiple gene expression datasets, to generate sets of correlated, and co-

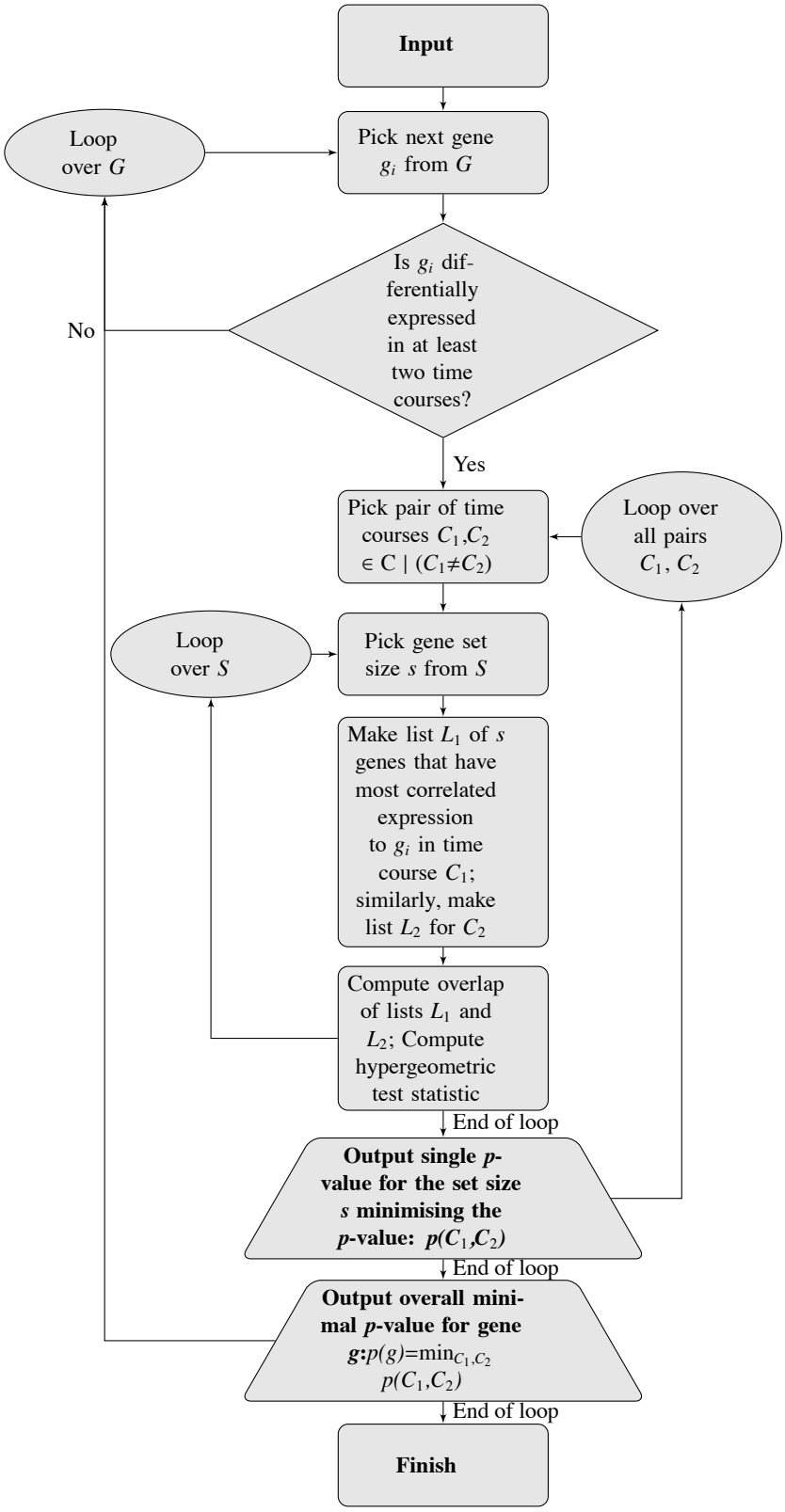



Figure 3.9: Flowchart illustrating the Wigwams pseudocode and methodology, complementing the screening method description given in Figures 3.3 to 3.6. By taking all time course experiments and all genes into consideration, the opportunity to detect all possible regulons arises. Refer to Section 3.2.1.1: Notation for definitions of the symbols used in this figure. A gene $g$ differentially expressed in two or more stresses is considered for testing. The user decides upon a range of values $S$ for the size $s$ of the sets of genes with the most correlated expression to the seed gene $g$. These sets of correlated genes are defined as $L$. The overlap in gene membership between lists $L$ of two stresses is used to calculate a $p$-value using the hypergeometric test.

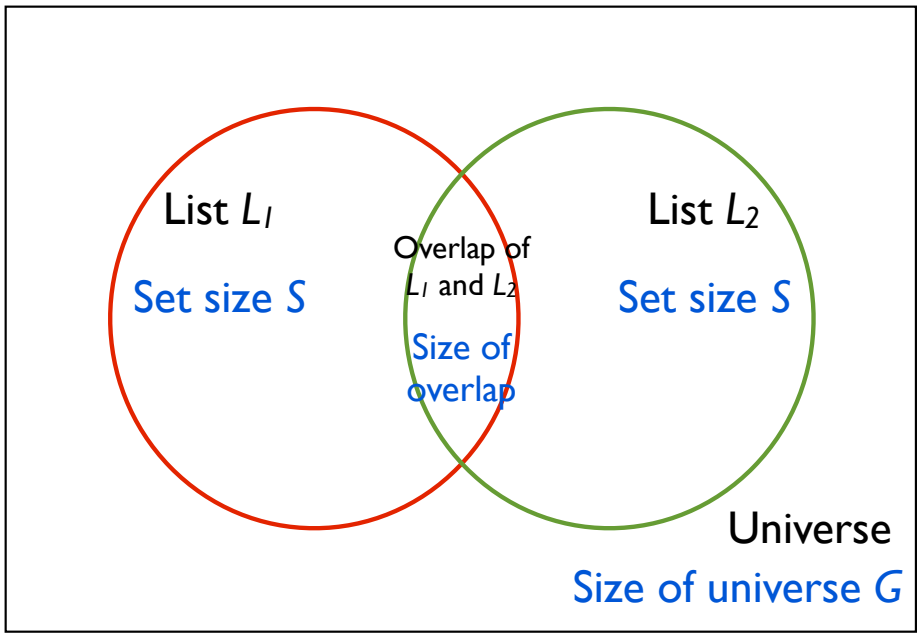



Figure 3.10: Calculating significantly large pairwise overlaps using the hypergeometric test. To calculate the whether the overlap of Lists $L_1$ and $L_2$ is significant, and could not have occurred by chance, the sizes of the variables are needed (labelled in blue).

expressed genes in pairwise stress combinations. At this stage, each set of pairwise co-expressed genes is associated with the gene that seeded this group. A list of these genes that have generated these groups of co-expressed genes is required for the subsequent stage of Wigwams.

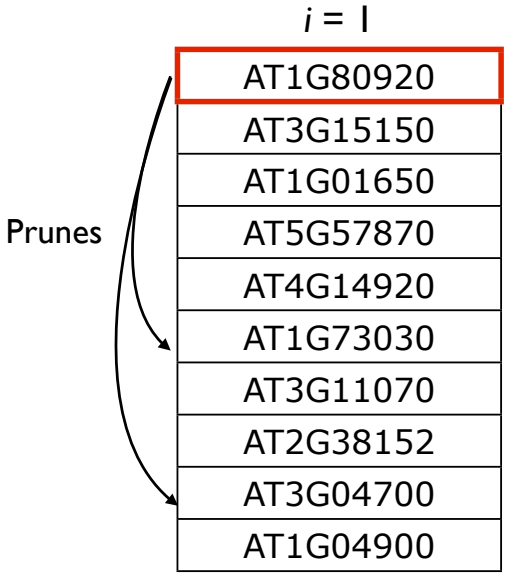
#### 3.2.1.3 Wigwams pruning

The list of genes that are associated with groups of co-expressed genes returned from the Wigwams screening method are subsequently subject to further assessment, in order to select the most biologically meaningful and significant pairwise regulons. Firstly, all non-significant pairwise regulons, based on their overall *p*-value, are removed.

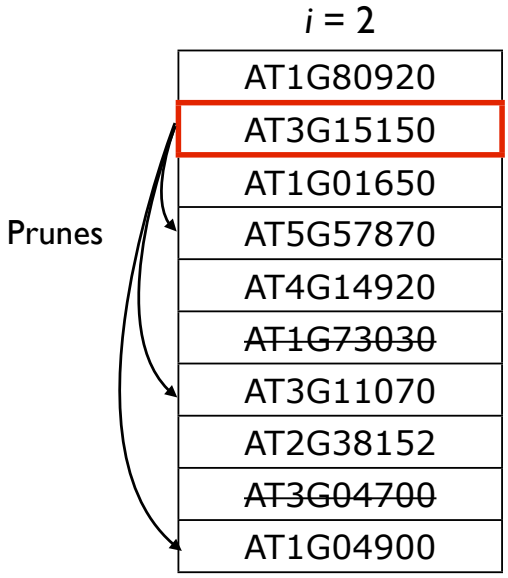
**Notation of inputs required for Wigwams pruning**

1. A list of genes $N$ returned by Wigwams screening stage

2. Set of time course expression measurements under $C$ conditions and for all genes in set $N$.




a)




b)

Figure 3.11: How the Wigwams pruning process compares pairwise potential regulons. All genes returned from the Wigwams screening method are ranked on overall *p*-value. The top ranked gene, highlighted in red in a) is compared to the next top ranked gene, AT3G15150. If all the criteria (which will be mentioned below) are fulfilled, then the top ranked gene is then compared to the next gene, AT1G01650, and so on. However, if criteria are not fulfilled, and the gene and its corresponding group of co-expressed genes share similar expression profiles to the top ranked gene, this gene and its group of co-expressed genes are removed from the list, as indicated by a strikethrough for genes AT1G73030 and AT3G04700 in b). Once the top ranked gene is compared to all the genes returned from the Wigwams screening method, the next top ranked gene becomes the 'top ranked gene', as shown in b), and this process is repeated.

3. A set of gene list sizes $S$.

4. A similarity measure $r$ for time course measurements for pairs of genes.

5. A single $p$-value for the set size $S$ minimising the $p$-value for each gene $p(n, C_1, C_2)$

6. The overall minimal $p$-value for each gene $n \in N \mid p(n)=\min_{C_1,C_2} p(C_1,C_2)$.

7. A user defined cut-off for significance, above which, $p$-values are deemed non-significant.

8. A user defined cut-off for similarity measure $r$, above which, expression profiles of the two genes being compared are considered similar.

9. A user defined value to state by how much a $p$-value has to be significantly stronger than another $p$-value before the potential regulon with the weaker $p$-value can be discarded.

**Pruning** Figure 3.11 shows that all genes which are returned from the Wigwams screening method are ranked on their overall $p$-value. The gene with the strongest $p$-value is the top ranked gene, and is compared to the next ranked gene (as illustrated in Figure 3.11 a). If the next ranked gene does not have a similar expression profile in all time courses, then this gene is not removed, and the top ranked gene is compared to the next ranked gene, and so on. Once the top ranked gene is compared to all the genes returned from the Wigwams screening method, the next top ranked gene becomes the top ranked gene, as shown in Figure 3.11 b), and this process is repeated.

As Wigwams has the ability to complete a comprehensive analysis, and detects all evidence of co-regulation within multiple datasets, it is probable that one gene may be a member of many pairwise regulons. For example, if gene $a$ had a similar expression profile to gene $b$ in time courses $C_1$ and $C_2$, then gene $b$ will be selected as a member of the pairwise regulon when gene $a$ is the seed gene, and vice versa. Therefore, there is an issue of the pairwise sets of correlated genes generated having redundant gene membership.

To address this, we consider similar gene expression profiles (as illustrated in Figure 3.12), which is computationally less intensive, as a substitute for similar gene membership. An example of this is shown in Figure 3.13: the datasets used for this figure show gene expression profiles in *A. thaliana* in response to stress conditions (see Section 4.1.3). The pairwise regulons have similar gene membership, as shown in Table 3.1. Fifteen of the 24 genes in the pairwise regulon shown in Figure 3.13 a) are identical to genes found in the pairwise regulon shown in Figure 3.13 b). If

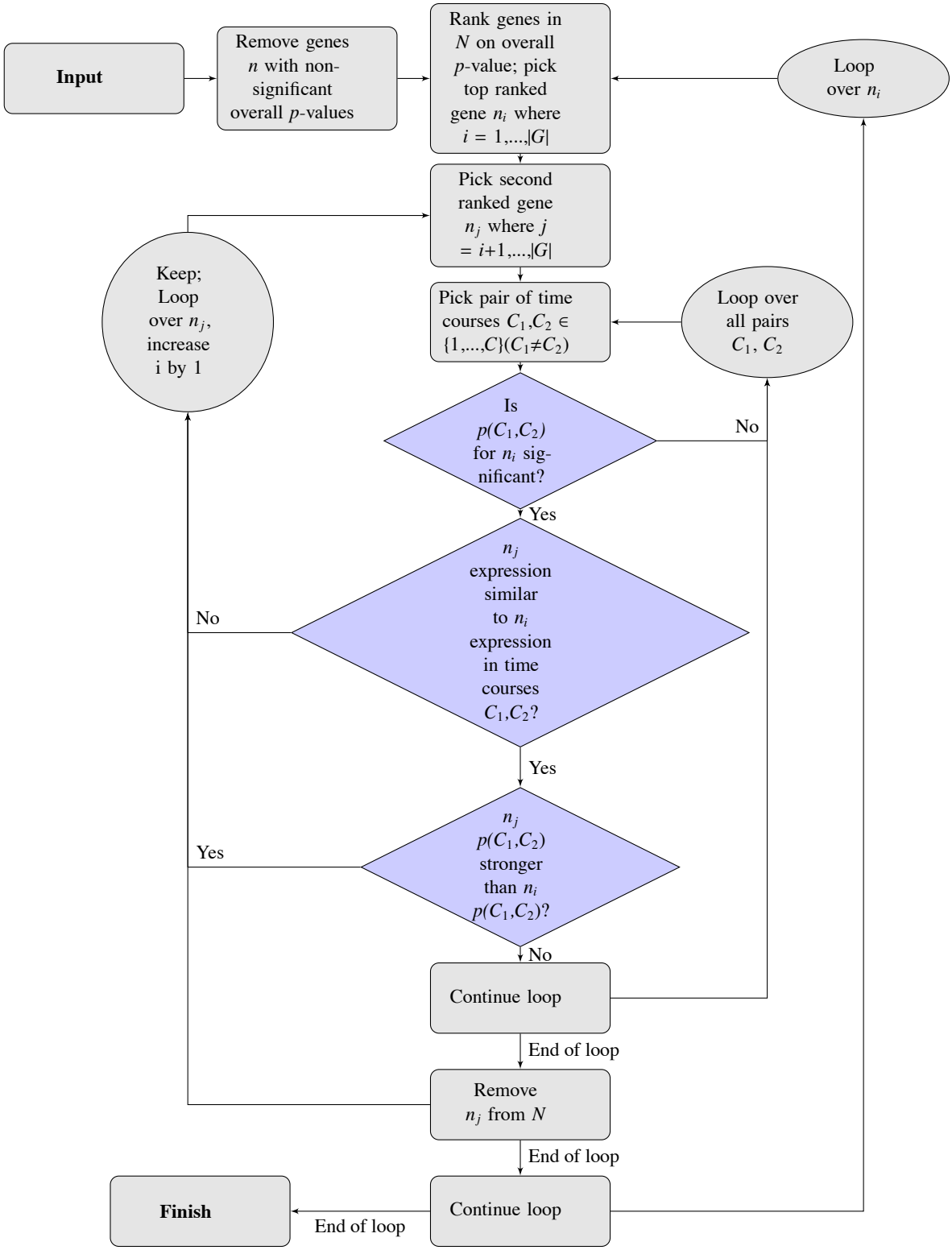
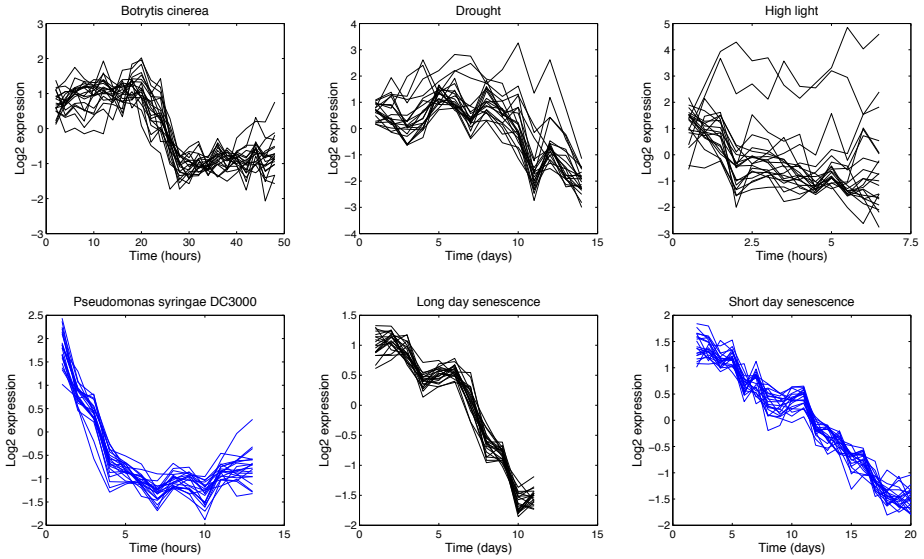



Figure 3.12: Selecting the most informative regulons by pruning. Wigwams provides a comprehensive output, however, in order to highlight the regulons that are most informative in terms of significance and expression pattern, this algorithm removes regulons that exhibit redundant expression profiles. Decision nodes in blue indicate where a user defined parameter is needed. Parameters to be defined include: $p$-value cutoff; correlation coefficient $r$ cutoff; a value by which one $p$-value is significantly stronger than another before the potential regulon with the weaker $p$-value can be discarded.

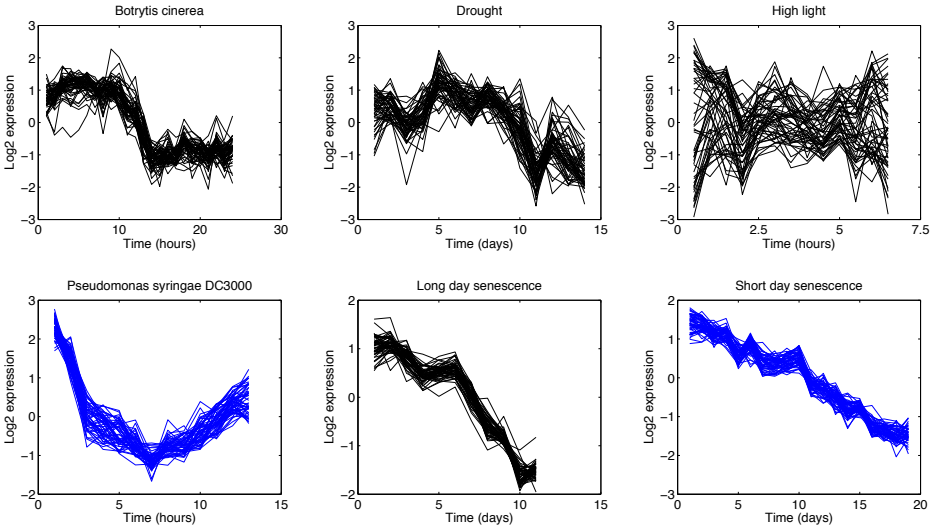
one considers the same condition in both pairwise regulons in Figure 3.13, then it can be observed that the expression profiles of the gene members in these pairwise regulons are similar in both conditions. The pairwise regulon in Figure 3.13 b) has a more significant $p$-value than the pairwise regulon in Figure 3.13 a), therefore, the pairwise regulon in Figure 3.13 b) will be retained in favour of the less significant pairwise regulon.

Referring to Figure 3.13, the gene which seeded the set of co-expressed genes shown in Figure 3.13 a) is defined as $n_j$, which is the notation used in Figure 3.12, and therefore, the gene which seeded the set of co-expressed genes shown in Figure 3.13 b) is defined as $n_i$. If $n_i$ and $n_j$ have similar expression profiles, we wish to keep the set of co-expressed genes with the more significant $p$-value. Therefore, if $n_i$ has a more significant $p$-value than $n_j$, then $n_j$ will be removed, as $n_i$ is more informative with regards the potential co-regulation of these particular genes.

In order to decide whether $n_j$ has a similar expression profile to $n_i$, a test to compare the expression profiles of $n_i$ and $n_j$ using Pearson's product moment correlation is performed between the expression values of $n_i$ and $n$j in each time course. The $r$ values generated by this correlation test are compared in a pairwise manner: the expression values of $n_i$ and $n_{i+1}$ are compared in time course $C_1$, a process which is repeated in $C_2$. If the pairwise $r$-values generated by comparing the expression values of $n_j$ to $n_i$ are above the cut-off for the similarity measure (Notation point 7), the expression profiles for $n_j$ and $n_i$ are deemed similar, and are likely to have redundant, or partially overlapping gene membership.




a)




b)

Figure 3.13: Removing pairwise regulons with redundant gene membership, using gene expression as a proxy. The expression profiles in both a) and b) shown in blue are similar in the same conditions. As shown in Table 3.1, the majority of genes in the pairwise regulon shown in a) are the same as the genes in the pairwise regulon shown in b). Therefore, we use the pruning algorithm described in 3.2.1.3 to remove regulons with redundant gene membership. a) A significant pairwise regulon, highlighting genes that are significantly working across *P. syringae* DC3000 infection and short day senescence. The *p*-value associated with this pairwise regulon is $3.16e^{-22}$. b) A significant pairwise regulon, highlighting genes that are significantly working across *P. syringae* DC3000 and short day senescence. The *p*-value associated with this pairwise regulon is $2.00e^{-59}$. As this pairwise regulon has a more significant *p*-value than the pairwise regulon shown in a), we wish to keep this regulon, and remove the pairwise regulon shown in a)

Table 3.1: Gene membership of pairwise regulons shown in Figure 3.13 a) and b) to illustrate that using similar expression profiles as a proxy for similar gene membership is valid. 15 out of 24 genes in pairwise regulon shown in Figure 3.13 a) are the same to genes in pairwise regulon shown in Figure 3.13 b)

| Genes of pairwise regulon shown in Figure 3.13 a) | | Genes of pairwise regulon shown in Figure 3.13 b) | |
|---|---|---|---|
| ATG identifier | Gene name | ATG identifier | Gene name |
| AT1G08390 | | AT1G01080 | |
| AT1G09340 | CRB | AT1G03130 | PSAD-2 |
| AT1G12900 | GAPA-2 | AT1G08390 | |
| AT1G14290 | SBH2 | AT1G09340 | CRB |
| AT1G15820 | LHCB6 | AT1G12900 | GAPA-2 |
| AT1G60600 | ABC4 | AT1G15820 | LHCB6 |
| AT1G62510 | | AT1G15980 | NDF1 |
| AT1G74730 | | AT1G18060 | |
| AT1G75690 | | AT1G21500 | |
| AT2G05100 | LHCB2.1 | AT1G22850 | |
| AT2G22800 | HAT9 | AT1G32550 | |
| AT2G22990 | SNG1 | AT1G42970 | GAPB |
| AT2G30790 | PSBP-2 | AT1G50730 | |
| AT3G08920 | | AT1G52230 | PSAH2 |
| AT3G27690 | LHCB2.3 | AT1G60600 | ABC4 |
| AT4G02680 | EOL1 | AT1G74730 | |
| AT4G03470 | | AT1G75690 | |
| AT4G38820 | | AT2G20890 | PSB29 |
| AT5G14740 | CA2 | AT2G29180 | |
| AT5G21920 | | AT2G30790 | PSBP-2 |
| AT5G45680 | | AT2G34860 | EDA3 |
| AT5G54270 | LHCB3 | AT2G35500 | |
| AT5G64460 | | AT2G43030 | |
| AT5G64470 | | AT2G48070 | RPH1 |
| | | AT3G08920 | |

| | | | |
|---|---|---|---|
| | | AT3G13120 | |
| | | AT3G21055 | PSBTN |
| | | AT3G27690 | LHCB2.3 |
| | | AT3G53190 | |
| | | AT3G63410 | APG1 |
| | | AT4G03470 | |
| | | AT4G15510 | |
| | | AT4G20760 | |
| | | AT4G21280 | PSBQA |
| | | AT4G25080 | CHLM |
| | | AT4G39710 | |
| | | AT5G08050 | |
| | | AT5G13510 | |
| | | AT5G14910 | |
| | | AT5G17870 | PSRP6 |
| | | AT5G21920 | |
| | | AT5G24314 | PTAC7 |
| | | AT5G45680 | |
| | | AT5G48790 | |
| | | AT5G52100 | crr1 |
| | | AT5G53490 | |
| | | AT5G57440 | GS1 |
| | | AT5G64460 | |
| | | AT5G64470 | |
| | | AT5G66055 | AKRP |
| | | | |

Observing both the pairwise regulons in Figure 3.13, it can clearly be seen that the expression profiles of the genes significantly co-expressed in *P. syringae* DC3000 and short day senescence are very similar. Also, by observing the gene membership of these two pairwise regulons, shown in

Table 3.1, it can be seen that similar gene expression provides a proxy for similar gene membership. However, in order to make this comparison automated, a correlation test between the expression profiles of the two pairwise regulons is performed, using the expression profile of both seed genes $g$, which were used to first generate both regulons (see Figure 3.9). If the correlation values $r$ for both conditions in which the genes are co-expressed is greater than the threshold value given, the expression profiles are deemed similar. Consequently, the pairwise regulon with the more significant $p$-value is retained, along with the lists of most correlation genes $L$ associated with this gene.

If the pairwise regulons are found to have similar expression profiles, the difference in the order of magnitude of the $p$-values is also compared, as one final check. For instance, a pairwise regulon with a $p$-value of order of magnitude -24 will not be considered more significant than a pairwise regulon with a $p$-value of order of magnitude -23. It can be hypothesised that this co-expression, and perhaps common regulatory mechanism, is demonstrated by both pairwise regulons, and therefore, they should both be retained.

#### 3.2.1.4 Extending beyond pairs of conditions

An extension of Wigwams allows the identification of a potential regulon working over three or more conditions. For example, if three potential regulons, all significant in the pairwise combinations over three conditions, could be consolidated into one potential regulon, it would be biologically interesting to keep the potential regulon significant over three conditions (as illustrated in Figure 3.14), in favour of the pairwise regulons. Also, by identifying potential regulons working over three or more stresses, this would aid our identifying of a core regulatory network. This process can be repeated for pairwise regulons significant over four, five, or more conditions, in order to identify one potential regulon significant over those conditions, in favour of the constituent pairwise regulons.

Some information will be lost, however, if this decision is made, as the overlap in gene membership over three conditions, for example, is likely to be smaller than the overlap in gene membership over pairs of conditions. To accommodate this, certain criteria, such as a minimum size of the overlap, are assigned when selecting a potential regulon working over three conditions for retention in favour of its constituent pairwise regulon.

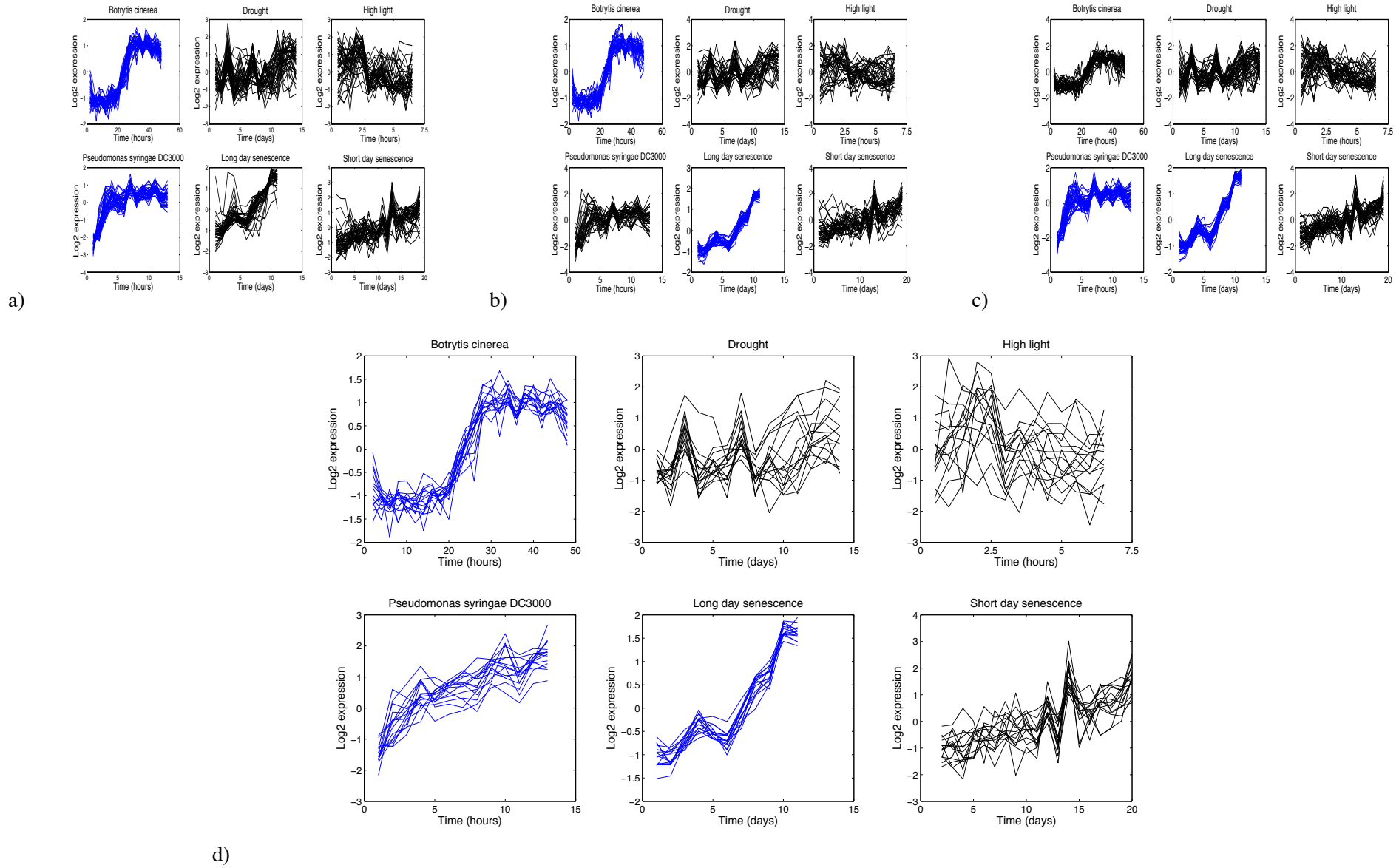



Figure 3.14: Extending beyond pairs of conditions. The pairwise regulons shown in a), b) and c) can be consolidated into one potential regulon working over three conditions, as shown in d), in favour of the constituent pairwise regulons. Criteria are applied before a potential regulon working over three conditions, for example, is kept in favour of its constituent regulons. Such criteria can be the size of the three-way overlap.

### 3.2.2 Algorithm summary

Previous sections have described the Wigwams algorithm in full. Wigwams is computationally inexpensive, taking approximately 5 seconds to compute the initial screening method (described in Section 3.2.1.1) per gene, per condition. Therefore, the greater the number of conditions (or the greater the resolution of time points for conditions), the more time is needed to complete the initial screening per gene.

#### 3.2.2.1 Standard output from Wigwams

Upon completion of extending beyond pairs of conditions, three files are created for the most significant and informative potential regulons, as described below.

**Graphical output** One file, containing encapsulated postscript figures of the most significant and informative potential regulons, is created. Each figure shows the expression profiles of the gene members of potential regulons in all datasets. Conditions in which these potential regulons are significantly co-expressed are highlighted in blue.

**Data output** Two text files are created for the most significant potential regulons. One text file contains the size of each potential regulon, along with the *p*-values associated with any potential regulons significant over two conditions. A second text file contains the gene membership of each potential regulon.

### 3.2.3 Wigwams identifies known biological examples of co-regulation

Gene expression time series were used to validate the ability of Wigwams to identify known co-expressed genes. In order to illustrate this, the following examples are given, where the biology behind the co-expression seen is relatively well understood.

**Time series datasets in multiple conditions used to uncover known examples of co-regulation** Kilian *et al*. (152) performed a series of microarray experiments of *A. thaliana* shoots grown under the same conditions but treated to different environmental stresses, including heat, cold, drought, genotoxic, salt, oxidative, osmotic stress, UV-B light and wounding. Control kinetics were generated from non-stressed plants which were grown in the same conditions but were not

subject to stress conditions. This resulted in a multiple datasets of gene expression over time for 2395 genes in nine conditions, all of which will be used in this section. For this analysis, the relative intensity of expression was calculated as described in the paper (152), by applying the following formula to all time points:

$$\text{Relative intensity} = \frac{\text{signal strength of gene A in sample X}}{\text{average signal strength of gene A in control samples}} \tag{3.2}$$

These datasets can be mined for known examples of co-regulation working over multiple conditions. Genes encoding components of metabolic pathways have been extensively studied, leading to the discovery that genes connected by a similar metabolic function are likely to show the same expression pattern, and therefore, be co-expressed (133, 315), and perhaps co-regulated. Therefore, genes whose products are components of the indole glucosinolate and flavanoid biosynthetic metabolic pathways were considered as seed genes, to identify other components of these pathways as co-expressed.

#### 3.2.3.1 Wigwams identifies the co-expression of genes encoding enzymes of metabolic pathways

We sought to investigate whether Wigwams could identify genes that were known to be co-expressed in the indole glucosinolate pathway (101), using *CYP83B1* as the seed gene $g$, and time series microarray datasets in various conditions (152).

The Trp-metabolising genes *CYP79B2*, *CYP79B3* and *CYP83B1* encode enzymes involved in the production of indolic glucosinolate, and are seen to be commonly regulated by *ATR1* (also known as *MYB34*), a MYB family transcription factor: an *atr1* mutant displayed elevated expression of *CYP79B2*, *CYP79B3* and *CYP83B1*, suggesting that downstream the Tryptophan metabolic pathway is transcriptionally regulated (53). Gachon *et al.* also saw *CYP79B2*, *CYP79B3* and *CYP83B1* cluster over multiple stress condition datasets (cold and heat stress, drought, osmotic stress, salt stress, oxidative stress, UV and wounding), suggesting co-expression (101).

Figure 3.15 shows that *CYP79B2*, *CYP79B3* and *CYP83B1* are significantly co-expressed (with a $p$-value of $2.19e^{-03}$ across all three conditions) in the conditions drought, wounding and cold stress, confirming the findings of Gachon *et al.* (101) and Celenza *et al.* (53). Interestingly, *CYP79B2*,

*CYP79B3* and *CYP83B1* are also co-expressed in these three conditions with *ASA1*, an anthranilate synthase, which was also found by Gachon *et al*. (101) to be co-expressed with these three cytochrome enzyme-encoding genes.

#### 3.2.3.2 Wigwams identifies co-expressed genes encoding members of the flavonoid biosynthetic pathway

Flavonoids are secondary metabolic products that are distinguished as the red, blue and purple anthocyanin pigments in plant tissue (324). Research over the past decade has pointed to flavonoids having varied functions in the stress response (301), particularly in the role of UV protection (324). The first research completed on the role of flavonoids in the response to UV came by cloning the flavonoid-biosynthetic enzyme chalcone flavanone isomerase (*CHI*) and flavanoid-specific chalcone synthase (*CHS*) from *Petunia hybrida* (301). Van Tunen *et al*. discovered that *CHI* and *CHS* were co-expressed and co-ordinately regulated in a light-dependant manner.

By using *CHS* as the seed gene *g* to complete the Wigwams analysis and search for potential co-regulated genes, *CHI* was found to be co-expressed, and potentially co-regulated with *CHS*. This confirms experimental data which shows these two genes to be co-expressed. Figure 3.16 shows 32 genes, including *CHS* and *CHI*, to be co-expressed under UV-B, oxidative, and heat stress (expression profiles are shown in blue in these two conditions). By identifying *CHS* and *CHI* as co-expressed in UV-B stress, this confirms findings by Winkel-Shirley *et al*. (324) and van Tunen *et al*. (301) that these two genes have roles in the UV stress response, as well as roles in the biosynthesis of flavonoids (84).

Mutants deficient in *CHS* and *CHI* have been found to have a protective role against oxidative stress (90), confirming that both of these genes remain co-expressed in oxidative stress, as predicted by Wigwams. Transcript levels of *CHS* have been shown to decrease by 50% after heat treatment (78). Although a similar reduction of transcript levels of *CHI* has not been published, the discovery of *CHS* and *CHI* co-expression by Wigwams in heat stress is novel.

Mehrtens *et al*. (199) discovered by using quantitative real time reverse transcription-PCR that *MYB12*, a member of the MYB transcription factor family, activates the promoters of *CHS* and *CHI*, along with flavanone 3-hydroxylase (*F3H*) and flavonol synthase (*FLS*) (335). Both *FLS* and *F3H* were found to be co-expressed with *CHS* and *CHI* in response to UV-B stress, as well

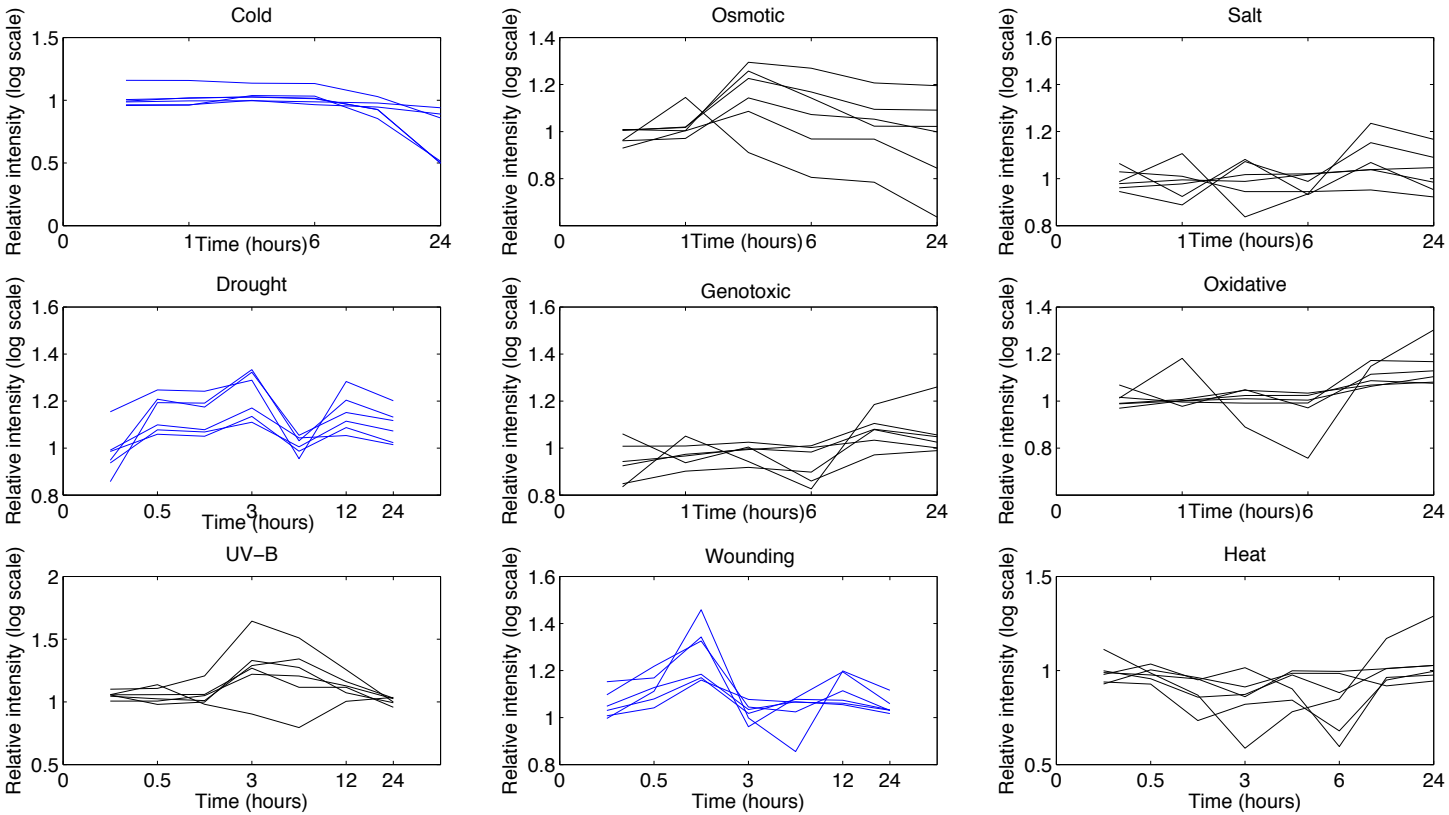



Figure 3.15: Wigwams identifies *CYP83B1*, *CYP79B2* and *CYP79B3* as co-expressed in drought, wounding and cold stress. The genes are significantly co-expressed with three additional genes (expression profiles shown in blue). This finding confirms previous results from Gachon *et al*. (101). Expression profiles shown in black are not significantly co-expressed in those particular stresses.

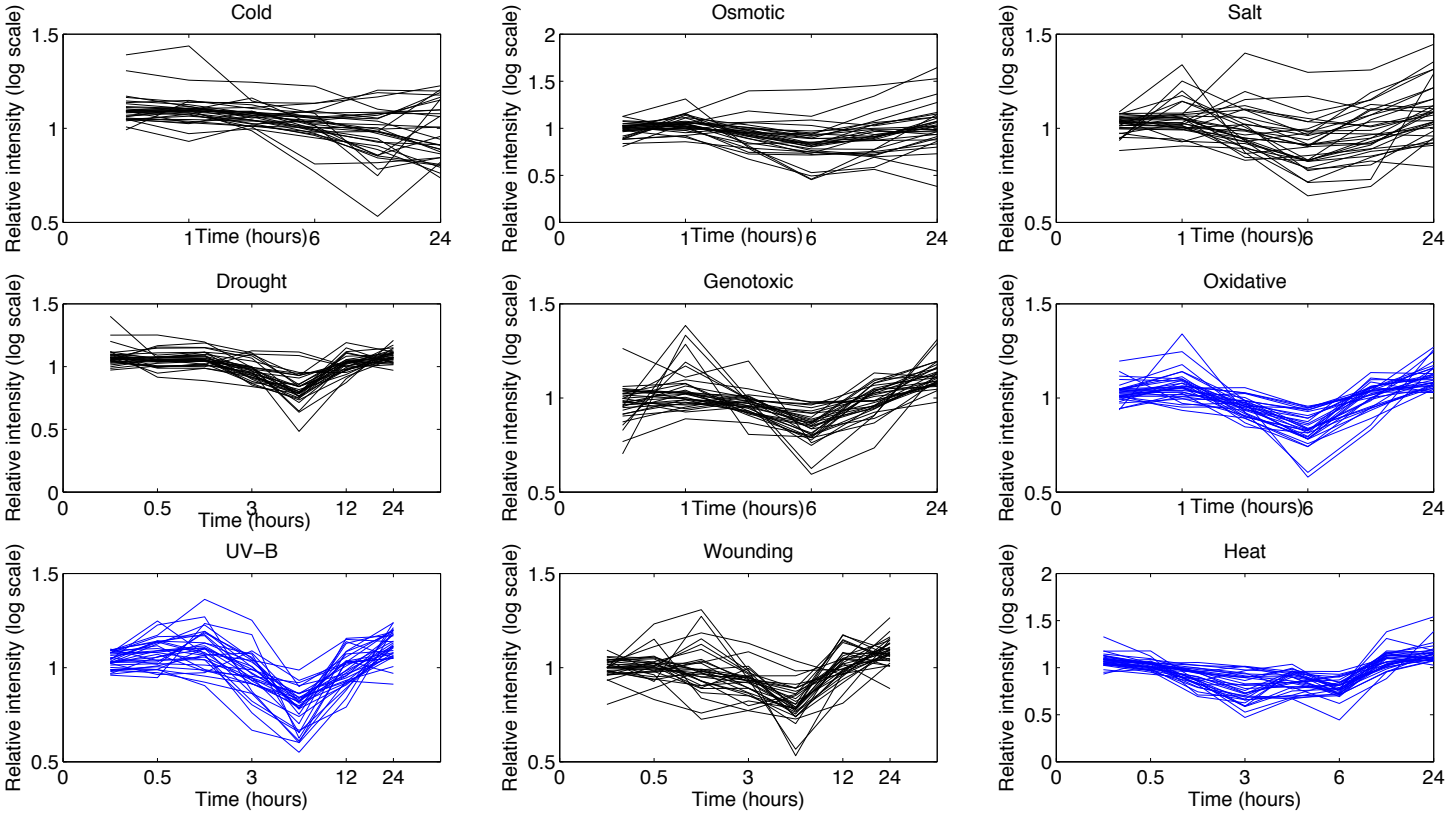



Figure 3.16: Wigwams identifies *CHS* and *CHI* as co-expressed in response to UV-B, oxidative, and heat stress. These genes are significantly co-expressed with 30 other genes (whose expression profiles are shown in blue). The *p*-value associated with this group of co-expressed genes in these three conditions is $2.349e^{-23}$.

as oxidative and heat stress, which substantiates Wigwams as a tool for discovering potentially co-regulated genes. Interestingly, a recent paper shows that the NAC family transcription factor *ANAC078* directly regulates the expression of *CHS*, *CHI*, and *F3H* in response to high light stress (212). Although high light stress is not investigated here, this finding does provide published evidence that *CHS* and *CHI* are co-expressed in an additional stress, as well as oxidative and heat stress.

### 3.2.4 Wigwams on simulated gene expression data

In order to demonstrate that Wigwams only detects evidence of co-expression if this evidence is really there, a simulated dataset was generated to confirm this. The hypothesis was that given a random set of gene expression data, Wigwams should not detect any significant potential regulons. Simulated, or '*in silico*' data, allows a researcher to inspect the performance of an algorithm: 'real' gene expression data is imperfect, due to a lack of control over noise levels (23).

#### 3.2.4.1 Generation of a simulated dataset

All variables were generated randomly. The dataset consisted of 15762 'genes'. This number was arrived at using a random number generator, and the identifiers for the 'genes' were randomly assigned. Scores stating which conditions these 'genes' were differentially expressed in were also randomly assigned, as shown in Table 3.2. Five conditions were randomly selected, along with randomly assigned time points for each condition. The gene expression data was simulated on a 'gene-by-gene-by-condition' basis, where the expression value was randomly selected between 0 and 1, in order to simulate normalised gene expression data.

Table 3.2: The number of differentially expressed genes unique to each simulated condition dataset. Numbers in parentheses state the number of genes for that stress which are included in the genes found to be differentially expressed in two or more stresses, and are then analysed using Wigwams

| Condition 1 | Condition 2 | Condition 3 | Condition 4 | Condition 5 |
|---|---|---|---|---|
| 7970 (7455) | 7845 (7380) | 7981 (7466) | 7825 (7340) | 7930 (7457) |

#### 3.2.4.2 Results

12808 potential pairwise regulons were generated in the Wigwams screening method. The pairwise potential regulon with the most significant overall $p$-value removed 12807 potential pairwise

regulons in the pruning step. This result validates that Wigwams is a critical clustering algorithm, and only provides evidence of co-expression across subsets of conditions if the evidence is really there to detect.

## 3.3 Discussion

In this chapter, the development of a novel tool, which aims to provide evidence of co-regulation, has been described. Limitation and suggested improvement, along with general conclusions will be discussed here.

### 3.3.1 Wigwams pruning

Whilst the pruning step is crucial for providing the most statistically significant and non-redundant potential regulons, the method itself is asymmetric in design: for example, the top ranked gene $n_i$ can prune the second ranked gene $n_j$, but $n_j$ cannot prune $n_i$. This issue is apparent when using the simulated '*in silico*' dataset to validate that Wigwams only detects real evidence of co-expression (as discussed in Section 3.2.4): it was hypothesised that given a random set of gene expression data, Wigwams should provide a critical output and not detect any significant potential regulons. Indeed, all pairwise regulons generated in the Wigwams screening method using the '*in silico*' were pruned in the subsequent pruning section, leaving only the 'top ranked gene' and its corresponding set of co-expressed genes. This one remaining potential regulon was not pruned due to the asymmetric nature in the design of the pruning method. This potential regulon is also expected not to be significantly co-expressed. However, there is no way to prune and remove this potential regulon, as all other evidence generated in the screening method has been removed. Therefore, this potential regulon would need to be critiqued by eye to discern whether it is truly showing evidence of co-expression.

### 3.3.2 Generalised hypergeometric distribution with three sets

As mentioned previously, the format in which Wigwams is currently written only allows the generation of *p*-values for overlaps between pairwise-compared sets of genes. Although Wigwams identifies potential regulons working across three or more stresses (as described in Section 3.2.1.4), the hypergeometric test does not calculate *p*-values for overlaps in gene membership in three or more

conditions. This presents a problem when we wish to assign *p*-values to overlaps of genes between three or more conditions. By using a generalised hypergeometric test, which allows the calculation of *p*-values of overlaps of genes for three or more conditions, this problem can be overcome.

The urn analogy is used as a classic metaphor for the hypergeometric distribution. If an urn contains black and white marbles, then drawing a white marble can be defined as a failure, and drawing a black marble can be defined as a success. *N* describes the number of all marbles placed in the urn, with *m* describing the number of black marbles in the urn, and *N - m* describing the number of white marbles in the urn. *k* then describes the number of balls drawn from the urn.

By using a generalised hypergeometric test, *p*-values can be calculated for overlaps in three or more conditions. Then, by selecting a *p*-value threshold, a potential regulon significant over three or more conditions can be picked in favour of its constituent pairwise regulons. We use the urn analogy again, with the white and black marbles in the urn. The generalised hypergeometric distribution calculates the probability of drawing at least *n* black balls twice. Here, the number of black balls drawn twice represents the overlap between the three sets. If you consider the Venn diagram shown in Figure 3.17 the generalised hypergeometric test works by first calculating the overlap of two lists of genes in a universe.

In order to calculate whether the overlap between the three lists is significant, the overlap between List 1 and 2 (shown in Figure 3.17 and described in Equation 3.6) is now considered as the 'List' which is compared against List 3 (shown as the black hashed area in Figure 3.18). This part of the generalised hypergeometric test is calculated in Equation 3.7, and is considered the 'second drawing' of black balls. Equation 3.7 aims to calculate the probability that out of the black balls drawn in the first draw, how many are drawn again in this second draw? In terms of the List of genes, how many of the genes found in the overlap of List 1 and 2, are also found in List 3?

#### 3.3.2.1 Notation

$$P^{(2)}(X \geq n) \text{ : the probability that at least } n \text{ black balls were drawn twice} \tag{3.3}$$

$$P(Y = o) \text{ : the probability that exactly o black balls were drawn in first draw} \tag{3.4}$$

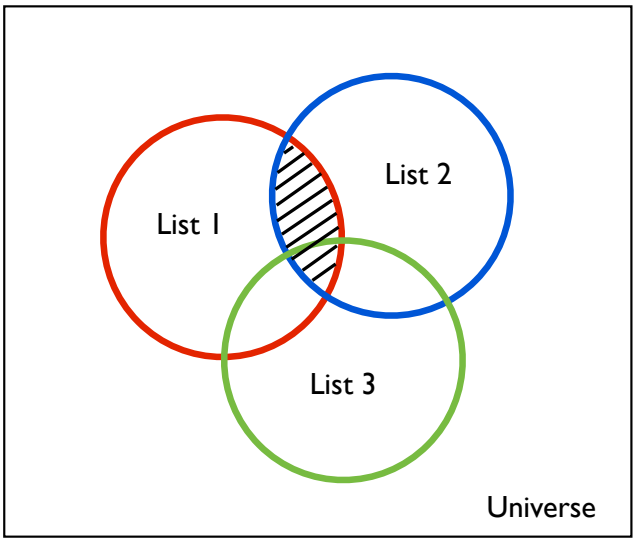



Figure 3.17: Calculating the overlap between three sets using the generalised hypergeometric test. The generalised hypergeometric test ultimately aims to calculate whether the overlap of the three lists is significantly large. In order to do this, the overlap of Lists 1 and 2 must first be calculated (indicated as the black hashed overlap between the two lists).

$P(X = n \mid Y = o)$ : the probability that out of o black balls drawn in the first draw, at least $n$ are drawn again

(3.5)

#### 3.3.2.2 Solution

$$P^{(2)}(X \geq n) = \sum_{o=n}^{\min\{k,l\}} P(Y = o) \cdot P(X \geq n \mid Y = o) \tag{3.6}$$

$$= \sum_{o=n}^{\min\{l,k\}} Hd(o, l, N - l, k) \cdot H(n - 1, o, N - o, m) \tag{3.7}$$

Note that Equation 3.6 will equal 1 for $n = 0$.

where $Hd$ refers to the density of the hypergeometric distribution ('dhyper' function in R (279)),

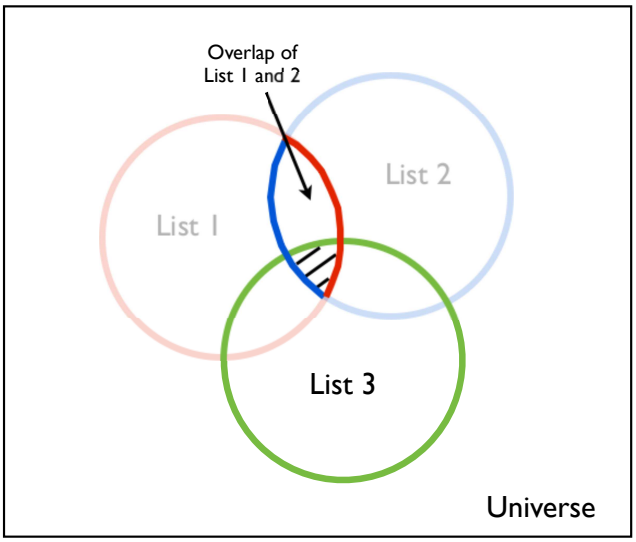



Figure 3.18: Calculating the overlap between three sets using the generalised hypergeometric test. In order the calculate whether the size of the overlap between three lists is significant, the overlap between List 1 and 2 is compared against List 3 (indicated by the black hashed area).

and $H$ refers to the hypergeometric distribution ('phyper' function in R (279)).

When dealing with overlaps over larger number of lists (i.e. four and above), the process is analogous to the one described here, adding an extra step for each additional list.

### 3.3.3 Distance metrics

Wigwams takes advantage of Pearson's product moment correlation coefficient, in order to calculate which genes have the most similar expression profile to the seed gene, in each stress individually. However, there are different methods of calculating the similarity of expression profiles. Yona *et al*. completed a study to analyse the usefulness and robustness of different distance metrics in detecting co-expression in time series datasets. Yona *et al*. evaluated the quality of various similarity measures to resolve which measure was the best for uncovering functional links, given gene expression data. The results showed that Pearson correlation was the most effective at detecting co-expression, closely followed by Spearman rank correlation, in time series datasets (334). There-

fore, it was justified to use the Pearson correlation measure for the correlation test in the Wigwams screening method. In spite of this, it is possible for the type of distance metric in Wigwams to be exchanged in favour of another distance metric.

### 3.3.4 Programming language selection

Whilst Matlab is a relatively easy programming language to use, and is quite powerful in terms of the built-in functions available, there are perhaps more appropriate languages that could improve on the computational time it takes for Wigwams to complete analysis. Part of Wigwams is written in Java, in order to communicate with the statistical programming language package R, as R calculates the *p*-value using the hypergeometric test to more significant figures than Matlab is capable. By re-writing Wigwams in Java, the computational time needed to complete the analysis could be reduced, as only two languages, instead of three, would be used. One advantage to using Matlab over Java, however, is the large library of mathematical functions Matlab possesses (225, 121), which Java does not possess.

### 3.3.5 Gene expression data types

Although Wigwams has been described in the context of using gene expression data generated by microarrays, Wigwams is also capable of searching for evidence of potential regulons in gene expression data generated by other methods. Recent developments in the technologies used to capture gene expression changes over time have meant that methods such as RNA-seq are now more commonplace, and are becoming increasing popular, in favour of microarray technology. RNA-seq improves on microarray limitations, by having low background signal and also greater sensitivity for genes expressed at low or very high levels (314).

mRNA abundance and gene expression changes can be measured using a variety of other methods, such as serial analysis of gene expression (SAGE) (304), nuclease protection (181) and differential display (175). All these techniques, including RNA-seq have the ability to measure gene expression changes in response to varying conditions. Since clustering methods have been used on gene expression data not produced by microarrays (232, 32), it is plausible to assume that Wigwams has the ability to utilise other forms of gene expression data.

### 3.3.6 Conclusions

In conclusion, Wigwams, a novel method developed to discover statistically significant genes that potentially share a common regulatory mechanism over multiple conditions, is capable of re-discovering known examples of co-expression, thus validating this method. Extensive research has been completed on metabolic pathways in *Arabidopsis*, and how these pathways differ in their function under certain conditions. Much experimental work and effort has been devoted to discovering how they differ, which has required substantial amounts of time. Wigwams was able to correctly uncover known examples of co-expression and co-regulation in a short space of time. This highlights the potential for uncovering new relationships between genes in stress conditions quickly, with the added bonus that experiments can be directed specifically at a small number of genes in specific conditions, rather than using large-scale reverse genetics approaches, which are time consuming and inefficient, as a means of discovering gene regulatory elements.

The main advantages of Wigwams are that it is capable of detecting evidence for potential regulons working across subsets of time series datasets, and does not require the potential regulon to fulfil the criteria of being significant in all datasets supplied. Wigwams also provides a concise output that is biologically meaningful, by only producing non-redundant and significant potential regulons. The method itself is computationally inexpensive. Limits on the size of datasets to be analysed will only be influenced by the practical restrictions of Matlab.

The generalised hypergeometric test, as described in Section 3.3.2, provides a means of extending, and therefore improving, Wigwams. As Wigwams currently stands, when extending pairwise regulons beyond two conditions (as described in Section 3.2.1.4), the size of the potential regulon must fulfil a specified criteria i.e. a potential regulon working over three conditions must contain at least 10 genes. By incorporating the generalised hypergeometric test into Wigwams, this approach to subsetting can be replaced by *p*-value criteria.

# Chapter 4

# Using multiple plant stress high resolution time series datasets to find potentially co-regulated genes

## 4.1 Introduction

Gene expression data, generated by tools such as microarrays, provide evidence for deciphering the regulatory codes that dominate expression changes under specific external conditions. By predicting co-regulated genes computationally, much time and effort can be saved from searching for these genes experimentally. Gene expression data can be mined for the identification of regulatory sub-networks that work across multiple conditions. By identifying genes whose expression is regulated by a common mechanism, and are therefore co-regulated, a common regulatory network that is active in the response to multiple conditions can be determined. A network with orthologous genes (i.e. genes which have the same function but occur in different species) (268, 100) to this common regulatory network can be identified in commercially important crop species. This network can be targeted in order to generate crops genetically engineered to withstand the effects of multiple stresses simultaneously.

Functionally alike genes are often co-expressed (133), therefore by considering genes with similar expression profiles, we can infer common functions of these genes, and common regulatory

mechanisms (82). However, these co-regulated genes must first somehow be discovered.

### 4.1.1 The role of transcriptional co-regulation in stress

Instances of transcriptional co-regulation have been extensively studied in other model organisms, such as *Drosophila melanogaster* and *Saccharomyces cerevisiae*. However, the modes of transcriptional co-regulation in *A. thaliana*, specifically in terms of stress responses, are less studied providing an opportunity to explore these responses in more depth. By integrating transcriptional regulatory information from motif analysis and network inference with gene expression data from time series microarrays, one can deduce the changes in network architecture, caused by transcription factors altering their interactions in certain conditions (188).

Co-regulation of gene expression is influenced by at least one common transcription factor binding to a *cis*-acting regulatory element (CARE) (333). For example, in *S. cerevisiae*, the transcription factors *ADR1* and *CAT8* co-regulate multiple pathways, mainly various forms of metabolism, such as ethanol and glycerol metabolic pathways (336): a small number (approximately 200) of *ADR1*-dependent genes are also *CAT8*-dependent, which include *ADH2* and *ACS1*, two genes involved in ethanol metabolism. Other genes that are co-regulated by *CAT8* and *ADR1* include *JEN1*, *ADY2*, *PUT4* and *ALP1*. Fourteen of the genes that are *ADR1*- and *CAT8*-dependent are also co-regulated during glucose depletion (274), confirming that the regulatory mechanism of *ADR1* and *CAT8* is shared across conditions, and fulfils the regulon criteria.

In *A. thaliana*, Clifton *et al*. found that the expression of alternative oxidase *AOX1a* and alternative NADH dehydrogenase *NDB2* was not only co-expressed, but was also co-regulated in *Arabidopsis* (71). Due to the presence of six similar sequence elements with a comparable arrangement within the promoter region of both *AOX1a* and *NDB2*, it was concluded that these two genes were likely to be co-regulated (125). Although the co-expression of *AOX1a* and *NDB2* has been seen to be maintained in response to various stresses, to date, no common regulator has been identified. It has been hypothesised that perhaps *ARR2* (71) or members of the *TCP* transcription factor family (206) could bind to CAREs found in the promoters of both *AOX1a* and *NDB2*, however, this has not been confirmed. Therefore, *AOX1a* and *NDB2* are potentially members of a regulon. Experimental techniques, such as Yeast-1-Hybrid, or network inference, would enable the identification of the transcription factor protein responsible for the co-regulation of *AOX1a* and *NDB2*.

### 4.1.2 Motivation

By applying the method described in Chapter 3 to high resolution time series data in *Arabidopsis*, potential regulons, which are genes under the control of the same transcription factor in multiple conditions, can be discovered. Gene members of regulons therefore have a common method of regulation. There is a lack of evidence for the role co-regulation has to play in the response to multiple stress in plants. By analysing these data using Wigwams, evidence for co-regulation in multiple plant stress responses can be found. However, Wigwams extracts information from gene expression data and provides information on which genes are co-regulated; Wigwams does not provide information on what is regulating these genes. Various methods are available to mine for examples of potential regulators.

If potential regulons are indeed co-regulated, there is likely to be a binding site motif within the promoter sequences for which the regulatory transcription factor can bind. The promoter sequences of gene members of potential regulons can be mined for statistically overrepresented promoter motifs, in order to identify a potential transcription factor binding site, as they are hypothesised to be under the control of the same regulatory mechanism. Motifs are conserved regions of DNA, consisting of a small number of nucleotides (dependant on the species), to which a regulatory protein can bind. Certain methods can be used, such as Multitple EM for Motif Elicitation (MEME) (19), which searches promoter sequences for overrepresented *de novo* or known motifs. If a motif is found to be statistically overrepresented in the promoter sequences of genes forming a potential regulon, this suggests co-regulation (82). If the motif is known in literature, this aids the identification of the regulatory protein binding to these promoters. Motif analysis results can be combined with results from network analysis tools, which are applied to time series gene expression data to predict regulators and infer connections between gene members of potential regulons (83). Network inference requires no prior knowledge of transcription factor binding motifs in the aforementioned genes, but can be combined to provide another facet of information, as shown in Figure 4.1. Potential regulons are analysed (using the hypergeometric analysis for motif finding in promoter sequences, as described in Section 2.7.5.1) for overrepresented plant promoter motifs. If a statistically significant motif is identified, and is a known binding site for the predicted regulatory transcription factor, then it is likely that this transcription factor is truly regulating the expression of this gene.

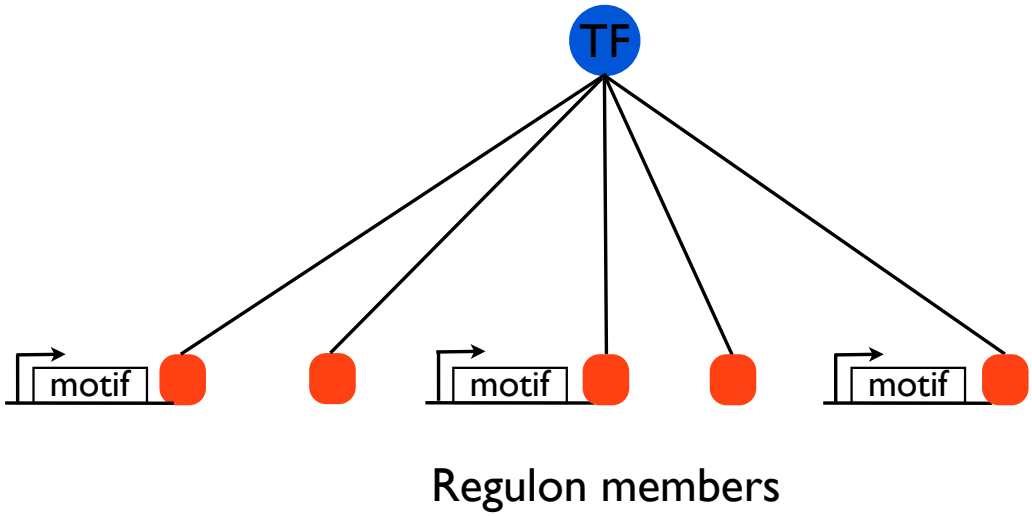



Figure 4.1: Motif analysis results are combined with network inference to provide further information on the likely transcription factor regulation of potential regulons. Here, network inference has predicted a transcription factor (blue circle) to co-regulate transcription factor members of a potential regulon (red squares). Motif analysis (using the hypergeometric test outlined in Section 2.7.5.1) which has been completed on the promoters of all potential regulon members has identified an overrepresented motif for this potential regulon. If the motif is a known binding site for the predicted regulatory transcription factor, then it is likely that this is a substantial method of regulation.

### 4.1.3 Datasets

The PRESTA group used microarray analysis to obtain high resolution time-course profiles of changes in gene expression during response to stress in *A. thaliana* leaves. These datasets were used to identify differentially expressed genes in response to each individual stress. Six stress responses were investigated in this way: responses to biotic stress caused by the pathogens *B. cinerea* (Windram *et al.* (322) and *P. syringae* pathovar *tomato* DC3000; responses to abiotic stresses drought and high light; and developmental stresses long (43) and short day senescence. For the biotic and abiotic stresses the microarray design included both treated (i.e. stressed) and mock (i.e. unstressed) samples. Both mock and treated experiments each included four biological replicates over the whole time course, with the exception of long day senescence, which had

eight replicates. In the case of senescence, where there were no mock samples, there were four biological replicates for the senescence samples. For the purposes of analysis using Wigwams, the MAANOVA (326) output that combined biological replicates by taking the average (mean) expression value at each time point over all four replicates was used for each treated dataset, at each timepoint.

CATMA ((8); `www.catma.org`) version 3 microarray slides were used for the *B. cinerea* and senescence time course experiments, whereas CATMA version 4 microarray slides were used for the remaining experiments. Due to the differences in the number of probes on version 3 and version 4 slides (version 4 has 2242 additional probes compared to version 3 slides), Wigwams was completed on the 30,366 probes common to both version 3 and version 4 slides in all datasets. Only stress-treated datasets for each stress were considered (i.e. the mock expression values were not considered in Wigwams analysis).

## 4.2 Application of Wigwams to multiple time-series of gene expression during stress

Wigwams (see Chapter 3) was applied to genes that were differentially expressed in two or more of six time series gene expression datasets (as explained in detail below in Section 4.1.3), in order to identify potential regulons sharing a common regulatory mechanism over a subset of stress conditions. By subsequently integrating motif analysis and predicted gene regulatory networks, more evidence that the gene members of potential regulons may be co-regulated can be inferred. GO term analysis, completed using BiNGO (191), identifies statistically overrepresented biological functions within the potential regulons, further lending weight to these genes being functionally similar.

A discussion on the parameters used for applying Wigwams to these datasets, followed by potential regulons found using this method on these datasets, and significant motifs and GO terms found to be overrepresented within these potential regulons will be presented. These findings will be integrated with literature available on motifs, in order to predict common upstream regulators of these potential regulons.

### 4.2.1 Wigwams considers only genes whose expression is differentially expressed in two or more conditions

For each dataset, a boolean vector of ones and zeros (where one means this gene is differentially expressed in this particular stress, and zero is false) was generated to identify genes whose expression is changing over time in response to these stresses. If a gene exhibits a differential expression profile in response to treatment (stress) compared to expression during mock treatment, then there is a significant difference in expression between treatment and mock, which cannot have occurred by chance. A gene must show a difference in expression, compared to mock expression, for two consecutive time points in order to be considered differentially expressed. This allows the inclusion of the curated lists of differentially expressed genes, generated for each dataset (see Sections 2.7.2 and 2.7.3 for details on how the lists were generated for high light and drought respectively). The curated lists of differentially expressed genes were generated as described in Table 4.1. Figure 4.2 describes the process of generating the time series datasets by means of a PERT chart, detailing the main stages in dataset construction in the project in relation to each other. Different analyses to identify lists of differentially expressed genes were used in some datasets. This is due to advances in analyses of time series dataset becoming available at different times. Also, due to the time at which the microarray experiments were carried out, different versions of CATMA (8) had been developed, as illustrated in Figure 4.2.

Table 4.2 shows the number of genes found to be differentially expressed in each stress individually. 11263 genes were found to be differentially expressed over time in response to two or more stresses, and only these genes were considered in the Wigwams analysis, as we are only interested in finding genes with a common regulatory mechanism across two or more stresses.

#### 4.2.1.1 Normalisation of gene expression data

Gene expression data for each stress, using the MAANOVA (326) output where all four biological replicates had been combined, was transformed in order to make the overall amplitude of expression comparable. This allows the comparison of shapes of gene expression, rather than the absolute expression values. In order to do this, the gene expression data per stress was normalised separately with a mean $\mu$ of zero, with a standard deviation of one.

Table 4.1: Comparative table detailing the generation of differentially expressed genes for each time series dataset, including: reference to paper and researcher who carried out analysis to identify differentially expressed genes; version of CATMA (8) slides; method used to identify differentially expressed genes

| | *P. syringae* pv. DC3000 | *B. cinerea* | Short day senescence | Long day senescence | High light | Drought |
|---|---|---|---|---|---|---|
| Reference | Dr. Laura Lewis (University of Warwick) | Dr. Oliver Windram (University of Warwick) (322) | Emily Breeze (University of Warwick) | Emily Breeze (University of Warwick) (43) | Myself - Section 2.7.2 | Myself - Section 2.7.3 |
| CATMA (8) version slides | v4 | v3 | v3 | v3 | v4 | v4 |
| Method of identifying DE genes | MAANOVA F-Test (326), BATS (12) and GP2S (270) | GP2S (270) and MAANOVA F-Test (326) | MAANOVA F-Test (326) | MAANOVA F-Test (326) and Time-Course (275) | GP2S (270) and MAANOVA F-Test (326) | GP2S (270) and MAANOVA F-Test (326) |
| Duplicated genes removed? | Yes | Yes | Yes | Yes | Yes | Yes |
| Non-hybridising probes removed? | Yes | Yes | Yes | Yes | Yes | Yes |

Table 4.2: The number of differentially expressed genes unique to each PRESTA stress condition dataset. Unless otherwise stated, lists of differentially expressed genes were generated by myself (as shown in Sections 2.7.2 and 2.7.3). Numbers in parentheses state the number of genes for that stress which are included in the 11263 genes found to be differentially expressed in two or more stresses, and are then analysed using Wigwams

| High light | *P. syringae* DC3000 (L. Lewis, University of Warwick) | *B. cinerea* (322) | Short day senescence (E. Breeze, University of Warwick) | Long day senescence (43) | Drought |
|---|---|---|---|---|---|
| 7036 (5624) | 5632 (4850) | 9838 (7536) | 6241 (5568) | 10258 (7780) | 1761 (1690) |

### 4.2.2 Parameters

As initially described in Section 3.2.1.1, and illustrated in Figure 3.9, Wigwams first screens a given list of genes for evidence of potential regulons, by firstly generating pairwise potential regulons. The inputs, in terms of using the datasets mentioned in Section 4.1.3 above, are as follows:

1. Set of time course expression measurements under $C$ conditions and for all genes in set $G$. The six datasets described in Section 4.1.3 were used for analysis.

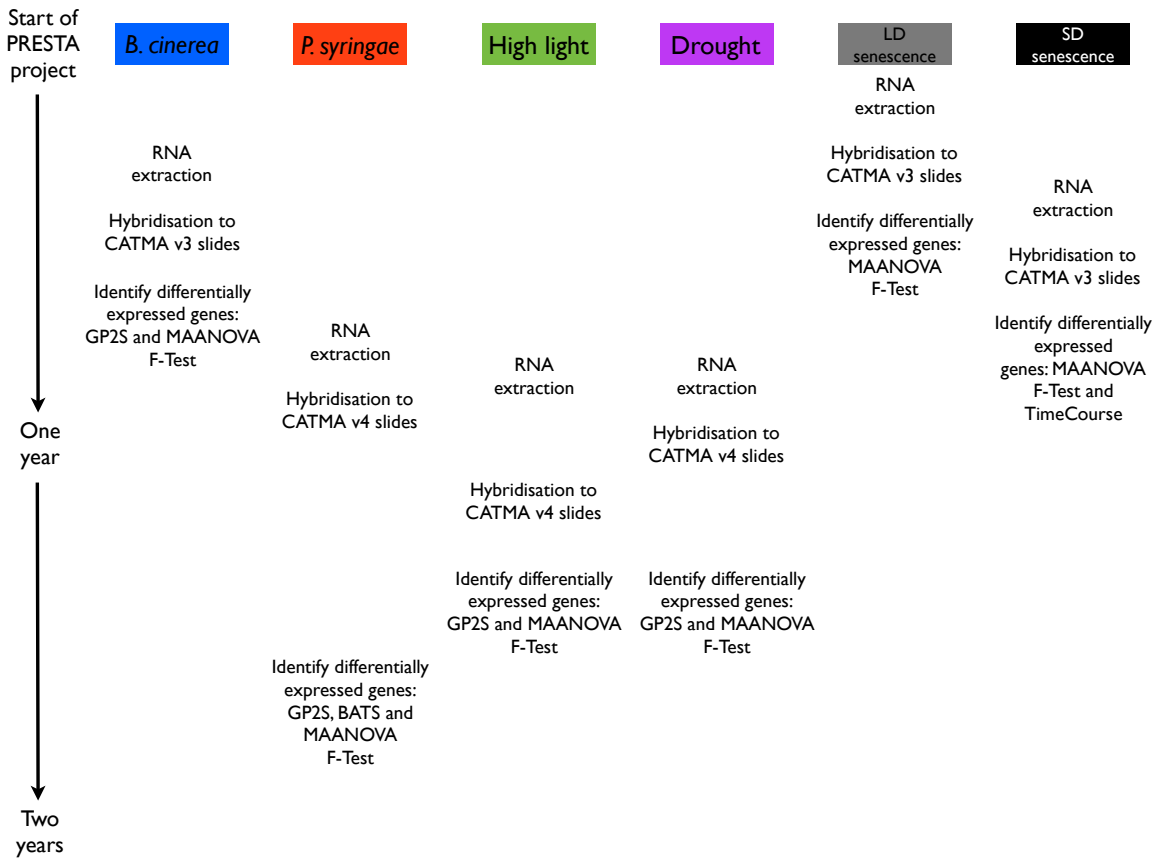



Figure 4.2: PERT chart for the construction of each of the PRESTA time series datasets, in relation to time. The major stages in generating the datasets are shown, along with the different versions of CATMA (8) slides used for each time series, and the statistical methods used to generate lists of differentially expressed genes in each dataset.

2. For each gene $g \in G$ and each time course $c \in C$ a boolean value indicating whether gene $g$ is differentially expressed in condition $C$. The lists described in Section 4.2.1 were used to manually create a matrix of Boolean values, indicating whether a gene was differentially expressed in a particular dataset.

3. A similarity measure of time course expression values for pairs of genes. Here, Pearson product moment correlation was used.

4. A set of gene list sizes $S$. A range of $S$ values were chosen between 50 and 250, at intervals of 50.

The lists of genes that are most correlated to $g$ in each time course $C$ were chosen to range between 50 and 250, at intervals of 50. Choosing smaller interval sizes would increase the computational intensity of Wigwams. It was anticipated that any regulatory network of genes working across multiple stresses was unlikely to be larger than 250 nodes (genes). However, intervals of 50 were

chosen as overlaps in gene membership may be more significant at certain sizes than others, which can be refined at a later stage. For instance, an overlap of 26 genes out of a top correlated gene list of 50 would provide a stronger *p*-value than those same 26 genes overlapping in a larger top correlated gene list of 100 genes. Therefore, we wish to capture this particular evidence of potential co-regulation, rather than it being discarded for being non-significant at larger sizes of top correlated gene lists.

In order to select for significant potential regulons (as illustrated in Figure 3.12), three parameters were required: A *p*-value cutoff of 1.000e$^{-05}$ was chosen, based on using Bonferroni Multiple Testing Correction of 30,336 genes over six datasets (as shown in Equation 4.1). Bonferroni MTC was chosen to counteract the problem of multiple comparisons: when using microarrays, expression levels of thousands of genes can be measured, which may not be found in a second, repeat experiment. Bonferroni MTC can counteract this. The significance level was required to be 0.05 at most, therefore, by considering each probe in the datasets (30,336), the threshold was decided as:

$$\left(\frac{0.05}{30,336}\right) \times 6 \approx 1.000e^{-05} \tag{4.1}$$

The correlation threshold was chosen to be 0.75, above which, the correlation coefficient *r* is deemed to represent two similar expression profiles. We arrived at this value by performing the following test, whereby a gene was chosen at random, and correlation tests using Pearson's product moment correlation (PPMC) were performed against the expression profile of this gene and the expression profile of other randomly selected genes, in the same stress condition dataset. An *r* value below 0.75 is deem to represent two dissimilar expression profiles. This is illustrated in Figure 4.3, which shows the *r*-values associated when comparing two expression profiles.

Figure 4.4 a) shows the expression profiles of the randomly selected probe, CATMA5a57110 (shown in red). The subsequent Figures 4.4 b) - d) show the expression profiles of randomly selected genes (in red). The Pearson's correlation test was performed per stress between CATMA5a57110 and each of the randomly selected genes shown in Figures 4.4 b) - d), and the resulting *r* values are shown in Table 4.3.

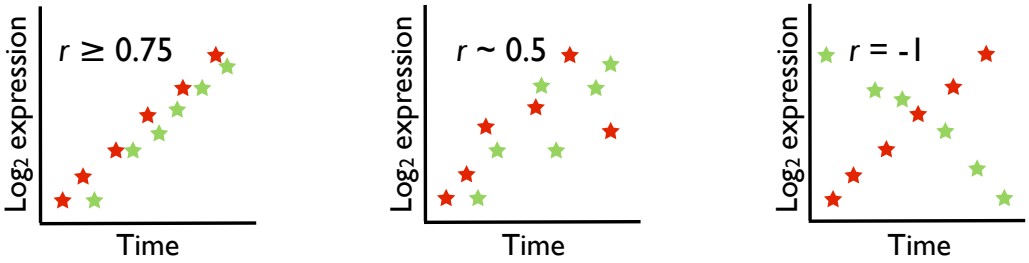
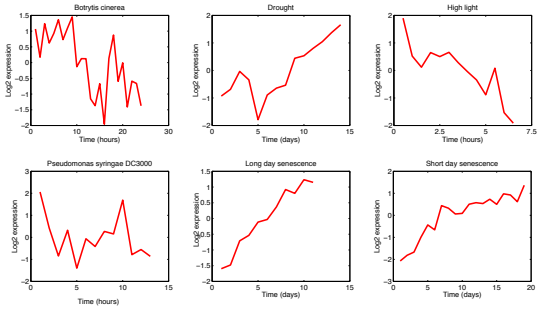



Figure 4.3: Pearson's product moment correlation coefficient *r*-values associated with different comparisons of two expression profiles. Here, two expression profiles are represented as red and green points on a scatter plot. When these two expression profiles are similar, the corresponding *r*-value is greater or equal to 0.75. Expression profiles calculated as having an *r*-value of around 0.5 are moderately correlated. Expression profiles that are highly anti-correlated generate an *r*-value of -1.

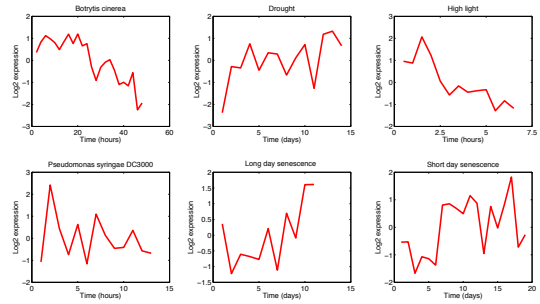
By comparing the expression profiles of the genes in the drought and high light conditions shown in Figure 4.4 b) to the gene expression profiles in the same conditions in Figure 4.4 a), it can be noted that the intra-condition expression profiles are dissimilar. This is supported by the correlation values shown in Table 4.3, generated by performing a correlation test of the gene expression of probe CATMA1a50250 to the expression of probe CATMA5a57110 in each condition, where the *r* values are 0.4918 and 0.5655 respectively. This can be interpreted as 'medium' correlation.

Comparing the expression profiles of the gene probes in Figure 4.4 c) in both Short Day and Long Day Senescence, with those of the expression profiles of the genes in the same conditions in Figure 4.4 a), you can see that the expression profiles are very similar in both Short Day Senescence profiles. This is supported by the correlation *r* value given in Table 4.3 (0.9183), which confirms that the gene expression profiles of the seed gene are very similar. Conversely, take into account

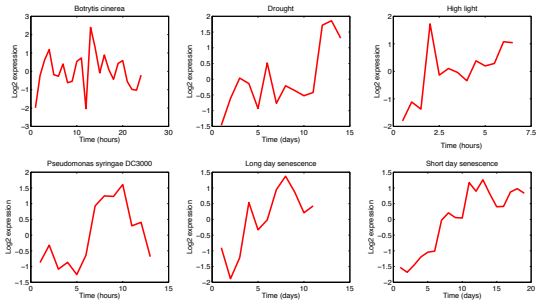
a)




b)




c)




d)

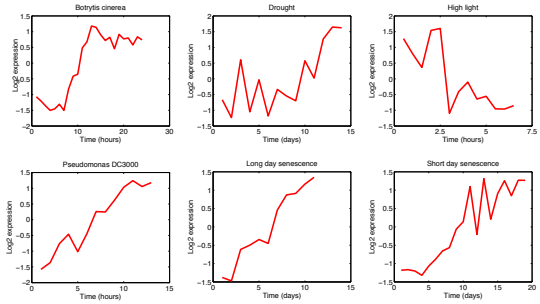



Figure 4.4: a) Expression profiles of randomly selected seed gene probe CATMA5a57110 (shown in red) in each of the six time course datasets, as described in Section 4.1.3. Pearson's product moment correlation (PPMC) test was performed per stress between these expression profiles and each of the randomly selected gene probes shown in b) - d), in order to decide upon a value for the correlation threshold. b) Expression profile of randomly selected gene probe CATMA1a50250 (here shown in red) in each of the six time course datasets, as described in Section 4.1.3. c) Expression profile of randomly selected gene probe CATMA1a50080 (here shown in red) in each of the six time course datasets, as described in Section 4.1.3. d) Expression profile of randomly selected gene probe CATMA1a50890 (here shown in red) in each of the six time course datasets, as described in Section 4.1.3.

the expression profiles of Long Day Senescence in both Figures 4.4 a) and c): the correlation *r* value given in Table 4.3 is 0.7892, suggesting the expression profiles are similar, but do not have as strong a correlation as witnessed in Short Day Senescence. By looking at these expression profiles by eye, we can see that there are differences in the profiles in Figure 4.4 a) and c) in Long Day Senescence.

Comparing the expression profiles of the genes in both Short Day and Long Day Senescence in Figure 4.4 d), with those of the expression profiles of the probe in the same conditions in Figure 4.4 a), you can see the profiles are very similar i.e. the expression profiles of the genes in Short Day Senescence in Figure 4.4 c) are very similar to those gene expression profiles in the Short Day Senescence in Figure 4.4 a). Also compare this to the correlation values associated with these two conditions (Table 4.3): the correlation *r* values between probes CATMA5a57110 and CATMA1a50890 in Short Day and Long Day senescence are 0.8188, and 0.9822 respectively. These values suggest the gene expression profiles of these two probes are very similar in their respective conditions.

From Figures 4.4 a) - d) and Table 4.3 we can conclude that the optimum value to use for the correlation cut-off parameter in the pruning algorithm is 0.75. This was chosen due to correlation values of 0.8 and higher corresponding to intra-condition gene expression profiles between sets of correlated genes being highly similar, as seen in Figure 4.4 d). Consider the expression profiles of the gene probe subject to long day senescence stress in both Figures 4.4 a) and d): both sets of expression profiles are highly similar, a fact reflected with the high *r* value shown in Table 4.3 (0.9822). However, correlation *r* values less than 0.75 show that intra-condition gene expression profiles between conditions can be dissimilar (shown in Figure 4.4 b)): for example, the expression profiles of genes subject to *B. cinerea* stress in both Figure 4.4 a) and b) are dissimilar, and the *r* value for the correlation between these two sets of expression profiles is 0.6676. This suggests that different regulatory processes are occurring in these two seed genes, and should be taken into consideration in the pruning step.

For the final parameter, in order for two *p*-values to be significantly different, and for both regulons to be kept, the *p*-value had to differ by an order of magnitude of 10 or more. If two potential regulons had a *p*-value that did not differ by an order of magnitude of 10, then it could be assumed that the potential common regulatory mechanism identified by these two regulons is the same or

Table 4.3: Correlation *r* values generated using PPMC for gene expression data of CATMA probe CATMA5a57110 in each condition against randomly selected probes.

| Randomly selected probe | *B. cinerea* | Drought | High Light | *P. syringae* DC3000 | Long Day Senescence | Short Day Senescence |
|---|---|---|---|---|---|---|
| CATMA1a50250 | 0.6676 | 0.4918 | 0.5655 | -0.2656 | 0.6105 | 0.6036 |
| CATMA1a50080 | 0.3087 | 0.7209 | -0.5804 | 0.2914 | 0.7892 | 0.9183 |
| CATMA1a50890 | -0.7619 | 0.7467 | 0.6329 | -0.2214 | 0.9822 | 0.8188 |

very similar. Therefore, the potential regulon with the less significant *p*-value will be discarded.

Potential regulons were subject to thresholds based on either the overall *p*-value (for regulons significant in two stresses) or the number of genes in the regulon (for regulons significant in three or more stresses). The thresholds were as follows:

1. Two conditions: *p*-value of $1.0000e^{-05}$, as discussed above and summarised in Equation 4.1.
2. Three conditions: minimum regulon size of 10 genes.
3. Four conditions: minimum regulon size of 8 genes.
4. Five conditions: minimum regulon of 5 genes.
5. Six conditions: minimum regulon size of 5 genes.

These numbers were chosen in order to prioritise potential regulons found in subsets of greater numbers of stresses, and identify greater numbers of genes co-expressed over many stresses. For example, identifying a potential regulon working over six conditions, but only having two gene members would provide less evidence, and a lower chance of identifying a core gene regulatory network than a potential regulon working over six conditions and having five gene members. This will provide greater gene candidates for identifying predicted regulatory transcription factors of potential regulon gene members via network inference.

## 4.3 Wigwams identifies 465 potential regulons co-expressed across subsets of stresses

2679 significant pairwise regulons were identified on completion of the Wigwams algorithm (as illustrated in Figure 3.9) and selection of informative regulons (as illustrated in Figure 3.12). By applying thresholds to extend beyond pairs of stresses, as described in Section 4.2.2 previously,

465 significant regulons were identified as working across subsets of the six datasets, represented by 3417 unique genes. *A. thaliana* allocates 5% of its genome to encode solely for transcription factors (243). It seems reasonable, therefore, that 5% of all gene members of all the potential regulons encode proteins annotated as transcription factors or to be involved in transcriptional regulation (according to annotation by TAIR9), demonstrating that transcription factors have been fully represented in these potential regulons.

The breakdown of how many potential regulons were found per number of stress combinations, along with mean gene membership size of potential regulons, is shown in Table 4.4.

Table 4.4: The number and average (mean) size of potential regulons found per number of stress combinations

| Number of conditions a singular potential regulon is co-expressed in | Number of potential regulons | Mean number of genes per potential regulon |
|---|---|---|
| Two | 412 | 22 |
| Three | 35 | 15 |
| Four | 17 | 14 |
| Five | 1 | 7 |
| Total | 465 | |

### 4.3.1 Analysis of potential regulons

Out of the 465 potential regulons identified by Wigwams, we wanted to identify any particular stress condition combinations that occurred more frequently than others, to discover if regulatory mechanisms were more likely to be shared between specific subsets of stresses. For example, it is reasonable to assume that regulatory mechanisms will be shared between the biotic stress responses *B. cinerea* and *P. syringae* pv. DC3000, due to both being infectious agents. However, we wish to address whether there is crosstalk between biotic, abiotic and developmental stresses: were any potential regulons identified as having a common regulatory mechanism in a biotic and abiotic stress response? Are potential regulons more likely to be identified as working within the confines of biotic or abiotic stress, for example? This is of particular interest, due to the limited amount of literature available on crosstalk between biotic and abiotic stresses (99): currently, literature is focused on identifying crosstalk within biotic (246) or abiotic stresses (261, 262, 47).

Table 4.5 shows the frequency of all combinations of stress conditions. Co-expression of genes in the conditions long and short day senescence provide the highest occurrence of pairwise-condition

potential regulons, with 18% of the occurences (85 occurrences out of the total 465). This finding is not surprising, as long and short day senescence induce similar stress responses, with subtle differences: for example, systems responsible within the plant for recognising day length interact with the regulation of redox signals. During short day senescence, redox-mediated acclimation signals are redirected, which enables efficient usage of light, whereas in long-day conditions, priority is given to systems required to prevent oxidative damage (31).

There were 75 occurrences of potential regulons with genes significantly working across both *B. cinerea* and *P. syringae* DC3000, which accounts for 16% of the total 465 potential regulons identified by Wigwams. Although *B. cinerea* and *P. syringae* DC3000 are both pathogens, their modes of infection are substantially different to each other, due to *B. cinerea* being a necrotrophic fungus and *P. syringae* DC3000 being a hemibiotrophic bacteria: *B. cinerea* actively decomposes the host tissue for its survival (300); *P. syringae* DC3000, on the other hand, being a biotroph, secretes effectors that signal the plant to direct nutrients to *P. syringae* lesions growing in the apoplast (283). Both mechanisms of infection can cause different responses within the plant, with downstream responses having opposite phenotypes (159). For example, overexpression of WRKY33 has been seen to enhance susceptibility to *P. syringae*, whereas *wrky33* mutant plants showed enhanced susceptibility to *B. cinerea* (343). However, there has been research that suggests that both infections are linked via the ABA signalling pathway (76), confirming that there is crosstalk between these two stresses. Other modes of crosstalk between *B. cinerea* and *P. syringae* include the changes in plant secondary metabolite accumulation. *B. cinerea* is thought to produce signals capable of regulating metabolite levels in the lesion, and the signal will disperse through uninfected tissue (156). This effect on metabolite levels has also been seen in *P. syringae* infections. The *P. syringae* pathovar *tomato* DC3000 strain uses its type III secretion system to block the secretion of antimicrobial secondary metabolites, however (21). Given this literature evidence that *B. cinerea* and *P. syringae* pv DC3000 are capable of influencing the levels of metabolites within the plant, it seems reasonable for Wigwams to discover a large number of significant potential regulons exhibiting crosstalk between these two stresses.

Potential regulons found at larger combinations of conditions have a lower frequency of occurrence than the frequency of pairwise regulons. This may be due to the unlikely event that plants have evolved a shared regulatory mechanism that is involved in the response to three or more stresses.

As the number of conditions increases, the smaller the number of genes differentially expressed in those conditions becomes. Therefore, fewer regulons will be found as the number of conditions in which they are significantly co-expressed over increases. This is to be expected, as Seki *et al*. found only 22 genes to be differentially expressed over cold, high salinity and drought, when analysing the expression of 7000 *A. thaliana* genes using microarrays (255).

The high occurrence of potential regulons in certain stresses highlights that the response to these particular conditions is a complex one: the pruning algorithm has been completed correctly, so all of these potential regulons exhibit none, or very little, gene membership redundancy. Therefore, certain stress responses involve more genes than others.

These results provide evidence for genes that are capable of operating in multiple stress responses, and provides evidence for a complex regulatory response system to abiotic and biotic stresses. However, it also highlights the possibility that the regulons in combinations of conditions with a low frequency have a less complex response to multiple conditions, with fewer genes in the regulatory network. Genes found to be potentially co-regulated using Wigwams provide one level of information in terms of a regulatory network capable of working across many stress responses. Wigwams provides information on sets of genes that are co-regulated, but does not provide information on the common regulator, or regulators, of these genes. These regulators need to be identified, using motif analysis and network inference, in order to infer a regulatory network around the gene members of potential regulons.

Table 4.6 describes the complexity of the various multi-stress combinations that Wigwams is capable of identifying. An interesting observation is the subtle differences between long day and short day senescence. Although, as previously stated, these two stresses are very similar in their responses, they do have small differences which lead to different genes being involved in the response to each senescence condition. By referring to Table 4.6 we can see that there are only seven significant potential regulons working over long day senescence and drought. Compare this to the 31 regulons working over short day senescence and drought, and it can be seen that there are differences between the two senescence processes. It is known that drought conditions can accelerate the onset of senescence (255, 31), however, these findings suggest that it is a specific type of senescence, which has a distinct set of genes dedicated to its stress response, rather than the genes shared between the short day and long day senescence response.

Table 4.5: Frequency of significant potential regulons. The number of conditions is listed in ascending order, and for each group of numbers of conditions are listed the frequency of potential regulons found for each multi-condition combination (highlighted in grey). Combinations with zero frequency are not listed.

| Number of conditions | *B. cinerea* | Drought | High Light | *P. syringae* DC3000 | Long Day Senescence | Short Day Senescence | Frequency |
|---|---|---|---|---|---|---|---|
| 2 | | | | | ■ | ■ | 85 |
| 2 | ■ | | | ■ | | | 75 |
| 2 | ■ | | | | ■ | | 54 |
| 2 | | | | ■ | ■ | | 49 |
| 2 | ■ | | | | | ■ | 39 |
| 2 | | ■ | | | | ■ | 31 |
| 2 | | | ■ | | | ■ | 17 |
| 2 | | | | ■ | | ■ | 12 |
| 2 | | ■ | ■ | | | | 10 |
| 2 | | ■ | | | ■ | | 7 |
| 2 | ■ | ■ | | | | | 6 |
| 2 | | | ■ | | ■ | | 4 |
| 2 | | ■ | | ■ | | | 2 |
| 2 | | | ■ | ■ | | | 2 |
| 3 | ■ | | | | ■ | ■ | 21 |
| 3 | ■ | | ■ | | | ■ | 19 |
| 3 | | | | ■ | ■ | ■ | 5 |
| 3 | ■ | | | ■ | ■ | | 3 |
| 3 | ■ | | | ■ | | ■ | 2 |
| 3 | | ■ | | | ■ | ■ | 2 |
| 3 | ■ | ■ | ■ | | | | 1 |
| 3 | ■ | ■ | | | | ■ | 1 |
| 4 | ■ | | | ■ | ■ | ■ | 16 |
| 4 | ■ | ■ | | | | ■ | 1 |
| 5 | ■ | | ■ | ■ | ■ | ■ | 1 |

Whilst considering the combinations of stresses in Table 4.6 we can see the frequency of certain combinations is influenced, to a certain extent, by the numbers of differentially expressed genes in Table 4.2. For example, there are more regulons containing long day senescence than any other combination of stresses. Long day senescence provided the greatest number of differentially expressed genes (10258), of which 7780 of these differentially expressed genes were also differentially expressed in at least one other stress. Therefore, it is to be expected that this stress should contribute the most to potential regulons.

Although drought stress generated the fewest differentially expressed genes over time, there were fewer regulons containing high light than there were drought: approximately five times more genes

Table 4.6: The combinations of conditions covered by significant regulons is complex. Here, the combinations of stresses are shown, with the frequency of regulons with this combination of stresses. Each stress condition is abbreviated in the first column, and in subsequent columns plus one, two, three and four other stress conditions. The total number of regulons significant in all quantities of conditions (i.e. co-expressed over two conditions, or three conditions etc.) is shown at the bottom of this table. The total number of potential regulons found to contain the common stress per row is given in the last column. Abbreviations are as follows: B - *B. cinerea*; D - Drought; H - High light; P - *P. syringae* DC3000; L - Long day senescence; S - Short day senescence.

| | +1 | +2 | +3 | +4 | Total |
|---|---|---|---|---|---|
| B | D - 6<br>P - 75<br>H - 19<br>L - 54<br>S - 39 | D, H - 1<br>D, L - 1<br>P, S - 3<br>P, L - 2<br>L, S - 21 | D, L, S - 1<br>P, L, S - 16 | H, P, L, S - 1 | 239 |
| P | B - 75<br>D - 2<br>H - 2<br>L - 49<br>S - 12 | B, S - 3<br>B, L - 2<br>L, S - 5 | B, L, S - 16 | B, H, L, S - 1 | 167 |
| H | B - 19<br>D- 10<br>P - 2<br>L - 4<br>S - 17 | B, D - 1 | | B, P, L, S - 1 | 54 |
| D | B - 6<br>H - 10<br>P - 2<br>L - 7<br>S - 31 | B, H - 1<br>B, L - 1<br>L, S - 2 | B, L, S - 1 | | 61 |
| L | B - 54<br>P - 49<br>D - 7<br>H - 4<br>S - 85 | B, D - 1<br>B, P - 2<br>B, S - 21<br>D, S - 2<br>P, S - 5 | B, D, S - 1<br>B, P, S - 16 | B, H, P, S - 1 | 248 |
| S | B - 39<br>P - 12<br>D - 31<br>H - 17<br>L - 85 | B, P - 3<br>B, L - 21<br>D, L - 2<br>P, L - 5 | B, D, L - 1<br>B, P, L - 16 | B, P, H, L - 1 | 233 |
| Total | 412 | 35 | 17 | 1 | |

were found to be differentially expressed in response to high light stress than drought stress over time. This finding may be due to high light not sharing a common regulatory mechanism with the other stresses being investigated here. Indeed, previous research has shown genes involved in high light stress to also be co-expressed in oxidative and freezing stresses (99), neither of which are investigated here. However, ten potential regulons were found to be co-expressed in drought and high light: crosstalk between these two stresses has been proven previously, providing confidence in the results seen here (154). Kimura *et al*. found, using microarray analysis on stresses plants, that a number of heat shock proteins and factors (HSPs and HSFs respectively) were induced by both high light and drought stress (154). Wigwams identified eight members of the HSF and HSP families to be co-expressed, along with other genes, in drought and high light.

Whilst there is a clear preference towards a developmental (long and short day senescence) stress split, there does not appear to be an abiotic and biotic stress split in terms of the frequency of potential regulons observed. 357 potential regulons are co-regulated in at least one of the senescence stresses, demonstrating that there is a developmental stress split. Despite 75 instances of potential regulons being co-regulated in *B. cinerea* and *P. syringae* DC3000, there is still a considerable number of potential regulons bridging the biotic, developmental, and abiotic stresses. There appears to be a lack of an abiotic stress split, as there are three times more potential regulons co-regulated in drought and short day senescence than there are co-regulated in drought and high light. Similarly, there are nineteen potential regulons co-regulated in high light and *B. cinerea* compared to ten potential regulons co-regulated in high light and drought. These numbers do not provide sufficient evidence to conclude there is an abiotic stress split in the number of potential regulons identified by Wigwams.

#### 4.3.1.1 Bioinformatical analysis of potential regulons to reveal functionally related genes

The identification of overrepresented motifs within promoters of genes in potential regulons aids the identification of co-regulated genes, and also the potential regulatory transcription factor (or the family of transcription factors). However, in order to discern whether these potential regulons are truly co-expressed, these genes can be analysed, using GO term analysis, to reveal whether they share a common function: genes that are truly co-expressed are more likely to share a regulatory mechanisms (and therefore, be co-regulated) (116). In order to determine whether Wigwams

identifies functionally related genes, a number of bioinformatical analyses were applied to the resulting 465 potential regulons. Wigwams provides evidence for a set of genes being co-expressed over multiple conditions, and to predict whether these co-expressed genes are co-regulation, various bioinformatical analyses can be performed.

**Gene Ontology analysis reveals potential functions of co-expressed genes** Functionally alike genes are often co-expressed (133). By using the GO term tool BiNGO (191), the potential regulons can be analysed to identify significantly overrepresented biological functions. If a potential regulon has a statistically significant overrepresented function, then it is more likely these genes are co-expressed than a potential regulon not identified to have any common function. A custom annotation file containing only genes found to be differentially expressed in two or more stresses was used for this GO term analysis. All *p*-values mentioned henceforth have been corrected using Benjamini and Hochberg false discovery rate, using a significance level of less than, or equal to, 0.05, as part of the BiNGO (191) analysis tool.

Table 4.7 shows only the most significantly overrepresented GO terms for each potential regulon. Only 219 potential regulons were found to have at least one significantly overrepresented GO term. The remaining 246 potential regulons may have gene members that did not match a GO term annotation, and were simply not significantly overrepresented for a given function. These potential regulons may not be co-expressed, due to a lack of evidence supporting the gene members being functionally alike. However, this may be due to the GO annotation of *A. thaliana* genes being out of date. Therefore, these potential regulons, which do not have gene members overrepresented for a GO term, may have been identified as false negatives for co-expression.

Table 4.7: Frequency of significantly overrepresented GO terms in 219 potential regulons, ranked in decreasing order. Only GO terms with most significant corrected *p*-value are considered.

| GO term function | Total potential regulons with function as most significantly overrepresented function |
|---|---|
| Structural constituent of ribosome | 28 |
| Plastid | 23 |
| Chloroplast thylakoid | 18 |

| | |
|---|---|
| Chloroplast | 13 |
| Chloroplast part | 11 |
| Plastid part | 11 |
| Thylakoid part | 7 |
| Photosystem | 6 |
| Chlorophyll binding | 5 |
| Chloroplast thylakoid membrane | 5 |
| Photosynthetic membrane | 5 |
| Thylakoid | 5 |
| DNA metabolic process | 4 |
| Gene expression | 4 |
| Iron ion binding | 4 |
| Light-harvesting complex | 4 |
| Organelle subcompartment | 4 |
| Plastid thylakoid | 4 |
| Amino acid activation | 3 |
| Autophagy | 3 |
| Chloroplast stroma | 3 |
| Mitochondrion | 3 |
| Photosynthesis | 3 |
| Plastid stroma | 3 |
| Translation | 3 |
| Anchored to membrane | 2 |
| Anchored to plasma membrane | 2 |
| Apoplast | 2 |
| Cellular catabolic process | 2 |
| Cellular macromolecule biosynthetic process | 2 |
| Cellulose biosynthetic process | 2 |
| Glucosinolate biosynthetic process | 2 |
| Glycosinolate biosynthetic process | 2 |

| | |
|---|---|
| Intrinsic to plasma membrane | 2 |
| Macromolecule metabolic process | 2 |
| Oxidoreductase activity | 2 |
| Oxygen binding | 2 |
| Pigment biosynthetic process | 2 |
| Plastid thylakoid membrane | 2 |
| Protein complex | 2 |
| Response to chitin | 2 |
| Response to heat | 2 |
| S-Glycoside biosynthetic process | 2 |
| Thylakoid lumen | 2 |
| Thylakoid membrane | 2 |
| tRNA aminoacylation for protein translation | 2 |
| tRNA metabolic process | 2 |
| tRNA processing | 2 |
| Zinc ion binding | 2 |
| Acetyl CoA catabolic process | 1 |
| Amine-lyase activity | 1 |
| Anti-apoptosis | 1 |
| ATP binding | 1 |
| Autophagy vacuole | 1 |
| Auxin homeostasis | 1 |
| Biosynthetic process | 1 |
| Carbon-sulphur lyase activity | 1 |
| Catalytic activity | 1 |
| Cell cycle arrest | 1 |
| Cellular biosynthetic process | 1 |
| Cellular protein metabolic process | 1 |
| Cellular response to JA stimulus | 1 |
| Cellulose metabolic process | 1 |

| | |
|---|---|
| Cellulose synthase activity | 1 |
| Chloroplast thylakoid lumen | 1 |
| Chromatin assembly | 1 |
| Chromosome organisation | 1 |
| Cysteine-type peptidase activity | 1 |
| Cytoplasm | 1 |
| diaminopimelate biosynthetic process | 1 |
| DNA replication | 1 |
| Electron transport chain | 1 |
| Extracellular region | 1 |
| Fructose-2,6-bisphosphate 2-phosphatase activity | 1 |
| Glutamine family amino acid biosynthetic process | 1 |
| Glutamine metabolic process | 1 |
| Glycerophospholipid metabolic process | 1 |
| Golgi apparatus | 1 |
| Heterocycle metabolic process | 1 |
| Hydrolase activity | 1 |
| Indolalkylamine biosynthetic process | 1 |
| Indolalkylamine metabolic process | 1 |
| Intracellular bound organelle | 1 |
| Intracellular lumen | 1 |
| Intrinsic to membrane | 1 |
| JA mediated signalling pathway | 1 |
| Large ribosomal subunit | 1 |
| Macromolecular complex | 1 |
| Macromolecule biosynthetic process | 1 |
| Membrane-enclosed lumen | 1 |
| NAD(P)H dehydrogenase complex assembly | 1 |
| NADH dehydrogenase complex | 1 |
| NADH dehydrogenase complex assembly | 1 |

| | |
|---|---|
| ncRNA metabolic process | 1 |
| Negative regulation of apoptosis | 1 |
| Negative regulation of cell cycle | 1 |
| Negative regulation of homeostatic process | 1 |
| Negative regulation of RNA metabolic process | 1 |
| Negative regulation of telomere maintenance via telomerase | 1 |
| Negative regulation of transcription, DNA-dependent | 1 |
| Non-membrane bound organelle | 1 |
| Nucleoid | 1 |
| Nucleotide-sugar metabolic process | 1 |
| Organelle lumen | 1 |
| Peptidyl-amino acid modification | 1 |
| Peroxiredoxin activity | 1 |
| Plastid envelope | 1 |
| Positive gravitropism | 1 |
| Promoter binding | 1 |
| Protein import into chloroplast stroma | 1 |
| Protein Serine/Threonine kinase inhibitor activity | 1 |
| Protein-DNA complex assembly | 1 |
| Receptor binding | 1 |
| Regulation of glucan biosynthetic process | 1 |
| Regulation of multi-organism process | 1 |
| Regulation of response to biotic stimulus | 1 |
| Regulation of systemic acquired response | 1 |
| Respiratory chain complex I | 1 |
| Response to carbohydrate stimulus | 1 |
| Response to dessication | 1 |
| Response to endogenous stimulus | 1 |
| Response to JA stimulus | 1 |
| Response to stress | 1 |

| | |
|---|---|
| Response to water | 1 |
| Ribonucleoprotein complex | 1 |
| Ribonucleotide binding | 1 |
| RNA metabolic process | 1 |
| Signal transmission | 1 |
| Signalling process | 1 |
| Small conjugating protein activity | 1 |
| Small ribosomal subunit | 1 |
| Sterol biosynthetic process | 1 |
| Sterol metabolic process | 1 |
| Strictosidine synthase activity | 1 |
| Sugar transmembrane transport activity | 1 |
| TCA cycle | 1 |
| Tryptophan biosynthetic process | 1 |
| Tryptophan metabolic process | 1 |
| Tryptophan synthase activity | 1 |
| Ubiquitin-protein ligase activity | 1 |
| Xyloglucan:xyloglucosyl transferase activity | 1 |

By observing the results in Table 4.7, 'Structural constituent of ribosome' is the most frequent significantly overrepresented GO term in the potential regulons. All of the potential regulons with this annotation as their most significant GO term have a majority of gene members which encode ribosomal subunits, and are annotated (TAIR9) as 'structural constituent of ribosome'. Upon further investigation of the potential regulons with this GO term as the most significantly overrepresented annotation, Table 4.8 shows potential regulons significantly co-expressed over senescence and high light are more likely to have this function overrepresented. This suggests that the process of translation, and possibly the regulation of translation, may have an important role in the response to high light and senescence stresses. Previous studies have shown that environmental stress on *A. thaliana* plants has an impact on the regulation of mRNA translation (145), suggesting

that the stress response is not limited to transcriptional regulation, as previously stated in this thesis.

Table 4.8: Frequency of potential regulons with 'Structural constituent of ribosome' as most significantly overrepresented GO term, in descending order, and the stress combinations the gene members are significantly co-expressed in.

| Stress combination | Frequency of potential regulons |
|---|---|
| High light and short day senescence | 8 |
| Long and short day senescence | 6 |
| Drought and short day senescence | 3 |
| *B. cinerea* and short day senescence | 3 |
| Drought and high light | 2 |
| *B. cinerea*, long and short day senescence | 2 |
| *B. cinerea*, *P. syringae* DC3000, long and short day senescence | 1 |
| *B. cinerea* and drought | 1 |
| *B. cinerea* and high light | 1 |

Considering only GO terms associated with stress responses in Table 4.7, Figure 4.5 shows the functional annotations of 20 potential regulons with significantly overrepresented functions referring to 'Response', 'Regulation', or a plant hormone signalling pathway. 'Response to chitin' and 'Response to heat' were the most frequent significantly overrepresented functions found in these potential regulons (represented as yellow and red sections in Figure 4.5, respectively).

The 'Response to chitin' function was found to be overrepresented in one potential regulon with gene members co-expressed in *P. syringae* DC3000 and long day senescence, and another potential regulon with gene members co-expressed in both long and short day senescence. Chitin is a PAMP, which is recognised through plants through an unknown receptor, to initialise PAMP-triggered immunity (PTI) (346). Therefore, it is likely that the potential regulon with gene members co-expressed in *P. syringae* DC3000 and long day senescence are involved in the same downstream response that can be triggered by chitin via an unknown receptor: a component in the long and short day senescence stress response may also trigger the co-expression of these genes via the unknown receptor. However, due to the influence of literature in creating GO annotations for genes (73, 51, 50), these genes may only be currently known to belong to a process involved in the response to chitin. Therefore, the identification of these genes co-expressed in long and short day senescence also potentially acting downstream of the chitin-binding unknown receptor suggests a point of crosstalk between these stresses.

Interestingly, both of the potential regulons found to be overrepresented for the function 'Response

to heat' are co-expressed in conditions drought and high light. Gene expression has been shown to change when tobacco plants are affected by drought and heat stress simultaneously, including altered expression of a homolog of *A. thaliana* WRKY4 (244). This suggests that there is a link, perhaps in the form of a regulatory network which is activated by both heat and drought stresses: the responses to these stresses have genes in common. *WRKY4* was not found in these potential regulons co-expressed in drought and high light conditions; however, co-expression of *WRKY4* in heat and drought stress may be identified by Wigwams if gene expression data for heat stress was included. Rizhsky *et al*. analysed the *Arabidopsis* transcriptome by dual stressing the plants with both drought and heat stress, subsequently discovering 454 differentially expressed transcripts in response to this combinatorial stress (245). A large number of transcripts encoding members of the HSP family were differentially expressed in this dataset: 39 members of the HSP family are also found to be differentially expressed in response to high light in the PRESTA dataset. Both potential regulons found to be overrepresented for this function contain different members of the HSP family, providing further evidence that these two groups of genes are co-expressed.

**Motif analysis of promoters of genes in potential regulons reveals transcription factor binding sites that suggest co-regulation** Transcriptional regulation is an important process in the response to stress in plants (67). Transcriptional regulation that occurs in non-stress conditions will most likely have to undergo change in order to provide the correct stress response. This change occurs via the interaction of transcription factors and *cis*-regulatory elements, or motifs (348), which are short, conserved sequences of DNA. Co-expressed genes may have similar regulatory control (277), such as the same transcription factor, or set of transcription factors, regulating their expression. Such co-expressed genes are said to be 'co-regulated'.

To identify groups of co-expressed genes as being co-regulated, the promoters of the gene members can be analysed for the presence of a conserved motif. If a motif is found to be significantly overrepresented in the promoters of genes in potential regulons, it is likely that these genes have a common transcription factor, or transcription factors, regulating their expression. It is possible, once the motif is known, to identify the transcription factor regulating the expression of these genes in the literature. If a potential regulon has a significantly overrepresented motif in the gene member promoter sequences, then it is likely that these genes share a common regulatory mechanism over multiple stresses, and are therefore co-regulated.

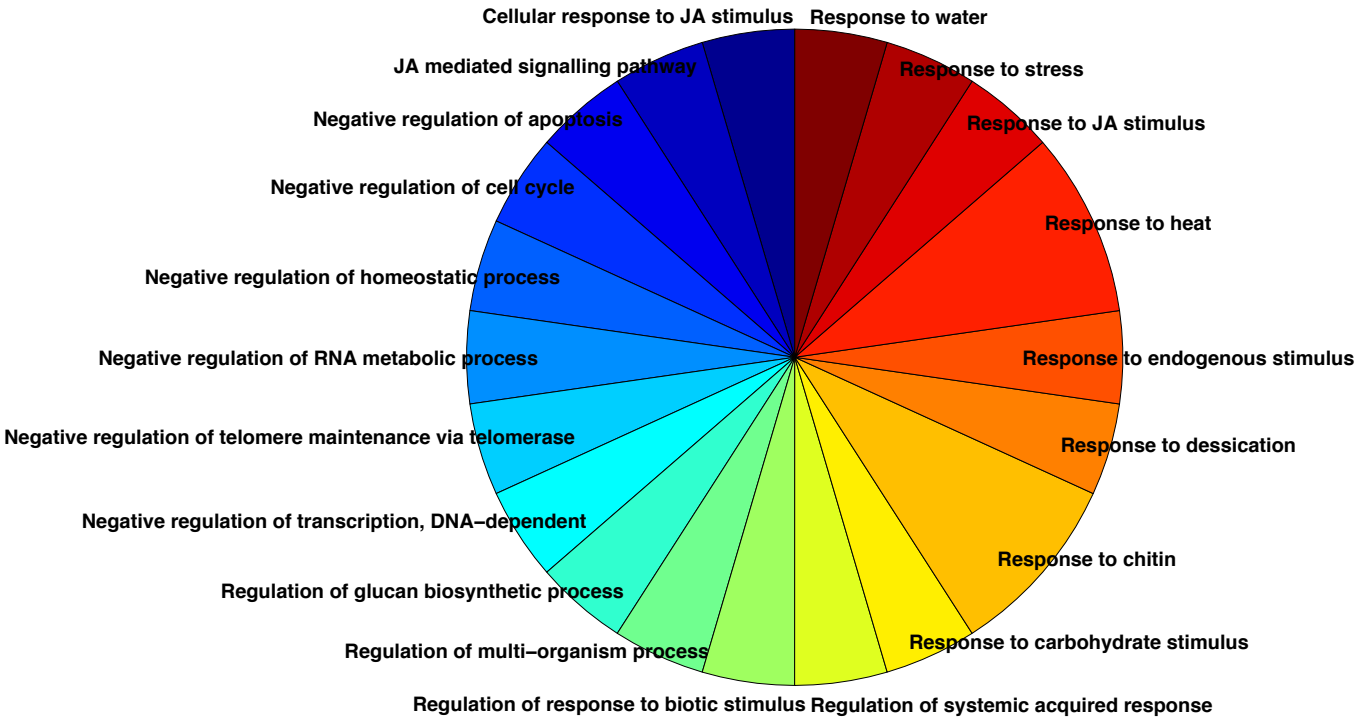
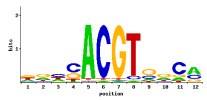
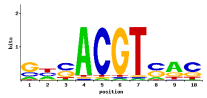
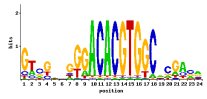
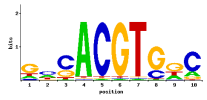
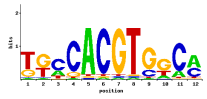



Figure 4.5: Functional annotations of 20 potential regulons with significantly overrepresented GO terms referring to 'Response', 'Regulation', or a plant hormone signalling pathway (i.e. Jasmonic Acid). 'Response to heat' and 'Response to chitin' were the most abundant GO terms found using this search criteria.

As Wigwams provides a facet of information for potential co-regulation, all significant potential regulons were submitted to the hypergeometric motif test (as mentioned in Section 2.7.5.1). This hypergeometric test scans the 500 bp sequence upstream of the transcriptional start site (retrieved from TAIR) of potential regulon members against 349 PSSMs. Each PSSM is a representative of that particular motif (43), after clustering of PSSMs from the TRANSFAC database (195), and the PLACE database (122). Using Bonferroni MTC, a *p*-value of less than, or equal to 0.0001 was deemed significant.

Using this *p*-value cut-off of $1.000e^{-04}$, 89 of the 465 (19%) potential regulons had known plant motifs significantly overrepresented within the promoters of the gene members. Motifs may have been present in the promoters of gene members of the remaining potential regulons, which have not been formally identified as a potential binding site. Another hypothesis is that the remaining potential regulons are regulated in another manner, that is not transcriptional: though the gene expression profiles appear co-expressed, this does not infer transcriptional regulation. For example, the plant hormones JA and MeJA are known to alter gene expression by inducing polypeptides called jasmonate-induced proteins (JIPs) (241). Of the 349 known plant promoter motifs, 61 non-redundant motifs were identified as significantly present and overrepresented within the promoters of gene members of the potential regulons.

Table 4.9: *p*-values of motifs found to be significantly overrepresented in the promoter regions of the genes present in potential regulons. Shading denotes the *p*-value cut-off used, based on Bonferroni MTC, to select these motifs i.e. the darker the shade, the more significant the *p*-value of that motif being statistically over-represented in the gene promoters in that particular potential regulon. No shading denotes a *p*-value cut-off of $\leq 1.000e^{-07}$. Grey shading denotes a *p*-value cut-off of $\leq 1.000e^{-09}$. Gene membership and expression profiles for each potential regulon presented, along with the conditions these genes are co-expressed in, are given in Appendix B. Abbreviations are as follows: B - *B. cinerea*; D - Drought; H - High light; P - *P. syringae* DC3000; L - Long day senescence; S - Short day senescence. PLACE database or TRANSFAC database identifiers for each motif are given.

| | Potential regulon number and stress combination | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Motif | 197<br>H, S | 168<br>B, P, L, S | 457<br>L, S | 365<br>B, P | 199<br>B, S | 320<br>L, S | 408<br>H, S | 117<br>L, S | 344<br>B, S | 280<br>B, H | 416<br>H, S | 456<br>H, S | 29<br>B, L, S | 166<br>L, S | 23<br>B, P, L, S |
|  M00182 | | | | | | | | | | | | | $1.84e^{-08}$ | | |
|  M00375 (135) | | | | | | | | | | $3.27e^{-08}$ | | | | | |
|  M00399 (69) | | | | $7.07e^{-08}$ | | | | | | $3.71e^{-08}$ | | | | | |
|  M00942 | | $3.36e^{-08}$ | | | | | | | | | | | $6.25e^{-10}$ | | $6.06e^{-08}$ |
|  M00367 (135) | | | | | | | | | | | | | $2.89e^{-08}$ | | |
|  M00441 | | $9.61e^{-08}$ | | | | | | | | | | | | | |

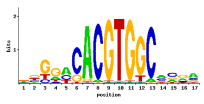
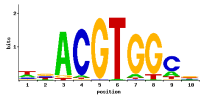
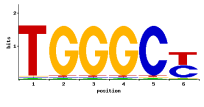
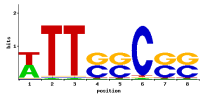
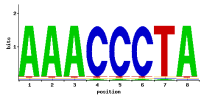
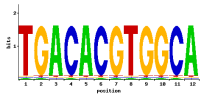
| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  M00442 | | | | | | | | | | $4.78e^{-08}$ | | | $7.23e^{-09}$ | | |
|  M01584 (166) | | | | | | | | | | | | | $2.17e^{-08}$ | | |
|  S-000474 (317) | | | $1.78e^{-09}$ | | $3.73e^{-09}$ | $1.83e^{-09}$ | | $3.28e^{-08}$ | $3.59e^{-11}$ | | | | | | |
|  S-000476 (303) | | | | | | | | | | | | | | $6.33e^{-09}$ | |
|  S-000472 (276) | $2.16e^{-09}$ | | $2.72e^{-11}$ | | | | $3.32e^{-10}$ | | | | $4.67e^{-09}$ | $1.03e^{-08}$ | | | |
|  S-000345 (106) | | $7.34e^{-08}$ | | | | | | | | | | | | | |

Table 4.9 shows which motifs (shown by their logos in the first column, which represent the PSSM) are significantly overrepresented in the promoters of gene members of potential regulons. Only the most stringent examples (with *p*-values of $1.000e^{-07}$ and increasing in stringency) of overrepresented motifs, and the regulons they were found in. The gene members of the potential regulons shown in Table 4.9 can be found in Tables B.1 - Table B.15 in Appendix B.

Two important observations are highlighted in Table 4.9: firstly, that potential regulon 29, has genes significantly acting in the *B. cinerea* and short and long day senescence responses, and has a high proportion of motifs significantly overrepresented in its promoters. One of the motifs found to be significantly overrepresented for potential regulon 29, with a corrected *p*-value of $1.84e^{-08}$, has

the conserved sequence of 'ACGT'. This symmetrical motif is recognised by bZIP proteins (294), and the flanking sequence to this core motif (i.e. the nucleotides present on either side of the core sequence) give this motif the specificity to bind certain transcription factors (93).

The other motifs found to be significantly overrepresented within the promoters of gene members of potential regulon 29 all have the 'ACGT' core, emphasising the previous statement which stated that the flanking sequences around this core are what give the motif its specificity to bind certain transcription factors. This is also the reason why so many apparently 'non-redundant' motifs have this 'ACGT' core: the sequences flanking this core affect DNA-binding specificity (93). Foster *et al*. investigated the limitations of binding specificity using different combinations of nucleotides flanking the 'ACGT' core. Two subfamilies of G-Box Binding Factors (GBFs) were discovered binding to 13 different flanking sequences (93), highlighting the importance of recognising the flanking sequences around the 'ACGT' core, as shown in the motifs in Table 4.9 with a core 'ACGT' sequence, but varying flanking sequences.

One motif found to be significantly overrepresented in potential regulon 29 is the palindromic, hexameric G-Box motif (5'-CACGTG-3') (202). Members of the bZIP transcription factor family in plants have been shown to bind to this promoter motif. Experimental investigation has led to the discovery of certain transcription factors that are capable of binding the G-Box motif: *ABI5* and *ABI3*, which encode ABA-insensitive transcription factors; *GBF1*, *GBF2* and *GBF3*, which encode G-Box Binding Factors, which have been shown to be induces by ABA during stress (184); and *HY5*, which encodes a transcription factor whose binding to the promoter of *ABI5* is enhanced by ABA (60). These transcription factors are potential co-regulators of this particular regulon.

Although the hypergeometric test used to discover overrepresentation of known plant promoter motifs within the potential regulons should correct for redundant motifs, the remaining motifs found to be significantly overrepresented in potential regulon 29 all have the conserved sequence of the ABF motif (5'-ACGTGGC-3'). Transcription factors identified as binding to the ABF motif (ABRE binding factors) belong to a subfamily of the bZIP transcription factor family (69), and have been shown to activate a large number of genes involved in the ABA response to stress.

Table 4.10, which is modified from a paper by Kim *et al*. (153), shows the phenotypes both plants over expressing members of the *ABF* family and *ABF* mutants have in response to abiotic stress.

Table 4.10: Abiotic stress phenotypes of plants over expressing *ABFs* or *ABF* mutants, as originally described in Kim *et al*. (153)

| | *ABF1* - At1g49720 | *ABF2* - At1g45249 | *ABF3* - At4g34000 | *ABF4* - At3g19290 |
|---|---|---|---|---|
| Expression | ABA- and cold-inducible | ABA- and salt-inducible | ABA- and salt-inducible | ABA-, cold-, drought- and salt-inducible |
| Overexpression | n/a | Drought (mixed), salt, cold, heat and oxidative tolerance | Salt (mixed), drought, cold, heat and oxidative tolerance | Salt hypersensitive, drought, cold, heat and oxidative tolerance |
| Knockout phenotypes | n/a | glucose insensitive | ABA, salt and drought insensitive | ABA, salt and drought insensitive |

Drought is not represented in Table 4.9 as a stress condition in which gene members of a potential regulon are co-expressed in, and also have overrepresented motifs in their promoters. It is possible, due to literature evidence, that ABFs are not responsible for the regulation of these particular potential regulons.

Table 4.11: Identifying possible transcription factor regulators of potential regulons overrepresented for 'ACGT' motifs in promoters of gene members. Members of the bZIP family that are differentially expressed in the same stress conditions as the potential regulon gene members are co-expressed in (indicated as abbreviations: B - *B. cinerea*; H - High light; P - *P. syringae* DC3000; L - Long day senescence; S - Short day senescence.)

| 168 - BPLS | 365 - BP | 280 - BH | 29 - BLS | 23 - BPLS |
|---|---|---|---|---|
| ABF1 | ABF1 | ABF1 | ABF1 | ABF1 |
| AtbZIP25 | AtbZIP18 | AtbZIP14 | AtbZIP25 | AtbZIP25 |
| AtbZIP51 | AtbZIP22 | AtbZIP19 | AtbZIP26 | AtbZIP51 |
| | AtbZIP25 | AtbZIP20 | AtbZIP34 | |
| | AtbZIP51 | AtbZIP22 | AtbZIP40 | |
| | AtbZIP9 | AtbZIP23 | AtbZIP51 | |
| | | AtbZIP24 | AtbZIP53 | |
| | | AtbZIP27 | AtbZIP54 | |
| | | AtbZIP37 | AtbZIP60 | |
| | | AtbZIP51 | | |
| | | AtbZIP56 | | |
| | | AtbZIP69 | | |

Table 4.11 shows the members of the bZIP family that are differentially expressed in the same stress conditions as the potential regulon gene members are co-expressed in; with the aim to identify possible transcription factors capable of binding to the 'ACGT' core, these groups of co-expressed genes have overrepresented within their promoters. The list of bZIPs was acquired from TAIR (`http://www.arabidopsis.org`), compiled using the AGRIS database. It can be seen from Table 4.11 that although literature evidence does not support *ABF1* regulating the ex-

pression of these regulons in stresses other than drought, *ABF1* is differentially expressed in all five stress conditions represented in this table. Therefore, *ABF1* is a likely regulator of the genes in these potential regulons in novel stress responses.

As the G-Box, ABF and 'ACGT' core motifs have been identified as significantly overrepresented within the potential regulon 29, and these motifs are known to have involvement with ABA signalling and bZIP proteins, we can assume that these two findings have repercussions for understanding the senescence response. Indeed, Breeze *et al*. (43) identified that senescence results in an increase in the levels of ABA in leaves. They also discovered that genes involved in the response to ABA stimulus were upregulated in response to senescence stress.

The role of ABA is somewhat controversial in terms of its role in the *B. cinerea* response, as it is poorly understood. However, ABA-related mutants were shown to be more resistant to *B. cinerea* by Adie *et al*. (6). The presence of these motifs in the promoters of the gene members of potential regulon 29 suggest a role for this potential regulon in ABA signalling. However, the gene expression profiles are down regulated during all three stress responses (see Figure B.13 in Appendix B), suggesting that their role is not protective against *B. cinerea* and long and short day senescence stress.

The second observation to note is that two motifs, the Site II motif and the TELO-box motif are the most abundant motif found to be statistically overrepresented in these potential regulons. Very little literature evidence exists for Site II motifs, despite being one of the most ubiquitous plant promoter motif, along with the G-box motif. However, the literature that does exist surrounding the Site II motif suggests that it is capable of binding TCP20 (292), a transcription factor member of the TCP family. It is also believed that Site II motifs are conserved promoter elements that control the co-ordination of ribosomal protein genes (292).

Table 4.9 shows that the Site II motif is found in potential regulons with genes co-expressed only in *B. cinerea* and short day senescence or long and short day senescence. Similarly, the TELO-box motif is only found in potential regulons with genes co-expressed in long and short day senescence or high light and short day senescence. It can be hypothesised, therefore, that these particular stress responses may be regulated through these motifs and the predicted regulators capable of binding to promoters of genes containing these motifs.

Potential regulon 457, which is identified as working across short day and long day senescence conditions, is the only group of co-expressed genes identified as having both the Site II motif and the TELO-box motif strongly overrepresented within the promoters of the gene members. A strong topological association between the Site II motif and the TELO-box motif is observed in *A. thaliana*, where the TELO-box enhances the expression of genes found to have a Site II motif in their promoter (292). Both of these promoters are commonly found in the promoters of ribosomal proteins: out of 216 promoters of genes annotated as encoding a ribosomal protein of the 40S or 60S ribosomal subunits, 174 contained at least one TELO-box. In 153 of these cases the TELO-box was associated with one or several Site II motifs (293). Out of the 58 gene members of potential regulon 457, fifteen genes were identified by Trémousaygue *et al*. (293) to contain the TELO-box motif in their promoters (as shown in Table 4.12). It may come as no surprise to discover that the majority of the genes found in this particular regulon encode ribosomal proteins: out of the 58 members constituting this regulon, only 17 were not annotated as being involved with any ribosomal process (shown in Table 4.12). The remainder were largely annotated as 60S and 40S ribosomal proteins. This potential regulon was also significantly overepresented for the biological function of structural constituent of ribosomes, with a corrected p-value of $3.29^{-06}$ when performing a GO term analysis using the plugin BiNGO in Cytoscape (see Section 2.7.4).

Table 4.12: Gene members of potential regulon 457 are involved in ribosomal processes. Genes annotated (using TAIR9) as being involved in any ribosomal process are highlighted. Gene members identified by Trémousaygue *et al*. (293) to contain the TELO-box motif in their promoters are shown using an asterisk.

| ATG Identifier | Gene description |
|---|---|
| AT1G07070 | 60S ribosomal protein L35a (RPL35aA) |
| AT1G14410 | WHIRLY 1 (WHY1) |
| AT1G18440 | peptidyl-tRNA hydrolase family protein |
| AT1G26880 | 60S ribosomal protein L34 (RPL34A) |
| AT1G31660 | FUNCTIONS IN: molecular_function unknown |
| AT1G48830 | 40S ribosomal protein S7 (RPS7A) |
| AT1G72370 | P40* |
| AT1G77750 | 30S ribosomal protein S13, chloroplast, putative |
| AT1G77940 | 60S ribosomal protein L30 (RPL30B)* |
| AT2G02450 | Arabidopsis NAC domain containing protein 35 (ANAC035) |

| | |
|---|---|
| AT3G09630 | 60S ribosomal protein L4/L1 (RPL4A) |
| AT2G31610 | 40S ribosomal protein S3 (RPS3A)* |
| AT2G31725 | unknown protein |
| AT2G32060 | 40S ribosomal protein S12 (RPS12C) |
| AT2G36620 | ribosomal protein L24 (RPL24A) |
| AT2G38300 | DNA binding / transcription factor |
| AT2G19670 | PROTEIN ARGININE METHYLTRANSFERASE 1A (PRMT1A) |
| AT2G27710 | 60S acidic ribosomal protein P2 (RPP2B) |
| AT2G37270 | RIBOSOMAL PROTEIN 5B (ATRPS5B) |
| AT3G04840 | 40S ribosomal protein S3A (RPS3aA)* |
| AT3G05560 | 60S ribosomal protein L22-2 (RPL22B) |
| AT3G14390 | diaminopimelate decarboxylase, putative / DAP carboxylase, putative |
| AT3G21300 | RNA methyltransferase family protein |
| AT3G23940 | dehydratase family |
| AT3G23990 | HEAT SHOCK PROTEIN 60 (HSP60) |
| AT3G47370 | 40S ribosomal protein S20 (RPS20B) |
| AT3G51190 | structural constituent of ribosome |
| AT3G56340 | 40S ribosomal protein S26 (RPS26C) * |
| AT3G60245 | 60S ribosomal protein L37a (RPL37aC) |
| AT3G06680 | 60S ribosomal protein L29 (RPL29B) |
| AT3G07110 | 60S ribosomal protein L13A (RPL13aA) |
| AT3G16780 | 60S ribosomal protein L19 (RPL19B)* |
| AT3G25520 | A. THALIANA RIBOSOMAL PROTEIN L5 (ATL5) |
| AT3G28900 | 60S ribosomal protein L34 (RPL34C) |
| AT4G10480 | nascent polypeptide associated complex alpha chain protein, putative / alpha-NAC, putative |
| AT4G12600 | ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein |
| AT4G13170 | 60S ribosomal protein L13A (RPL13aC) |
| AT4G16141 | sequence-specific DNA binding / transcription factor/ zinc ion binding |
| AT4G17390 | 60S ribosomal protein L15 (RPL15B)* |
| AT4G25890 | 60S acidic ribosomal protein P3 (RPP3A) |

| | |
|---|---|
| AT4G31700 | RIBOSOMAL PROTEIN S6 (RPS6) |
| AT4G31710 | ATGLR2.4 |
| AT4G16720 | 60S ribosomal protein L15 (RPL15A)* |
| AT5G02870 | 60S ribosomal protein L4/L1 (RPL4D) |
| AT5G09510 | 40S ribosomal protein S15 (RPS15D) * |
| AT5G10920 | argininosuccinate lyase, putative / arginosuccinase, putative |
| AT5G16130 | 40S ribosomal protein S7 (RPS7C) |
| AT5G20720 | CHAPERONIN 20 (CPN20) |
| AT5G22440 | 60S ribosomal protein L10A (RPL10aC) |
| AT5G23535 | KOW domain-containing protein |
| AT5G47700 | 60S acidic ribosomal protein P1 (RPP1C)* |
| AT5G52650 | 40S ribosomal protein S10 (RPS10C) * |
| AT5G58420 | 40S ribosomal protein S4 (RPS4D)* |
| AT5G59850 | 40S ribosomal protein S15A (RPS15aF)* |
| AT5G60390 | elongation factor 1-alpha / EF-1-alpha* |
| AT5G60670 | 60S ribosomal protein L12 (RPL12C)* |
| AT5G61170 | 40S ribosomal protein S19 (RPS19C) |
| AT5G63050 | embryo defective 2759 (EMB2759) |

#### 4.3.1.2 Wigwams identifies potential regulons involved in novel multiple stress responses

Much literature is available on the crosstalk of responses to multiple stresses. However, certain multiple plant stress responses have very little literary evidence supporting the existence of crosstalk. Here, we discuss the findings of Wigwams, which has identified potential regulons involved in the novel responses to multiple plant stresses.

***B. cinerea* and drought stress response** Six potential regulons were discovered as having gene members significantly co-expressed in the *B. cinerea* and drought stress responses. There is little literary evidence supporting crosstalk between these two stresses, apart from single gene studies: *bos1* knockout plants exhibit reduced susceptibility to *B. cinerea* and decreased tolerance

to drought (189). Mittler described some cases where plants exhibit an increased tolerance to pathogens upon exposure to abiotic stress (208).

These six potential regulons are significantly overrepresented for the 'Plastid' GO term, except for one, which is significantly overrepresented for 'Translation'. Neither functions have been associated with the response to *B. cinerea* or drought stress before. The majority of the gene members in the potential regulon overrepresented for translation encode structural components of the ribosome, with the exception of one gene, which encodes the *ANAC035* transcription factor. *ANAC035* does not have any known literary evidence to support a role in the *B. cinerea* and drought responses, however, this novel finding presents an avenue of investigation.

The expression of the genes in all six potential regulons is down regulated in response to both *B. cinerea* and drought stress. Since these potential regulons are significantly enriched for genes linked to the plastid and translation, this suggests that these genes and functions are required in an unstressed leaf. However, during the response to *B. cinerea* and drought, these functions are down regulated, perhaps in favour of redirecting energy to pathways important in the stress response. Drought is known to affect chloroplasts via photosynthesis, by limiting the opening of the stomata, which is mediated by hormones, and also by general alterations of the leaf photochemistry (57). In contrast, *B. cinerea* promotes the generation of reactive oxygen species, which target the electron transport chain, thus affecting chloroplast function (201).

***B. cinerea* and high light stress response** Nineteen potential regulons were discovered as having gene members significantly co-expressed in the *B. cinerea* and high light stress responses. Crosstalk has been hypothesised between these two stresses via the production of reactive oxygen species (201, 13), however, this is the only link, with circumstantial evidence behind it.

The expression profiles for gene members in all potential regulons are down regulated in the *B. cinerea* response. In the high light response, gene members of ten of the potential regulons have down regulated expression profiles, and the other nine have upregulated expression profiles. There does not seem to be an obvious pattern to explain why groups of genes exhibit these differences. However, potential regulons identified as having genes enriched for 'Photosystem' or 'Photosynthesis' GO term functions all have down regulated gene expression profiles in both stress responses.

Interestingly, the two potential regulons with contradictory expression profiles (i.e down regulated expression profiles in response to *B. cinerea* and upregulated expression profiles in high light) are significantly overrepresented for the GO terms 'Mitochondria' and 'Respiratory chain complex I'. Since biotic stresses are known to cause disruptions in respiratory homeostasis via the induction of SA (10), it seems logical for the plant to down regulate genes which may function in respiration, in order to preserve the respiration process. However, respiration, by its own merit, produces reactive oxygen species in mitochondria, which ultimately cause damage to protein and DNA (13). Since high light stress causes the production of reactive oxygen species by its method of stress, the purpose of these genes having upregulated expression during the high light stress response may be as a direct downstream activation of the stress, in order to promote itself further.

**High light and *P. syringae* DC3000 stress response** Only two potential regulons were found to be co-expressed in the high light and *P. syringae* DC3000 stress responses. Lesion in *lsd1* mutant plants can be induced by both high light environments and infection by *Pseudomonas syringae* pathovar *maculicola* (Psm) (155). *LSD1* is involved in the signalling pathway that induces CuZnSOD proteins, a member of the superoxide dismutase protein family, upon perception of SA (157).

The expression profiles for gene members were down regulated in the high light response, but were down regulated, and were immediately upregulated after a few hours of infection by *P. syringae* DC3000. Only one potential regulon was found to be enriched for GO terms, specifically, 'Chloroplast thylakoid' function. The *P. syringae* effector HopI1 is targeted to the chloroplast, and causes the structure of the chloroplast thylakoid to change, which suppresses the accumulation of SA (137). This process could explain the down regulation and subsequent up regulation of these genes found to be co-expressed in both high light and *P. syringae* DC3000 infection: the effector HopI1 causes the down regulation of these genes, which are enriched for term 'chloroplast thylakoid', in *P. syringae* DC3000 infection, in order to promote the re-structuring of the thylakoid. However, as the expression of these genes subsequently becomes upregulated, it is plausible that this effect of HopI1 is suppressed by the plant in a bid to return the chloroplast thylakoid to its 'normal' state.

## 4.4 Discussion

In this chapter, the application of Wigwams, a novel tool that mines gene expression datasets in multiple conditions for evidence of potential co-regulation, has been applied to six high resolution time course datasets. These data, which have captured the gene expression changes in response to biotic, abiotic and developmental stresses, have provided the means to discover evidence of shared regulatory mechanisms working across multiple plant stresses. Here, improvements and possible limitations on these approaches will be discussed.

465 potential regulons were identified by Wigwams as co-expressed, and possibly co-regulated, across these datasets. We discovered that, as expected, there was significant crosstalk within the biotic, abiotic and developmental stresses. However, there was also significant evidence to suggest that crosstalk between biotic and abiotic stresses exist.

GO term analysis, using the BiNGO (191) plugin for Cytoscape (72) allowed the investigation of which of the 465 potential regulons had gene members enriched for similar functions, and consequently suggesting these genes were co-expressed. 219 potential regulons had gene members which were enriched for a specific GO term, whereas the remaining 246 potential regulons returned no GO terms. A possible reason for this may be due to the custom annotation file used for completing the GO term analysis: the file contained only genes found to be differentially expressed in two or more stresses. It is likely that some of these 246 potential regulons may have returned significant GO terms had the analysis been completed using the standard whole annotation. However, using the whole annotation as the reference set would provide a different universe of genes than was used in the initial Wigwams analysis to generate the potential regulons. This would skew the resulting *p*-values, as all the genes in the genome are considered when calculating the *p*-values.

An attractive feature of BiNGO is the ability of the researcher to choose the correction method for *p*-values: BiNGO supports not only Benjamini and Hochberg FDR (which was used in this analysis), but also Bonferroni Family Wise Error Rate. Whilst investigation of this different method of correction to provide significantly different *p*-values would be an avenue of exploration, it is known that Bonferroni correction can become conservative if more than 50 functional categories are involved (149). Since 59 of the potential regulons, out of a total of 465, were identified as having over 50 functional categories in GO term analysis using BiNGO, another form of correction

should be considered, for comparison. For example, Šidàk correction, which is often confused with Bonferroni correction due their similarity, does not experience the same conservativeness at which Bonferroni suffers (126).

To identify groups of co-expressed genes as being co-regulated, the promoters of the gene members were analysed for the presence of a conserved motif. 89 of the 465 (19%) potential regulons had known plant motifs significantly overrepresented within the promoters of the gene members. Motifs may have been present in the promoters of gene members of the remaining potential regulons, which have not been formally identified as a potential binding site. To address this issue, the 500 bp sequence upstream of the transcriptional start site of potential regulon members could be submitted to a *de novo* motif search algorithm, such as MEME (18). This would allow the discovery of overrepresented motifs that were not included in the hypergeometric test used for the initial motif analysis.

# Chapter 5

# Generation and confirmation of a predicted gene regulatory network operating in multiple stress responses

Literature is available on the co-expression of genes and whether this co-expression is maintained during the response to single stresses (101, 53, 301), however, literature on whether the co-regulation of genes is maintained during specific stress responses is more difficult to come by, due to the lack of context-specific (i.e. stress-specific) nature of transcriptional regulation. However, understanding the transcription regulation of genes which have been identified as significant to the stress response is important, due to the very nature of this sparse knowledge. By using modelling techniques to predict a gene regulatory network that is working over multiple stress responses, these predictions can be used as a basis of experimental validation in non-stress conditions. If these predictions are substantiated in non-stress conditions, the methods of transcriptional regulation can be transferred to infer the dynamics of regulation during the stress response.

During this chapter, a model of gene regulation will be developed using evidence of potential regulons from Wigwams. The aim is to identify a gene regulatory network that is involved in the responses to multiple plant stresses. Therefore, to fulfil this aim, all transcriptionally regulated components of the network need to be identified, using evidence from Wigwams. Transcriptional regulation of the genes in potential regulons identified as co-expressed across subsets of stresses

by Wigwams will be inferred using modelling techniques. Literature on current knowledge of any regulation within this predicted network will be collected to validate components of the network. Subsequently, validation of novel predicted regulation within this predicted network will be carried out using experimental techniques in non-stress conditions, which will aim to confirm the regulation of genes in potential shared regulons. This network will then be investigated using bioinformatical methods to demonstrate this model is one possible common gene regulatory network working in response to multiple plant stresses.

## 5.1 Experimental techniques for the confirmation or elucidation of transcriptional regulation in GRNs

In Chapter 3 Section 3.1, various mechanisms of regulation, including transcriptional regulation, were discussed. There are several techniques available for the elucidation, or confirmation of predicted, transcriptional regulation in gene regulatory networks.

### 5.1.1 High throughput Yeast-1-Hybrid assay to detect transcription factor binding to promoters of interest

Y1H, as discussed in Section 2.4, is primarily used to investigate protein-DNA interactions, rather than protein-protein interactions, which are found using Y2H. Y1H allows the investigation of which transcription factors bind to a particular section of a promoter. Additionally, Y1H is capable of discovering regulatory elements within the promoter regions of genes that may have been predicted bioinformatically using motif analysis, for example. In the last decade, Y1H has been made compatible with Gateway technology (81), which reduces the need for a transcription factor library biased towards highly expressed genes, seen in libraries constructed using cDNA from total RNA. Also, by only using a library containing transcription factor cDNA clones reduces the background noise experienced when using a library consisting of transcription factors and non-transcription factors. Using a cDNA library consisting solely of transcription factors also reduces the cost of detecting protein-DNA interactions (305). Cloned libraries using the Gateway technology can be pooled so that multiple cDNA clones occupy one well in order make the process high-throughput (307).

Y1H screens have been used extensively in the study of protein binding in *A. thaliana*: by using Y1H, *CHE*, also known as *TCP11* (At5g08330), was discovered to directly interact with *TOC1*, adding a new component to the *A. thaliana* circadian clock model (236). Liu *et al.* created custom cDNA libraries using drought-stressed, cold-stressed and unstressed *A. thaliana* leaves for use in Y1H screening. Using the Y1H screening method and these libraries created from drought-stressed, cold-stressed plants, and unstressed plants, two cDNAs, *DREB1A* and *DREB2A* were identified as directly interacting with the DRE sequence involved in dehydration and low temperature response gene expression (179). The DRE (dehydration-responsive element) element is a conserved sequence (TACCGACAT), which was found to be essential for the regulation of dehydration responsive gene expression (179). This study, along with others, has made Y1H a viable choice when studying transcriptional regulation in terms of stress responses. However, these studies did not use the matrix Y1H method mentioned above, and also required extensive analysis of the promoter regions before Y1H could be carried out to identify which regions the transcription factor could bind to.

### 5.1.2 *In vivo* detection of transcription factor binding using Chromatin Immunoprecipitation (ChIP)

The chromatin immunoprecipitation assay is used to investigate the binding of transcription factors to DNA *in vivo* (316). Cells are treated with formaldehyde to cross link proteins to DNA, and then an antibody against a transcription factor of choice is used to immunoprecipitate the chromatin fragments (141). By subsequently using PCR, the immunoprecipitated DNA can be amplified (141), hybridised to a microarray (ChIP-chip) (48), or sequenced (ChIP-Seq) (140) to reveal the sequence occupied by the transcription factor.

However, ChIP-based methods require high-quality antibody, and since the *A. thaliana* genome contains over 2000 transcription factors (113), ChIP-based methods for the elucidation of transcription regulation on the genome-scale would be expensive and far-fetched. However, ChIP analysis using antibodies to single transcription factors have been reported in *A. thaliana*: Zheng *et al.* used ChIP-chip to map all binding sites of *AGL15*, identifying many downstream targets, such as *WRKY18* and *MYB4* (342). To investigate the effect of added salicylic acid to the binding of *TGA2*, Thibaud-Nissen *et al.* used ChIP-chip and ChIP-PCR to confirm downstream targets of *TGA2*.

Many salicylic acid induced genes were found to be significantly overrepresented among genes downstream of *TGA2* binding sites, suggesting involvement of *TGA2* in the stress response (282).

The maturation of ChIP-based technology has yielded a methodology with numerous advantages: high base-pair resolution of transcription factor binding sites can be achieved using ChIP-Seq (27); also ChIP-Seq generates data with lower quantities of noise seen in other methods (226).

However, there are a number of disadvantages to ChIP-based methods: firstly, antibodies are expensive to produce, and this cost is passed on to researchers (327). For high-resolution profiling, a customised microarray would yield as much biological data for a substantially lower cost. The amount of starting material needed is also considerably larger than material needed for array and Y1H methods, and requires significant amounts of amplification of ChIP-enriched DNA (134).

Therefore, due to the limitations of ChIP methods, Y1H will be used in this chapter: Y1H is easily an easily automated method (310), and with the development of the matrix-Y1H method (305), allows for high-throughput analysis of many promoters against many transcription factors.

## 5.2 Integrating potential regulons, gene expression data and predicted gene regulatory networks

Section 5.1 described two experimental techniques used to elucidate and confirm transcriptional regulation. However, these methods are time-consuming, expensive, and often do not take the dynamics of the whole system into account, potentially missing crucial interactions. Indeed, investigating the *A. thaliana* system as a whole would prove to be too complex using modern day experimental techniques. For example, the identification of the regulation of microRNA (miRNAs) transcription alone in *A. thaliana* is experimentally challenging, due to tissue- and time-specific expression, and the low abundance of some miRNAs (313). Despite intensive laboratory research to determine miRNA function, little is known about miRNA regulation. However, by using computational methods, miRNAs and their targets mRNAs can be predicted, and subsequently confirmed using experimental techniques (313), limiting the amount of laboratory-based work, thus reducing time and effort taken, needed to find such results.

Since cDNA microarrays are powerful techniques for profiling genome-wide mRNA expression, large amounts of gene expression data can be generated relatively quickly (62). However, as a

source to study transcriptional regulation, gene expression data is not sufficient: gene expression is controlled by a number of factors, such as transcription factors, but also post transcriptional, translational regulation and protein degradation (33) (as described in Section 3.1). These variables can be modelled as 'hidden' state variables, as their values are not known (or we simply do not have the measurements for these variables) (214).

### 5.2.1 Predicting gene regulatory networks using gene expression data

A number of models generated using time series gene expression data in response to single stresses are described below. These models use only transcription factors that are differentially expressed in response to a stress: it is more likely that these transcription factors are involved in transcriptional regulation in this stress, being as they have altered expression in response to the stress, than transcription factors not differentially expressed in response to a stress.

#### 5.2.1.1 Network inference

Variational Bayesian State-Space Modelling (VBSSM) (30) is a quantitative modelling approach which outperforms other network inference tools that do not incorporate hidden states when analysing time-series data (231). However, this modelling approach is limited by the amount of gene expression data it can model: the greater the number of time points or genes, the more computationally expensive the modelling process becomes. Therefore, the question of how to select genes for modelling is raised.

Christopher Penfold (University of Warwick, unpublished) developed a Metropolis-Hastings version of the VBSSM modelling software (30): this Metropolis-Hastings version probabilistically selects subsets of genes to model from a gene list, as shown in Figure 5.1. *N* genes are randomly swapped from the gene list, and one VBSSM model (30) is generated for each set of genes. The marginal likelihood is used to determine if the updated set of genes is better than the previous random selection of genes. This model can be accepted or rejected based on this marginal likelihood. Another *N* genes are swapped randomly from the gene list, and this process continues.

A series of networks, where each gene in a dataset is kept systematically in turn in order to generate a model, was produced for each of the PRESTA time series datasets (described in full below). Multiple networks were then combined into one single stress-specific network, which will now be

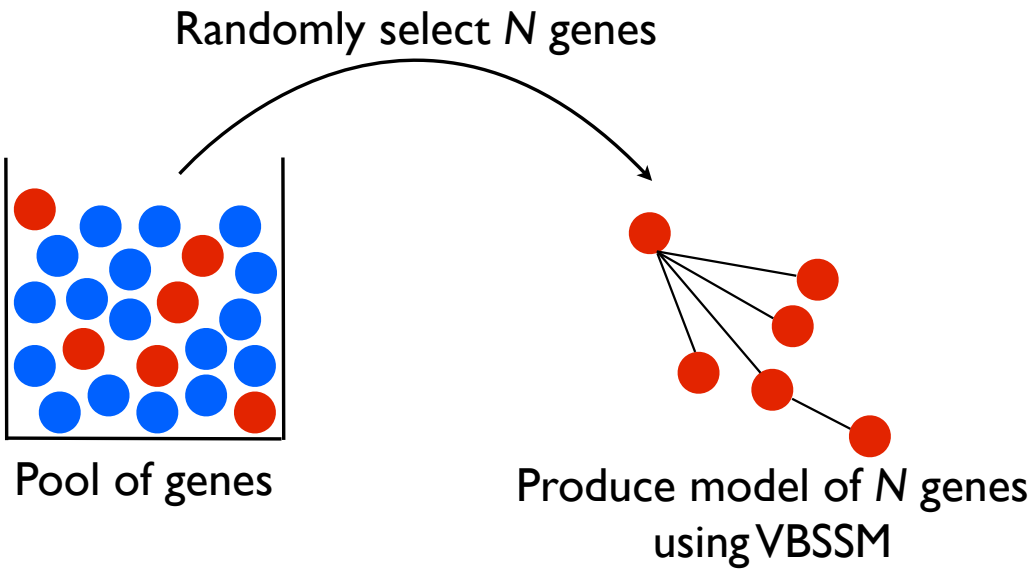



Figure 5.1: Describing the Metropolis-Hastings extension of the VBSSM modelling software (30). A pool of genes (for example, all differentially expressed genes in response to a singular stress) is represented here as blue circles. *N* genes (red circles) are randomly selected from this pool of genes, and VBSSM (30) is used to generate a predicted model of gene regulation for these *N* genes (here represented as the red circles joined via black lines, which represent edges, or inferred regulation). This process is repeated for another set of *N* randomly selected genes. The marginal likelihood is used to determine if the updated set of genes is better than the previous random selection of genes. This model can be accepted or rejected based on this marginal likelihood.

termed as a 'consensus network', by considering the pairwise intersection of genes. The z-score, which indicates how many standard deviations an observation is above or below the mean, was set to 1.65 for each network. A z-score of 1.65 equates to a threshold with 95% confidence in the interactions between nodes detected. The modelling for each stress-specific dataset was carried out for approximately 2000 iterations the ensure that the marginal likelihood converged, and a final consensus model was chosen. Table 5.1 details the number of nodes (i.e. the number of genes encoding transcription factors) and the number of edges (i.e. the number of interactions between nodes) for each consensus model. For the purpose of this integrative analysis with Wigwams potential regulons, consensus networks were filtered on genes which appeared in two or more stress-specific models.

Table 5.1: The number of nodes (genes encoding transcription factors) edges (interactions between genes) included in each stress-specific consensus model

| | *B. cinerea* | Drought | High light | *P. syringae* pv. tomato DC3000 | Senescence |
|---|---|---|---|---|---|
| Nodes | 572 | 165 | 538 | 634 | 501 |
| Edges | 10565 | 1907 | 9511 | 18572 | 10208 |

The PRESTA consensus networks were only modelled on transcription factors differentially expressed in a particular stress, in order to reduce the computational time taken to generate the models. Therefore, although these networks predict transcriptional regulation, they do not consider non-transcription factors in the models. However, Wigwams do include both transcription factors and non-transcription factors, making these two approaches complementary. By observing the number of nodes in each consensus model in Table 5.1, it can also be seen that each network model is relatively large. Therefore, by integrating the information from Wigwams, one can limit which interactions predicted by the modelling to validate experimentally.

**PRESTA time-series datasets** The PRESTA group used microarray analysis to obtain high resolution time-course profiles of changes in gene expression during response to stress in *A. thaliana* leaves. These datasets were used to identify differentially expressed genes in response to each individual stress. Six stress responses were investigated in this way: responses to biotic stress caused by the pathogens *B. cinerea* (Windram *et al.* (322)) and *P. syringae* pathovar *tomato* DC3000 (unpublished); responses to abiotic stresses drought (unpublished) and high light (unpublished); and developmental stresses long (43) and short day senescence (unpublished). A summary of the time points taken for each experiment, along with the number of replicates, is summarised in Table 2.15.

For the purposes of analysis using Wigwams, the MAANOVA (326) output that combined biological replicates was used for each treated dataset, at each timepoint.

### 5.2.2 Combining network inference findings with experimental results confirms co-regulation of Wigwams potential regulons

As stated in Section 4.3, 5% of gene members of the total 465 potential regulons were transcription factors. As 5% of the *A. thaliana* genome is dedicated to encoding transcription factors (243), this highlights that transcriptional regulation plays an important in the adaptation and response to plant

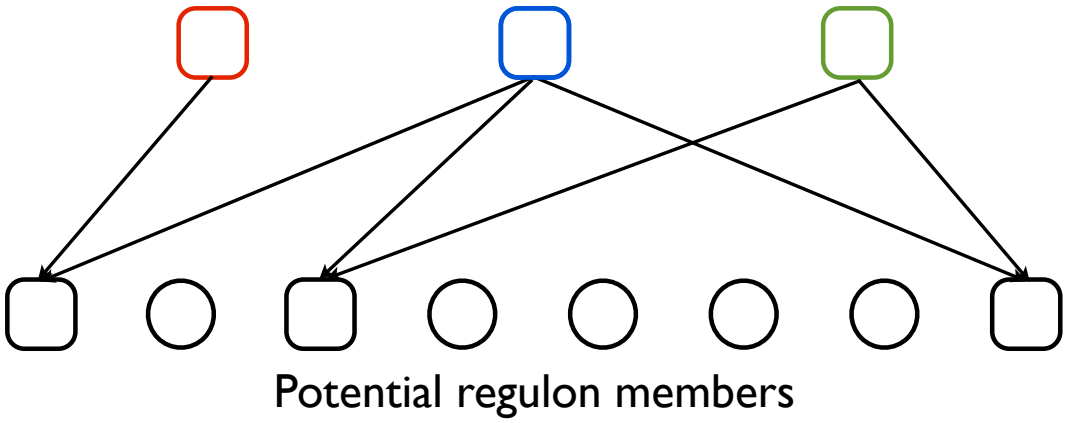



Figure 5.2: Illustrating common regulation of transcription factor members of potential regulons. Nodes outlined in black are potential regulon members, whereas nodes in colour are genes in consensus models. Circular nodes are non-transcription factors, whereas square nodes are transcription factors. The blue node is a common regulator of the transcription factor members of the potential regulon, whereas the red and green nodes are not common regulators: they only regulate one or two transcription factor members of the potential regulon, respectively. Therefore, only the blue node, which represents a transcription factor also, is a common regulator of this potential regulon, and will be considered for further analysis.

stress, and that Wigwams is capable of detecting this response proportionally. By identifying the common regulatory transcription factor of the gene members of a potential regulon, the network architecture for a shared response to plant stress can be deciphered.

The construction of the single stress response consensus networks, as described in Section 5.2.1.1, were limited by the amount of gene expression data as an input. It is hypothesised that transcription factors form a large component of the stress response (43, 63, 161, 265). Therefore, only transcription factors differentially expressed in response to a particular stress were considered for modelling.

#### 5.2.2.1 The integration of predicted transcriptional regulation from modelling and Wigwams potential regulons

The output from applying the VBSSM modelling algorithm to singular time series datasets consists of a Cytoscape file, in order to visualise the nodes and edges, and a text file, which consists of two columns: the left-hand column contains the ATG identifier of a transcription factor, and the right-hand column contains the ATG identifier of the target gene (also a transcription factor) of the aforementioned transcription factor. The gene members of Wigwams potential regulons are filtered for transcription factors only. Identifying predicted regulators of transcription factor members of potential regulons consists of a simple use of the 'vlookup' function in Excel (325).

Since only transcription factors were used for generating the consensus models, only regulation of transcription factor members of potential regulons can be inferred from the modelling. In order to investigate whether co-expressed gene members of potential regulons were indeed co-regulated, the predicted regulators of transcription factor members were considered. Figure 5.2 shows that if any regulator was predicted to regulate all transcription factor members of a potential regulon, this was a 'common' regulator. This analysis was completed for all 186 potential regulons that contained transcription factors, out of an original 465 potential regulons detected, as mentioned previously in Section 4.3. 72 potential regulons were found to have a common transcription factor regulator, identified from the consensus models. Table 5.2 shows the common regulators of transcription factor members of all potential regulons, as inferred from consensus models. Potential regulons with no common regulators of transcription factor members, or with no transcription factor members, are not included.

Table 5.2: Predicted upstream regulators of transcription factor members of potential regulons. Abbreviations for which conditions potential regulons are co-expressed are as follows: B - *B. cinerea*; D - Drought; H - High light; P - *P. syringae* DC3000; L - Long day senescence; S - Short day senescence. Regulation of transcription factor members of potential regulons is also within the same conditions as stated.

| Potential regulon number | Conditions | Regulon TF member | Common predicted regulator(s) |
|---|---|---|---|
| 5 | P, L | *MYB2*<br>*ERF1*<br>*ANAC055* | *WRKY45*, *RHL41*, *DREB2A* |
| 6 | B, S | *At2g28200* | *ERF1*, *WRKY45* |

| | | | |
|---|---|---|---|
| | | *At1g62370* | |
| 8 | B, L | *DREB19*<br>*At1g62370* | *WRKY45* |
| 15 | B, L | *STOP1*<br>*At5g55970*<br>*ANAC092*<br>*ANAC083* | *ANAC092* |
| 17 | B, P | *MYB2*<br>*ANAC055* | *AXR3*, *WRKY45*, *ANAC047*, *ANAC046*, *SIGA*, *RHL41*, *ANAC092*, *MYB2*, *Rap2.6L*, *PDF2*, *MYB15*, *DREB2A*, *HY5*, *bZIP1*, *bZIP25*, *At4g32800*, *At3g46080*, *At4g17810*, *At3g16350* |
| 18 | P, L | *WRKY6*<br>*TCP13* | *ANAC092* |
| 35 | B, P | *Rap2.6L*<br>*ANAC046* | *PMZ*, *AGL18*, *bZIP1*, *DREB2A*, *ERF1*, *ANAC046*, *ANAC052*, *PDF2*, *Rap2.6L*, *SIGA*, *WRKY45*, *WRKY75*, *At4g17810* |
| 36 | L, S | *At4g13110*<br>*At1g21000* | *AXR3*, *bZIP25*, *At3g46080* |
| 43 | P, L | *ANAC046*<br>*ATHB-7*<br>*TGA1* | *Rap2.6L* |
| 44 | B, L | *ANAC029*<br>*At3g51960* | *LIL3:1*, *ANAC046*, *At2g28200* |
| 64 | B, P | *WRKY28*<br>*ANAC072* | *AXR3*, *HY5*, *MYB112*, *MYB7*, *PDF2*, *WRKY45* |
| 87 | B, P | *ANAC029*<br>*At-HSFB2A*<br>*AGD12*<br>*ANAC055* | *RHL41*, *WRKY45* |

| | | | |
|---|---|---|---|
| 101 | B, L | *At5g07580*<br>*At1g25440* | *SIGA* |
| 102 | B, P | *IAA11*<br>*LIL3:1* | *AGL18*, *ANAC046* |
| 117 | L, S | *WRKY4*<br>*HDA3* | *ANAC092* |
| 127 | P, L, S | *WRKY45*<br>*ANAC047* | *At1g10586*, *At3g46080*, *bZIP25*, *HB-7*, *MYB112*, *ANAC047*, *APRR5*, *Rap2.6L* |
| 132 | B, L | *LIL3:1*<br>*ATML1* | *HAM4*, *bZIP1*, *DREB2A*, *ANAC092*, *RHL41*, *SIGA* |
| 134 | B, L | *MYB3*<br>*At1g21000* | *AXR3* |
| 147 | L, S | *WRKY45*<br>*MYB2*<br>*At2g28200*<br>*AGL18*<br>*ANAC055*<br>*At1g62370* | *WRKY45* |
| 178 | L, S | *MYB15*<br>*BLH1* | *WRKY75* |
| 198 | L, S | *At2g38300*<br>*HDA3* | *ANAC092* |
| 208 | P, L | *PMZ*<br>*WRKY6*<br>*TCP13* | *ANAC092* |
| 216 | B, L | *WHY3*<br>*ATAUX2-11* | *DREB2A* |
| 228 | B, L | *Rap2.6L*<br>*ANAC003* | *PMZ*, *PDF2*, *Rap2.6L*, *SIGA*, *WRKY45*, *WRKY75* |
| 255 | L, S | *AtIDD7* | *SCL3* |

| | | | |
|---|---|---|---|
| | | *ATHB-7*<br>*MYB7* | |
| 271 | B, P | *LIIL3:1*<br>*SIGA* | *PMZ*, *bZIP1*, *At4g17810*, *JAZ8 (TIFY5A)*, *ANAC092*, *ANAC046*, *SIGA* |
| 276 | B, L | *ANAC046*<br>*ANAC002*<br>*At3g52800* | *ANAC092*, *PDF2*, *SCL3*, *WRKY45* |
| 285 | L, S | *HDA3*<br>*WRKY75* | *ANAC092*, *APRR5*, *SCL3* |
| 296 | B, P | *DREB19*<br>*ANAC003* | *PDF2* |
| 304 | B, S | *ANAC002*<br>*At3g52800* | *ANAC092*, *PDF2*, *SCL3*, *WRKY45* |
| 309 | L, S | *MYB2*<br>*AZF2*<br>*ANAC055* | *WRKY45*, *PDF2* |
| 310 | B, L | *MYB2*<br>*ANAC055* | *At4g17810*, *At3g16350*, *At3g46080*, *At4g32800*, *AXR3*, *bZIP25*, *bZIP1*, *DREB2A*, *HY5*, *MYB15*, *MYB2*, *ANAC092*, *ANAC046*, *ANAC047*, *PDF2*, *Rap2.6L*, *RHL41*, *SIGA*, *WRKY45* |
| 312 | B, P | *ERF1*<br>*WRKY75* | *PMZ*, *ANAC092*, *ANAC046*, *SIGA* |
| 336 | B, P | *WRKY28*<br>*ANAC019*<br>*At4g13110* | *At1g73870*, *AXR3*, *HY5*, *HDA3*, *PDF2* |
| 345 | L, S | *HDA3*<br>*At2g38300* | *ANAC092* |
| 378 | B, S | *LIL3:1* | *HAM4*, *bZIP1*, *DREB2A*, *ANAC092*, *SIGA*, *RHL41* |

| | | | |
|---|---|---|---|
| | | *ATML1* | |
| 401 | L, S | *TRFL8*<br>*At4g16141* | *Rap2.6L* |
| 418 | L, S | *At3g04930*<br>*ANAC019*<br>*At-HSFB2A* | *ANAC092*, *DREB2A* |
| 446 | B, P | *STOP1*<br>*WRKY28*<br>*JAZ1 (TIFY10A)*<br>*At2g42350*<br>*ANAC019* | *WRKY45* |
| 450 | B, L | *ANAC046*<br>*MYB3*<br>*bZIP24* | *Rap2.6L*, *At2g28200* |
| 451 | L, S | *ANAC046*<br>*At1g21000*<br>*At4g13110* | *bZIP25* |

71 unique nodes in the consensus models were identified as common regulators of transcription factors in potential regulons. The transcription factor *ANAC092* was predicted to regulate the most gene members of potential regulons (43 genes). The number of target genes predicted to the regulated by each transcription factor is shown in Table 5.3.

As can be seen from Table 5.2, the stresses drought and high light are not represented. Although potential regulons found to have gene members co-expressed in drought and high light conditions did have transcription factor members, they did not share a common regulatory protein, which would infer they are not co-regulated, or were not predicted to have any upstream regulators in the consensus models. An obvious implication of the lack of representation of drought and high light is that no candidate genes (via the detection of potential regulons) have been identified for further experimental analysis. Whilst drought and high light have not been seen to be co-regulated with the

other stresses tested here, this does not mean they are not co-regulated: this analysis relies on the assumption that the network inference has identified all regulatory behaviour between transcription factors and their targets. Therefore, if the network inference method is subject to false-negatives, then it is possible that predicted regulators of potential regulons working in drought and high light stress will not have been identified.

Table 5.3: The frequency of occurrences a predicted regulated is seen to interact with a target gene

| Predicted regulator | Occurrences of regulation of targets |
|---|---|
| ANAC092 | 43 |
| PDF2 | 40 |
| WRKY45 | 38 |
| ANAC046 | 36 |
| AXR3 | 31 |
| SIGA | 31 |
| bZIP1 | 30 |
| DREB2A | 29 |
| Rap2.6L | 28 |
| bZIP25 | 22 |
| RHL41 | 22 |
| ANAC047 | 17 |
| HY5 | 16 |
| PMZ | 15 |
| SCL3 | 15 |
| MYB112 | 14 |
| AGL18 | 13 |
| NF-YB4 | 12 |
| JAZ10 | 11 |
| LIL3:1 | 11 |
| MYB15 | 9 |
| MYB2 | 8 |
| TBP1 | 7 |

| | |
|---|---|
| ANAC052 | 6 |
| APRR5 | 6 |
| ATHB7 | 6 |
| WRKY66 | 6 |
| WRKY75 | 6 |
| ERF1 | 5 |
| ANAC078 | 4 |
| HDA3 | 4 |
| HSFB2A | 4 |
| PTF1 | 4 |
| SIG5 | 4 |
| STZ | 4 |
| AGL24 | 3 |
| ANAC091 | 3 |
| ARF5 | 3 |
| ATHB16 | 3 |
| HAM4 | 3 |
| TCP14 | 3 |
| TCP3 | 3 |
| TCP4 | 3 |
| TCP8 | 3 |
| WHY1 | 3 |
| ANAC070 | 2 |
| bHLH093 | 2 |
| JAZ8 | 2 |
| MYB7 | 2 |
| SCL13 | 2 |
| LIIL3:1 | 1 |
| RGL1 | 1 |
| TBP1 | 1 |

| | |
|---|---|
| WRKY28 | 1 |

Figure 5.3 illustrates the hypothesis of a transcription factor, predicted to regulate other transcription factors, which are members of a potential regulon, also regulating the non-transcription factor members of the same potential regulon. Publicly available microarray data generated on plants with the predicted regulatory transcription factor either knocked out or over expressed will identify downstream targets due to their altered expression. If the transcription factor members of the potential regulon are seen to have altered expression in the microarray data for the mutant predicted regulator, it is likely the edges of the model (i.e. the transcriptional regulation) is correct. Furthermore, if the non-transcription factor members of the potential regulons can be also be seen to have altered expression, then it is likely the predicted regulatory transcription factor is a common regulator of all genes in that particular potential regulon. Therefore, the gene members of this potential regulon are said to be co-regulated.

#### 5.2.2.2 Summary of combining network inference with experimental results to confirm upstream regulation of Wigwams potential regulons

Seventy-two potential regulons have transcription factor gene members. Network models constructed to predict the transcriptional regulation occurring in each stress were used to identify common regulators of transcription factor members of potential regulons. 71 unique transcription factors were identified in the consensus models as common regulators of transcription factor members of potential regulons. Validation of predicted regulation can now be attempted.

## 5.3 Integrating experimental results from gene expression data with network inference models confirms co-regulation of co-expressed genes identified by Wigwams

The edges of a network models, which are the interaction between two nodes (in this case, genes), can be confirmed via experimental data, such as microarray gene expression data, Y1H and ChIP. By using publicly available data, these interactions shown in Table 5.2 can be confirmed. The

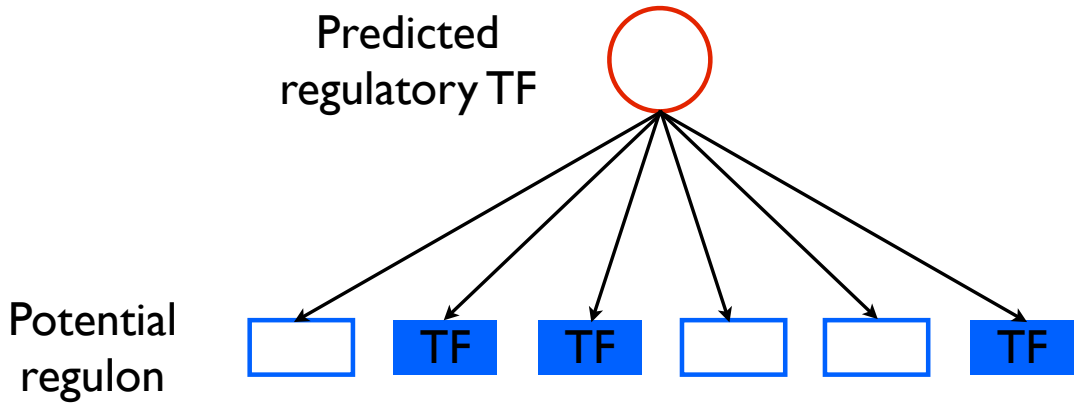



Figure 5.3: Hypothesising that a transcription factor predicted in the consensus model to regulate the transcription factor members of a potential regulon can also regulate the non-transcription factor members of the potential regulon. The regulatory transcription factor (in red) is predicted to regulate the transcription factor members (solid blue box) of a potential regulon. Does the predicted regulatory transcription factor also regulate the non-transcription factor members (blue outline) of the potential regulon? By using microarray and Y1H data, the edges (i.e. arrows), which represent transcription regulation, can be confirmed.

data may also confirm the regulation of the non-transcription factor members of the potential regulons, as hypothesised in Figure 5.3, which were not included in the consensus models. Data may not be available to confirm these interactions in the specific stress, or stresses, they have been predicted in, however. By also searching for plant promoter motifs enriched in gene members (both transcription factors and non-transcription factors) of potential regulons (see Methods 2.7.5.1), potential binding sites for predicted regulators can be identified. If the motif does not match the predicted regulator, this data also provides means for identifying a potential common regulator from known examples in literature.

Hub genes (genes with high connectivity (185)) are likely to be more important for stress tolerance than genes exhibiting less connectivity (218). The top three largest hubs, found to regulate a large proportion of transcription factor members of potential regulons, as shown in Table 5.2 were

considered for further analysis. These three hub genes, *WRKY45*, *ANAC092*, and *Rap2.6L*, for which microarray gene expression data was available for, were predicted to regulate 38, 43 and 28 unique transcription factors found in potential regulons respectively.

### 5.3.1 Confirming *WRKY45* as a common regulator of potential regulons

*WRKY45* was predicted to be a common regulator of the transcription factors of 38 potential regulons co-expressed in combinations of the stresses *B. cinerea*, *P. syringae* pv. tomato DC3000, long day senescence and short day senescence. It is hypothesised that if *WRKY45* does indeed regulate the transcription factor members of these potential regulons, then *WRKY45* is also capable of co-regulating non-transcription factor members of these potential regulons. By integrating data from a microarray experiment carried out on *A. thaliana* plant over expressing *WRKY45*, these potential regulons may be confirmed as direct or indirect targets of *WRKY45*. These potential regulons are also more likely to be direct targets of *WRKY45* if gene members are enriched for the presence of a W-Box motif in their promoters.

Table 5.4: Potential regulons with the highest percentage of predicted *WRKY45* targets confirmed in *35S:WRKY45* microarray. Abbreviations for which conditions potential regulons are co-expressed are as follows: B - *B. cinerea*; P - *P. syringae* DC3000; L - Long day senescence; S - Short day senescence.

| Potential regulon number | Conditions | Percentage of gene members confirmed by microarray experiments |
|---|---|---|
| 275 | B, L | 33% (10/30) |
| 446 | B, P | 24% (9/38) |
| 35 | B, P | 19% (6/32) |
| 147 | L, S | 13% (6/46) |
| 304 | B, S | 13% (2/16) |

Table 5.4 shows the percentage of gene members of potential regulons confirmed as regulated by *WRKY45* using microarray data (unpublished). The microarray experiments were completed by Thomas Vigrass (University of Warwick), to find genes differentially expressed in a *35S:WRKY45* line compared to wild type. The microarray was carried out on six week old *A. thaliana* leaves, using CATMA v4 arrays (8). The expression data was analysed by Dr. Katherine Denby (University of Warwick) using LimmaGUI (318). Using Benjamini and Hochberg FDR to correct $p$-values, a cutoff of $p < 0.05$ was used to acquire a list of 1332 differentially expressed genes in response to *35S:WRKY45* compared to wild type. Whilst this is a reasonable number of differentially expressed genes (around 5% of all *A. thaliana* genes), using FC-ranking as well as a non-stringent

*p*-value cutoff would generate a more reproducible list of differentially expressed genes (259).

*WRKY45* was predicted to regulate 38 potential regulons. However, from Table 5.4, it can be seen that only five of these potential regulons were confirmed, using the *35S:WRKY45* microarray results, as having gene members under the regulation of *WRKY45*. Nine of the 38 genes in potential regulon 446 were confirmed as being regulated by *WRKY45* using the *35S:WRKY45* microarray results, after being predicted to have this transcription factor as a common regulator in the consensus network models.

The W-Box (PSSM ID S-000390 (337)) was found to be significantly overrepresented in the promoters of gene members in potential regulon 446, with a *p*-value of $7.955e^{-04}$. However, only one of the nine genes was found to contain a W-Box in its 500 bp promoter sequence upstream of the transcriptional start site, suggesting that *WRKY45* is an indirect regulator of this potential regulon. The remaining potential regulons shown in Table 5.4 were overrepresented for either the G-Box motif (PSSM ID M00942 (202)) or the ABF motif (PSSM ID M00442 (112)), neither of which are preferred binding sites of the predicted regulator of these potential regulons, *WRKY45*. This suggests that *WRKY45* is an indirect regulator of these potential regulons, and the direct regulator was unidentified by modelling. Due to the way in which this particular modelling approach works, direct and indirect regulation cannot be distinguished.

From Figure 5.4 it can be assumed that given the slight time delay of expression of gene members of potential regulon 446, compared to the expression profile of *WRKY45*, these genes are indeed regulated by *WRKY45* in response to *B. cinerea*. However, the expression profiles of the gene members of potential regulon 446 in response to *P. syringae* DC3000 infection is more complex: the initial peak in the expression of the potential regulon gene members may not be differentially expression due to infection by *P. syringae* DC3000, but rather a wounding response: infection by *P. syringae* primarily triggers a wound response, and it is difficult to tease apart wounding and pathogen infection (228, 61). Therefore, it can be seen that *WRKY45* drives the decrease in expression of the potential regulon members, suggesting *WRKY45* is a negative regulator of these genes in response to *P. syringae* DC3000 infection.

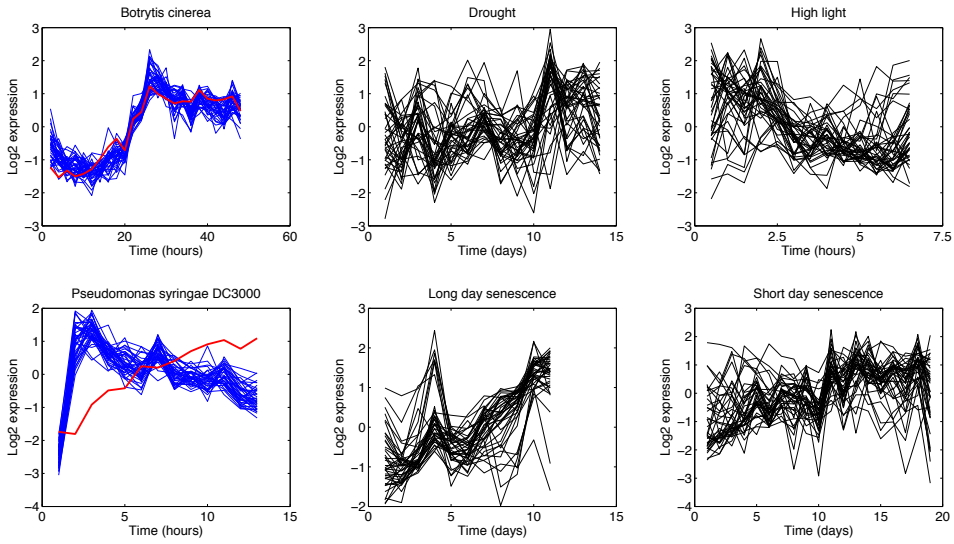



Figure 5.4: Expression profiles of gene members of potential regulon 446 significantly co-expressed in *B. cinerea* and *P. syringae* DC3000 (in blue) and predicted regulator *WRKY45* (in red). Nine of the 38 members of potential regulon 446 were found to have altered expression in response to *35S:WRKY45* compared to wild type.

Table 5.5: Potential regulons with the highest percentage of predicted *ANAC092* targets confirmed in *anac092* and *Est:ANAC092* microarray. Abbreviations for which conditions potential regulons are co-expressed are as follows: B - *B. cinerea*; P - *P. syringae* DC3000; L - Long day senescence; S - Short day senescence.

| Potential regulon number | Conditions | Percentage of gene members confirmed by microarray experiments |
|---|---|---|
| 15 | B, S | 38% (5/13) |
| 310 | B, S | 27% (8/30) |
| 18 | P, L | 8% (3/36) |
| 208 | P, L | 7% (2/27) |
| 304 | B, S | 6% (1/16) |
| 202 | L, S | 6% (1/18) |
| 17 | B, P | 4% (1/25) |
| 271 | B, P | 3% (1/39) |

### 5.3.2 Confirming *ANAC092* as a common regulator of potential regulons

Table 5.5 shows the percentage of gene members of potential regulons confirmed by two independent, publicly available, microarray datasets. An estradiol-inducible *Est:ANAC092* seedling overexpressor line was analysed five hours post-induction, compared to Col-0 wild type using Affymetrix ATH1 microarrays (22). Balazadeh *et al*. (22) analysed the expression data using LimmaGUI (318) to identify 170 differentially expressed genes. Balazadeh *et al*. (22) also performed expression analysis on leaf 11 of *anac092* T-DNA mutant (SALK_090154) knockout line,

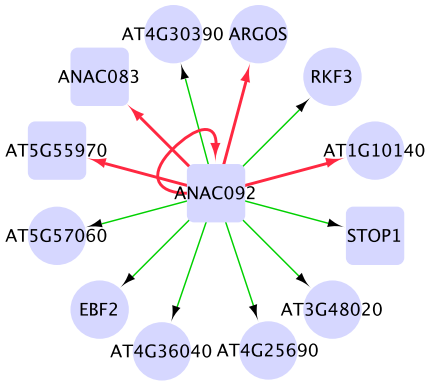
compared to Col-0 wild type, 38 days after sowing. These sample were analysed in an identical manner to the estradiol-inducible *Est:ANAC092* overexpressor line expression analysis. Using both overexpressor and knockout lines enables the researcher to gain a true understanding of the role of *ANAC092*: because of gene redundancy, is a gene is knocked out using T-DNA insertion, for example, the effect of this may be masked by the redundancy in gene function of another gene. Using overexpressor lines overcomes this issue.

*ANAC092* was predicted to regulate 43 potential regulons. However, from Table 5.5, it can be seen that only eight of these potential regulons were confirmed, using the *Est:ANAC092* and *anac092* microarray results, as having gene members under the regulation of *ANAC092*.

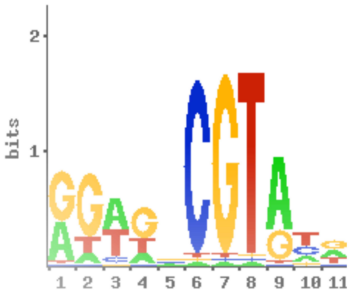
Potential regulon 15, which had the highest percentage of gene members confirmed as targets of *ANAC092* using both microarray datasets (22), was found to be significantly overrepresented for a NAC binding motif in the promoters of gene members. The NAC binding motif (PSSM ID M01055), with conserved core sequence CGT(GA) as shown in Figure 5.5 b), had a corrected *p*-value of $9.366e^{-03}$. Olsen *et al*. confirmed the presence of this motif in the promoter of *ANAC092* using EMSA (223). In addition, the core sequence of the NAC binding site was also found in the promoters of *At2g48010*, *At4g25690* and *At4g30390*, which, although predicted to be regulated by *ANAC092*, were not confirmed as such in the microarray data.

Given the correlated nature of the expression profiles of the predicted regulator *ANAC092* to its potential targets, it is difficult to decipher whether this transcription factor does indeed regulate the gene members of potential regulon 15.

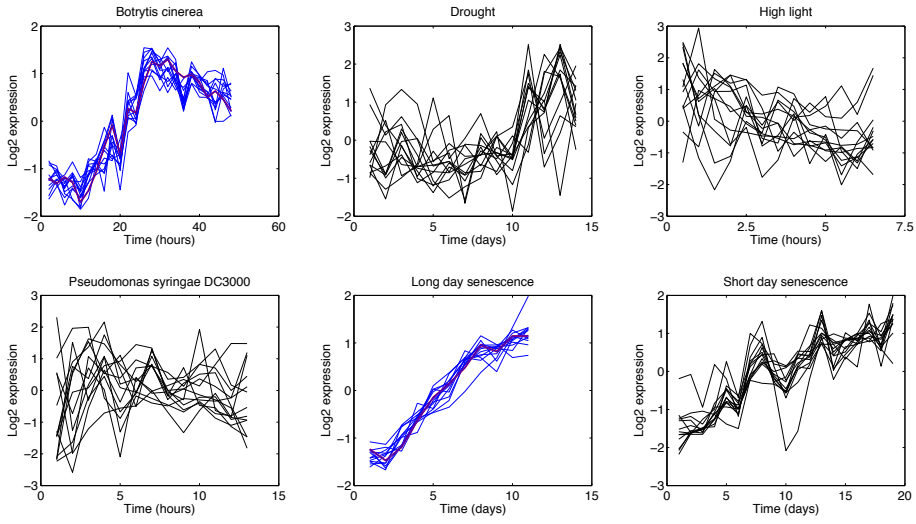
Potential regulons 18 and 310, which had three and 8 gene members confirmed as targets of *ANAC092* in both microarray experiments respectively, were both found to be significantly over-represented for the W-Box motif in the promoters of their gene members. Potential regulon 310 was predicted to be co-regulated by *WRKY45*, providing an alternative, and perhaps more likely regulator, given the presence of the W-Box motif in the promoters of gene members. Potential regulon 18, however, is only predicted to be regulated by *ANAC092*; given the low percentage of confirmed targets of *ANAC092* in this regulon, this transcription factor may not be the correct regulator, therefore, invalidating this modelling prediction.




a)




b)




c)

Figure 5.5: *ANAC092* regulates a potential regulon co-expressed in *B. cinerea* and long day senescence. a) Members of potential regulon 15, with predicted regulator *ANAC092*. Transcription factor members of potential regulon as shown as squares, whereas non-transcription factors are shown as circles. Genes whose predicted regulation by *ANAC092* has been confirmed by *anac092* or *Est:ANAC092* microarrays are shown as red arrows. b) NAC binding motif (223) found to be significantly overrepresented in promoters of gene members of potential regulon 15, with a corrected *p*-value of $9.366e^{-03}$. c) Expression profiles of gene members of potential regulon 15 significantly co-expressed in *B. cinerea* and long day senescence (in blue) and predicted regulator *ANAC092*, also a gene members of potential regulon 15 (in red). Five of the 13 members of potential regulon 15 were found to have altered expression in response to *Est:ANAC092* compared to wild type, and *anac092* mutant compared to wild type.

### 5.3.3 Confirming *Rap2.6L* as a common regulator of potential regulons

Table 5.6: Potential regulons with the highest percentage of predicted *Rap2.6L* targets confirmed in *rap2.6L* microarray (58). Only potential regulons with at least 50% of gene members confirmed as targets of *Rap2.6L* by microarray data are shown. Abbreviations for which conditions potential regulons are co-expressed are as follows: B - *B. cinerea*; P - *P. syringae* DC3000; L - Long day senescence; S - Short day senescence.

| Potential regulon number | Conditions | Percentage of gene members confirmed by microarray experiments |
|---|---|---|
| 378 | B, S | 72.88% (43/59) |
| 296 | B, P | 66.67% (10/15) |
| 132 | B, L | 63.41% (26/41) |
| 271 | B, P | 61.54% (24/39) |
| 310 | B, L | 60% (18/30) |
| 309 | L, S | 56% (13/23) |
| 126 | B, L, S | 53.33% (8/15) |
| 297 | B, L | 52.77% (19/36) |
| 48 | B, L | 51.72% (15/29) |
| 216 | B, L | 51.52% (17/33) |
| 133 | B, S | 50% (7/14) |
| 102 | B, P | 50% (16/32) |

Table 5.6 shows the percentage of gene members of potential regulons confirmed by a *rap2.6L* mutant array (58). *Rap2.6L* was predicted to regulate transcription factors in 27 potential regulons; Table 5.6 shows the top 12 potential regulons with 50% or more of gene members confirmed as *Rap2.6L* targets. Two independent mutant lines, generated by T-DNA insertion into the *Rap2.6L* gene, were compared against three independent wild type lines using Affymetric ATH1 microarrays. Che *et al*. (58) analysed the expression data to identify 5744 differentially expressed genes.

*Rap2.6L* was predicted to regulate 28 potential regulons: 27 of these potential regulons were confirmed, using the *rap2.6L* microarray results, as having gene members under the regulation of *Rap2.6L*.

APETALA2 (AP2) transcription factors, which include *Rap2.6L*, are predicted to bind to GCC-box motifs (197, 97) (PSSM ID S000430 (44)) or the CACCTG sequence (219) (PSSM ID S000315 (143)). However, none of the potential regulons listed in Table 5.6 were found to have either of these motifs significantly overrepresented within the promoters of gene members. Due to the lack of a known binding site identified for *Rap2.6L* in the promoters of gene members of these potential regulons it can be concluded that either *Rap2.6L* binds to an uncharacterised binding domain, or that *Rap2.6L* is an indirect regulator of these potential regulons.

### 5.3.4 Summary of validating regulation inferred through network inference using microarray gene expression data

To summarise, the edges of network models seen to show a common regulator of transcription factor members of potential regulon were attempted to be validated using microarray gene expression data: if these potential regulons are indeed targets of a common regulator, as predicted, the gene members should be differentially expressed in response to altered expression of the regulator. By searching for plant promoter motifs enriched in gene members (both transcription factors and non-transcription factors) of potential regulons, binding sites for predicted regulators could also be identified.

The three largest hub genes, *WRKY45*, *ANAC092* and *Rap2.6L* were considered in this analysis, as they were seen to regulate a larger than average number of targets, and array data for these genes was available. 33% of gene members in potential regulon 275, which is co-expressed in conditions *B. cinerea* and long day senescence, were differentially expressed in a microarray completed on *35S:WRKY45 A. thaliana* leaves compared to control. On closer analysis, none of these genes were found to contain the W-Box motif in their promoters. However, potential regulon 446, which is co-expressed in conditions *B. cinerea* infection and *P. syringae* DC3000 infection was found to be enriched for the W-Box motif. Comparing the expression profiles of *WRKY45* to those of the potential regulon gene members in *B. cinerea* infection and *P. syringae* DC3000 infection suggested it was plausible *WRKY45* could be a common regulator. The gene members of the potential regulon were enriched for the W-Box, however, out of the nine found to be differentially expressed in response to the *35S:WRKY45* microarray, only one gene had the W-Box present in its promoter. This suggests that *WRKY45* is an indirect regulator of these potential regulons: the modelling technique used does not distinguish between direct and indirect regulation, making this a plausible conclusion.

38% of gene members of potential regulon 15, which is co-expressed in conditions *B. cinerea* and short day senescence, were differentially expressed in both a microarray completed on *Est:ANAC092* overexpressor lines and *anac092* T-DNA mutant lines, compared to control. The gene members were also significantly enriched for a NAC binding motif: this motif was also identified in three additional gene members of the potential regulon which were not identified as differentially ex-

pression in either microarray. However, the expression profiles of *ANAC092* and of its predicted targets are tightly correlated in both conditions, making it difficult to confirm this regulation.

72 potential regulons were found to have gene members differentially expressed in response to *rap2.6L* in microarray analysis. However, none of the potential regulons were found to have a suitable binding site for the AP2 transcription factor, such as a GCC-box, in the promoters of the gene members. Althought *Rap2.6L* could bind to an unknown site, it is possible that *Rap2.6L* is an indirect regulator of these potential regulons.

## 5.4 The role of the ERF transcription factor *Rap2.6L* as a regulator of genes involved in multiple stress responses

*Rap2.6L* has been implicated as having a role in the drought and salt stress responses (163), but previous studies do not suggest any role in crosstalk between biotic, abiotic and developmental stress responses. However, using qRT-PCR after independent treatment of *A. thaliana* plants with stress hormones SA, JA, ABA and ET showed that *Rap2.6L* transcript abundance was significantly increased 6 hours post-exposure (163). Transcript abundance of *Rap2.6L* remained significantly increased 24 hours post-exposure to JA and ET, although transcript abundance exhibited no significant change 24 hours post-exposure to SA and ABA (163). Although ABA is known to have a role in abiotic stress response, JA, SA and ET are involved in the biotic stress response (99), suggesting contribution of *Rap2.6L* in both abiotic and biotic stress responses.

In the previous section, combining potential regulons found using Wigwams on time course gene expression data with inferred network models had shown *Rap2.6L* to be a predicted regulator of 28 groups of co-expressed genes (see Table 5.3). Microarray analysis on *rap2.6L* mutant lines provided over five thousand differentially expressed genes, identifying direct and indirect targets of *Rap2.6L* (58). Despite a high frequency of potential regulon gene members being identified as down regulated in *rap2.6L* mutant lines, a potential binding site for *Rap2.6L* in the promoters of the gene members could not be identified.

Of the 28 potential regulons predicted to be co-regulated by *Rap2.6L*, none were co-expressed in the stress conditions drought or high light. Previous studies have shown that over expression of *Rap2.6L* can prevent premature senescence due to water logging (178), and also enhances drought

tolerance (163). In terms of biotic stress responses, *rap2.6L* mutants have been seen to have reduced susceptibility to *P. syringae* pv. DC3000 (271). However, the role and downstream targets of *Rap2.6L* in a gene regulatory network, and the influence of *Rap2.6L* on biotic and senescence stresses have not been investigated in great depth. Since all 27 potential regulons predicted by network modelling to be co-regulated by *Rap2.6L* were found to be co-expressed in combinations of *B. cinerea* infection, *P. syringae* pv. DC3000 infection, short day and long day senescence (see Table 5.6), it seems prudent to investigate the role of *Rap2.6L* in these stresses.

A *35S:Rap2.6L* line with WS ecotype background was acquired from Dr. Nataraj Kav (University of Alberta). This line was used in Krishnaswamy *et al*. (163) to investigate the effect over expression of *Rap2.6L* on the response to abiotic stresses.

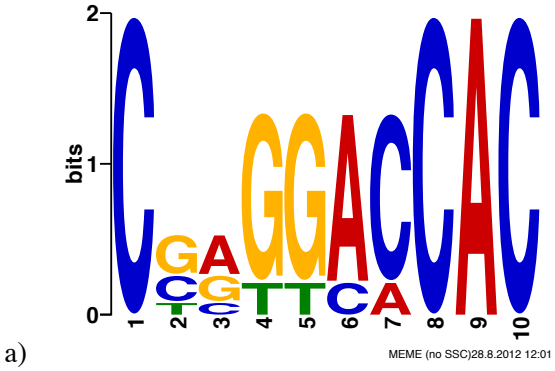
### 5.4.1 Network inference predicts a gene regulatory network around *Rap2.6L*

Since the 28 potential regulons in Section 5.3 were not significantly overrepresented for a known AP2 binding motif in the promoters of regulon gene members, three additional potential regulons, which were initially discarded in the pruning stage (see Chapter 4) were selected for further analysis. These potential regulons were initially discarded due to another potential regulon having similar expression profiles, and was overall more significant than the potential regulons we wish to study now. These new potential regulons also yielded additional evidence for a possible binding site for the predicted regulator *Rap2.6L*, as shown in the section below.

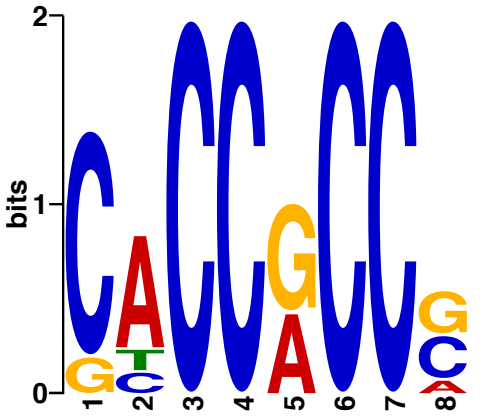
#### 5.4.1.1 *De novo* motif analysis yields potential binding sites for transcription factors

In order to identify a possible binding site for the *Rap2.6L* transcription factor in the promoters of gene members of potential regulons shown in Table 5.7, a motif analysis was performed to identify conserved motifs. By identifying the presence of a conserved motif, this substantiates the choice of using these potential regulons in further analysis, and will increase the likelihood of these genes being co-regulated by a common transcription factor, or factors.

Using the hypergeometric test (outlined in Section 2.7.5.1) as described in Section 4.3.1.1, no significantly overrepresented motifs were discovered in the promoters of gene members of the potential regulons. Therefore, MEME (19), which does not use PSSMs representing known plant promoter motifs, was used to identify a potential binding site for a common regulatory transcription




a)




b)

Figure 5.6: Promoter motifs significantly overrepresented in the promoters of all gene members of potential regulons listed in Table 5.7. a) The (G)GACCAC motif is the preferred binding site of *TCP4* (253). The *p*-value associated with this motif is $1.51e^{-07}$. b) The ACCGCC motif is a variant of the GCCGCC motif, which is the preferential binding site of AP2-domain transcription factors. AP2-domain proteins, such as *Rap2.6L* are also found to bind to this motif (329). The *p*-value associated with this motif is $1.000e^{-06}$.

factor of the gene members of the potential regulons. The 500bp promoter sequences upstream of the transcriptional start site of the gene members were acquired from TAIR and submitted to MEME, which searches the sequences for similarities, and produces a motif for each pattern MEME discovers.

The MEME analysis was completed twice using the promoter sequences of the gene members of the potential regulons: once searching for motifs with an optimum width between four and eight nucleotides, and second searching for motifs with an optimum width between six and 12 nucleotides. Plant motifs are believed to have an optimal width of between four and 12 nucleotides (136, 328, 176).

One motif per search was identified as a possible binding site for a common regulatory transcription factor in the promoters of the gene in the potential regulons listed in Table 5.7. These motifs,

shown in Figure 5.6, were likely to be true motifs significantly overrepresented in the promoters of these gene members due to the location of the motif being skewed towards the transcriptional start site (130, 198): McGrath *et al*. noted that DNA binding motifs were more likely to be identified near the transcriptional start sites of genes, and as the positioning of motifs is precise (130), it is more likely that motifs close to transcriptional start sites are responsible for the regulatory transcription factor binding there and initiating transcription of that particular gene (198).

Figure 5.6 a) shows the (G)GACCAC motif, which is the preferred binding site of *TCP4* (253). Schommer *et al*. found that miR319, a plant microRNA, regulates the translation of TCP4, which was discovered to have a role in negatively regulating leaf growth, and positively regulating leaf senescence (253). Due to *TCP4* being identified as capable of binding to the promoters of transcription factor members of the potential regulons, this is a likely candidate for a direct regulator.

Figure 5.6 b) shows a variant of the GCC-box promoter motif, an ethylene responsive element, which is the preferred binding site of AP2-domain transcription factors (260), such as *Rap2.6L*. The ACCGCC motif has been shown to bind AP2 protein transcription factors in barley: the binding of these proteins to this ACCGCC motif was enhanced by ABA and drought treatment (329). Yang *et al*. showed that *AtERF4* was capable of binding to the ACCGCC motif using electrophoretic mobility shift assays, hypothesising that these regulon genes may also have a role in the ethylene response (332). The presence of this motif in the promoters of gene members of these potential regulons suggests *Rap2.6L* is capable of directly regulating these genes, and substantiates the need to investigate the regulation of these potential regulons further.

#### 5.4.1.2 Predicted regulation of potential regulons working across multiple stresses

The gene members of these potential regulons are shown in Table 5.7. Regulators of the transcription factor members of the three potential regulons were inferred from the consensus network models, originally described in Section 5.2.1.1. By identifying likely co-regulators of the transcription factor members of the potential regulons, we can hypothesise that these co-regulators will also regulate the non-transcription factor members. Figure 5.7 illustrates the predicted gene regulatory network around the transcription factor members of these three potential regulons.

Figure 5.7 shows the predicted gene regulatory network around genes encoding transcription factors which are members of three potential regulons (see Table 5.7) found to be co-expressed in *B*.

Table 5.7: Gene members of potential regulons predicted to be co-regulated by *Rap2.6L*. Abbreviations for conditions potential regulons are co-expressed are as follows: B - *B. cinerea*; P - *P. syringae* DC3000; L - Long day senescence.

| Stress combination | ATG identifier | Gene name |
|---|---|---|
| B, P | AT1G01480 | ACS2 |
| B, P | AT1G63720 | |
| B, P | AT1G70300 | KUP6 |
| B, P | AT1G71520 | |
| B, P | AT1G22400 | UGT85A1 |
| B, P | AT2G33710 | |
| B, P | AT2G47190 | MYB2 |
| B, P | AT3G50310 | MAPKKK20 |
| B, P | AT3G06490 | MYB108 |
| B, P | AT3G21700 | SGP2 |
| B, P | AT3G25250 | AGC2-1 |
| B, P | AT5G14700 | |
| B, P | AT5G15160 | BANQUO 2 |
| B, P | AT5G24600 | |
| B, P | AT5G63130 | |
| B, P | AT5G39920 | |
| P, L | AT1G01480 | ACS2 |
| P, L | AT1G01725 | |
| P, L | AT1G05100 | MAPKKK18 |
| P, L | AT1G14420 | AT59 |
| P, L | AT1G15430 | |
| P, L | AT1G71520 | |
| P, L | AT1G52565 | |
| P, L | AT2G25460 | |
| P, L | AT3G02150 | PTF1 |
| P, L | AT3G06490 | MYB108 |
| P, L | AT3G21700 | SGP2 |
| P, L | AT3G53600 | |
| P, L | AT4G21440 | ATMYB102 |
| P, L | AT4G39670 | |
| P, L | AT5G04540 | |
| P, L | AT5G13880 | |
| P, L | AT5G15160 | BANQUO 2 |
| P, L | AT5G24600 | |
| P, L | AT5G34930 | |
| P, L | AT5G45900 | APG7 |
| P, L | AT5G16830 | SYP21 |
| B, P, L | AT1G01480 | ACS2 |
| B, P, L | AT1G71520 | |
| B, P, L | AT3G06490 | MYB108 |
| B, P, L | AT3G21700 | SGP2 |
| B, P, L | AT5G15160 | BANQUO 2 |
| B, P, L | AT5G24600 | |

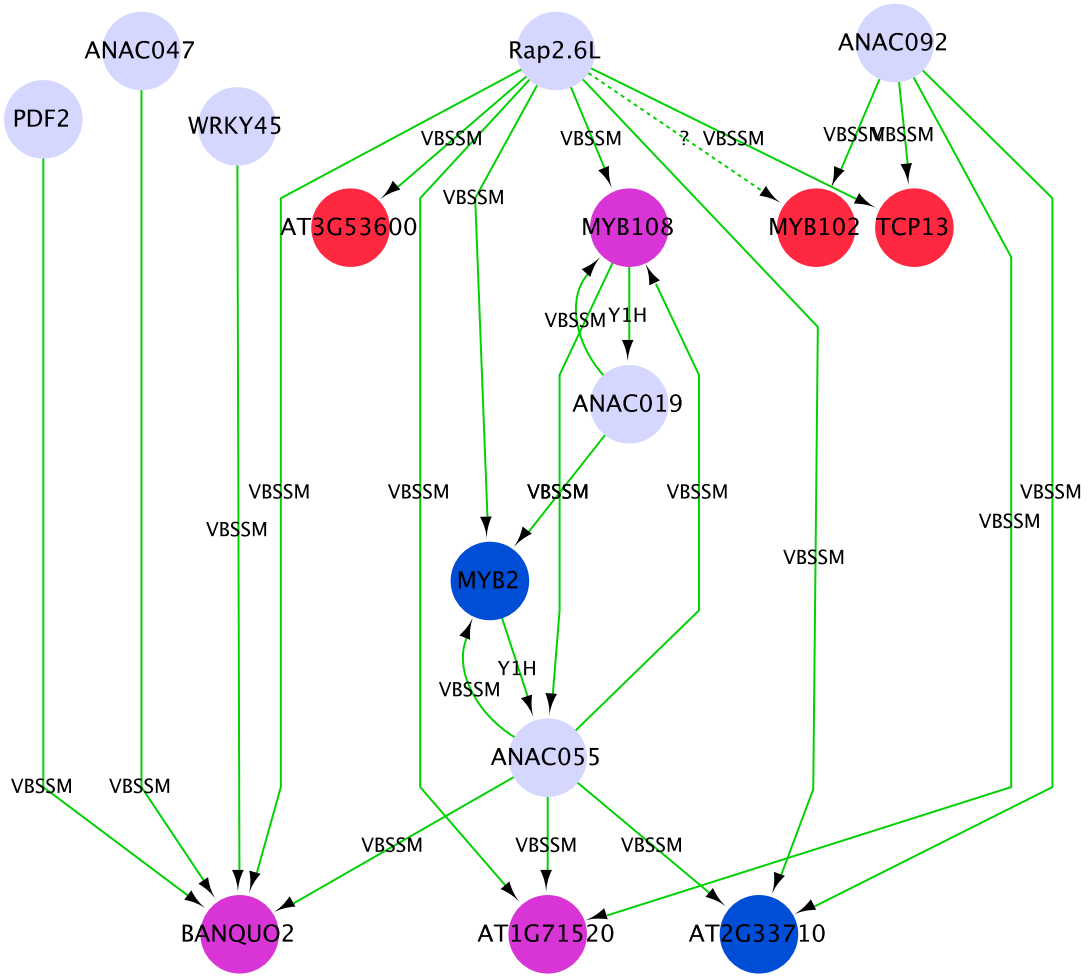



Figure 5.7: Predicted gene regulatory network around genes encoding transcription factors in potential regulons found to be co-expressed in *B. cinerea*, *P. syringae* DC3000 and long day senescence. Genes highlighted in blue are co-expressed in *B. cinerea* and *P. syringae* DC3000. Genes highlighted in red are co-expressed in *P. syringae* DC3000 and long day senescence. Genes highlighted in purple are co-expressed in all three stresses. Other genes present encode transcription factors predicted by VBSSM modelling to regulate transcription factors in potential regulons presented here. Edges of regulation are labelled as 'VBSSM', which represents predicted regulation, or 'Y1H', where this regulation has been confirmed by Y1H experimentation.

*cinerea*, *P. syringae* DC3000 and long day senescence. The edges, i.e. the arrows, represent transcriptional regulation, and are predicted by the consensus network modelling. However, the arrow does not infer direction, for example, *MYB108* → *ANAC019*, which has been confirmed by Yeast 1-Hybrid, was predicted to have the opposite regulation. VBSSM infers regulation between two genes, but not necessarily the correct direction. Therefore, these regulations need to be confirmed experimentally. Two edges in the network in Figure 5.7 (*MYB108* → *ANAC019*, and *MYB2* → *ANAC055*) have been confirmed previously by Richard Hickman (University of Warwick) using Y1H, supporting the notion to investigate this particular set of potential regulons.

Figure 5.7 also shows predicted regulators of transcription factors which are members of the potential regulons (Table 5.7), in addition to *Rap2.6L*. These genes, *WRKY45*, *ANAC092*, *PDF2* and *ANAC047* were included in this model due to the availability of microarray data for *ANAC092* and *WRKY45*, which can confirm downstream predicted targets, and phenotype evidence to suggest roles for *PDF2* and *ANAC047* in stress responses: *ANAC047* knockout transgenic lines showed reduced susceptibility to *B. cinerea* infection, whilst *A. thaliana* plant over expressing *ANAC047* exhibited increased susceptibility to *B. cinerea* and an early senescence phenotype; *PDF2* is crucial for the induction of the JA-dependent response (174), which is crucial for the response to necrotrophic pathogens, such as *B. cinerea*. *ANAC019* and *ANAC055* were included due to their mediator-role of regulation between *Rap2.6L* and the transcription factor members of the potential regulons. Microarray data is available for both *ANAC055* and *ANAC019*. These two NAC genes may also be identified as differentially expressed in *Rap2.6L* microarray data, which would confirm *Rap2.6L* indirect regulation of potential regulons via *ANAC055* and *ANAC019*.

Figure 5.8 shows the hypothesis that the non-transcription factor members of the potential regulons (shown in Table 5.7) are also regulated by *Rap2.6L*: this prediction was not included in the consensus models, as due to computational constraints, only transcription factors were included in the modelling. Therefore, it is hypothesised that if *Rap2.6L* is predicted to regulate transcription factor members of the potential regulons, it may also be a common regulator for all gene members of the potential regulons, including non-transcription factors.

By validating this quantitative model experimentally, a network underpinning the response to multiple stresses can be identified. This has considerable impact on the development of commercially important crop species to be resistant to multiple stress responses by the incorporation, or alteration

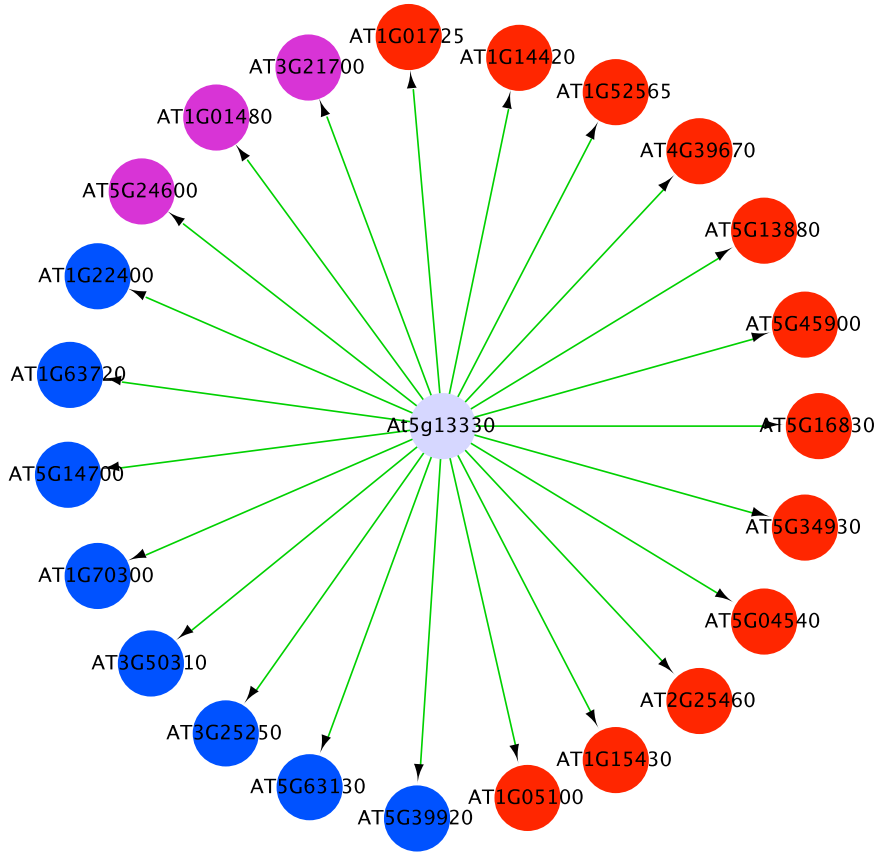



Figure 5.8: Hypothesising that Rap2.6L is the upstream regulator of non transcription factors found in potential regulons co-expressed in *B. cinerea*, *P. syringae* DC3000 and long day senescence. Since the VBSSM models only included transcription factors, we hypothesise that Rap2.6L is also a common regulator of these genes encoding non transcription factors found in potential regulons, along with genes encoding transcription factors in Figure 5.7. Genes highlighted in blue are co-expressed in *B. cinerea* and *P. syringae* DC3000. Genes highlighted in red are co-expressed in *P. syringae* DC3000 and long day senescence. Genes highlighted in purple are co-expressed in all three stresses.

of key components of this network.

Figure 5.9 shows the expression profiles of the gene members of the potential regulons shown in Table 5.7, which are significantly co-expressed in a combination of the stress *B. cinerea* infection, *P. syringae* DC3000 infection and long day senescence, along with the expression profile of *Rap2.6L*, which is predicted to regulate the transcription factor members of these potential regulons. As can be seen from Figure 5.9, the expression profile of *Rap2.6L* is correlated to the expression profiles of the potential regulons in each stress condition: the rapid increase in expression of the potential regulons' gene members in *P. syringae* DC3000 infection may be due to wounding, and not necessarily due to differential expression because of infection. Therefore, *Rap2.6L* expression positively regulates the expression of potential regulon members in *P. syringae* DC3000 infection; the expression of the potential regulons in response to *B. cinerea* infection seems to drive the expression of *Rap2.6L*, due to the slight time delay of the predicted regulators expression, compared to the expression of gene members of the potential regulons; the expression profiles of gene members of the potential regulons exhibit a time delay compared to the predicted regulator *Rap2.6L* in long day senescence, suggesting *Rap2.6L* is indeed a potential positive regulator of these genes in this particular stress.

Confirmation of co-regulation of these potential regulons in Table 5.7 by *Rap2.6L* will be two-fold: firstly, using microarray analysis of *35S:Rap2.6L* to observe whether the gene members, both transcription factors and non-transcription factors (shown in Figure 5.8) of the potential regulons are differentially expressed. One would expect to see the a majority of predicted transcription factor targets and non-transcription factor targets as differentially expressed in response to *35S:Rap2.6L*. However, microarray experiments are capable of generating false negatives (234), and therefore not all truly differentially expressed genes will be identified as such. Therefore, it is expected that not all genes identified as targets of *Rap2.6L* will be confirmed using microarray technology alone. Secondly, using Y1H to screen the promoters of transcription factor members of the potential regulons for binding of *Rap2.6L*. The gene members of the potential regulons which encode transcription factors will also be analysed using phenotype testing to confirm their importance in the defence response. Knockout and overexpressor mutant lines of genes encoding transcription factor members of the potential regulons will be screened for phenotypes against the stress conditions the genes are predicted to be co-expressed.

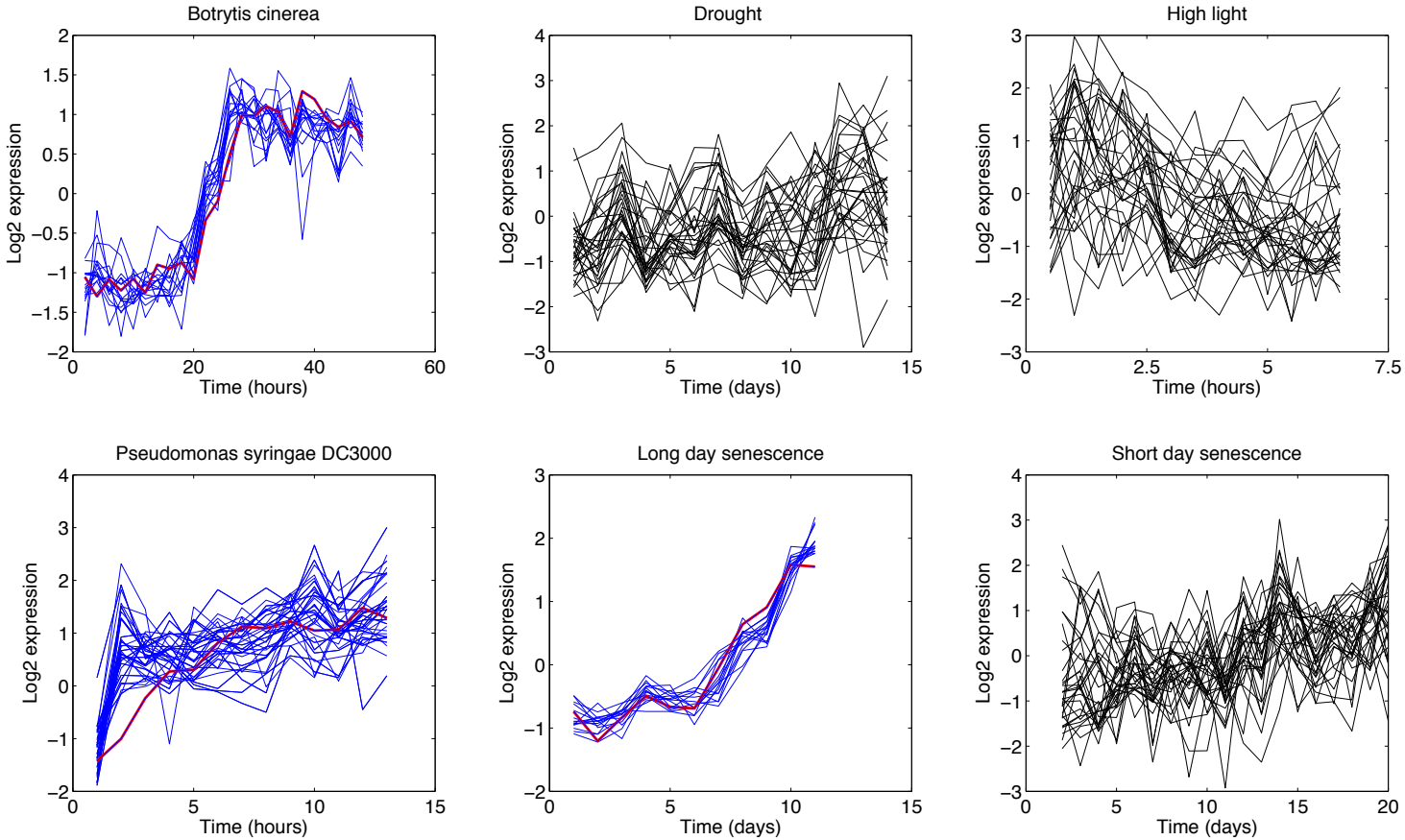



Figure 5.9: Gene expression profiles of potential regulon members in the conditions *B. cinerea* infection, *P. syringae* pv. DC3000 infection and long day senescence, along with the expression profile of predicted regulator *Rap2.6L* (shown in red). Gene expression profiles in blue are significantly co-expressed in those conditions. Gene expression profiles in black are not significantly co-expressed in those conditions.

#### 5.4.1.3 Identifying genes important in the response to multiple stresses by phenotype testing

If the expression of a gene encoding a transcription factor is altered by mutation (such as knockout or over expression), then a phenotype may be seen when studying a specific biological process. If a phenotype is observed, then it can be hypothesised the transcription factor is important in the response to that particular biological process. Therefore, by altering the expression of the transcription factor, the downstream targets are no longer regulated in the same manner as observed in wild type plants.

However, since many components of signalling pathways are functionally redundant, by using knockout mutants (75), a phenotype based on the altered expression of the gene of interest may not be seen. Overexpressor mutations aim to overcome this redundancy (338).

Figure 5.10 shows a developmental and growth phenotype of the *35S:Rap2.6L* plants compared to their background control *A. thaliana* ecotype WS plants. As can be clearly seen, the *35S:Rap2.6L* have bolted and flowered much earlier than their control counterpart. Krishnaswamy *et al.* investigated the role of *Rap2.6L* in plant growth and development, and saw earlier flowering phenotypes in the *35S:Rap2.6L* line compared to wild type, suggesting a role for *Rap2.6L* in flower development (163). Literature has suggested, however, that AP2 gene family members negatively regulate flowering time: the microRNA *miR172* has been shown to regulate a subfamily of AP2 genes. Overexpression of *miR172* down regulates the target AP2 genes, and subsequently encourages early flowering (16). Despite this, early flowering phenotypes have been observed in plants over expressing important stress-related genes, suggesting that the early flowering phenotype of *35S:Rap2.6L* may be due to the up regulation of stress-related genes (163).

Table 5.8 shows the results of phenotyping experiments, on mutant and overexpressor *A. thaliana* lines with the appropriate background (see Section 2.2 for methods), during the application of various stresses. Table 5.8 also includes results obtained from literature. Unless stated as a phenotype observed in the literature, all phenotyping screens were carried out as initial described in Methods section 2.2.

Both the *Rap2.6L* mutant line and the *35S:Rap2.6L* transgenic line provide novel phenotypes in response to *B. cinerea* and dark induced senescence, as previous studies have focused on the role of *Rap2.6L* in abiotic stress responses (163). The *B. cinerea* and *P. syringae* pv. DC3000 phenotypes

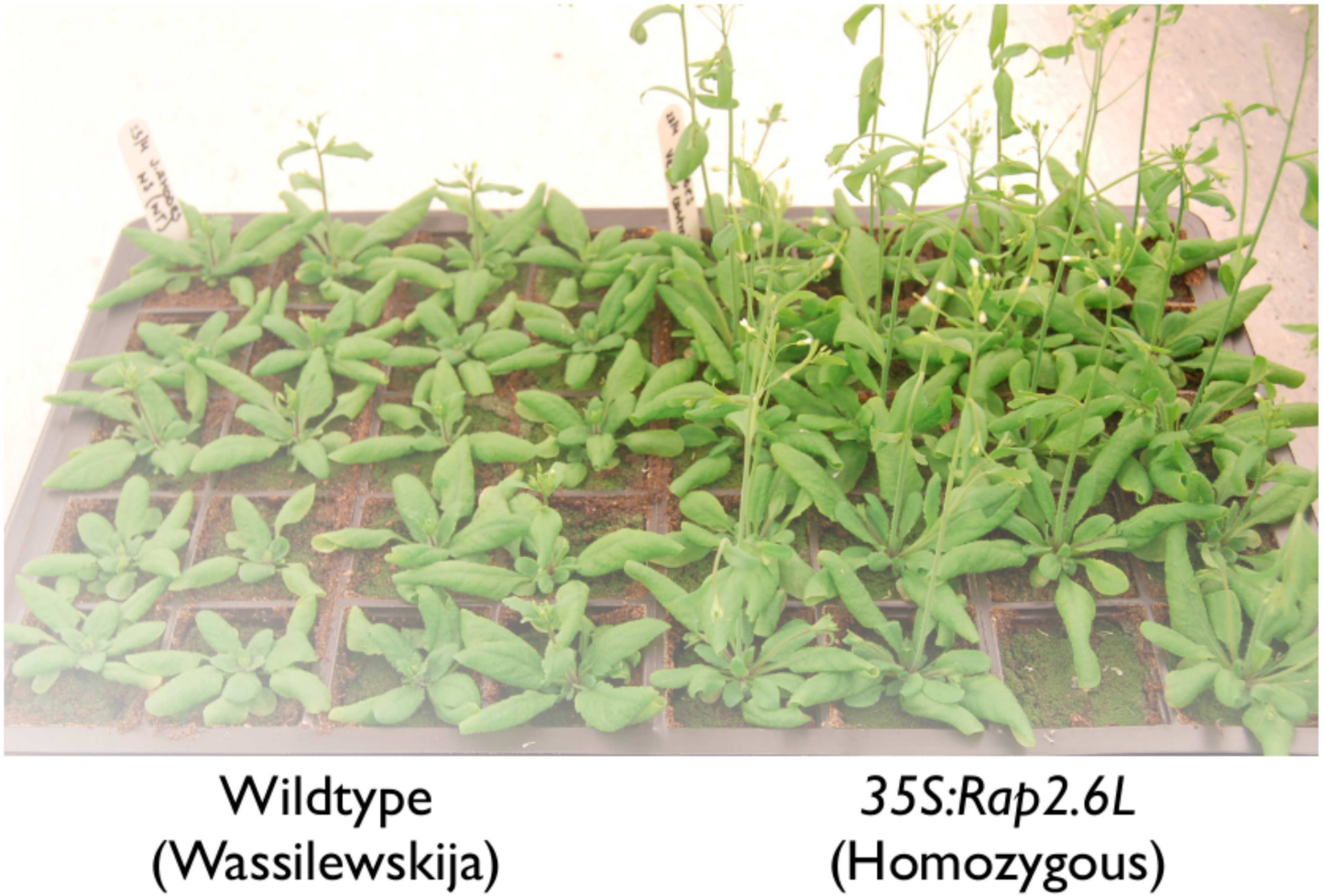



Figure 5.10: Developemental and early flowering phenotypes of *35S:Rap2.6L* (right) compared to its background control, Wassilewskija (left). *35S:Rap2.6L* showed accelerated growth compared to its background control, *A. thaliana* ecotype WS.

Table 5.8: Phenotypes of mutant plants (either knockout or plants over expressing) genes encoding transcription factors in network described in Figure 5.7. Phenotypes with * represent experiment carried out by myself. The remaining results were found by a member of the PRESTA group or through literature.

| Gene | Mutant type | *B. cinerea* | *P. syringae* DC3000 | Dark induced senescence |
|---|---|---|---|---|
| *Rap2.6L* | Overexpressor | Reduced susceptibility* | | |
| *Rap2.6L* | Knockout | Increased susceptibility* | Reduced susceptibility (271) | Mixed phenotype |
| *MYB108* | Knockout | Reduced susceptibility (201) | Wildtype (201) | Early |
| *ANAC019* | Overexpressor | Increased susceptibility (46) | | |
| *ANAC019* | Knockout | Wildtype* | | Late |
| *ANAC055* | Overexpressor | Reduced susceptibility* | | Late* |
| *ANAC055* | Knockout | Reduced susceptibility* | | Late |
| *ANAC092* | Overexpressor | Reduced susceptibility | Reduced susceptibility | |
| *ANAC092* | Knockout | Reduced susceptibility | Reduced susceptibility | |
| *ANAC047* | Overexpressor | Increased susceptibility* | | Early |
| *ANAC047* | Knockout | Reduced susceptibility* | | |
| *WRKY45* | Knockout | Increased susceptibility | | Late |
| *PDF2* | Knockout | | | Wildtype |
| *At1g17520* | Knockout | Wildtype | | |

observed in the *myb108* mutant line have previously been seen by Mengiste *et al*., and reproduced by members of the PRESTA group. However, the phenotypes observed by the *MYB108* knockout mutant is at odds with literature phenotypes: Mandaokar and Browse observed delayed senescence in *myb108* mutant lines (192).

NAC transcription factors have been heavily implemented in the response to abiotic stresses (255, 238, 224), meaning the majority of the phenotypes observed by NAC family member mutant lines during phenotyping experiments (Section 2.2), excluding the phenotypes from literature, are novel. The phenotypes observed for both *WRKY45* and *PDF2* are also novel.

By observing these phenotypes of components of the network shown in Figure 5.7, this suggests

that this network does have a role in the stresses *B. cinerea*, *P. syringae* DC3000 and long day senescence, which the potential regulon components shown in Table 5.7 are predicted to be co-expressed and co-regulated in. The regulators are also predicted in the modelling to regulate these potential regulons in the same combinations of stresses.

#### 5.4.1.4 Analysis of *35S:Rap2.6L* by microarray to identify differentially expressed genes

As previously stated, the microarray analysis of *35S:Rap2.6L* is used to observe whether the gene members, both transcription factors and non-transcription factors (shown in Figures 5.7 and 5.8 respectively) of the potential regulons shown in Table 5.7 are differentially expressed. Transgenic plants over expressing *Rap2.6L* were not subject to stress treatment, as *Rap2.6L* is already expressed, and expression does not need inducing via application of a stress.

Three biological replicates of amplified RNA (aRNA) were pooled for *35S:Rap2.6L* leaf samples and wild type WS leaf samples. Both the leaf samples for the *35S:Rap2.6L* line and the WS control were 28 days old at harvesting. Details of plant growth conditions were originally described in Methods section 2.1. This comparison between *35S:Rap2.6L* and WS included three technical replicates, including dye swaps. Therefore, six arrays were completed in total. All protocols for labelling, hybridisation and scanning of array slides are discussed in Section 2.5.

Raw expression data was processed using LimmaGUI (318), using no background correction, and normalised both within (using print tip loess normalisation) and between arrays (using quantile normalisation) for each *35S:Rap2.6L* versus control WS comparison. A least squares linear model was used to identify 94 differentially expressed genes, using a Benjamini and Hochberg FDR, and considering all genes with a corrected $p$-value $\leq 0.05$.

None of the gene members of the three potential regulons shown in Table 5.7 were differentially expressed in the *35S:Rap2.6L* array. This suggests that *Rap2.6L* may be an indirect regulator of these potential regulons, with a direct regulator driving the expression of the gene members of these potential regulons. However, 13 unique genes (10 non-transcription factors, three transcription factors) were identified as differentially expressed in the *rap2.6L* microarrays conducted by Che *et al*. (58), which are shown in Table 5.9.

The original network models shown in Figures 5.7 and 5.8, illustrating the predicted and hypothe-

Table 5.9: Gene members of all three potential regulons identified as differentially expressed in *rap2.6L* microarray (58). Abbreviations for conditions potential regulons are co-expressed are as follows: B - *B. cinerea*; P - *P. syringae* DC3000; L - Long day senescence.

| Stresses | ATG identifier | Gene name |
|---|---|---|
| B,P | AT1G63720 | |
| B,P | AT1G70300 | *KUP6* |
| B,P | AT2G33710 | |
| B,P | AT3G50310 | *MAPKKK20* |
| B,P | AT5G63130 | |
| P,L | AT1G15430 | |
| P,L | AT4G21440 | *ATMYB102* |
| P,L | AT5G04540 | |
| P,L | AT5G13880 | |
| P,L | AT5G16830 | *SYP21* |
| B,P,L | AT1G01480 | *ACS2* |
| B,P,L | AT5G15160 | *BNQ 2* |
| B,P,L | AT5G24600 | |

sised regulation of transcription factor and non-transcription factor members of potential regulons shown in Table 5.7, respectively, by *Rap2.6L* have been modified to show the edges which have been confirmed by *rap2.6L* array data (58). Figure 5.11 illustrates the confirmed predicted edges of regulation of transcription factor members of potential regulons by *Rap2.6L*, and Figure 5.12 illustrates the confirmed hypothesised edges of regulation of non-transcription factor members of potential regulons by *Rap2.6L*.

#### 5.4.1.5 Combining publicly available microarray data to confirm further edges of a gene regulatory network

The predicted gene regulatory network shown in Figure 5.7 also contains predicted regulators *WRKY45*, *ANAC047*, *PDF2*, *ANAC055*, *ANAC019* and *ANAC092*. These genes were predicted to regulate the expression of transcription factor members of potential regulons co-expressed in combinations of the stresses *B. cinerea*, *P. syringae* DC3000 and long day senescence (shown in Table 5.7). They were included in this network due to mutant lines and transgenic lines over expressing these genes exhibiting phenotypes in response to *B. cinerea*, *P. syringae* DC3000 and long day senescence, and also the availability of array data for *WRKY45*, *ANAC092*, *ANAC019* and *ANAC055*. It is hoped that Y1H on transcription factor members of the three potential regulons will identify *PDF2* and *ANAC047* as direct regulators, confirming this predicted regulation.

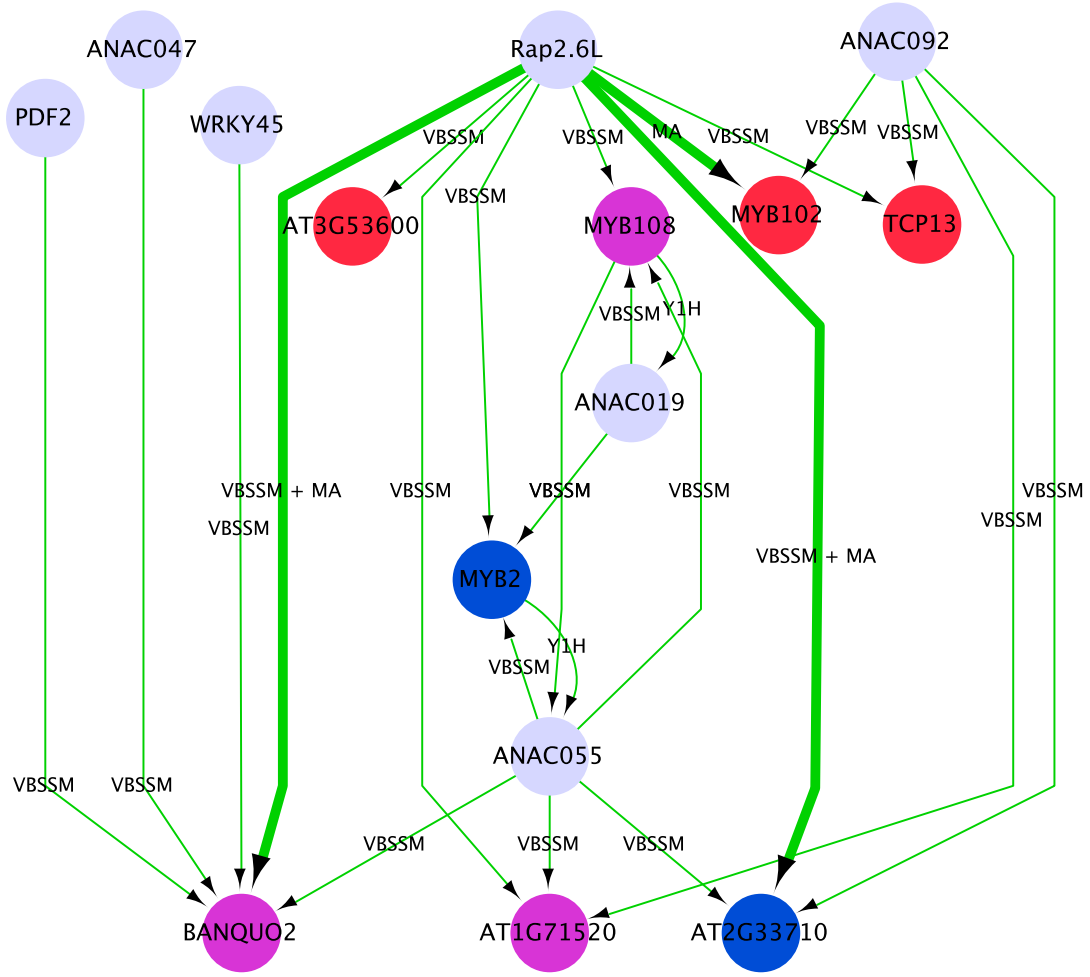



Figure 5.11: Edges of regulation of transcription factor members of potential regulons confirmed using *rap2.6L* array data (58). Based on the original network shown in Figure 5.7, where a predicted gene regulatory network around genes encoding transcription factors in potential regulons co-expressed in combinations of *B. cinerea*, *P. syringae* DC3000 and long day senescence. Genes highlighted in blue are co-expressed in *B. cinerea* and *P. syringae* DC3000. Genes highlighted in red are co-expressed in *P. syringae* DC3000 and long day senescence. Genes highlighted in purple are co-expressed in all three stresses. Edges confirmed by *rap2.6L* array data are shown as a thick green line, and labelled 'MA'.

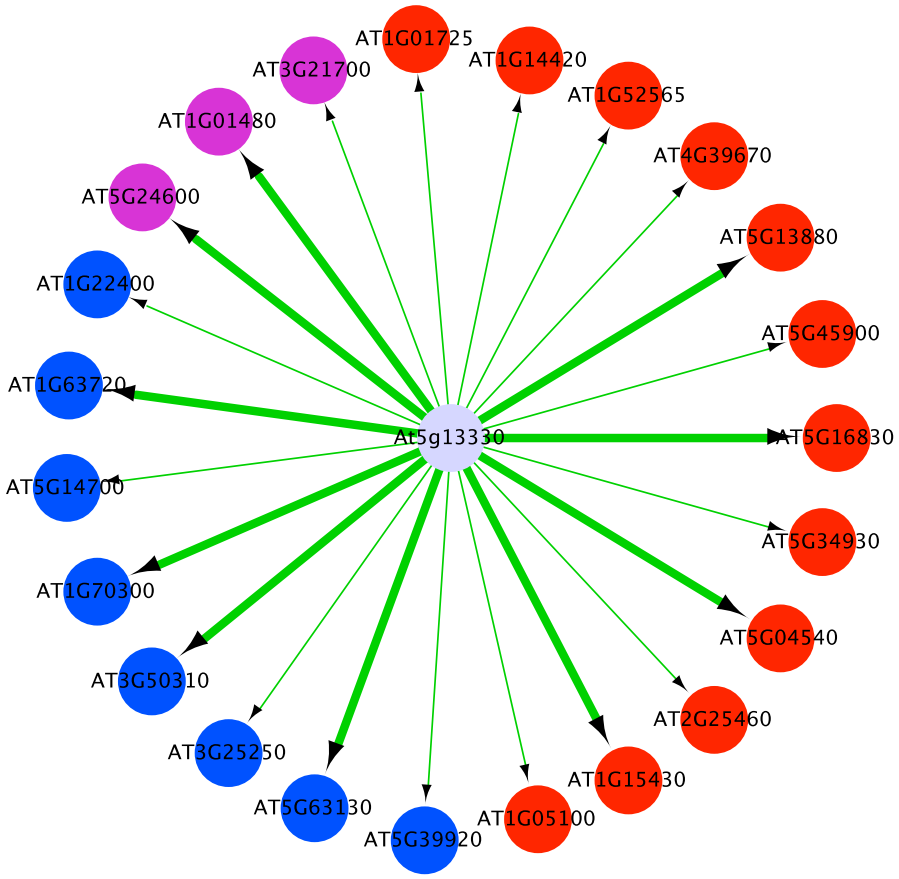



Figure 5.12: Edges of regulation of non-transcription factor members of potential regulons confirmed using *rap2.6L* array data (58). Based on the original network shown in Figure 5.8, where it was hypothesised that *Rap2.6L* regulated non-transcription factor members of potential regulons co-expressed in combinations of *B. cinerea*, *P. syringae* DC3000 and long day senescence. Genes highlighted in blue are co-expressed in *B. cinerea* and *P. syringae* DC3000. Genes highlighted in red are co-expressed in *P. syringae* DC3000 and long day senescence. Genes highlighted in purple are co-expressed in all three stresses. Edges confirmed by *rap2.6L* array data are shown as a thick green line.

Array data is available for four of the predicted transcription factor regulators shown in Figure 5.7: *35S:WRKY45* (unpublished, Thomas Vigrass, University of Warwick), *anac092* and *Est:ANAC092* (22), *35S:ANAC055* (291) and *35S:ANAC019* (291), which provide a list of differentially expressed genes in response to one of these transcription factors having they expression knocked out or over expressed. Therefore, these differentially expressed genes are directly, or indirectly, regulated by the transcription factor in question. By integrating these differentially expressed genes with the predicted edges observed in Figures 5.11 and 5.12, additional edges of regulation can be confirmed.

The *anac092* array data (22) did not identify the four predicted transcription factor targets, *MYB102*, *TCP13*, *At2g33710*, and *At1g71520*, which are shown in Figure 5.7, as differentially expressed. The *Est:ANAC092* array data (22) only identified one non-transcription factor member, *At4g39670*, of the potential regulon co-expressed in stresses *P. syringae* DC3000 and long day senescence, as differentially expressed. This regulation was not predicted by the modelling, as only transcription factors were included in the network models. The regulation was not hypothesised either, as *ANAC092* was not identified as a common regulator of all transcription factor members of the potential regulons. A NAC binding site (223) was not identified in the 500 bp upstream of the transcriptional start site of *At4g39670* however, indicating *ANAC092* is an indirect regulator of this gene.

Two non-transcription factor members, *At1g22400* and *At5g14700*, of the potential regulon co-expressed in *B. cinerea* and *P. syringae* DC3000 infection (shown in Table 5.7) were identified as differentially expressed in the *35S:WRKY45* array data (unpublished, Thomas Vigrass, University of Warwick). Again, this regulation was not predicted by the modelling, as only transcription factors were included in the network models, and was not hypothesised, as *WRKY45* was not identified as a common regulator of all transcription factor members of the potential regulons. W-Box motifs (337) were not identified in the 500 bp upstream of the transcriptional start site of *At1g22400* and *At5g14700*, suggesting *WRKY45* is an indirect regulator of these genes.

The predicted regulation of *MYB2* by both *ANAC019* and *ANAC092* was validated in the *35S:ANAC019* and *35S:ANAC055* array data (291). The confirmation of the predicted regulation of *MYB2* is shown in Figure 5.13. However, the regulation of *MYB108* by *ANAC019* and *ANAC055* was not confirmed, which can be explained by a number of reasons: there could be errors in the designs of

the microarray experiment (such as flaws in the collection of samples, analysis, or interpretation of the resulting data), which would lead to the wrong genes being identified as differentially expressed (286). It is also possible that the regulation predicted by the modelling approach used was incorrect.

The predicted regulation of *BANQUO 2*, *At1g71520* and *At2g33710* by *ANAC055* was also not confirmed in the *35S:ANAC055* array, suggesting that this predicted regulation may be indirect. An interesting observation is the confirmed regulation of *MYB2* by *ANAC055*, which is the opposite interaction seen in the Y1H results (where *MYB2* is seen to bind to the promoter of *ANAC055*). This suggests a feedback loop in the regulation in this GRN.

#### 5.4.1.6 Summary of the integration of microarray data to confirm predicted gene regulatory networks

Figure 5.13 shows the edges which have been confirmed by microarray analyses from literature. Microarray analysis completed on *35S:Rap2.6L* leaves yielded 94 differentially expressed genes. However, none of these differentially expressed genes were predicted targets of *Rap2.6L*. Therefore, microarray data on a *rap2.6L* mutant line was used to validate thirteen edges of regulation of *Rap2.6L* to predicted transcription factor targets (see Figure 5.11) and hypothesised non-transcription factor targets (see Figure 5.12). Microarray data available on *ANAC019* and *ANAC055* also confirmed their regulation of the potential regulon member *MYB2*.

The *35S:Rap2.6L* and *rap2.6L* (58) arrays have not confirmed the regulation of *MYB2*, *MYB108*, *TCP13*, *At1g71520* or *AT3G53600* transcription factors by *Rap2.6L*; however, Y1H may confirm this regulation. Therefore, Y1H will be performed, testing the promoters of the genes encoding *MYB2*, *MYB108*, *TCP13*, *At1g71520* and *AT3G53600* against a library of transcription factors. This Y1H experiment aims to identify *Rap2.6L* as a direct regulator of these genes, or identify an alternative common transcription factor regulator of these genes.

The arrays also did not confirm the regulation of thirteen non-transcription factor members (shown in Table 1.7) of the potential regulons shown in Table 5.7 by *Rap2.6L* either. These predicted regulations may still be real, yet were not identified on the microarrays as the regulation of the expression of these genes may require additional mechanisms or genes other than *Rap2.6L* (e.g. combinatorial regulation). Therefore, the expression levels of these genes will have stayed con-

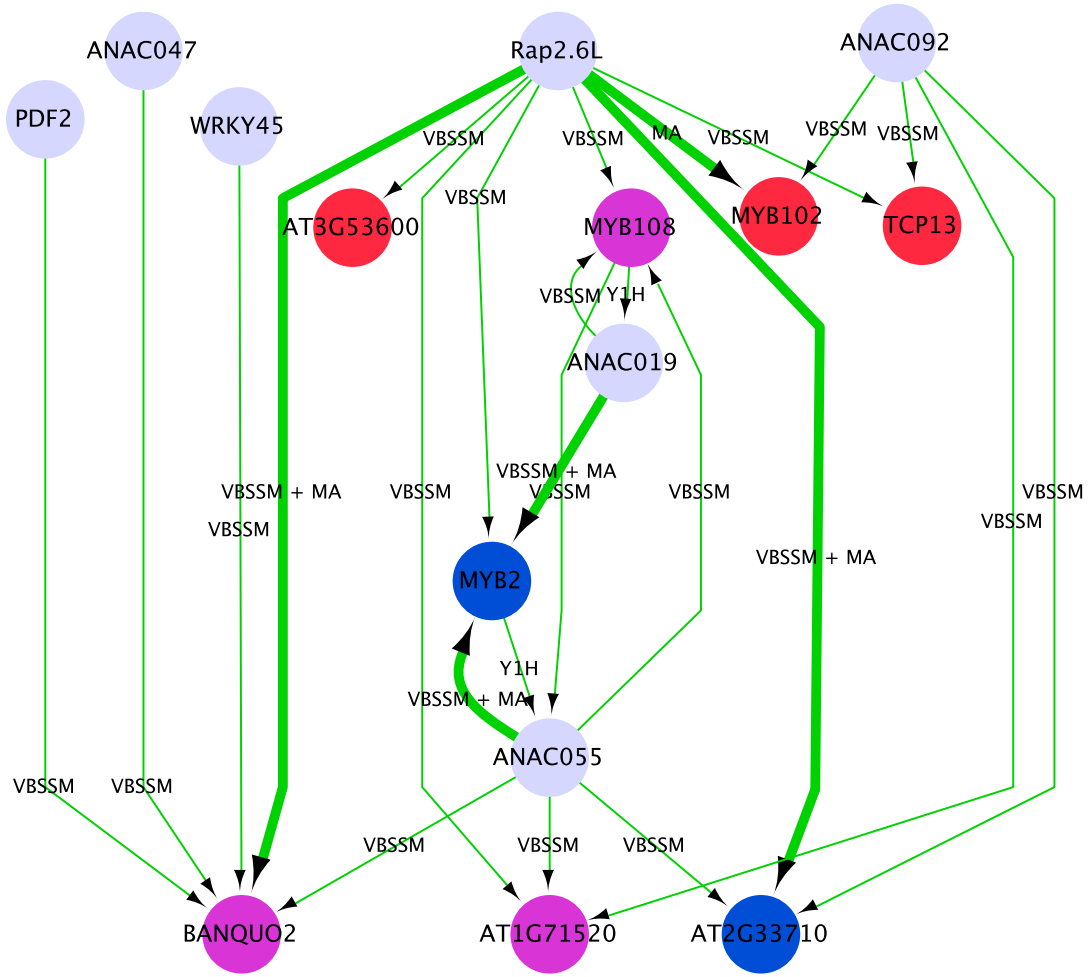



Figure 5.13: Edges of regulation of transcription factor members of potential regulons confirmed using *rap2.6L* (58), *35S:ANAC019* and *35S:ANAC055* array data (291). Based on the original network shown in Figure 5.7, where it was predicted that *Rap2.6L* regulated transcription factor members of potential regulons co-expressed in combinations of *B. cinerea*, *P. syringae* DC3000 and long day senescence. Genes highlighted in blue are co-expressed in *B. cinerea* and *P. syringae* DC3000. Genes highlighted in red are co-expressed in *P. syringae* DC3000 and long day senescence. Genes highlighted in purple are co-expressed in all three stresses. Edges confirmed by *rap2.6L*, *35S:ANAC019* and *35S:ANAC055* array data are shown as a thick green line.

sistently expressed despite the over expression of Rap2.6L. Alternatively, the statistical power of the microarray experiments may not have been sufficient to identify these genes as differentially expressed.

#### 5.4.1.7 Y1H analysis to identify regulator(s) of transcription factor members of potential shared regulons

Y1H was used to identify transcription factors capable of directly binding to the promoters of transcription factor members of the potential regulons shown in Figure 5.7. Previously, it was shown that a common transcription factor regulator of the genes *MYB2*, *MYB108*, *TCP13*, *At1g71520* and *AT3G53600* has not been identified using microarray experiments. However, network inference has predicted *Rap2.6L* to be a common regulator of these genes. Therefore, Y1H will be performed to identify *Rap2.6L* as a direct regulator of these genes. Promoter fragments of these genes of interest were screened against a cloned library of transcription factors (described in Section 2.4.2), and colonies were grown on SD-LT, SD-LTH, and three levels of 3AT, in order to eliminate auto-activation and select for an interaction.

***At1g71520*** Three promoter fragments, each 400 bp in length, with 100 bp overlap between the fragments, were designed to cover 1000 bp upstream of the TSS of *At1g71520*. Transcription factors confirmed to interact with these fragments are summarised in Table 5.10, and illustrated in Figure 5.14.

Table 5.10: Matrix high-throughput Y1H screen of *At1g71520* promoter fragments using pooled transcription factor library and 3AT selection. Fragment 1 covers the promoter sequence closest to the TSS. Interactors labelled with * were found in both arrangements of the library.

| Fragment number | Transcription factor confirmed to interact with fragment |
|---|---|
| 1 | *TCP1*, *TCP2*, *TCP3*, *TCP4*, *TCP5*, *TCP6*, *TCP7*, *TCP9*, *TCP10*, *TCP11*, *TCP12*, *TCP13*, *TCP14*, *TCP15*, *TCP16*, *TCP17*, *TCP18*, *TCP19*, *TCP20*, *TCP21*, *TCP22*, *TCP23*, *TCP24* |
| 2 | *TCP1*, *TCP3**, *TCP4*, *TCP7*, *TCP8**, *AGL88*, *ERF15*, *At5g18450*, *bZIP4*, *ILR3*, *ANAC102*, *BPE* |
| 3 | *TCP1*, *TCP2*, *TCP3*, *TCP4*, *TCP8**, *TCP14*, *TCP15*, *TCP24*, *At4g35280*, *At5g18450*, *ARF17*, *WRKY8*, *WRKY21*, *WRKY53*, *ABF3*, *MYB26*, *GBF1*, *WRKY15* |

Figure 5.14 illustrates the presence of a TCP4 motif (253) in Fragment 2 and a W-Box (337) in Fragment 3 of the promoter of *At1g71520*. Y1H identified that *TCP4* is capable of binding to

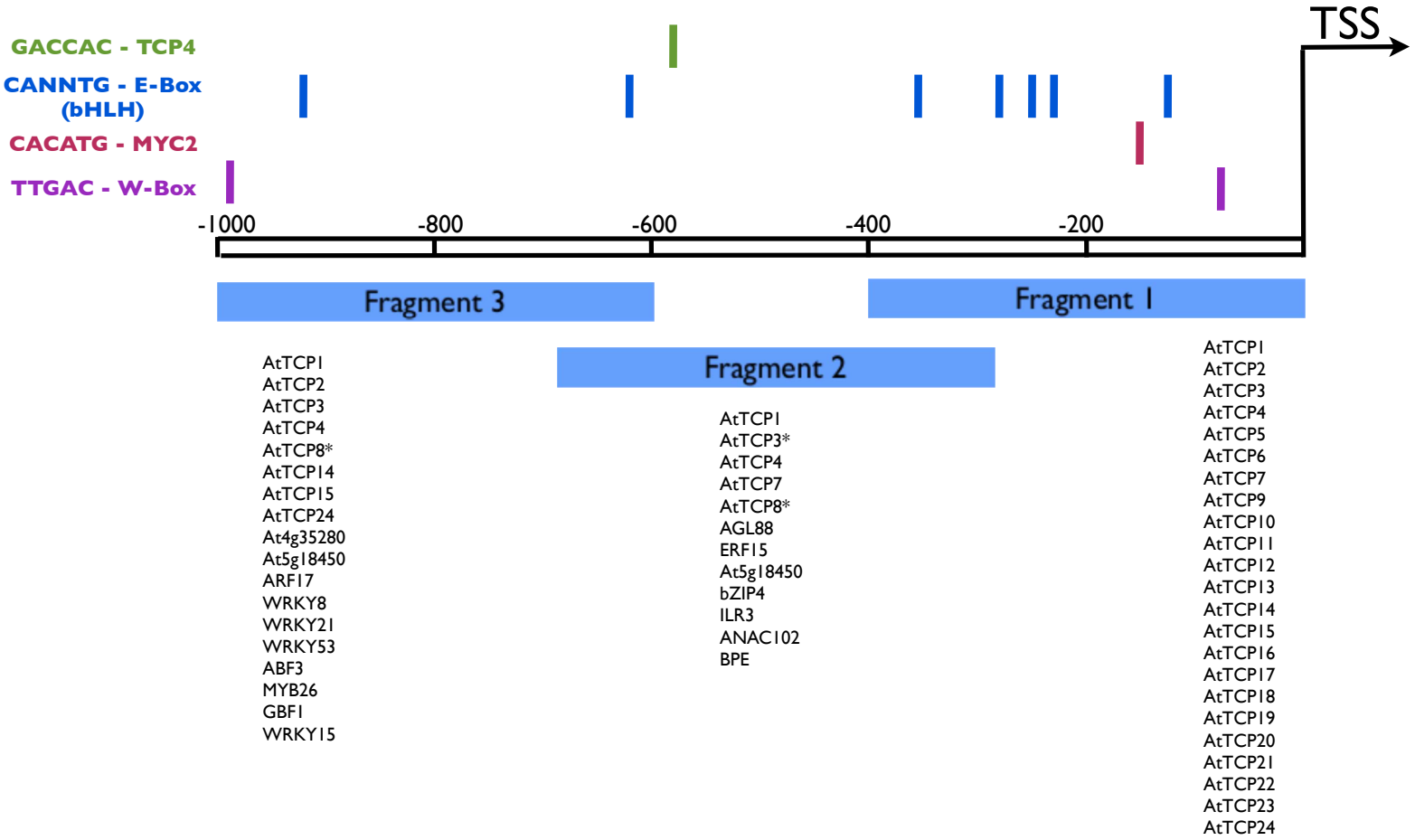



Figure 5.14: 1 kb promoter of *At1g71520* upstream of the transcriptional start site, with the interactors identified by Y1H for each promoter fragment, and motifs present. Transcription factor binding motifs: WRKY-box (337); MYC2 (3, 291); E-box (bHLH motif) (288), where N can be any nucleotide; TCP4 binding site (253).

Fragment 2. With the presence of the TCP4 motif, it is likely the binding takes place at this site. Y1H also identified *WRKY15*, *WRKY8*, *WRKY21*, and *WRKY53* binding to Fragment 3, suggesting these WRKY transcription factors are capable of binding to the W-Box motif identified in this fragment.

***At2g33710*** Three promoter fragments, each 400 bp in length, with 100 bp overlap between the fragments, were designed to cover 800 bp upstream of the TSS of *At2g33710*. Transcription factors confirmed to interact with these fragments are summarised in Table 5.11, and illustrated in Figure 5.15.

Table 5.11: Matrix high-throughput Y1H screen of *At2g33710* promoter fragments using pooled transcription factor library and 3AT selection. Fragment 1 covers the promoter sequence closest to the TSS. Interactors labelled with * were found in both arrangements of the library.

| Fragment number | Transcription factor confirmed to interact with fragment |
|---|---|
| 1 | *TCP1*, *TCP2*, *TCP3*, *TCP4*, *TCP13*, *TCP14*, *TCP20*, *TCP23* |
| 2 | *TCP1**, *TCP2**, *TCP3**, *TCP4**, *TCP5*, *TCP6*, *TCP8*, *TCP13**, *TCP14**, *TCP15**, *TCP16*, *TCP18*, *TCP20*, *TCP23**, *TCP24**, *WRKY8*, *WRKY28*, *WRKY41*, *ATHB13*, *At1g76110*, *GLABRA 2*, *HB40*, *HB23** |
| 3 | *TCP1**, *TCP3**, *TCP4*, *TCP7*, *TCP14**, *TCP15**, *TCP16*, *TCP20*, *AGL78*, *bZIP5*, *At3g24120*, *At2g16210*, *ASML2*, *ERF15**, *At5g53420*, *At4g31060*, *At5g06250*, *At1g63840**, *At2g35310*, *At4g01580*, *At1g80580*, *TGA1*, *DOF1.5*, *GATA17*, *ATXR5*, *bZIP61*, *MYB55*, *ANAC003*, *KNAT7*, *bHLH25*, *At1g31310*, *At4g38900*, *PI*, *SAL2*, *At4g18650*, *At3g49950*, *At3g56220**, *DOF4.6*, *ARR5*, *DOF4.4*, *IAA7*, *CBF4*, *IAA19*, *ATK5*, *ATHB5*, *At5g44180*, *CCA1*, *APRR2*, *At1g69580*, *ANAC069*, *ERF11*, *MYB96*, *ATHB14*, *AGL61*, *TBP1*, *HMGB5*, *MYB45*, *MYB51*, *AS1*, *MYB78*, *CBF1*, *MYB19*, *MYB17*, *WRKY43*, *WRKY31*, *LFY*, *ARF4*, *REM1*, *IAA7*, *IAA12*, *IAA13*, *RMR1*, *bZIP12*, *SVP*, *bZIP75*, *At2g17600*, *ERF2*, *NFY-C4*, *MYB26*, *GATA14*, *AGL24*, *TBP2*, *At5g25475*, *ERF9*, *RMA1*, *WOX12*, *ANAC097*, *WRKY49*, *ATL41*, *ATAF2*, *At3g12730*, *ATK5*, *GBF2*, *ANAC036*, *RHA2B*, *ANAC006* |

Figure 5.15 shows the presence of a W-Box (337) in the 100 bp sequence shared by both Fragments 1 and 2 in the promoter of *At2g33710*. Three WRKY transcription factors, *WRKY8*, *WRKY28* and *WRKY41* were identified by Y1H as capable of binding to Fragment 2, suggesting they are binding at the W-Box present in this sequence. Although no WRKY transcription factors were seen binding to the same sequence in Fragment 1, this may be due to a weak interaction which was not observed on stringent selection (e.g. concentrations of 3AT).

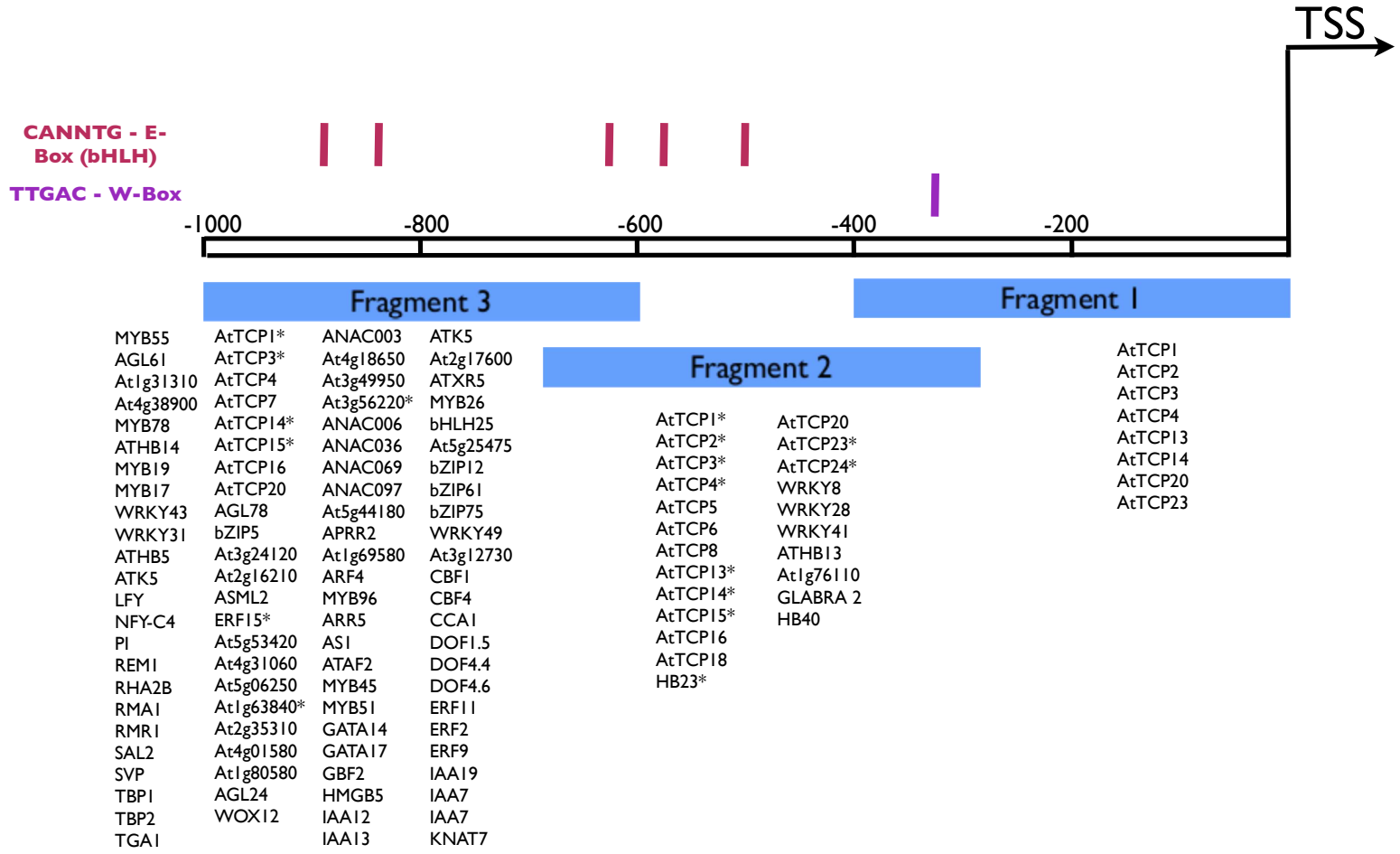



Figure 5.15: 1 kb promoter of *At2g33710* upstream of the transcriptional start site, with the interactors identified by Y1H for each promoter fragment, and motifs present. Transcription factor binding motifs: WRKY-box (337); E-box (bHLH motif) (288), where N can be any nucleotide.

***At3g53600*** Three promoter fragments, each 400 bp in length, with 100 bp overlap between the fragments, were designed to cover 800 bp upstream of the TSS of *At3g53600*. Transcription factors confirmed to interact with these fragments are summarised in Table 5.12, and illustrated in Figure 5.16.

Table 5.12: Matrix high-throughput Y1H screen of *At3g53600* promoter fragments using pooled transcription factor library and 3AT selection. Fragment 1 covers the promoter sequence closest to the TSS. Interactors labelled with * were found in both arrangements of the library.

| Fragment number | Transcription factor confirmed to interact with fragment |
|---|---|
| 1 | *TCP3**, *TCP4**, *ERF11*, *PIF3*, *SOC1*, *ERF10**, *TBP1*, *MYB53*, *MYB92*, *HMGB1*, *MYB58*, *LAF1*, *MYB19*, *MYB118*, *MYB26**, *TT2*, *AGL96*, *WRKY40*, *AGL56*, *ANAC038*, *IAA4*, *IAA13*, *GBF4*, *RMR1*, *DREB1A*, *IAA18*, *At1g63040*, *bZIP75*, *ERF15*, *ATL68*, *At3g43430*, *ERF84*, *DOF1.5*, *CBF1*, *GATA17*, *ATL18*, *HEC1*, *HSFB2A*, *YABBY1*, *DDF1*, *WRKY49*, *TFIIIA*, *SHL1*, *At2g39900*, *WOX7*, *WOX5*, *ATHB21*, *ATHB20*, *HAT2*, *AGL14*, *bZIP8*, *DOF4.6*, *PAT1*, *At3g56220*, *TAF6*, *NFY-B6*, *At5g41920*, *WRKY24*, *GATA13*, *At1g16640*, *WRKY59*, *At2g34450*, *HAT9*, *NGA2*, *WRKY75*, *ATHB17*, *HAM3*, *ANAC088*, *AIP3*, *MBF1C*, *ATL5*, *ANAC029*, *ATCTH*, *MYB14*, *At2g34000*, *RMA1*, *GI*, *bHLH51*, *AGL36*, *At2g33720*, *ARF3*, *bHLH83* |
| 2 | *TCP4*, *TCP13* |
| 3 | *TCP3**, *TCP4*, *TCP8*, *ANAC038*, *YABBY1* |

From Figure 5.16 it can be seen that a GATA promoter motif (278) and a NAC core binding site (291) were identified in Fragment 1 of the promoter of *At3g53600*. *GATA17* was identified by Y1H to be capable of binding to Fragment 1, suggesting binding takes place at the GAT promoter motif present in this fragment. Three NAC transcription factors were also seen to bind to Fragment 1, ANAC088, *ANAC029* and *ANAC038*, suggesting they are capable of binding to the NAC core binding motif present in this fragment. Two NAC core binding motifs were also identified in Fragment 3, with one of the motifs located in the 100 bp sequence shared with Fragment 2. *ANAC038* was also identified as binding to Fragment 3, but not Fragment 2, by Y1H. The Site II motif, which has been identified as a potential binding site the TCP transcription factor family (105), was identified in Fragment 3, where Y1H also confirmed the direct binding of *TCP3*, *TCP4* and *TCP8*. The Site II motif was not located in Fragments 1 and 2, despite TCP binding to these fragments identified by Y1H.

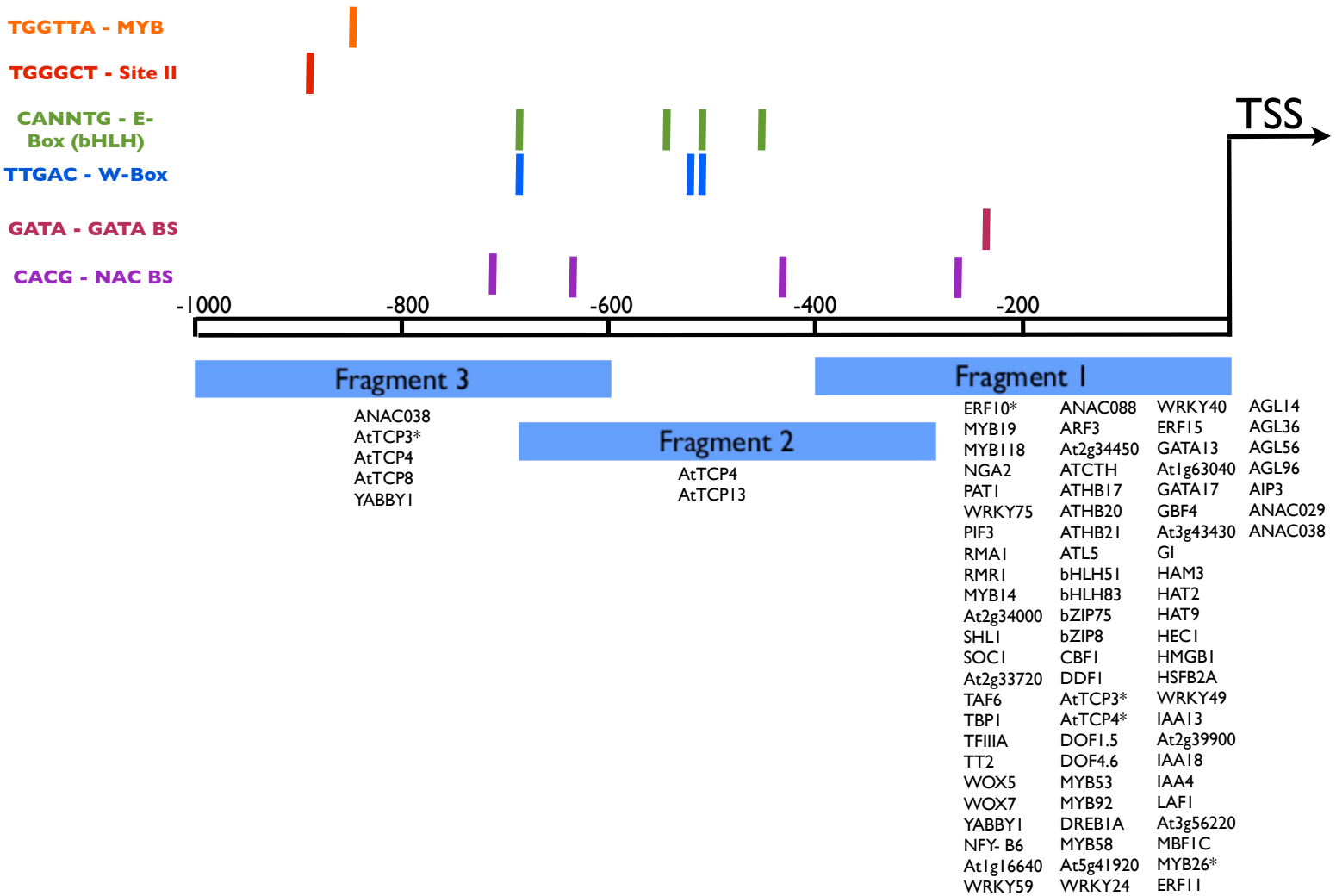



Figure 5.16: 1 kb promoter of *At3g53600* upstream of the transcriptional start site, with the interactors identified by Y1H for each promoter fragment, and motifs present. Transcription factor binding motifs: NAC core binding site (291); GATA promoter motif (278); WRKY-box (337); E-box (bHLH motif) (288), where N can be any nucleotide; Site II motif (105); MYB binding motif (2)

***TCP13*** Three promoter fragments, each 400 bp in length, with 100 bp overlap between the fragments, were designed to cover 800 bp upstream of the TSS of *TCP13*. However, only one of these fragments, the fragment closest to the TSS, was successfully cloned in yeast strain Y187. Transcription factors confirmed to interact with this fragment are summarised in Table 5.13, and illustrated in Figure 5.17.

Table 5.13: Matrix high-throughput Y1H screen of the *TCP13* promoter fragment using pooled transcription factor library and 3AT selection. Fragment 1 covers the promoter sequence closest to the TSS. Interactors labelled with * were found in both arrangements of the library.

| Fragment number | Transcription factor confirmed to interact with fragment |
|---|---|
| 1 | *TCP2*, *TCP3*, *TCP4*, *TCP5*, *TCP10*, *TCP13* |

Whilst three different motifs, the W-Box, the NAC core binding site and the E-Box, were identified in Fragment 1 of the promoter of *TCP13*, none of the transcription factors identified as directly binding to this fragment are known to be capable of binding to these motifs.

***MYB2*** Four promoter fragments, each 100 bp in length, with 30 bp overlap between the fragments, were designed to cover 310 bp upstream of the TSS of *MYB2*. A fifth fragment, of 120 bp in length, and overlapping 30 bp with the previous fragment, was designed to cover 281 bp to 401 bp upstream of the TSS. However, only two of these fragments, the fragments closest to the TSS and a fragment covering 141 bp to 241 bp upstream of the TSS were successfully cloned in yeast strain Y187. Transcription factors confirmed to interact with these fragments are summarised in Table 5.14, and illustrated in Figure 5.18.

Table 5.14: Matrix high-throughput Y1H screen of *MYB2* promoter fragments using pooled transcription factor library and 3AT selection. Fragment 1 covers the promoter sequence closest to the TSS. Interactors labelled with * were found in both arrangements of the library.

| Fragment number | Transcription factor confirmed to interact with fragment |
|---|---|
| 1 | *TCP2*, *TCP4*, *TCP10*, *TCP13* |
| 3 | *TCP1*, *TCP2*, *TCP3*, *TCP4*, *TCP8*, *TCP13*, *TCP14*, *TCP15*, *TCP16*, *TCP20*, *TCP23*, *TCP24* |

Of the two promoter fragments of *MYB2* successfully cloned and transformed into yeast, neither had known transcription factor binding motifs present in their sequence. Therefore, although a number of TCP transcription factors were seen to directly bind to Fragments 1 and 3 of the promoter of *MYB2*, a potential binding site could not be identified.

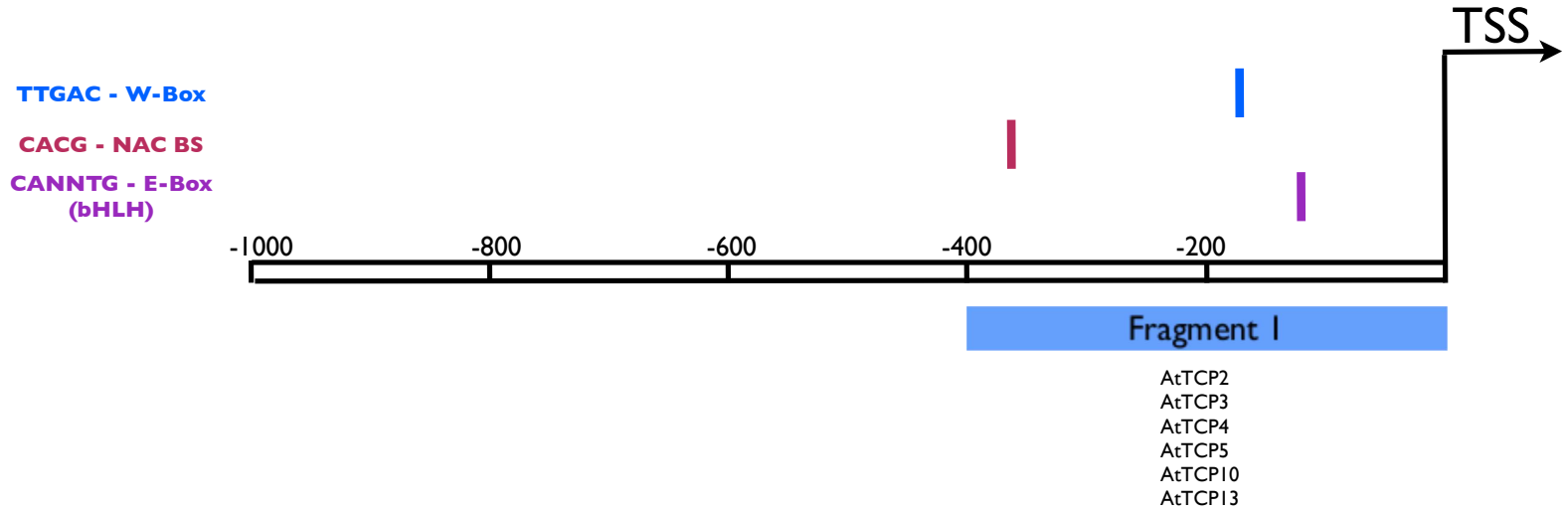



Figure 5.17: 1 kb promoter of *TCP13* upstream of the transcriptional start site, with the interactors identified by Y1H for the promoter fragment, and motifs present. Transcription factor binding motifs: E-box (bHLH motif) (288), where N can be any nucleotide; NAC core binding site (291); WRKY-box (337).

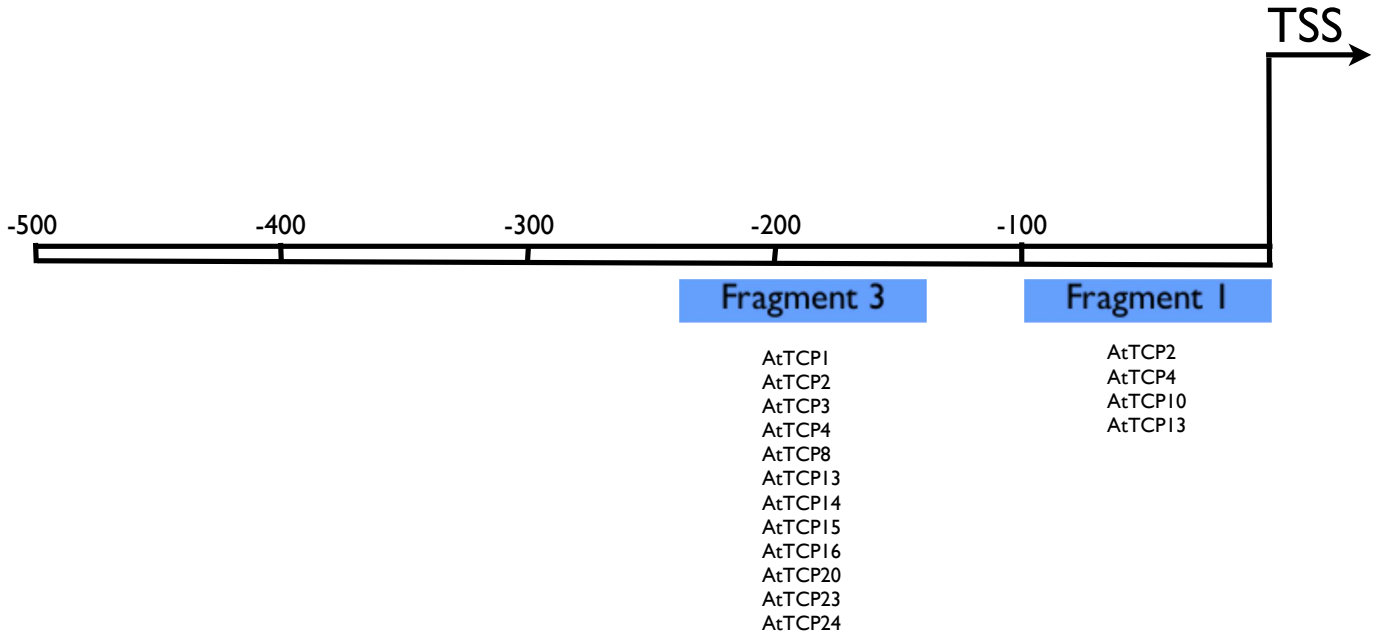



Figure 5.18: 500 bp promoter of *MYB2* upstream of the transcriptional start site, with the interactors identified by Y1H for each promoter fragment. No transcription factor binding motifs were identified in these two fragments.

***MYB108*** Four promoter fragments, each approximately 400 bp in length, with 100 bp overlap between the fragments, were designed to cover 1000 bp upstream of the TSS of *MYB2*. However, only two of these fragments, the fragments closest to and furthest away from the TSS, and the fragment covering nucleotides 601 to 1000 upstream of the TSS, were successfully cloned in yeast strain Y187. Transcription factors confirmed to interact with these fragments are summarised in Table 5.15, and illustrated in Figure 5.19.

Table 5.15: Matrix high-throughput Y1H screen of *MYB108* promoter fragments using pooled transcription factor library and 3AT selection. Fragment 1 covers the promoter sequence closest to the TSS. Interactors labelled with * were found in both arrangements of the library.

| Fragment number | Transcription factor confirmed to interact with fragment |
|---|---|
| 1 | *TCP4*, *TCP14*, *TCP15**, *ANAC102*, *MYC2*, *AHL1*, *BEE2*, *At3g57800*, *ILR3*, *AGL56*, *ATHB52*, *ZCW32*, *At4g33280*, *TBP2*, *Rap2.6L*, *VRN1*, *At3g23220*, *ZFP2*, *At4g33280* |
| 3 | *TCP3*, *TCP14*, *TCP15*, *TCP16*, *GATA14* |

From Figure 5.19 it can be seen that the MYC2 binding site (3, 291) was identified in Fragment 1 of the promoter of *MYB108*, where *MYC2* was also seen to be directly binding by Y1H. The GATA binding site (278) was also identified in the 100 bp overlap in sequence between Fragment 1 and 2, and was also identified twice in Fragment 2. Y1H identified *GATA14* as directly binding to Fragment 2, suggesting it binds preferentially to the two binding sites located solely in Fragment 2.

#### 5.4.1.8 Comparison of expression profiles of TF-target pairs in multiple stresses

Y1H generated a list of regulators seen to be capable of directly binding to the promoters of the genes *MYB2*, *MYB108*, *TCP13*, *At1g71520*, *At2g33710* and *At3g53600* (as shown in Tables 5.10 to 5.15). However, these interactions were not observed in the context of a stress response; therefore, by observing the expression profiles of the interactor and its target during *B. cinerea* infection, *P. syringae* DC3000 infection and long day senescence, which are the stresses these six genes are predicted to be co-expressed in, one can identify which targets are the most likely to be regulated in these particular stresses. Figure 5.20 to 5.23 show the expression profiles of interactors that would most likely to binding to the promoters of the genes *MYB108*, *At1g71520*, *At2g33710* and *At3g53600*, due to presence of a motif in the promoters, which is recognised by the interactors.

The expression profiles of *At1g71520*, and the transcription factors *TCP4*, *WRKY8*, *WRKY21*,

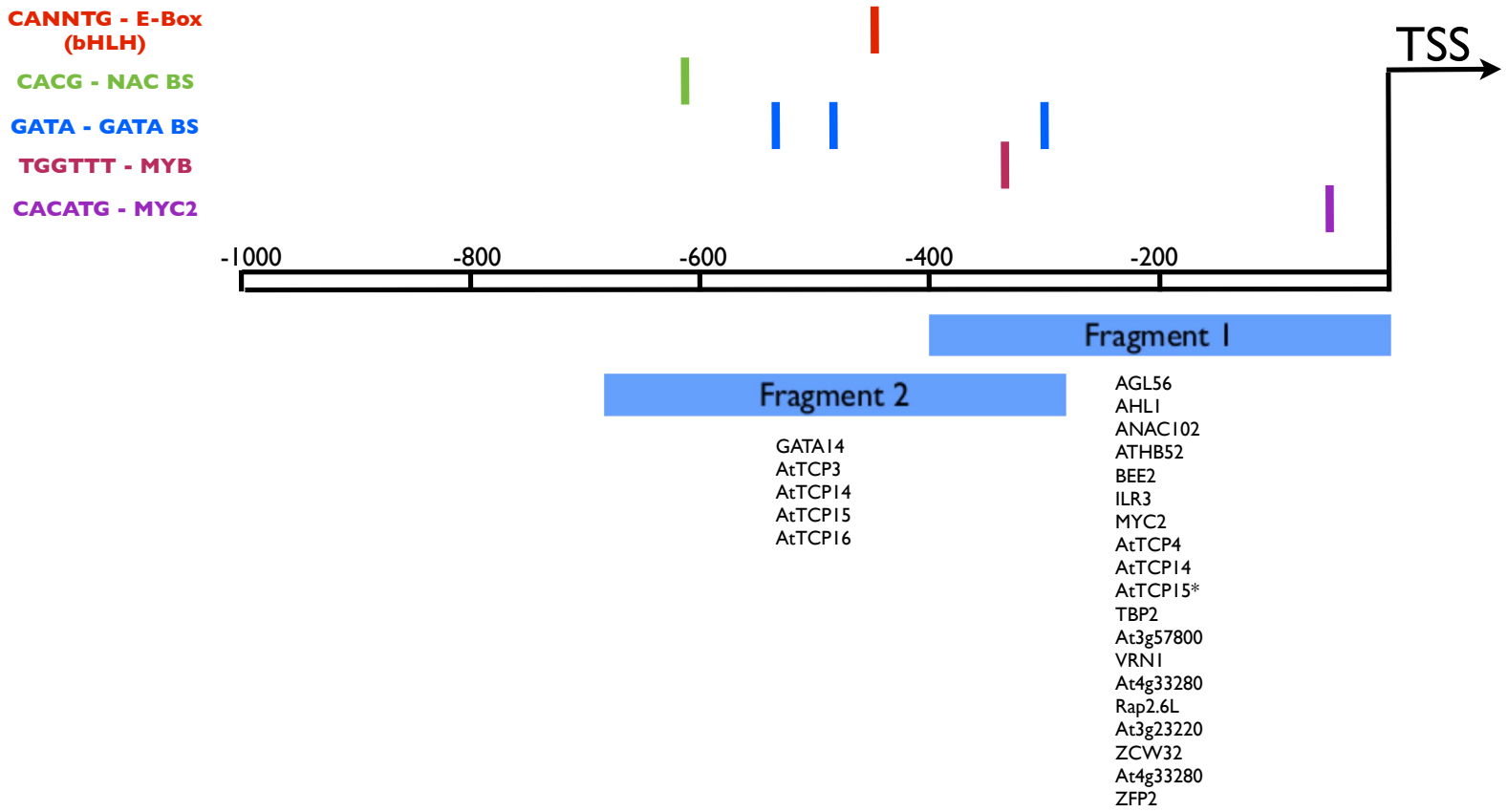





Figure 5.19: 1 kb promoter of *MYB108* upstream of the transcriptional start site, with the interactors identified by Y1H for each promoter fragment. Transcription factor binding sites: MYC2 (3, 291); MYB binding motif (2); GATA promoter motif (278); NAC core binding site (291); E-box (bHLH motif) (288), where N can be any nucleotide.

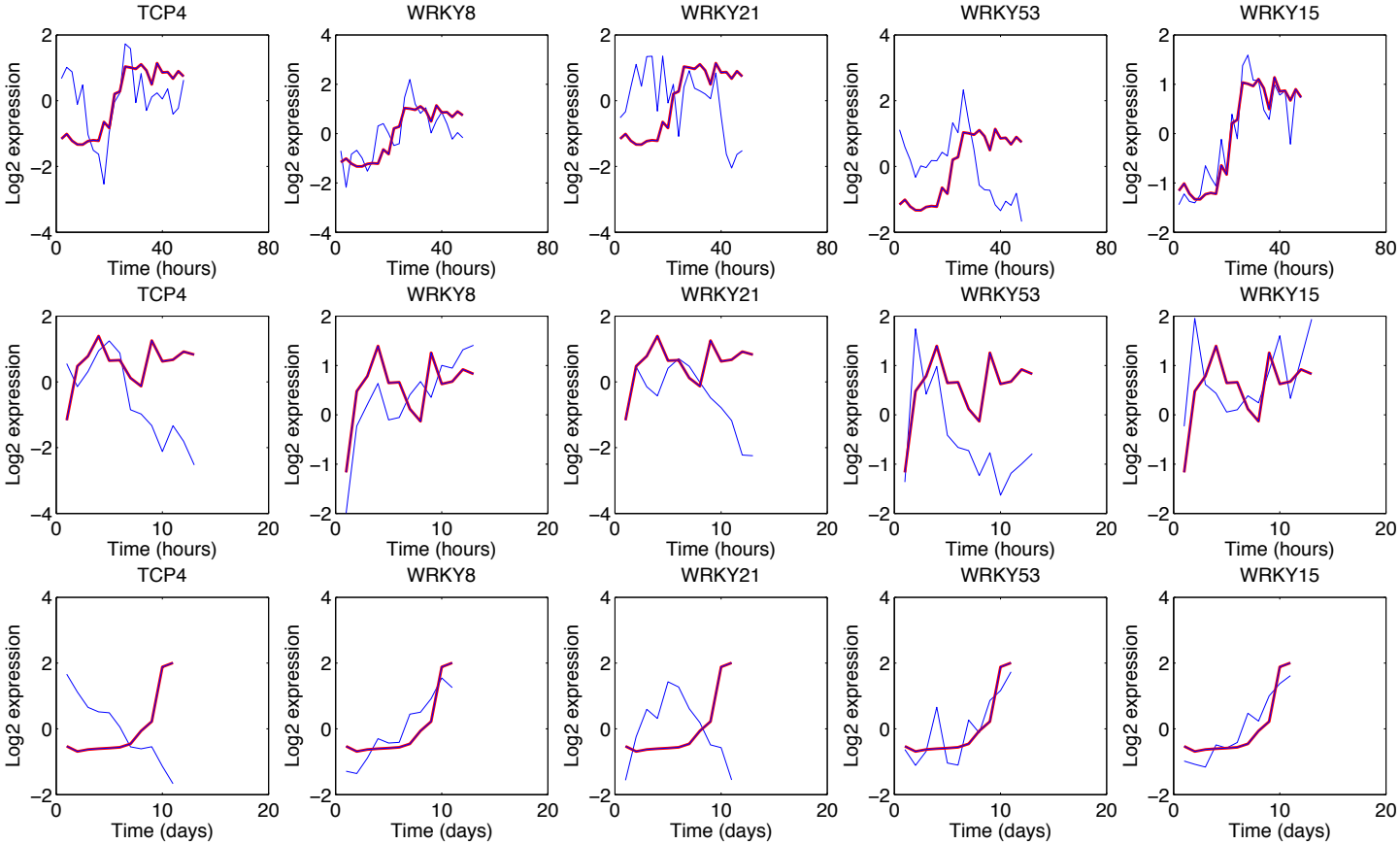



Figure 5.20: Expression profiles of Y1H interactors, that had a confirmed binding site motif present in the promoter of *At1g71520*, during multiple stress responses. The top row shows expression profiles in response to *B. cinerea*; the middle row shows expression profiles in response to *P. syringae* DC3000; the bottom row shows expression profiles in response to long day senescence. Expression profiles of *At1g71520* in each stress response is shown in red, with the interactor expression profile (the name of which is given as the title of each plot) is shown in blue.

*WRKY53* and *WRKY15*, which were found to directly bind to its promoter in a Y1H screen, are shown in Figure 5.20, in the stresses *B. cinerea* infection, *P. syringae* DC3000 infection and long day senescence. A potential TCP4 binding motif (253) and a W-Box motif (337) were located in the promoter of *At1g71520*, so only the expression profiles of these interactors are shown, as they are the most likely candidates for 'real' interaction. The possible regulation by each interactor will be discussed in the context of each stress response.

***B. cinerea* infection** By observing the top row of Figure 5.20, it appears that the regulation of *At1g71520* in response to *B. cinerea* infection is positively regulated by *WRKY8*, *WRKY53* and *WRKY15*. *TCP4* appears to be a negatively regulator of expression of *At1g71520*: as expression of *TCP4* increases around 22 hours, the expression of *TCP4* levels off. Observing the expression profiles of *WRKY21* and *At1g71520*, it is difficult to infer regulation during this stress response.

***P. syringae* DC3000 infection** By observing the middle row of Figure 5.20, the regulation of *At1g71520* in response to *P. syringae* DC3000 infection appears to be negatively regulated by *TCP4*, *WRKY21*, and *WRKY53*: ignoring the initial increase in expression, which could be due to wounding, the expression of *At1g71520* increases as the expression of *TCP4*, *WRKY21*, and *WRKY53* decreases. The expression profiles of *At1g71520* and interactors *WRKY8* and *WRKY15* appear positively correlated with each other, respectively, suggesting positive regulation of *At1g71520*.

**Long day senescence** By observing the bottom row of Figure 5.20, the expression of *At1g71520* appears to be positively regulated by *WRKY8*, *WRKY53* and *WRKY15*. The expression profile of *WRKY21* in relation to the expression profile of *At1g71520* is interesting: whilst an initial increase in expression of *WRKY21* appears to positively regulate the expression of *At1g71520*, the expression profile of *WRKY21* decreases, suggesting *WRKY21* itself is regulated by another transcription factor. However, the expression of *At1g71520* does not appear affected by this decrease in expression of *WRKY21*.

Figure 5.21 shows the expression profiles of *At2g33710* and the transcription factors *WRKY8*, *WRKY28*, and *WRKY41* in response to *B. cinerea* infection, *P. syringae* DC3000 infection and long day senescence stresses. These transcription factors were found to directly bind to the promoter of *At2g33710* in a Y1H screen (see Table 5.10). A W-Box motif (337) was located in the promoter of *At2g33710*, so only the expression profiles of these interactors are shown, as they are

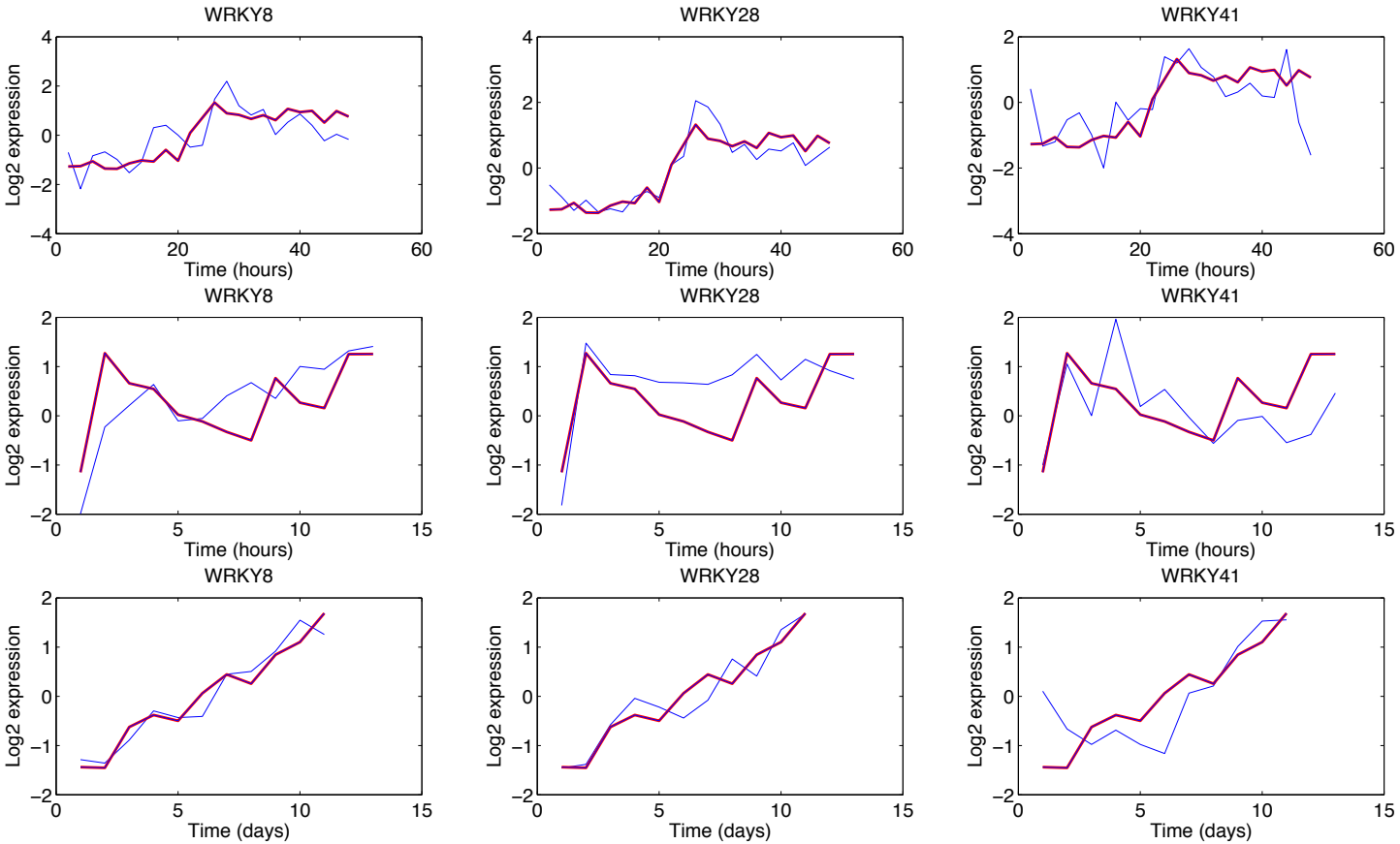



Figure 5.21: Expression profiles of Y1H interactors, that had a confirmed binding site motif present in the promoter of *At2g33710*, during multiple stress responses. The top row shows expression profiles in response to *B. cinerea*; the middle row shows expression profiles in response to *P. syringae* DC3000; the bottom row shows expression profiles in response to long day senescence. Expression profiles of *At2g33710* in each stress response is shown in red, with the interactor expression profile (the name of which is given as the title of each plot) is shown in blue.

the most likely candidates for 'real' interaction. However, by observing the expression profiles of *WRKY8*, *WRKY28*, and *WRKY41*, compared to their target, *At2g33710*, it is difficult to tease apart evidence for likely regulation of *At2g33710* in each stress. The expression profiles of all three interactors and *At2g33710* are correlated, and exhibit no time delay in expression, which could infer regulation.

The expression profiles of the Y1H-identified interactors and *At3g53600* during *B. cinerea* infection are difficult to infer regulation from in Figure 5.22. An interesting observation is the expression profile of Y1H-identified interactor *TCP3* with its target *At3g53600* during *P. syringae* DC3000 infection: the expression profiles are highly anti-correlated to each other, suggesting negative regulation of *At3g53600*.

The expression profiles of *MYB108*, and the transcription factors *MYC2*, *GATA14*, and *Rap2.6L*, which were found to directly bind to its promoter in a Y1H screen, are shown in Figure 5.23, in the stresses *B. cinerea* infection, *P. syringae* DC3000 infection and long day senescence. A MYC2 motif (2), and a GATA binding site motif (278) were located in the promoter of *MYB108*, so only the expression profiles of these interactors are shown. *Rap2.6L* was included due to the prediction of its regulating *MYB108*. However, by observing the expression profiles of *MYC2*, *GATA14*, and *Rap2.6L*, compared to their target, *MYB108*, it is difficult to tease apart evidence for likely regulation of *MYB108* in each stress, with the exception of the regulation of *MYB108* by *MYC2* in response to *B. cinerea* infection: the expression of *MYC2* appears to decrease as the expression of *MYB108* increases.

#### 5.4.1.9 Summary of validation of TF-DNA interactions identified by Y1H

As can be seen from Tables 5.10 to 5.15, *Rap2.6L* is only found to bind to the promoter of *MYB108*. However, *TCP4* was found to bind to the promoters of the transcription factor gene members tested here, providing an alternative, albeit non-predicted by the network modelling, regulator of the potential regulons. By observing Figure 5.24, we can see that the expression of *TCP4* is down regulated in comparison to the expression of the transcription factor members of the potential regulons. Therefore, *TCP4* may negatively regulate the expression of these genes under non-stress conditions, only to be itself repressed in stress conditions, to allowed the expression of the potential regulons to occur. A potential *TCP4* binding site motif was located in the promoters

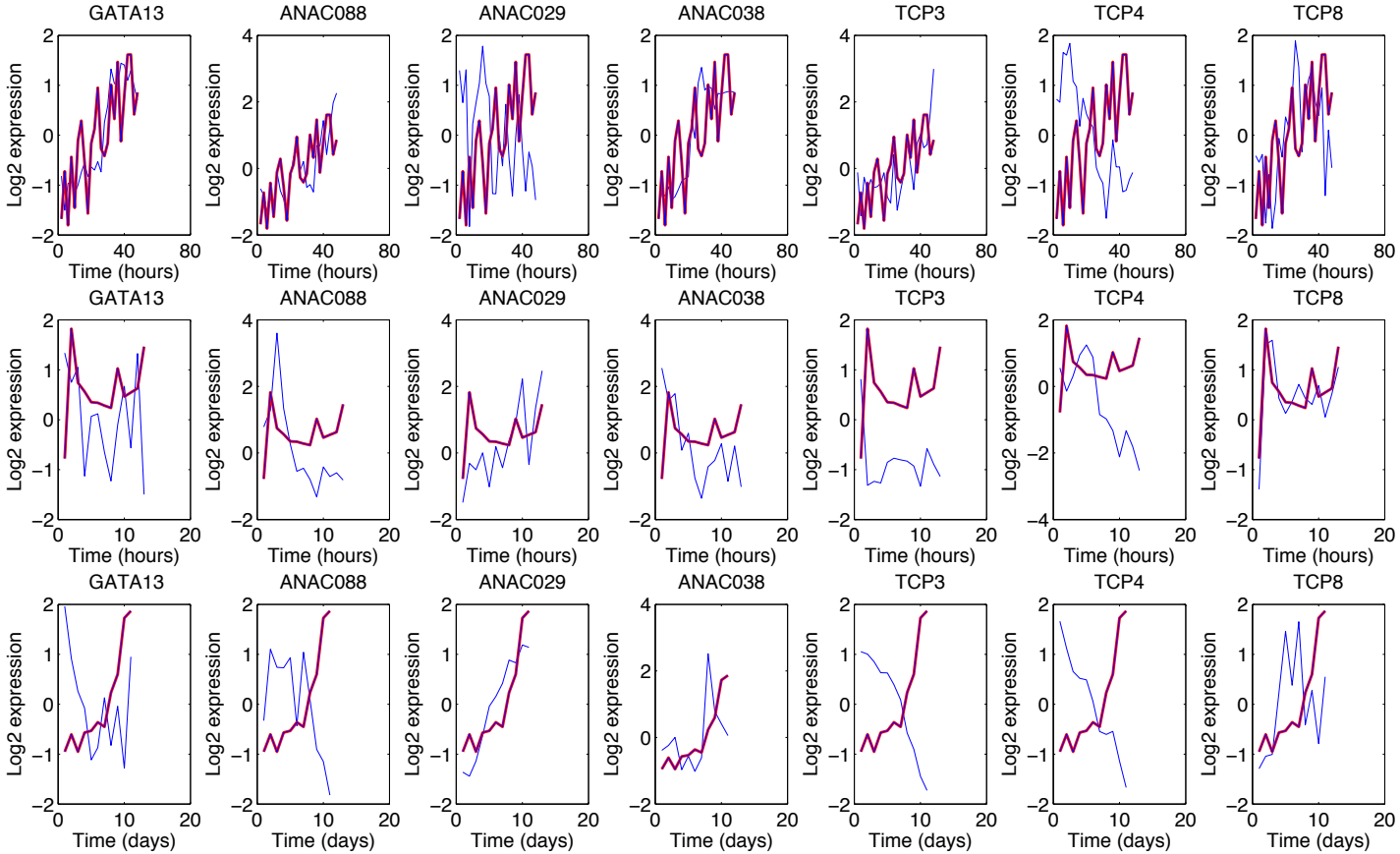



Figure 5.22: Expression profiles of Y1H interactors, that had a confirmed binding site motif present in the promoter of *At3g53600*, during multiple stress responses. The top row shows expression profiles in response to *B. cinerea*; the middle row shows expression profiles in response to *P. syringae* DC3000; the bottom row shows expression profiles in response to long day senescence. Expression profiles of *At3g53600* in each stress response is shown in red, with the interactor expression profile (the name of which is given as the title of each plot) is shown in blue.

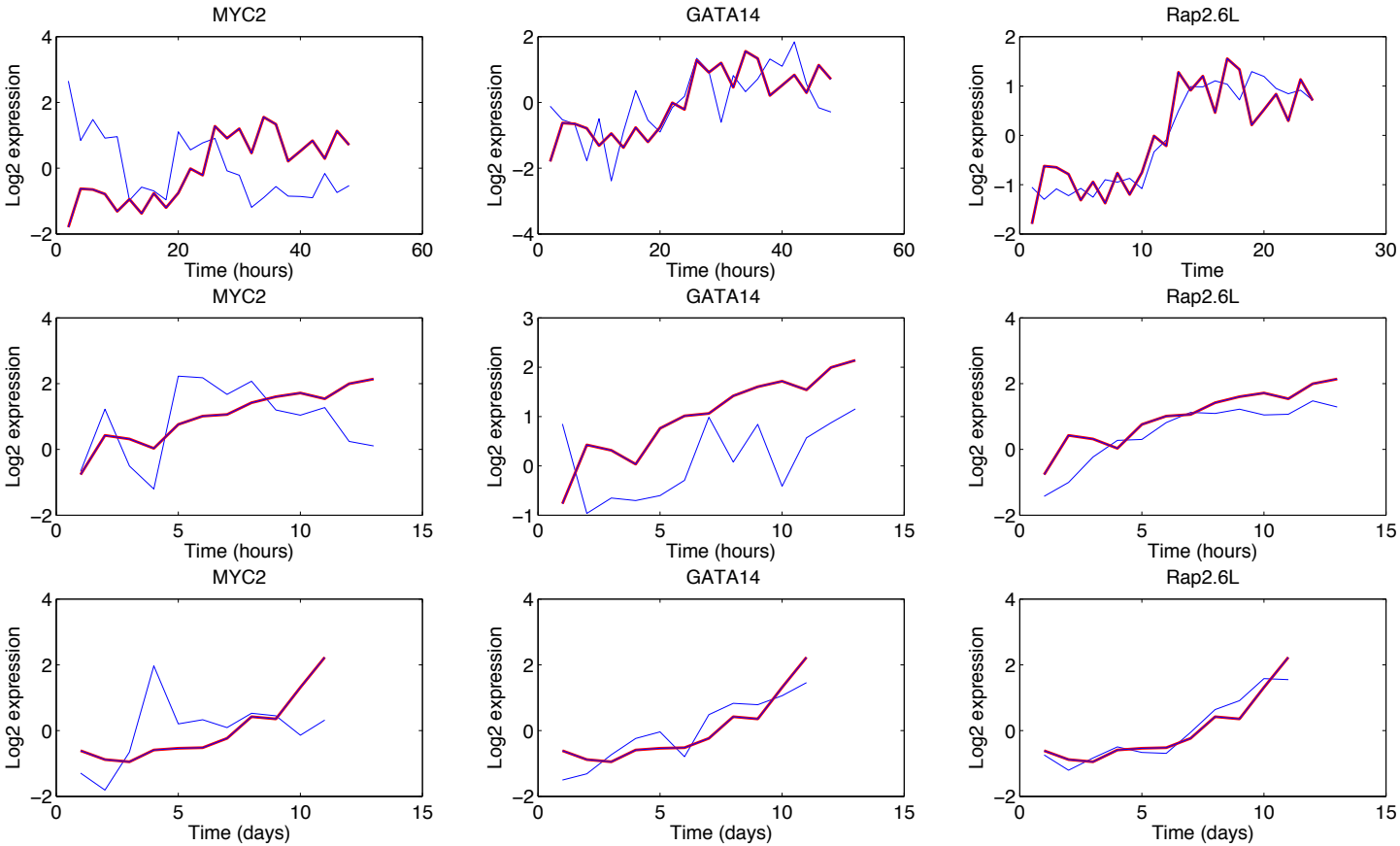



Figure 5.23: Expression profiles of Y1H interactors, that had a confirmed binding site motif present in the promoter of *MYB108*, during multiple stress responses. The top row shows expression profiles in response to *B. cinerea*; the middle row shows expression profiles in response to *P. syringae* DC3000; the bottom row shows expression profiles in response to long day senescence. Expression profiles of *MYB108* in each stress response is shown in red, with the interactor expression profile (the name of which is given as the title of each plot) is shown in blue.

of one transcription member of the potential regulons *At1g71520*, along with five non-transcription factor members of the potential regulons.

*ERF15* was found to bind to the promoters of three of the six transcription factors tested using Y1H. Although there is current literature on a stress response role for *ERF15*, it contains an AP2 domain, and its closest homologs are *ORA59* and *ERF1*, based on sequence similarity (197, 54). Given its ability to bind to promoters of three members of the potential regulons, and a potential binding site significantly overrepresented within the gene members, *ERF15* is a potential regulator of these potential regulons. Figure 5.25 shows the expression profiles of gene members of the potential regulons and that of *ERF15* (shown in red). In *B. cinerea*, and possibly in long day senescence, the slight time delay of the potential regulon gene members (shown in blue) compared to the expression profile of *ERF15* (shown in red) suggests positive transcription regulation of the potential target genes. However, in *P. syringae* pv. DC3000 the opposite in true: the expression of *ERF15* is sharply down regulated within the first few hours of infection, whilst the expression of the potential regulon gene members are upregulated. This suggests that *ERF15* negatively regulates the expression of these genes under non-stress conditions. In *P. syringae* pv. DC3000, however, the expression of *ERF15* appears to be repressed, allowing the expression of the potential regulons to occur.

Y1H provided a context-free, that is to say, stress-free idea of regulation of transcription factor members of potential regulons co-expressed in *B. cinerea* infection, *P. syringae* DC3000 infection and long day senescence. By observing the expression profiles of transcription factors identified by Y1H to directly bind to the promoters of gene members of potential regulons (Table 5.7) during stress responses (in Figures 5.20 to 5.25), it can be seen which of the regulations are still occurring in *B. cinerea* infection, *P. syringae* DC3000 infection and long day senescence. The potential of regulation during the stress response appears to be higher when a binding site motif for the interactor is present within the promoter of the target gene.

#### 5.4.1.10 Integrating Y1H results into the predicted network model

Figure 5.26 details the confirmed edges of the original network described in Figure 5.7. *Rap2.6L* was only identified as directly regulating two transcription factor members, *MYB102* and *MYB108* using Y1H. Other edges have been confirmed using microarray data, which suggests that *Rap2.6L*

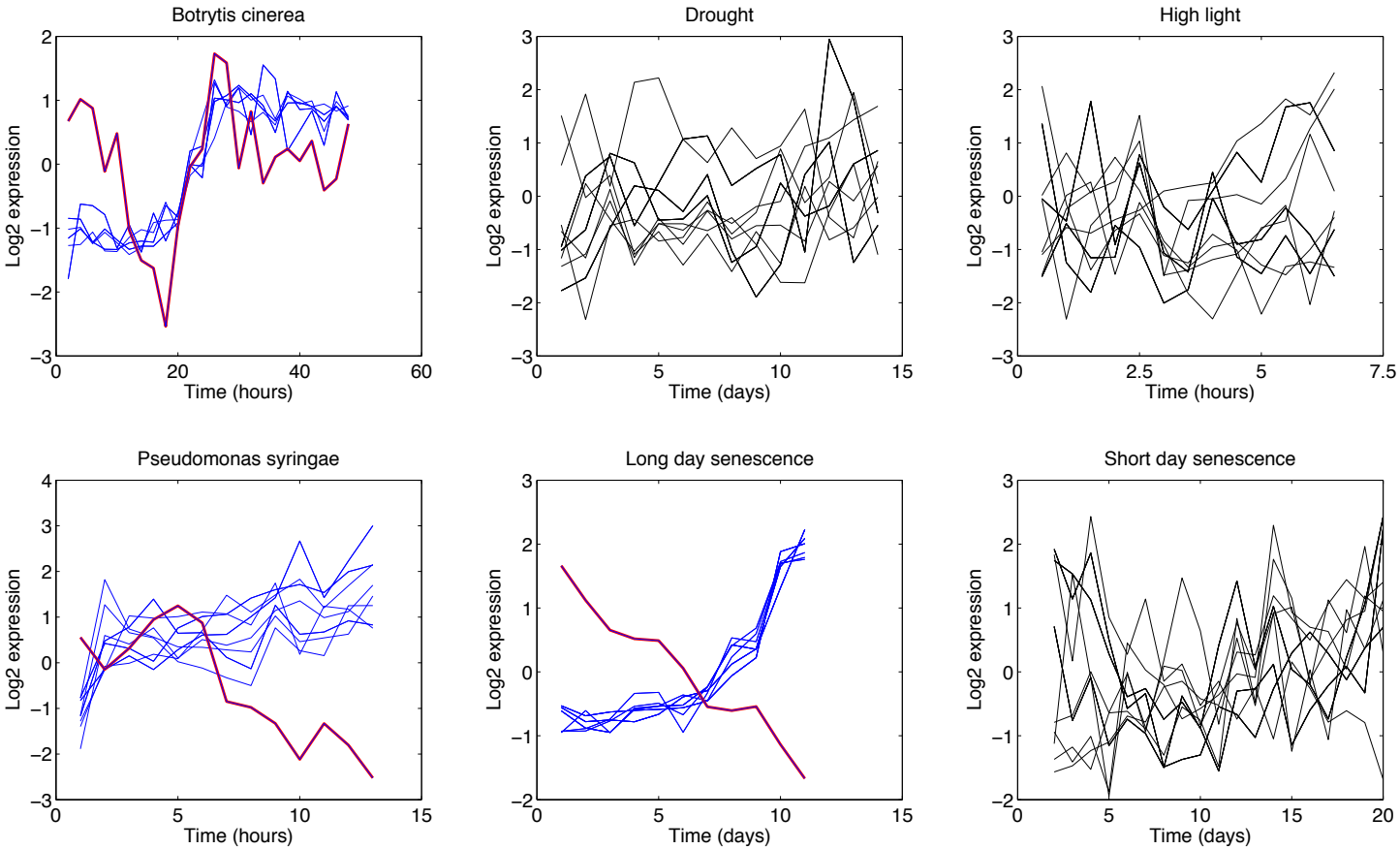



Figure 5.24: Expression profiles of potential regulon gene members and regulator *TCP4*. Y1H has shown *TCP4* (shown in red) to be a common regulator of the transcription factor members of the potential regulons (shown in blue) shown in Table 5.7. Expression profiles shown in black are those of potential regulon members, but are not significantly co-expressed in those particular stress conditions.

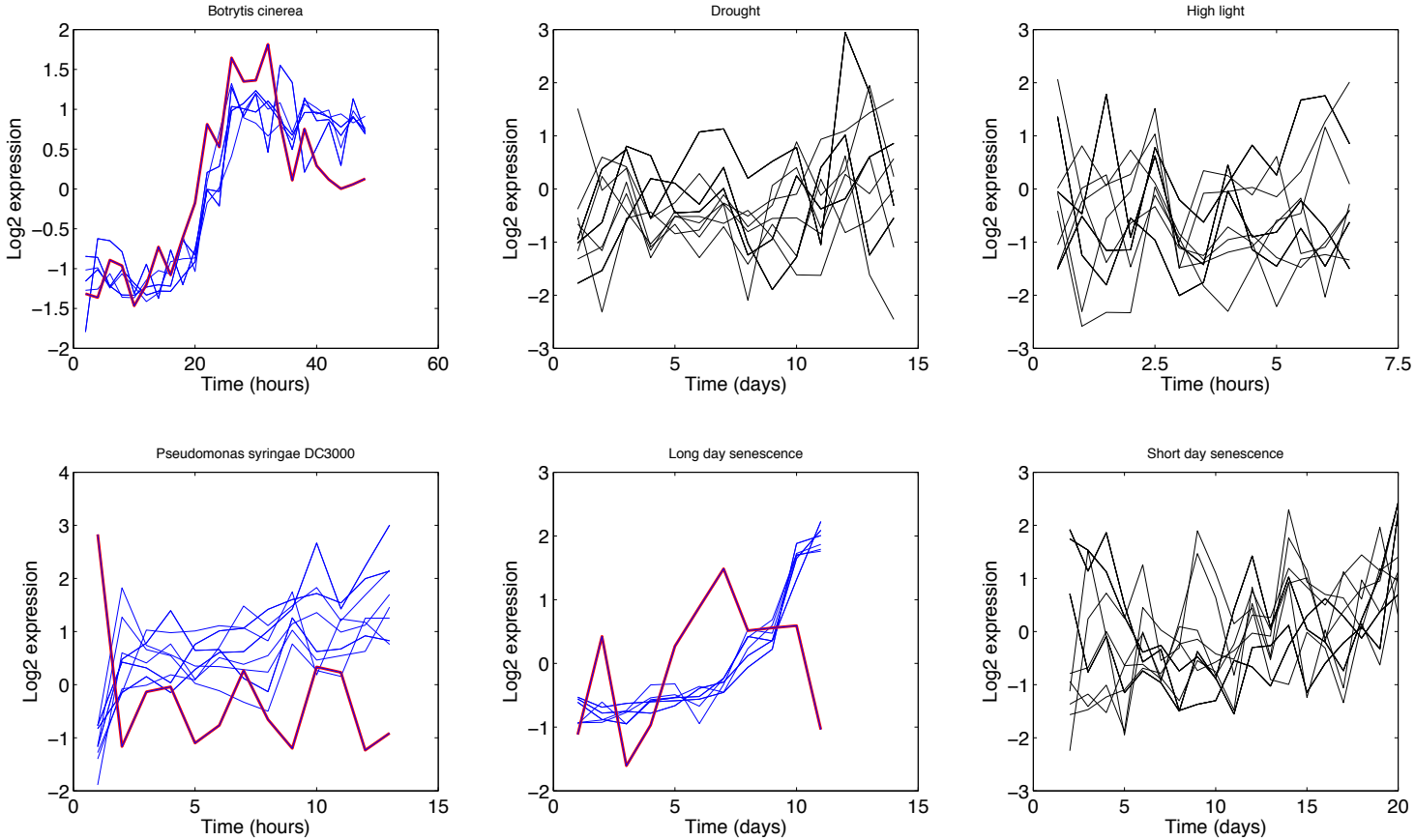



Figure 5.25: Expression profiles of potential regulon gene members and potential common regulator *ERF15*. Y1H has shown *ERF15* (shown in red) to be a common regulator of three transcription factor members of the potential regulons. The expression profiles of these transcription factors, along with other members of the potential regulon that *ERF15* may regulate (shown in Table 5.7) are shown in blue. Expression profiles shown in black are those of potential regulon members, but are not significantly co-expressed in those particular stress conditions.

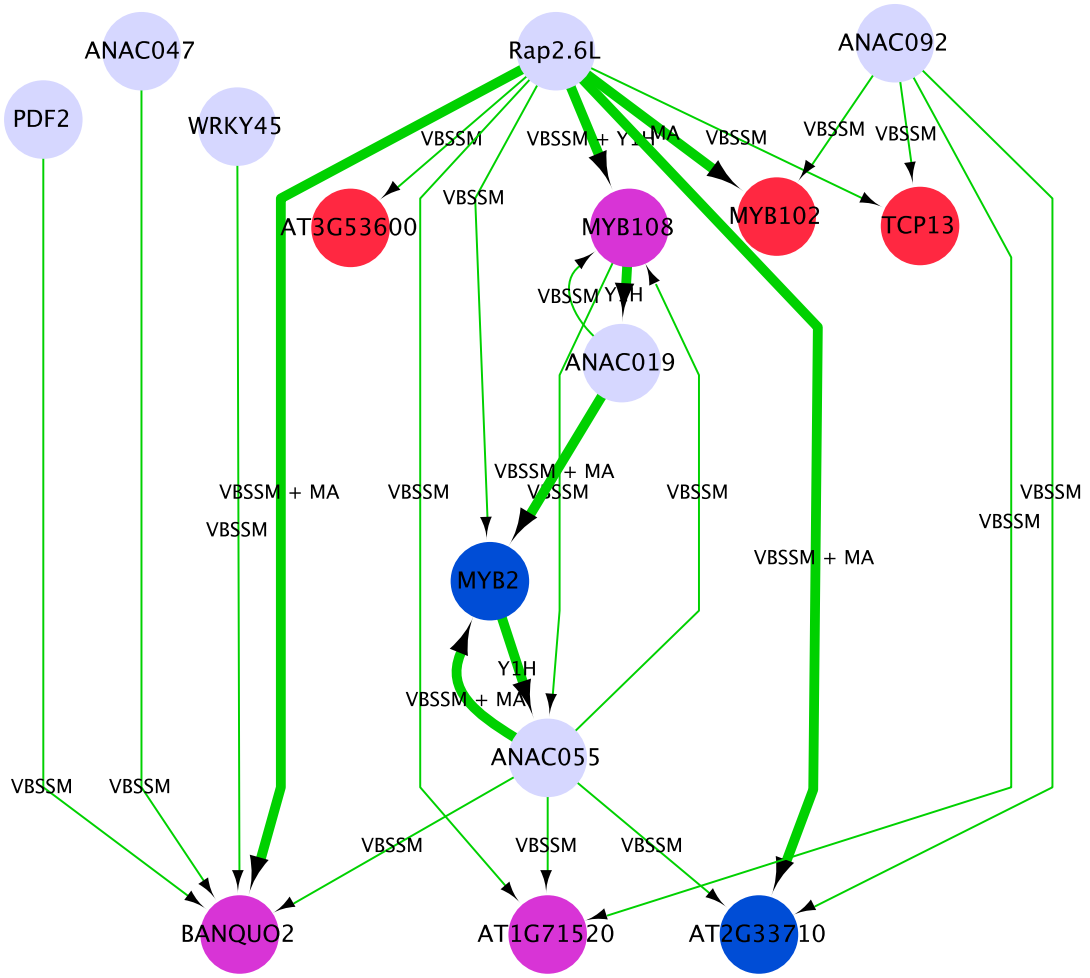



Figure 5.26: Incorporating data acquired from Y1H screens on transcription factor member of potential regulons, with the data summarised in Figure 5.13, to yield a partial gene regulatory network operating in response to multiple stresses. Genes highlighted in blue are co-expressed in *B. cinerea* and *P. syringae* DC3000. Genes highlighted in red are co-expressed in *P. syringae* DC3000 and long day senescence. Genes highlighted in purple are co-expressed in all three stresses. Edges confirmed by array data or Y1H are shown as a thick green line.

could indirectly regulate the potential regulons which are co-expressed in combination of *B. cinerea* infection, long day senescence and *P. syringae* DC3000 infection.

If transcription factors identified as binding to the potential regulon members *MYB2*, *MYB108*, *TCP13*, *At1g71520*, *At2g33710* and *At3g53600* by Y1H were differentially expressed in the *rap2.6L* (58) and *35S:Rap2.6L* array data, these interactors could be the intermediates that allow *Rap2.6L* to indirectly regulate the potential regulons. Table 5.16 shows the interactors, which were identified in Y1H screens as directly binding to the promoters of potential regulon gene members, which are differentially expressed in the *rap2.6L* array (58).

Table 5.16: Identifying genes differentially expression in *rap2.6L* array (58), which were also found to directly bind to the promoters of transcription factor members of potential regulons

| Interactor identified as differentially expressed | Interaction promoter |
|---|---|
| *TCP5* | *At1g71520* |
| *TCP5* | *At2g33710* |
| *At1g63040* | *At3g53600* |
| *TCP5* | *TCP13* |
| *TCP10* | *MYB2* |
| *TCP15* | *MYB108* |

As can be seen from Table 5.16, all the transcription factor members of the potential regulons co-expressed in *B. cinerea* infection, long day senescence and *P. syringae* DC3000 stresses (Table 5.7) are seen to be indirectly regulated by *Rap2.6L*, by a transcription factor which was identified via Y1H. This data has been combined with the original predicted network shown in Figure 5.7, along with edges confirmed by Y1H and microarray data, to yield a final network operating in response to multiple plant stresses, as shown in Figure 5.27.

The network shown in Figure 5.27 incorporates the finding in Table 5.16, which shows *Rap2.6L* indirectly regulates transcription factor members of potential regulons via *TCP5*, *TCP10*, *TCP15*, and *At1g63040*, a member of TINY transcription factor family. These four transcription factors were found to be differentially expressed in the *rap2.6L* array (58), suggesting that their expression is regulated by *Rap2.6L*. *TCP5*, *TCP10*, *TCP15*, and *At1g63040* also bound directly to promoters of potential regulon members (as shown in Tables 5.10 to 5.15), highlighting their role as an intermediary between *Rap2.6L* and its predicted targets, the potential regulons originally described in Table 5.7.

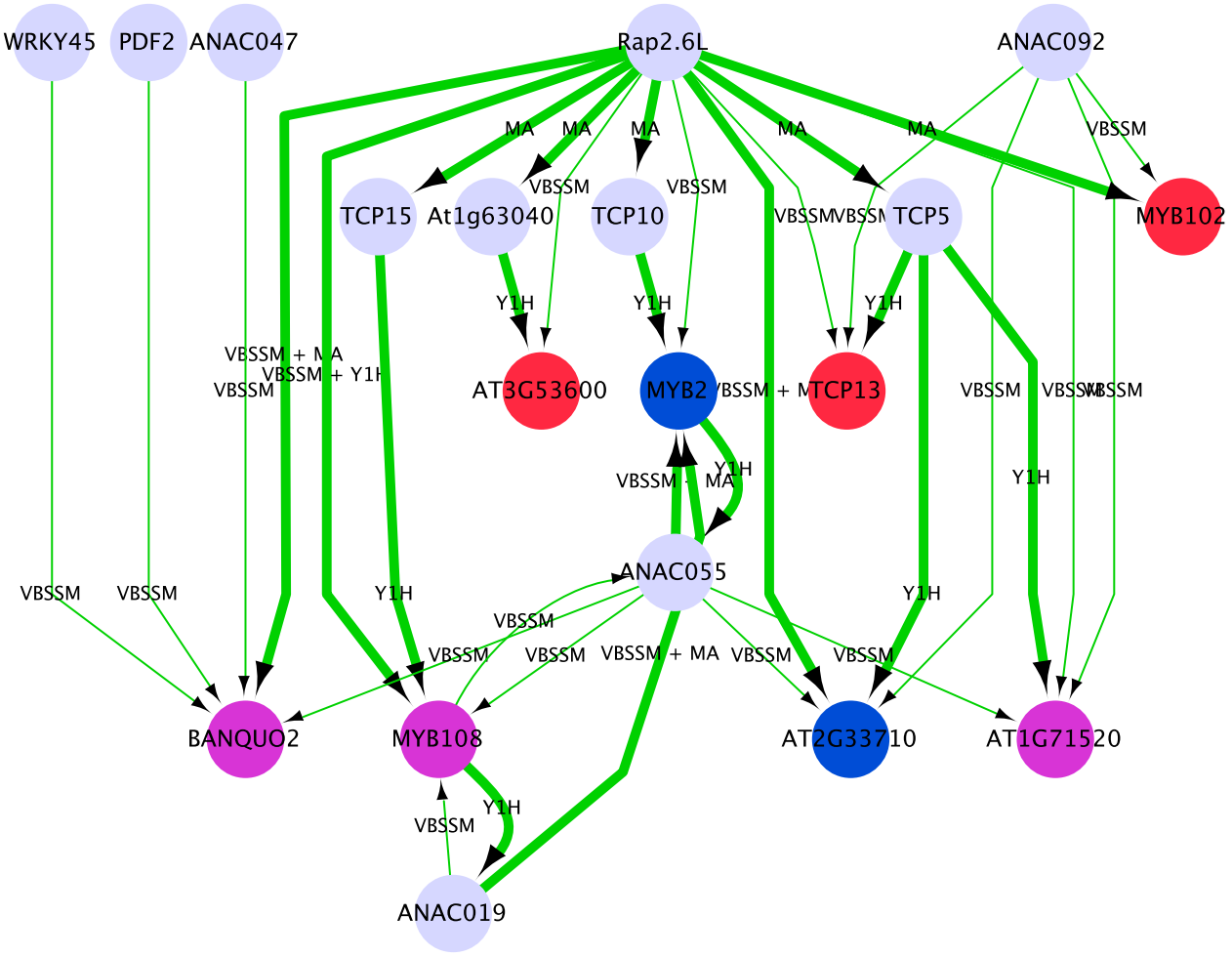



Figure 5.27: A gene regulatory network operating in response to the plant stresses *B. cinerea* infection, long day senescence and *P. syringae* DC3000 infection. This network incorporates the interactions shown in Table 5.16, which shows *Rap2.6L* indirectly regulates transcription factor members of potential regulons via *TCP5*, *TCP10*, *TCP15*, and *At1g63040*. Genes highlighted in blue are co-expressed in *B. cinerea* and *P. syringae* DC3000. Genes highlighted in red are co-expressed in *P. syringae* DC3000 and long day senescence. Genes highlighted in purple are co-expressed in all three stresses. Edges confirmed by array data or Y1H are shown as a thick green line.

## 5.5 Discussion

In this chapter, a network model was tested experimentally using expression and binding assays to confirm transcriptional regulation of genes predicted to be co-expressed in *B. cinerea* infection, *P. syringae* pv. tomato DC3000 infection, and long day senescence. The transcription factor members of three potential regulons co-expressed in combinations of the stresses *B. cinerea* infection, *P. syringae* pv. tomato DC3000 infection, and long day senescence were predicted by modelling to be commonly regulated by *Rap2.6L*. This regulation was tested using publicly available array data (58) using a knockout line of *rap2.6L* to identify differentially expressed genes, and therefore, genes regulated by *Rap2.6L*; discovering differentially expressed genes in a *35S:Rap2.6L* array; and elucidating direct regulators using a 'bottom-up' approach in a Y1H screen of transcription factor members of potential regulons.

The *35S:Rap2.6L* array yielded 94 differentially expressed genes, none of which were predicted targets of *Rap2.6L*. Although, by integrating *rap2.6L* array data (58) and confirmed interactions of potential regulon members from Y1H, it could be seen that *Rap2.6l* indirectly regulates the transcription factor members of the potential regulons through *TCP5*, *TCP10*, *TCP15*, and *At1g63040*.

However, these approaches confirmed the regulation of these potential regulons in a stress-free context. By observing the expression profiles of interactors of potential regulon members identified via Y1H in the stresses *B. cinerea* infection, *P. syringae* pv. tomato DC3000 infection, and long day senescence, one can infer which regulations are real. It was observed that if a binding site motif for the interactor was located in the promoter of the potential regulon member, the regulation was more likely to be real than if the motif was not present.

### 5.5.1 Binding assays detect transcriptional regulators

In Sections 5.4.1.4 and 5.4.1.7, microarray analysis and Y1H were used to confirm predicted edges of transcription regulation inferred by network modelling. *Ra2.6L* was predicted to regulate the expression of potential regulons (Table 5.7) identified by Wigwams as being co-expressed in *B. cinerea*, *P. syringae* pv. DC3000 and long day senescence.

The Y1H screen of the potential regulon members *MYB108*, *TCP13*, *MYB2*, *At3g53600*, *At1g71520* and *At2g33710* identified many different transcription factors capable of binding to their promot-

ers. Members of the TCP transcription factor family were found to interact with all of the promoter fragments tested, suggesting that their role in regulation is ubiquitous. However, transcription factors associated with stress responses were found to interact with specific promoter fragments, such as members of the ERF transcription factor family binding to single promoter fragments (the exception being *ERF15* and *ERF12* which bound to three and two fragments respectively). This data provides specific information about direct gene regulation that may be involved in stress responses for the genes *MYB108*, *TCP13*, *MYB2*, *At3g53600*, *At1g71520* and *At2g33710*.

A novel protein-DNA interaction of *TCP13* binding to its own promoter is strengthened by the literature: it was thought that TCP family members could form homo- or heterodimers via one of two consensus binding sequences (160). Using interactome data generated in the *A. thaliana* using Y2H, *TCP13* was also found to form a homodimer (15, 213). However, literature on exactly which TCP family members form dimers is scarce.

### 5.5.2 *TCP4* and *Rap2.6L* are common regulators of genes important to the response to multiple plant stresses.

Combining microarray completed on *35S:Rap2.6L* with publicly available array data on *rap2.6L* confirmed the regulation of fifteen predicted targets of *Rap2.6L*, out of thirty unique gene members across the three potential regulons, shown in Table 5.7. The Y1H screen only confirmed the regulation of *MYB108* by *Rap2.6L* which was predicted by the network modelling. Since the modelling predicts both direct and indirect regulation, the Y1H screen may not have identified the potential regulons as direct targets of *Rap2.6L* due to *Rap2.6L* being an indirect regulator. This indirect regulation of the potential regulons by *Rap2.6L* was confirmed when transcription factors identified as directly binding to promoters of potential regulon members were differentially expressed in the *rap2.6L* array data (58). This indirect regulation was summered in Figure 5.27.

The Y1H screen identified *TCP4* as a common regulator of six transcriptional factor members of the three potential regulons: *MYB108*, *TCP13*, *MYB2*, *At3g53600*, *At1g71520* and *At2g33710*. The presence of a preferred binding site for *TCP4*, with the sequence (G)GACCAC (253), further adds to the likelihood of *TCP4* potential being a common regulator of these regulons. Further testing of regulon members by Y1H, to confirm direct regulation, or microarray analysis of *TCP4* mutants would confirm this.

Figure 5.27 shows the final gene regulatory network predicted to be involved in the response to *B. cinerea*, *P. syringae* DC3000 and long day senescence. The network presents a complex set of transcriptional regulation, with microarray analysis also confirming the co-regulation of potential regulon members by *ANAC055* and *ANAC019* also. This network may also potentially be involved in hormone signalling pathways: *Rap2.6L* and *ERF15*, members of the ERF family, are responsive to ethylene; *ANAC019* and *ANAC055* have been shown to be induced by methyl-jasmonate (46); and a number of WRKY transcription factors have been shown to have altered expression upon treatment of *A. thaliana* plants with methyl-jasmonate (197).

### 5.5.3 Conclusion

Binding assays, such as Y1H, and microarrays are 'bottom up' and 'top down' approaches, respectively, and have the capability of providing a detailed view of transcriptional regulation. Microarrays allow the detection of direct and indirect targets, whilst Y1H screens allow the identification of direct regulators of genes of interest. However, Y1H does not allow the investigation of direct binding under stress conditions; therefore, the regulation on the potential regulons in Table 5.7 may alter during the stress response. Microarrays allow the researcher to study transcriptional regulation in the context of stress, providing a supportive method to Y1H to study regulation both in terms of stress and non-stress responses.

*Rap2.6L* was identified as an indirect regulator of the transcription factor members of three potential regulons co-expressed in the stresses *B. cinerea*, *P. syringae* pv. tomato DC3000 and long day senescence: the transcription factors *TCP5*, *TCP10*, *TCP15* and *At1g63040*, a members of the TINY family, were identified as differentially expression in the *rap2.6L* array data (58), and were also seen to directly bind to the promoters of the potential regulon members *TCP13*, *At3g53600*, *MYB2*, *MYB108*, *At1g71520* and *At2g33710*.

Microarray analysis also confirmed the regulation of ten out of 23 non-transcription factor members of the three potential regulons also. Although, due to the indirect regulation of the transcription factor members by *Rap2.6L*, it is likely that the regulation of the non-transcription factor members are also indirectly regulated by *Rap2.6L* also.

In light of these results, it can be concluded that Wigwams contributes greatly to the transcription factor-only modelling approach used: the non-transcription factor targets of *Rap2.6L*, although

they may be indirect targets, were not predicted by the modelling. Given the confirmation of the indirect regulation of the transcription factor members of the potential regulons by *Rap2.6L*, Wigwams is a computational less-intensive approach which can be used to incorporate the non-transcription factor targets of genes included in the consensus models.

Although the basis of this work was to identify a common regulator of the potential regulons, the elucidation of a common *indirect* regulator is not a disadvantage against Wigwams: the consensus models generated were used to infer regulation of transcription factor members of potential regulons. Due to the modelling approach not distinguishing between direct and indirect regulation, this was not discovered until further experimental analysis was completed.

# Chapter 6

# General conclusions

The overall aim of the research carried out within this thesis was to generate and validate a gene regulatory network operating in multiple plant stress responses. Here, this and other aims will be addressed to provide conclusions on whether they have been successfully met.

## 6.1 Genome-wide inference of shared regulatory mechanisms from multiple gene expression time-series

Due to the shortcomings of other predictive methods, such as modelling, standard clustering and existing biclustering methods, Wigwams was developed to provide clear statistical evidence for groups of co-expressed genes that were hypothesised to have a common regulatory mechanism. The development and results of Wigwams will be summarised below.

### 6.1.1 Wigwams

Wigwams was developed to provide evidence that a core gene regulatory network could exist, by inferring co-expression and co-regulation across subsets of high resolution time series datasets generated under multiple stress conditions in *A. thaliana*.

Benchmarking exercises were completed to confirm Wigwams is capable of uncovering known examples of co-expression and co-regulation from the literature. Wigwams analysed the publicly available AtGenExpress array data (152), which investigated the changes in gene expression in

response to multiple environmental stresses, including heat, cold, drought, genotoxic, salt, oxidative, osmotic stress, UV-B light and wounding, to identify known examples of co-expression and co-regulation. Wigwams identified the Tryptophan-metabolishing genes *CYP79B2*, *CYP79B3* and *CYP83B1* as co-expressed and potentially sharing a common regulatory mechanism. Indeed, mutant analysis of *ATR1* demonstrated the CYP genes were targets of this MYB family transcription factor (53). Wigwams also correctly identified components of the flavonoid biosynthesis pathway, *CHS*, *CHI*, *F3H* and *FLS* as co-expressed in the response to UV-B, heat, and oxidative stress. Literature has identified two possible transcription factors, *ANAC078* and *MYB12*, as common regulators of these genes, further substantiating Wigwams as a tools for discovering co-expressed genes in subsets of stresses which are hypothesised to have a share regulatory mechanism. Wigwams was applied to multiple time series expression datasets of genes differentially expressed over time in response to two or more of the following conditions: *B. cinerea* infection (322); drought stress; high light stress; *P. syringae* pv. tomato DC3000 infection, long day senescence (43); and short day senescence. 465 potential regulons were identified as co-expressed across subsets of stresses. Upon analysis of these potential regulons using GO term analysis tool BiNGO (191), 219 potential regulons were found to have at least one significantly overrepresented GO term. 89 of the 465 potential regulons also had known plant motifs significantly overrepresented within the promoters of the gene members. The results from the GO term and motif analyses show that Wigwams is capable of detecting functionally similar, and therefore co-expressed, and potentially co-regulated groups of genes.

There are many extensions which could be made to improve Wigwams: making the tool available as either a graphical user interface (GUI) or web tool would prove beneficial and more user-friendly, especially to pure biologists. However, making the tool available on the web could prove difficult, as the website would be required to hold multiple datasets of time series gene expression data. As mentioned previously, a method for generating $p$-values for three or more sets of genes needs to be integrated into the Wigwams code to make this automated. Currently, Wigwams uses Pearsons's correlation as a similarity measure to find groups of co-expressed genes. This method, however, is interchangeable with other methods, such as a time-course model. Wigwams could also be altered to detect time delayed correlation, in order to infer transcriptional regulation across subsets of stresses.

These extensions would serve to improve Wigwams. However, Wigwams in its current form does provide a valuable contribution to the field of research into crosstalk over multiple stresses in plants.

### 6.1.2 Wigwams is based on identifying similarly expressed genes, which are likely to be functionally similar

The basis of Wigwams is identifying genes with similar expression profiles, which are likely to be functionally similar and share a common regulatory mechanism. Genes with similar functions and similar expression profiles are hypothesised to be regulated by a common mechanism, via a *cis*--regulatory element in their promoter (9). However, a study by Allocco *et al*. discovered that this only occurs at high levels of expression similarity, and calculated that for co-expressed genes to have a greater than 50% chance of being regulated by the same transcription factor, or transcription factors, the correlation between expression profiles must be greater than 0.84 (9).

When Wigwams was applied to the PRESTA time-series datasets (see Chapter 4), although an extensive investigation was carried out to decide the optimum value to use for the correlation cut-off in the pruning algorithm (see Section 4.2.2 of Chapter 4), ultimately, the value chosen was 0.75. This value is considerably less than advised by Allocco *et al*.. Therefore, some of the potential regulons identified by Wigwams are less likely of being co-regulated than potential regulons whose intra-condition correlation values of gene expression profiles are higher than 0.84. By re-analysing the PRESTA datasets using Wigwams, and using the advised correlation cut-off of 0.84, although fewer potential regulons will be identified post-pruning, these potential regulons are more likely to truly be co-regulated.

### 6.1.3 The importance of correlation methods

An assumption of Wigwams is that similarly expressed genes are the consequence of an underlying biological process, namely a common regulatory mechanism, and that co-expression of a set of genes is fundamental to that process (334). Yona *et al*. completed a study to determine which of the different similarity measures was most effective for detecting functional links: it was found that combining similarity measures (especially Euclidean and Pearson correlation) performed better than distance metrics used on their own (334).

Due to design of Wigwams, different distance metrics for identifying similarly expressed genes can be used in place of Pearson correlation (which was used in this application of Wigwams). Also, correlation-based methods can be substituted for time-series clustering, such as the time-series model used in SplineCluster (118): the SplineCluster algorithm uses Bayesian model-based hierarchical clustering to investigate mechanisms of regulation by fitting splines to expression profiles in time series datasets, and takes time dependancies of the observations into account, unlike standard correlation-based methods (118).

## 6.2 Experimental validation of a predicted gene regulatory network during multiple stress responses

Results from Wigwams were integrated with predicted network models of multiple stress responses, and motif analysis results, which identify transcription factor binding sites for regulators of potential regulons.

### 6.2.1 Microarray analysis

In Chapter 5 microarray analysis of *A. thaliana* plants over expressing the gene encoding the ERF transcription factor family member *Rap2.6L*, which has been identified as being involved in the response to drought and salt stress (163). *Rap2.6L* was predicted to be a common regulator of transcription factor members of potential regulons co-expressed in the *B. cinerea* infection, *P. syringae* pv. tomato DC3000 infection, and long day senescence stress responses.

The microarray analysis identified 94 differentially expressed genes in response to the over expression of *Rap2.6L*. However, none of the predicted targets of *Rap2.6L* were differentially expressed. A microarray analysis on an *A. thaliana* transgenic *rap2.6L* line was carried out by Che *et al.* (58), does identify 13 of the potential regulon members as targets of *Rap2.6L*.

The modelling approach used to predict the regulation of the potential regulons does not distinguish between direct and indirect interactions, however. Therefore, *Rap2.6L* may be a direct or indirect regulator of the potential regulons.

### 6.2.2 Yeast 1-Hybrid

In Chapter 5 Yeast 1-Hybrid (Y1H) was used to confirm the binding of transcription factors to the promoters of potential regulon gene members. Transcription factors were predicted to bind to potential regulon gene members using an extended version of the variational Bayesian state space modelling approach, which was used to generate a consensus model of transcriptional regulation of the response to each of the following stresses: *B. cinerea* infection (Windram *et al.* (322)); drought stress; high light stress; *P. syringae* pv. tomato DC3000 infection, long day senescence (43); and short day senescence. These models utilised the time series gene expression data generated in response to each of these stress by the PRESTA project group.

Y1H identified that *TCP4* was a direct regulator of the transcription factors *MYB108*, *TCP13*, *MYB2*, *At3g53600*, *At1g71520* and *At2g33710*. To confirm these interactions *in plant*, the binding of *TCP4* to the promoters of these genes could be tested using ChIP, or during the *B. cinerea* infection, *P. syringae* pv. tomato DC3000 infection, or long day senescence, to confirm the co-regulation of these genes during multiple plant stress responses.

By incorporating the Y1H and *rap2.6L* array data, *Rap2.6L* was identified as an indirect regulator of the transcription factor members of three potential regulons co-expressed in the stresses *B. cinerea*, *P. syringae* pv. tomato DC3000 and long day senescence: the transcription factors *TCP5*, *TCP10*, *TCP15* and *At1g63040*, a members of the TINY family, were identified as differentially expression in the *rap2.6L* array data (58), and were also seen to directly bind to the promoters of the potential regulon members *TCP13*, *At3g53600*, *MYB2*, *MYB108*, *At1g71520* and *At2g33710*.

Microarray analysis also confirmed the regulation of ten out of 23 non-transcription factor members of the three potential regulons also. Although, due to the indirect regulation of the transcription factor members by *Rap2.6L*, it is likely that the regulation of the non-transcription factor members are also indirectly regulated by *Rap2.6L* also.

### 6.2.3 Integration of many data types is required to uncover GRNs

Chapter 5 has shown that the integration of many different sources of data is required to uncover mechanisms of transcriptional regulation: the integration of Wigwams potential regulons, network models, and experimental data from microarrays and Y1H uncovered evidence of transcriptional

regulation. This confirms that researchers today cannot rely on one data analysis tool alone to analyse data, and uncover the complex biological processes occurring in response to stress (273).

However, with the recent maturation of sequencing technology, the incorporation of additional datasets, such as genome sequence and protein-DNA interaction datasets would prove advantageous for the inference of GRNs (119). Microarrays are popular for measuring the abundance of mRNA, however, the data produced can be inherently noisy and subject to a high degree of variability (286). This can be overcome by incorporating real-time quantitative PCR assays to obtain precise transcript levels of genes (119).

## 6.3 Overall conclusions

This thesis presents various bioinformatical, modelling and experimental approaches to investigating the gene regulation events involved in the response to multiple plant stresses in *A. thaliana*. A multi-clustering method, Wigwams, was developed to analyse multiple time series gene expression datasets, to mine for potential regulons, where the gene members were co-expressed across subsets of stresses. These potential regulons were hypothesised to share a regulatory mechanism. Due to the inability to infer transcription regulation using Wigwams, the potential regulons were integrated with predicted network models generated for a number of stress responses in *A. thaliana*. Wigwams and the VBSSM models were complementary: VBSSM was limited by the number of genes which could be modelled, therefore, only differentially expressed transcription factors were used. However, Wigwams has no such limitations, so the potential regulons generated contained both transcription factors and non-transcription factors. From this integration of network models and potential regulons, predicted transcription factor regulators of potential regulons were identified, and a quantitative model of regulation in response to the stress *B. cinerea*, *P. syringae* pv. tomato DC3000 and long day senescence was constructed.

This quantitative model of transcriptional regulation during multiple stress responses was validated experimentally using Yeast 1-Hybrid and microarray analysis. The microarray analysis confirmed that *Rap2.6L* was an indirect regulator of potential regulons co-expressed in the *B. cinerea*, *P. syringae* pv. tomato DC3000 and long day senescence stress responses. Additionally, Y1H identified *TCP4* are a direct regulator of the transcription factor members of the potential regulons. Both the microarray and Y1H analyses proved to work well in combination to validate predicted regulation

with a GRN and explained upon the predicted structure.

# Bibliography

[1] Aarts, N., M. Metz, E. Holub, B. J. Staskawicz, M. J. Daniels, and J. E. Parker (1998, August). Different requirements for EDS1 and NDR1 by disease resistance genes define at least two R gene-mediated signaling pathways in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America 95*(17), 10306–10311.

[2] Abe, H. (2002, December). Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) Function as Transcriptional Activators in Abscisic Acid Signaling. *The Plant cell 15*(1), 63–78.

[3] Abe, H., K. Yamaguchi-Shinozaki, and T. Urao (1997). Role of Arabidopsis MYC and MYB homologs in drought-and abscisic acid-regulated gene expression. *The Plant cell 9*(10), 1859–1868.

[4] Abràmoff, M., P. Magalhães, and S. Ram (2004). Image processing with ImageJ. *Biophotonics international 11*(7), 36–42.

[5] Abuqamar, S., X. Chen, R. Dhawan, B. Bluhm, J. Salmeron, S. Lam, R. A. Dietrich, and T. Mengiste (2006, October). Expression profiling and mutant analysis reveals complex regulatory networks involved in Arabidopsis response to Botrytisinfection. *The Plant Journal 48*(1), 28–44.

[6] Adie, B. A. T., J. Perez-Perez, M. M. Perez-Perez, M. Godoy, J. J. Sanchez-Serrano, E. A. Schmelz, and R. Solano (2007, May). ABA Is an Essential Signal for Plant Resistance to Pathogens Affecting JA Biosynthesis and the Activation of Defenses in Arabidopsis. *The Plant cell 19*(5), 1665–1681.

[7] Agarwal, M. L., W. R. Taylor, M. V. Chernov, O. B. Chernova, and G. R. Stark (1998). The p53 network. *The Journal of biological chemistry 273*(1), 1–4.

[8] Allemeersch, J., S. Durinck, and R. Vanderhaeghen (2005). Benchmarking the CATMA Microarray. A Novel Tool forArabidopsis Transcriptome Analysis. *PLANT PHYSIOLOGY 137*(2), 588–601.

[9] Allocco, D. J., I. S. Kohane, and A. J. Butte (2004, February). Quantifying the relationship between co-expression, co-regulation and gene function. *BMC bioinformatics 5*, 18.

[10] Amirsadeghi, S. and C. A. Robson (2007). The role of the mitochondrion in plant responses to biotic stress. *Physiologia Plantarum 129*(1), 253–266.

[11] Anderson, J. P. (2004, December). 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.

[12] Angelini, C., L. Cutillo, D. De Canditiis, M. Mutarelli, and M. Pensky (2008). BATS: a Bayesian user-friendly software for analyzing time series microarray experiments. *BMC bioinformatics 9*, 415.

[13] Apel, K. and H. Hirt (2004, June). REACTIVE OXYGEN SPECIES: Metabolism, Oxidative Stress, and Signal Transduction. *Annual Review of Plant Biology 55*(1), 373–399.

[14] Arabidopsis, G. I. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature 408*(6814), 796.

[15] Arabidopsis Interactome Mapping Consortium, M. Dreze, A.-R. Carvunis, B. Charloteaux, M. Galli, S. J. Pevzner, M. Tasan, Y.-Y. Ahn, P. Balumuri, A.-L. Barabasi, V. Bautista, P. Braun, D. Byrdsong, H. Chen, J. D. Chesnut, M. E. Cusick, J. L. Dangl, C. De Los Reyes, A. Dricot, M. Duarte, J. R. Ecker, C. Fan, L. Gai, F. Gebreab, G. Ghoshal, P. Gilles, B. J. Gutierrez, T. Hao, D. E. Hill, C. J. Kim, R. C. Kim, C. Lurin, A. Macwilliams, U. Matrubutham, T. Milenkovic, J. Mirchandani, D. Monachello, J. Moore, M. S. Mukhtar, E. Olivares, S. Patnaik, M. M. Poulin, N. Przulj, R. Quan, S. Rabello, G. Ramaswamy, P. Reichert, E. A. Rietman, T. Rolland, V. Romero, F. P. Roth, B. Santhanam, R. J. Schmitz, P. Shinn, W. Spooner, J. Stein, G. M. Swamilingiah, S. Tam, J. Vandenhaute, M. Vidal, S. Waaijers, D. Ware, E. M. Weiner, S. Wu, and J. Yazaki (2011, July). Evidence for Network Evolution in an Arabidopsis Interactome Map. *Science 333*(6042), 601–607.

[16] Aukerman, M. J. (2003, October). Regulation of Flowering Time and Floral Organ Identity by a MicroRNA and Its APETALA2-Like Target Genes. *The Plant cell 15*(11), 2730–2741.

[17] Babu, M. M., N. M. Luscombe, L. Aravind, M. Gerstein, and S. A. Teichmann (2004, June). Structure and evolution of transcriptional regulatory networks. *Current Opinion in Structural Biology 14*(3), 283–291.

[18] Bailey, T. L., M. Boden, F. A. Buske, and M. Frith (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic acids research 37*(suppl 2), W202–W208.

[19] Bailey, T. L. and C. Elkan (2005, April). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, 28–36.

[20] Bailey-Serres, J., R. Sorenson, and P. Juntawong (2009, August). Getting the message across: cytoplasmic ribonucleoprotein complexes. *Trends in plant science 14*(8), 443–453.

[21] Bais, H. P., B. Prithiviraj, A. K. Jha, F. M. Ausubel, and J. M. Vivanco (2005, March). Mediation of pathogen resistance by exudation of antimicrobials from roots. *Nature 434*(7030), 217–221.

[22] Balazadeh, S., H. Siddiqui, A. D. Allu, L. P. Matallana-Ramirez, C. Caldana, M. Mehrnia, M.-I. Zanor, B. Köhler, and B. Mueller-Roeber (2010, January). A gene regulatory network controlled by the NAC transcription factor ANAC092/AtNAC2/ORE1 during salt-promoted senescence. *The Plant Journal 62*(2), 250–264.

[23] Bansal, M., V. Belcastro, A. Ambesi-Impiombato, and D. di Bernardo (2007, February). How to infer gene networks from expression profiles. *Molecular Systems Biology 3*, 1–10.

[24] Bar-Joseph, Z. (2004, October). Analyzing time series gene expression data. *Bioinformatics 20*(16), 2493–2503.

[25] Barabási, A.-L. and Z. N. Oltvai (2004, February). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics 5*(2), 101–113.

[26] Bari, R. and J. D. G. Jones (2008, December). Role of plant hormones in plant defence responses. *Plant Molecular Biology 69*(4), 473–488.

[27] Barski, A., S. Cuddapah, K. Cui, T.-Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao (2007, May). High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell 129*(4), 823–837.

[28] Baxter, L., A. Jironkin, R. Hickman, J. Moore, C. Barrington, P. Krusche, N. P. Dyer, V. Buchanan-Wollaston, A. Tiskin, J. Beynon, K. Denby, and S. Ott (2012, October). Conserved Noncoding Sequences Highlight Shared Components of Regulatory Networks in Dicotyledonous Plants. *The Plant cell 24*(10), 3949–3965.

[29] Beal, M. (2003, July). Variational Bayesian Linear Dynamical Systems. *Thesis*, 1–47.

[30] Beal, M., F. Falciani, Z. Ghahramani, C. Rangel, and D. Wild (2005). A Bayesian approach to reconstructing genetic regulatory networks with hidden factors. *Bioinformatics 21*(3), 349.

[31] Becker, B., S. Holtgrefe, S. Jung, C. Wunrau, A. Kandlbinder, M. Baier, K.-J. Dietz, J. E. Backhausen, and R. Scheibe (2006, January). Influence of the photoperiod on redox regulation and stress responses in Arabidopsis thaliana L. (Heynh.) plants under long- and short-day conditions. *Planta 224*(2), 380–393.

[32] Becquet, C., S. Blachon, B. Jeudy, J.-F. Boulicaut, and O. Gandrillon (2002). Strong-association-rule mining for large-scale gene-expression data analysis: a case study on human SAGE data. *Genome Biology 3*(12), 1–16.

[33] Beelman, C. A. and R. Parker (1995). Degradation of mRNA in eukaryotes. *Cell 81*(2), 179–183.

[34] Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological) 57*(1), 289–300.

[35] Berger, B., R. Stracke, R. Yatusevich, B. Weisshaar, U.-I. Flügge, and T. Gigolashvili (2007, April). A simplified method for the analysis of transcription factor-promoter interactions that allows high-throughput data generation. *The Plant Journal 50*(5), 911–916.

[36] Berri, S., P. Abbruscato, O. Faivre-Rampant, A. C. Brasileiro, I. Fumasoni, K. Satoh, S. Kikuchi, L. Mizzi, P. Morandini, M. Pè, and P. Piffanelli (2009). Characterization of WRKY co-regulatory networks in rice and Arabidopsis. *BMC Plant Biology 9*(1), 2–22.

[37] Berrocal-Lobo, M., A. Molina, and R. Solano (2002, January). Constitutive expression of ETHYLENE-RESPONSE-FACTOR1 in Arabidopsis confers resistance to several necrotrophic fungi. *The Plant journal : for cell and molecular biology 29*(1), 23–32.

[38] Birkenbihl, R. P., C. Diezel, and I. E. Somssich (2012, May). Arabidopsis WRKY33 is a key transcriptional regulator of hormonal and metabolic responses toward Botrytis cinerea infection. *PLANT PHYSIOLOGY 159*(1), 266–285.

[39] Blackwell, T. S. and J. W. Christman (1997). The role of nuclear factor-kappa B in cytokine gene regulation. *American journal of respiratory cell and molecular biology 17*(1), 3–9.

[40] Boller, T. and S. Y. He (2009, May). Innate Immunity in Plants: An Arms Race Between Pattern Recognition Receptors in Plants and Effectors in Microbial Pathogens. *Science 324*(5928), 742–744.

[41] Bostock, R. M. (2005, September). Signal Crosstalk and Induced Resistance: Straddling the Line Between Cost and Benefit. *Annual Review of Phytopathology 43*(1), 545–580.

[42] Bradbury, J. F. (1986). *Guide to plant pathogenic bacteria*. CAB International.

[43] Breeze, E., E. Harrison, S. Mchattie, L. Hughes, R. Hickman, C. Hill, S. Kiddle, Y.-S. Kim, C. A. Penfold, D. Jenkins, C. Zhang, K. Morris, C. Jenner, S. Jackson, B. Thomas, A. Tabrett, R. Legaie, J. D. Moore, D. L. Wild, S. Ott, D. Rand, J. Beynon, K. Denby, A. Mead, and V. Buchanan-Wollaston (2011). High-Resolution Temporal Profiling of Transcripts during Arabidopsis Leaf Senescence Reveals a Distinct Chronology of Processes and Regulation. *The Plant cell 23*(3), 873–894.

[44] Brown, R. L. (2003, May). A Role for the GCC-Box in Jasmonate-Mediated Activation of the PDF1.2 Gene of Arabidopsis. *PLANT PHYSIOLOGY 132*(2), 1020–1032.

[45] Bruce, W. B., G. O. Edmeades, and T. C. Barker (2002, January). Molecular and physiological approaches to maize improvement for drought tolerance. *Journal of Experimental Botany 53*(366), 13–25.

[46] Bu, Q., H. Jiang, C.-B. Li, Q. Zhai, J. Zhang, X. Wu, J. Sun, Q. Xie, and C. Li (2008). Role of the Arabidopsis thaliana NAC transcription factors ANAC019 and ANAC055 in regulating jasmonic acid-signaled defense responses. *Cell Research 18*(7), 756–767.

[47] Buchanan-Wollaston, V., T. Page, E. Harrison, E. Breeze, P. O. Lim, H. G. Nam, J.-F. Lin, S.-H. Wu, J. Swidzinski, K. Ishizaki, and C. J. Leaver (2005, May). Comparative transcriptome analysis reveals significant differences in gene expression and signalling pathways between developmental and dark/starvation-induced senescence in Arabidopsis. *The Plant Journal 42*(4), 567–585.

[48] Buck, M. J. and J. D. Lieb (2004, March). ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics 83*(3), 349–360.

[49] Busygin, S., O. Prokopyev, and P. M. Pardalos (2008, September). Biclustering in data mining. *Computers & Operations Research 35*(9), 2964–2987.

[50] Camon, E. (2003, March). The Gene Ontology Annotation (GOA) Project: Implementation of GO in SWISS-PROT, TrEMBL, and InterPro. *Genome Research 13*(4), 662–672.

[51] Camon, E. (2004, January). The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic acids research 32*(90001), 262D–266.

[52] Carmona-Saez, P., R. D. Pascual-Marqui, F. Tirado, J. M. Carazo, and A. Pascual-Montano (2006). Biclustering of gene expression data by non-smooth non-negative matrix factorization . *BMC bioinformatics 7*(1), 78.

[53] Celenza, J. L. (2005, January). The Arabidopsis ATR1 Myb Transcription Factor Controls Indolic Glucosinolate Homeostasis. *PLANT PHYSIOLOGY 137*(1), 253–262.

[54] Cevik, V., B. N. Kidd, P. Zhang, C. Hill, S. Kiddle, K. J. Denby, E. B. Holub, D. M. Cahill, J. M. Manners, P. M. Schenk, J. Beynon, and K. Kazan (2012). MED25 Acts as an Integrative Hub for the Regulation of Jasmonate-Responsive Gene Expression in Arabidopsis. *PLANT PHYSIOLOGY 160*, 541–555.

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

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

[57] Chaves, M. M., J. Flexas, and C. Pinheiro (2008, May). Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. *Annals of Botany 103*(4), 551–560.

[58] Che, P., S. Lall, D. Nettleton, and S. H. Howell (2006, June). Gene expression programs during shoot, root, and callus development in Arabidopsis tissue culture. *PLANT PHYSIOLOGY 141*(2), 620–637.

[59] Chen, H., L. Xue, S. Chintamanani, H. Germain, H. Lin, H. Cui, R. Cai, J. Zuo, X. Tang, X. Li, H. Guo, and J.-M. Zhou (2009). ETHYLENE INSENSITIVE3 and ETHYLENE INSENSITIVE3-LIKE1 repress SALICYLIC ACID INDUCTION DEFICIENT2 expression to negatively regulate plant innate immunity in Arabidopsis. *The Plant cell 21*(8), 2527–2540.

[60] Chen, H., J. Zhang, M. Neff, S. Hong, H. Zhang, X. Deng, and L. XIONG (2008). Integration of light and abscisic acid signaling during seed germination and early seedling development. *Proceedings of the National Academy of Sciences of the United States of America 105*(11), 4495.

[61] Chen, L., L. Zhang, and D. Yu (2010). Wounding-induced WRKY8 is involved in basal defense in Arabidopsis. *Molecular Plant-Microbe Interactions 23*(5), 558–565.

[62] Chen, T., H. L. He, and G. Church (1999). Modeling gene expression with differential equations. *Pacific Symposium on Biocomputing*. *Pacific Symposium on Biocomputing*, 29–40.

[63] Chen, W. (2002, March). Expression Profile Matrix of Arabidopsis Transcription Factor Genes Suggests Their Putative Functions in Response to Environmental Stresses. *The Plant cell 14*(3), 559–574.

[64] Chen, W. J. and T. Zhu (2004, December). Networks of transcription factors with roles in environmental stress response. *Trends in plant science 9*(12), 591–596.

[65] Cheng, Y. and G. Church (2000). Biclustering of expression data. *Proceedings of the eighth international conference on intelligent systems for molecular biology 1*, 93–103.

[66] Cheong, Y. H., H.-S. Chang, R. Gupta, X. Wang, T. Zhu, and S. Luan (2002, June). Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in Arabidopsis. *PLANT PHYSIOLOGY 129*(2), 661–677.

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

[68] Cho, R., M. Campbell, E. A. Winzeler, L. Steinmetz, A. Conway, L. Wodicka, T. G. Wolfsberg, A. E. Gabrielian, D. Landsman, and D. J. Lockhart (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. *Molecular cell 2*(1), 65–74.

[69] Choi, H., J. Hong, J. Ha, J. Kang, and S. Y. Kim (2000, January). ABFs, a family of ABA-responsive element binding factors. *The Journal of biological chemistry 275*(3), 1723–1730.

[70] Chu, S. (1998, October). The Transcriptional Program of Sporulation in Budding Yeast. *Science 282*(5389), 699–705.

[71] Clifton, R., R. Lister, K. L. Parker, P. G. Sappl, D. Elhafez, A. H. Millar, D. A. Day, and J. Whelan (2005, May). Stress-induced co-expression of alternative respiratory chain components in Arabidopsis thaliana. *Plant Molecular Biology 58*(2), 193–212.

[72] Cline, M. S., M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, K. Hanspers, R. Isserlin, R. Kelley, S. Killcoyne, S. Lotia, S. Maere, J. Morris, K. Ono, V. Pavlovic, A. R. Pico, A. Vailaya, P.-L. Wang, A. Adler, B. R. Conklin, L. Hood, M. Kuiper, C. Sander, I. Schmulevich, B. Schwikowski, G. J. Warner, T. Ideker, and G. D. Bader (2007, October). Integration of biological networks and gene expression data using Cytoscape. *Nature Protocols 2*(10), 2366–2382.

[73] Consortium, G. O. (2004, January). The Gene Ontology (GO) database and informatics resource. *Nucleic acids research 32*(Database Issue), D258–D261.

[74] Crombach, A. and P. Hogeweg (2008). Evolution of evolvability in gene regulatory networks. *PLoS Computational Biology 4*(7), e1000112.

[75] Cutler, S. (2005, May). Dude, Where's My Phenotype? Dealing with Redundancy in Signaling Networks. *PLANT PHYSIOLOGY 138*(2), 558–559.

[76] De Torres-Zabala, M., W. Truman, M. H. Bennett, G. Lafforgue, J. W. Mansfield, P. Rodriguez Egea, L. Bögre, and M. Grant (2007, February). Pseudomonas syringae pv. tomato hijacks the Arabidopsis abscisic acid signalling pathway to cause disease. *The EMBO Journal 26*(5), 1434–1443.

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

[78] Dela, G., E. Or, R. Ovadia, A. Nissim-Levi, D. Weiss, and M. Oren-Shamir (2003, March). Changes in anthocyanin concentration and composition in 'Jaguar' rose flowers due to transient high-temperature conditions. *Plant Science 164*(3), 333–340.

[79] Dembinsky, D., K. Woll, M. Saleem, Y. Liu, Y. Fu, L. A. Borsuk, T. Lamkemeyer, C. Fladerer, J. Madlung, B. Barbazuk, A. Nordheim, D. Nettleton, P. S. Schnable, and F. Hochholdinger (2007, November). Transcriptomic and proteomic analyses of pericycle cells of the maize primary root. *PLANT PHYSIOLOGY 145*(3), 575–588.

[80] Denby, K., P. Kumar, and D. Kliebenstein (2004). Identification of Botrytis cinerea susceptibility loci in Arabidopsis thaliana. *The Plant Journal 38*(3), 473–486.

[81] Deplancke, B. (2004, October). A Gateway-Compatible Yeast One-Hybrid System. *Genome Research 14*(10b), 2093–2101.

[82] DeRisi, J. L., V. R. Iyer, and P. O. Brown (1997, October). Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. *Science 278*(5338), 680–686.

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

[84] Dixon, R. A. and N. L. Paiva (1995). Stress-induced phenylpropanoid metabolism. *The Plant cell 7*(7), 1085.

[85] Dombrecht, B., G. P. Xue, S. J. Sprague, J. A. Kirkegaard, J. J. Ross, J. B. Reid, G. P. Fitt, N. Sewelam, P. M. Schenk, J. M. Manners, and K. Kazan (2007, July). MYC2 Differentially

Modulates Diverse Jasmonate-Dependent Functions in Arabidopsis. *The Plant cell 19*(7), 2225–2245.

[86] Dopazo, J. and J. M. Carazo (1997, February). Phylogenetic reconstruction using an unsupervised growing neural network that adopts the topology of a phylogenetic tree. *Journal of molecular evolution 44*(2), 226–233.

[87] Dudoit, S., Y. Yang, M. Callow, and T. Speed (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica sinica 12*(1), 111–140.

[88] Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein (1998, December). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America 95*(25), 14863–14868.

[89] Ernst, J., G. J. Nau, and Z. Bar-Joseph (2005, June). Clustering short time series gene expression data. *Bioinformatics 21*(Suppl 1), i159–i168.

[90] Filkowski, J., O. Kovalchuk, and I. Kovalchuk (2004, April). Genome stability of vtc1, tt4, and tt5 Arabidopsis thalianamutants impaired in protection against oxidative stress. *The Plant Journal 38*(1), 60–69.

[91] Flavell, R. B. (2005). Model plants with special emphasis on Arabidopsis thaliana, and crop improvement. *Proceedings of the International Congress "In the Wake of the Double Helix: From the Green Revolution to the Gene Revolution"*, 365–378.

[92] Forcat, S., M. H. Bennett, J. W. Mansfield, and M. R. Grant (2008). A rapid and robust method for simultaneously measuring changes in the phytohormones ABA, JA and SA in plants following biotic and abiotic stress. *Plant methods 4*(1), 16.

[93] Foster, R., T. Izawa, and N. H. Chua (1994, February). Plant bZIP proteins gather at ACGT elements. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology 8*(2), 192–200.

[94] Frey, B. J. and D. Dueck (2007, February). Clustering by passing messages between data points. *Science 315*(5814), 972–976.

[95] Friedman, N., M. Linial, I. Nachman, and D. Pe'er (2000). Using Bayesian networks to analyze expression data. *Journal of computational biology : a journal of computational molecular cell biology 7*(3-4), 601–620.

[96] Fritzke, B. (1994). Growing cell structures—a self-organizing network for unsupervised and supervised learning. *Neural networks 7*(9), 1441–1460.

[97] Fujimoto, S. Y., M. Ohta, A. Usui, H. Shinshi, and M. Ohme-Takagi (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.

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

[99] Fujita, M., Y. Fujita, Y. Noutoshi, F. Takahashi, Y. Narusaka, K. Yamaguchi-Shinozaki, and K. Shinozaki (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.

[100] Fulton, T. M. (2002, July). Identification, Analysis, and Utilization of Conserved Ortholog Set Markers for Comparative Genomics in Higher Plants. *The Plant cell 14*(7), 1457–1467.

[101] Gachon, C. M. M., M. Langlois-Meurinne, Y. Henry, and P. Saindrenan (2005, May). Transcriptional co-regulation of secondary metabolism enzymes in Arabidopsis: functional and evolutionary implications. *Plant Molecular Biology 58*(2), 229–245.

[102] GAZZARRINI, S. (2003, March). Cross-talk in Plant Hormone Signalling: What Arabidopsis Mutants Are Telling Us. *Annals of Botany 91*(6), 605–612.

[103] Gilchrist, M., V. Thorsson, B. Li, A. G. Rust, M. Korb, K. Kennedy, T. Hai, H. Bolouri, and A. Aderem (2006, May). Systems biology approaches identify ATF3 as a negative regulator of Toll-like receptor 4. *Nature 441*(7090), 173–178.

[104] Gilmour, S. J., A. M. Sebolt, M. P. Salazar, and J. D. Everard (2000). Overexpression of the Arabidopsis CBF3transcriptional activator mimics multiple biochemical changes associated with cold acclimation. *PLANT PHYSIOLOGY 124*(2), 1854–1865.

[105] Giraud, E., S. Ng, C. Carrie, O. Duncan, J. Low, C. P. Lee, O. Van Aken, A. H. Millar, M. Murcha, and J. Whelan (2011, January). TCP Transcription Factors Link the Regulation of Genes Encoding Mitochondrial Proteins with the Circadian Clock in Arabidopsis thaliana. *The Plant cell 22*(12), 3921–3934.

[106] Giuliano, G., E. Pichersky, V. S. Malik, M. P. Timko, P. A. Scolnik, and A. R. Cashmore (1988, October). An evolutionarily conserved protein binding sequence upstream of a plant light-regulated gene. *Proceedings of the National Academy of Sciences of the United States of America 85*(19), 7089–7093.

[107] Glazebrook, J. (2005, September). Contrasting Mechanisms of Defense Against Biotrophic and Necrotrophic Pathogens. *Annual Review of Phytopathology 43*(1), 205–227.

[108] Glazebrook, J., W. Chen, B. Estes, H. Chang, C. Nawrath, J. Metraux, T. Zhu, and F. Katagiri (2003). Topology of the network integrating salicylate and jasmonate signal transduction derived from global expression phenotyping. *Plant Journal 34*(2), 1–12.

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

[110] Gómez-Gómez, L. and T. Boller (2002, June). Flagellin perception: a paradigm for innate immunity. *Trends in plant science 7*(6), 251–256.

[111] Gonzalez, N., G. T. Beemster, and D. Inzé (2009, April). David and Goliath: what can the tiny weed Arabidopsis teach us to improve biomass production in crops? *Current opinion in plant biology 12*(2), 157–164.

[112] Guiltinan, M. J., W. R. Marcotte, and R. S. Quatrano (1990, October). A plant leucine zipper protein that recognizes an abscisic acid response element. *Science 250*(4978), 267–271.

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

[114] Guo, H. and J. R. Ecker (2004, February). The ethylene signaling pathway: new insights. *Current opinion in plant biology 7*(1), 40–49.

[115] Haas, B. J., S. Kamoun, M. C. Zody, R. H. Y. Jiang, R. E. Handsaker, L. M. Cano, M. Grabherr, C. D. Kodira, S. Raffaele, T. Torto-Alalibo, T. O. Bozkurt, A. M. V. Ah-Fong, L. Alvarado, V. L. Anderson, M. R. Armstrong, A. Avrova, L. Baxter, J. Beynon, P. C. Boevink, S. R. Bollmann, J. I. B. Bos, V. Bulone, G. Cai, C. Cakir, J. C. Carrington, M. Chawner, L. Conti, S. Costanzo, R. Ewan, N. Fahlgren, M. A. Fischbach, J. Fugelstad, E. M. Gilroy, S. Gnerre, P. J. Green, L. J. Grenville-Briggs, J. Griffith, N. J. Grünwald, K. Horn, N. R. Horner, C.-H. Hu, E. Huitema, D.-H. Jeong, A. M. E. Jones, J. D. G. Jones, R. W. Jones, E. K. Karlsson, S. G. Kunjeti, K. Lamour, Z. Liu, L. Ma, D. Maclean, M. C. Chibucos, H. McDonald, J. McWalters, H. J. G. Meijer, W. Morgan, P. F. Morris, C. A. Munro, K. O'Neill, M. Ospina-Giraldo, A. Pinzón, L. Pritchard, B. Ramsahoye, Q. Ren, S. Restrepo, S. Roy, A. Sadanandom, A. Savidor, S. Schornack, D. C. Schwartz, U. D. Schumann, B. Schwessinger, L. Seyer, T. Sharpe, C. Silvar, J. Song, D. J. Studholme, S. Sykes, M. Thines, P. J. I. van de Vondervoort, V. Phuntumart, S. Wawra, R. Weide, J. Win, C. Young, S. Zhou, W. Fry, B. C. Meyers, P. van West, J. Ristaino, F. Govers, P. R. J. Birch, S. C. Whisson, H. S. Judelson, and C. Nusbaum (2009, September). Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. *Nature 461*(7262), 393–398.

[116] Hanisch, D., A. Zien, R. Zimmer, and T. Lengauer (2002). Co-clustering of biological networks and gene expression data. *Bioinformatics 18 Suppl 1*, S145–54.

[117] Hao, D., M. Ohme-Takagi, and A. Sarai (1998, October). Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. *The Journal of biological chemistry 273*(41), 26857–26861.

[118] Heard, N. A., C. C. Holmes, and D. A. Stephens (2006, March). A Quantitative Study of Gene Regulation Involved in the Immune Response of Anopheline Mosquitoes. *Journal of the American Statistical Association 101*(473), 18–29.

[119] Hecker, M., S. Lambeck, S. Toepfer, E. van Someren, and R. Guthke (2009, April). Gene regulatory network inference: Data integration in dynamic models—A review. *Biosystems 96*(1), 86–103.

[120] Heil, M. and I. T. Baldwin (2002, February). Fitness costs of induced resistance: emerging experimental support for a slippery concept. *Trends in plant science 7*(2), 61–67.

[121] Higham, D. J. and N. J. Higham (2000, August). MATLAB Guide. *SIAM*, 1–302.

[122] Higo, K., Y. Ugawa, M. Iwamoto, and H. Higo (1998, January). PLACE: a database of plant cis-acting regulatory DNA elements. *Nucleic acids research 26*(1), 358–359.

[123] Hinrichs Berger, J. (2004). Epidemiology of Pseudomonas syringae pathovars associated with decline of plum trees in the southwest of Germany. *Journal of Phytopathology 152*(3), 153–160.

[124] Hirayama, T. and K. Shinozaki (2010, March). Research on plant abiotic stress responses in the post-genome era: past, present and future. *The Plant Journal 61*(6), 1041–1052.

[125] Ho, L. H. M., E. Giraud, V. Uggalla, R. Lister, R. Clifton, A. Glen, D. Thirkettle-Watts, O. Van Aken, and J. Whelan (2008). Identification of regulatory pathways controlling gene expression of stress-responsive mitochondrial proteins in Arabidopsis. *PLANT PHYSIOLOGY 147*(4), 1858–1873.

[126] Holland, B. S. and M. D. Copenhaver (1988). Improved Bonferroni-type multiple testing procedures. *Psychological Bulletin 104*(1), 145–149.

[127] Hu, X. and I. Yoo (2004, January). Cluster ensemble and its applications in gene expression analysis. *Proceedings of the second conference on Asia-Pacific bioinformatics 29*, 297–302.

[128] Hua, J., C. Chang, Q. Sun, and E. M. Meyerowitz (1995, September). Ethylene insensitivity conferred by Arabidopsis ERS gene. *Science 269*(5231), 1712–1714.

[129] Hua, J., H. Sakai, S. Nourizadeh, Q. G. Chen, A. B. Bleecker, J. R. Ecker, and E. M. Meyerowitz (1998, August). EIN4 and ERS2 are members of the putative ethylene receptor gene family in Arabidopsis. *The Plant cell 10*(8), 1321–1332.

[130] Hughes, J. D., P. W. Estep, S. Tavazoie, and G. M. Church (2000, March). Computational identification of Cis-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae. *Journal of molecular biology 296*(5), 1205–1214.

[131] Ihmels, J., S. Bergmann, and N. Barkai (2004, August). Defining transcription modules using large-scale gene expression data. *Bioinformatics 20*(13), 1993–2003.

[132] Ihmels, J., G. Friedlander, S. Bergmann, O. Sarig, Y. Ziv, and N. Barkai (2002, July). Revealing modular organization in the yeast transcriptional network. *Nature genetics 31*, 370–377.

[133] Ihmels, J., R. Levy, and N. Barkai (2003, November). Principles of transcriptional control in the metabolic network of Saccharomyces cerevisiae. *Nature Biotechnology 22*(1), 86–92.

[134] Impey, S., S. R. McCorkle, H. Cha-Molstad, J. M. Dwyer, G. S. Yochum, J. M. Boss, S. McWeeney, J. J. Dunn, G. Mandel, and R. H. Goodman (2004, December). Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. *Cell 119*(7), 1041–1054.

[135] Izawa, T., R. Foster, and N. Chua (1999, June). Plant bZIP protein DNA binding specificity. *Journal of molecular biology 230*(4), 1131–1144.

[136] Janaki, C. and R. R. Joshi (2004). Motif detection in Arabidopsis: Correlation with gene expression data. *Silico Biol 4*(2), 149–161.

[137] Jelenska, J., N. Yao, B. A. Vinatzer, C. M. Wright, J. L. Brodsky, and J. T. Greenberg (2007, March). A J Domain Virulence Effector of Pseudomonas syringae Remodels Host Chloroplasts and Suppresses Defenses. *Current Biology 17*(6), 499–508.

[138] Jiang, D., C. Tang, and A. Zhang (2004). Cluster analysis for gene expression data: A survey. *Knowledge and Data Engineering, IEEE Transactions on 16*(11), 1370–1386.

[139] Jiao, Y., S. L. Tausta, N. Gandotra, N. Sun, T. Liu, N. K. Clay, T. Ceserani, M. Chen, L. Ma, M. Holford, H.-y. Zhang, H. Zhao, X. W. Deng, and T. Nelson (2009, February). A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. *Nature genetics 41*(2), 258–263.

[140] Johnson, D. S., A. Mortazavi, R. M. Myers, and B. Wold (2007, June). Genome-Wide Mapping of in Vivo Protein-DNA Interactions. *Science 316*(5830), 1497–1502.

[141] Johnson, K. D. and E. H. Bresnick (2002, January). Dissecting long-range transcriptional mechanisms by chromatin immunoprecipitation. *Methods (San Diego, Calif.) 26*(1), 27–36.

[142] Jones, J. D. G. and J. L. Dangl (2006, November). The plant immune system. *Nature 444*(7117), 323–329.

[143] Kagaya, Y., K. Ohmiya, and T. Hattori (1999, January). RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic acids research 27*(2), 470–478.

[144] Katsir, L., H. S. Chung, A. J. K. Koo, and G. A. Howe (2008, August). Jasmonate signaling: a conserved mechanism of hormone sensing. *Current opinion in plant biology 11*(4), 428–435.

[145] Kawaguchi, R., T. Girke, E. A. Bray, and J. Bailey-Serres (2004, June). Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in Arabidopsis thaliana. *The Plant Journal 38*(5), 823–839.

[146] Kazan, K. and J. M. Manners (2012, January). JAZ repressors and the orchestration of phytohormone crosstalk. *Trends in plant science 17*(1), 22–31.

[147] Kel, A., E. Gößling, and I. Reuter (2003). MATCHTM: a tool for searching transcription factor binding sites in DNA sequences. *Nucleic acids research 31*(13), 3576–3579.

[148] Kennelly, M. M., F. M. Cazorla, A. de Vicente, C. Ramos, and G. W. Sundin (2007, January). Pseudomonas syringae Diseases of Fruit Trees: Progress Toward Understanding and Control. *Plant Disease 91*(1), 4–17.

[149] Khatri, P. and S. Draghici (2005, August). Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics 21*(18), 3587–3595.

[150] Kiddle, S. J. (2011). Systems biology approaches to model and experimentally validate eukaryotic gene regulation. *Thesis*, 1–270.

[151] Kiddle, S. J., O. P. F. Windram, S. Mchattie, A. Mead, J. Beynon, V. Buchanan-Wollaston, K. J. Denby, and S. Mukherjee (2010). Temporal clustering by affinity propagation reveals transcriptional modules in Arabidopsis thaliana. *Bioinformatics 26*(3), 355–362.

[152] Kilian, J., D. Whitehead, J. Horak, D. Wanke, S. Weinl, O. Batistic, C. D'Angelo, E. Bornberg-Bauer, J. Kudla, and K. Harter (2007, March). 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.

[153] Kim, S. Y. (2005, November). The role of ABF family bZIP class transcription factors in stress response. *Physiologia Plantarum 0*(0), 519–527.

[154] Kimura, M., Y. Y. Yamamoto, M. Seki, T. Sakurai, M. Sato, T. Abe, S. Yoshida, K. Manabe, K. Shinozaki, and M. Matsui (2003). Identification of Arabidopsis Genes Regulated by High Light–Stress Using cDNA Microarray. *Photochemistry and photobiology 77*(2), 226–233.

[155] Kliebenstein, D., R. A. Dietrich, A. C. Martin, R. L. Last, and J. L. Dangl (1999, November). LSD1 regulates salicylic acid induction of copper zinc superoxide dismutase in Arabidopsis thaliana. *Molecular Plant-Microbe Interactions 12*(11), 1022–1026.

[156] Kliebenstein, D., H. Rowe, and K. Denby (2005). Secondary metabolites influence Arabidopsis/Botrytis interactions: variation in host production and pathogen sensitivity. *The Plant Journal 44*(1), 25–36.

[157] Kliebenstein, D. J., R.-A. Monde, and R. L. Last (1998). Superoxide dismutase in Arabidopsis: an eclectic enzyme family with disparate regulation and protein localization. *PLANT PHYSIOLOGY 118*(2), 637–650.

[158] Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE 78*(9), 1464–1480.

[159] Koornneef, A. and C. M. J. Pieterse (2008, March). Cross Talk in Defense Signaling. *PLANT PHYSIOLOGY 146*(3), 839–844.

[160] Kosugi, S. and Y. Ohashi (2002). DNA binding and dimerization specificity and potential targets for the TCP protein family. *The Plant Journal 30*(3), 337–348.

[161] Kreps, J. A. (2002, December). Transcriptome Changes for Arabidopsis in Response to Salt, Osmotic, and Cold Stress. *PLANT PHYSIOLOGY 130*(4), 2129–2141.

[162] Kretschmer, M. and M. Hahn (2011). Fungicide resistance and genetic diversity of Botrytis cinerea isolates from a vineyard in Germany. *Journal of Plant Diseases and Plant Protection (JPDP) 5*(115), 214–219.

[163] Krishnaswamy, S., S. Verma, M. H. Rahman, and N. N. V. Kav (2010, November). Functional characterization of four APETALA2-family genes (RAP2.6, RAP2.6L, DREB19 and DREB26) in Arabidopsis. *Plant Molecular Biology 75*(1-2), 107–127.

[164] KUNKEL, B. N. and D. M. Brooks (2002, August). Cross talk between signaling pathways in pathogen defense. *Current opinion in plant biology 5*(4), 325–331.

[165] Lavagi, I., M. Estelle, W. Weckwerth, J. Beynon, and R. Bastow (2012, July). From Bench to Bountiful Harvests: A Road Map for the Next Decade of Arabidopsis Research. *The Plant cell 24*(6), 2240–2247.

[166] Lee, J., K. He, V. Stolc, H. Lee, P. Figueroa, Y. Gao, W. Tongprasit, H. Zhao, I. Lee, and X. W. Deng (2007, March). Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. *The Plant cell 19*(3), 731–749.

[167] Lenhard, B. and W. W. Wasserman (2002, August). TFBS: Computational framework for transcription factor binding site analysis. *Bioinformatics 18*(8), 1135–1136.

[168] Leon-Reyes, A., Y. Du, A. Koornneef, S. Proietti, A. P. Körbes, J. Memelink, C. M. J. Pieterse, and T. Ritsema (2010, February). Ethylene signaling renders the jasmonate response of Arabidopsis insensitive to future suppression by salicylic Acid. *Molecular Plant-Microbe Interactions 23*(2), 187–197.

[169] Leon-Reyes, A., D. Van der Does, E. S. De Lange, C. Delker, C. Wasternack, S. C. M. Van Wees, T. Ritsema, and C. M. J. Pieterse (2010, November). Salicylate-mediated suppression of jasmonate-responsive gene expression in Arabidopsis is targeted downstream of the jasmonate biosynthesis pathway. *Planta 232*(6), 1423–1432.

[170] Leroch, M., C. Plesken, R. W. S. Weber, F. Kauff, G. Scalliet, and M. Hahn (2013, January). Gray mold populations in german strawberry fields are resistant to multiple fungicides and dominated by a novel clade closely related to Botrytis cinerea. *Applied and environmental microbiology 79*(1), 159–167.

[171] Leroux, P., R. Fritz, D. l. Debieu, C. Albertini, C. Lanen, J. Bach, M. Gredt, and F. Chapeland (2002). Mechanisms of resistance to fungicides in field strains ofBotrytis cinerea. *Pest Management Science 58*(9), 876–888.

[172] Levin, M. (1998). Journal of Global Optimization, Volume 12, Number 1 - SpringerLink. *Journal of Global Optimization 12*(1), 105–108.

[173] Li, J. (2004, February). The WRKY70 Transcription Factor: A Node of Convergence for Jasmonate-Mediated and Salicylate-Mediated Signals in Plant Defense. *The Plant cell 16*(2), 319–331.

[174] Li, J., G. Brader, T. Kariola, and E. Tapio Palva (2006, May). WRKY70 modulates the selection of signaling pathways in plant defense. *The Plant Journal 46*(3), 477–491.

[175] Liang, P. and A. B. Pardee (1992, August). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science 257*(5072), 967–971.

[176] Lim, L. P., N. C. Lau, P. Garrett-Engele, A. Grimson, J. M. Schelter, J. Castle, D. P. Bartel, P. S. Linsley, and J. M. Johnson (2005, February). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature 433*(7027), 769–773.

[177] Litvak, V., S. A. Ramsey, A. G. Rust, D. E. Zak, K. A. Kennedy, A. E. Lampano, M. Nykter, I. Shmulevich, and A. Aderem (2009, March). Function of C/EBP in a regulatory circuit that discriminates between transient and persistent TLR4-induced signals. *Nature Immunology 10*(4), 437–443.

[178] Liu, P., F. Sun, R. Gao, and H. Dong (2012, June). RAP2.6L overexpression delays waterlogging induced premature senescence by increasing stomatal closure more than antioxidant enzyme activity. *Plant Molecular Biology 79*(6), 609–622.

[179] Liu, Q., M. Kasuga, Y. Sakuma, H. Abe, S. Miura, K. Yamaguchi-Shinozaki, and K. Shinozaki (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), 1391–1406.

[180] Locke, J. C. W., L. Kozma-Bognár, P. D. Gould, B. Fehér, É. Kevei, F. Nagy, M. S. Turner, A. Hall, and A. J. Millar (2006, November). Experimental validation of a predicted feedback loop in the multi-oscillator clock of Arabidopsis thaliana. *Molecular Systems Biology 2*, 1–6.

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

[182] Lopato, S., N. Bazanova, S. Morran, A. S. Milligan, N. Shirley, and P. Langridge (2006). Isolation of plant transcription factors using a modified yeast one-hybrid system. *Plant methods 2*, 3.

[183] Lorenzo, O. (2004, July). 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.

[184] Lu, G., A. L. Paul, D. R. McCarty, and R. J. Ferl (1996, May). Transcription factor veracity: is GBF3 responsible for ABA-regulated expression of Arabidopsis Adh? *The Plant cell 8*(5), 847–857.

[185] Lu, X., V. V. Jain, P. W. Finn, and D. L. Perkins (2007, April). Hubs in biological interaction networks exhibit low changes in expression in experimental asthma. *Molecular Systems Biology 3*, 1–6.

[186] Luo, F., L. Khan, F. Bastani, I.-L. Yen, and J. Zhou (2004, October). A dynamically growing self-organizing tree (DGSOT) for hierarchical clustering gene expression profiles. *Bioinformatics 20*(16), 2605–2617.

[187] Luo, F., K. Tang, and L. Khan (2003). Hierarchical clustering of gene expression data. *Bioinformatics and Bioengineering, 2003. Proceedings. Third IEEE Symposium on 1*, 328–335.

[188] Luscombe, N., M. Babu, H. Yu, M. Snyder, S. Teichmann, and M. Gerstein (2004). Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature 431*(7006), 308–312.

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

[190] Madeira, S. C. and A. L. Oliveira (2004, January). Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM 1*(1), 24–45.

[191] Maere, S., K. Heymans, and M. Kuiper (2005, August). BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics 21*(16), 3448–3449.

[192] Mandaokar, A. and J. Browse (2008, December). MYB108 Acts Together with MYB24 to Regulate Jasmonate-Mediated Stamen Maturation in Arabidopsis. *PLANT PHYSIOLOGY 149*(2), 851–862.

[193] Mase, K., T. Mizuno, N. Ishihama, T. Fujii, H. Mori, M. Kodama, and H. Yoshioka (2012). Ethylene Signaling Pathway and MAPK Cascades Are Required for AAL Toxin-Induced Programmed Cell Death. *Molecular Plant-Microbe Interactions 25*(8), 1015–1025.

[194] Matys, V., E. Fricke, R. Geffers, E. Goessling, M. Haubrock, R. Hehl, K. Hornischer, D. Karas, A. E. Kel, and O. V. Kel-Margoulis (2003). TRANSFAC®: transcriptional regulation, from patterns to profiles. *Nucleic acids research 31*(1), 374–378.

[195] Matys, V., O. V. Kel-Margoulis, E. Fricke, I. Liebich, S. Land, A. Barre-Dirrie, I. Reuter, D. Chekmenev, M. Krull, K. Hornischer, N. Voss, P. Stegmaier, B. Lewicki-Potapov, H. Saxel, A. E. Kel, and E. Wingender (2006, January). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic acids research 34*(Database issue), D108–10.

[196] McClung, C. R. (2006, April). Plant circadian rhythms. *The Plant cell 18*(4), 792–803.

[197] McGrath, K. C. (2005, September). Repressor- and Activator-Type Ethylene Response Factors Functioning in Jasmonate Signaling and Disease Resistance Identified via a Genome-Wide Screen of Arabidopsis Transcription Factor Gene Expression. *PLANT PHYSIOLOGY 139*(2), 949–959.

[198] McGrath, P. T., H. Lee, L. Zhang, A. A. Iniesta, A. K. Hottes, M. H. Tan, N. J. Hillson, P. Hu, L. Shapiro, and H. H. McAdams (2007, April). High-throughput identification of transcription start sites, conserved promoter motifs and predicted regulons. *Nature Biotechnology 25*(5), 584–592.

[199] Mehrtens, F., H. Kranz, P. Bednarek, and B. Weisshaar (2005, June). The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. *PLANT PHYSIOLOGY 138*(2), 1083–1096.

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

[201] Mengiste, T., X. Chen, J. Salmeron, and R. Dietrich (2003, November). The BOTRYTIS SUSCEPTIBLE1 gene encodes an R2R3MYB transcription factor protein that is required for biotic and abiotic stress responses in Arabidopsis. *The Plant cell 15*(11), 2551–2565.

[202] Menkens, A. E. and A. R. Cashmore (1994, March). Isolation and characterization of a fourth Arabidopsis thaliana G-box-binding factor, which has similarities to Fos oncoprotein. *Proceedings of the National Academy of Sciences of the United States of America 91*(7), 2522–2526.

[203] Mewes, H. W., J. Hani, F. Pfeiffer, and D. Frishman (1998, January). MIPS: a database for protein sequences and complete genomes. *Nucleic acids research 26*(1), 33–37.

[204] Miflin, B. (2000, January). Crop improvement in the 21st century. *Journal of Experimental Botany 51*(342), 1–8.

[205] Millar, A., I. A. Carré, C. A. Strayer, N. H. Chua, and S. A. Kay (1995, February). Circadian clock mutants in Arabidopsis identified by luciferase imaging. *Science 267*(5201), 1161–1163.

[206] Millar, A. H., J. Whelan, K. L. Soole, and D. A. Day (2011, June). Organization and Regulation of Mitochondrial Respiration in Plants. *Annual Review of Plant Biology 62*(1), 79–104.

[207] Mitchell-Olds, T. (2001). Arabidopsis thaliana and its wild relatives: a model system for ecology and evolution. *Trends in Ecology & Evolution 16*(12), 693–700.

[208] Mittler, R. (2006, January). Abiotic stress, the field environment and stress combination. *Trends in plant science 11*(1), 15–19.

[209] Mjolsness, E., T. Mann, R. Castano, and B. Wold (1999). From coexpression to coregulation: an approach to inferring transcriptional regulation among gene classes from large-scale expression data. *Neural Information Processing Systems*.

[210] Mohr, P. G. and D. M. Cahill (2003). Abscisic acid influences the susceptibility of Arabidopsis thaliana to Pseudomonas syringae pv. tomato and Peronospora parasitica. *Functional Plant Biology 30*(4), 461–469.

[211] Moreno-Risueno, M. A., W. Busch, and P. N. Benfey (2010, April). Omics meet networks - using systems approaches to infer regulatory networks in plants. *Current opinion in plant biology 13*(2), 126–131.

[212] Morishita, T., Y. Kojima, T. Maruta, A. Nishizawa-Yokoi, Y. Yabuta, and S. Shigeoka (2009, December). Arabidopsis NAC Transcription Factor, ANAC078, Regulates Flavonoid Biosynthesis under High-light. *Plant and Cell Physiology 50*(12), 2210–2222.

[213] Mukhtar, M. S., A.-R. Carvunis, M. Dreze, P. Epple, J. Steinbrenner, J. Moore, M. Tasan, M. Galli, T. Hao, M. T. Nishimura, S. J. Pevzner, S. E. Donovan, L. Ghamsari, B. Santhanam, V. Romero, M. M. Poulin, F. Gebreab, B. J. Gutierrez, S. Tam, D. Monachello, M. Boxem, C. J. Harbort, N. McDonald, L. Gai, H. Chen, Y. He, European Union Effectoromics Consortium, J. Vandenhaute, F. P. Roth, D. E. Hill, J. R. Ecker, M. Vidal, J. Beynon, P. Braun, and J. L. Dangl (2011, July). Independently Evolved Virulence Effectors Converge onto Hubs in a Plant Immune System Network. *Science 333*(6042), 596–601.

[214] Murphy, K. and S. Mian (1999). Modelling gene expression data using dynamic Bayesian networks. *Technical report, Computer Science Division, Univerity of California*.

[215] Mysore, K. S., R. P. Tuori, and G. B. Martin (2001). Arabidopsis genome sequence as a tool for functional genomics in tomato. *Genome Biology 2*(1), Reviews 1003.1–1003.4.

[216] Narusaka, Y., M. Narusaka, M. Seki, T. Umezawa, J. Ishida, M. Nakajima, A. Enju, and K. Shinozaki (2004). Crosstalk in the responses to abiotic and biotic stresses in Arabidopsis: analysis of gene expression in cytochrome P450 gene superfamily by cDNA microarray. *Plant Molecular Biology 55*(3), 327–342.

[217] Nawrath, C. and J. P. Métraux (1999, August). Salicylic acid induction-deficient mutants of Arabidopsis express PR-2 and PR-5 and accumulate high levels of camalexin after pathogen inoculation. *The Plant cell 11*(8), 1393–1404.

[218] Nikiforova, V. J. (2005, April). Integrative gene-metabolite network with implemented causality deciphers informational fluxes of sulphur stress response. *Journal of Experimental Botany 56*(417), 1887–1896.

[219] Nole-Wilson, S. and B. A. Krizek (2000, October). DNA binding properties of the Arabidopsis floral development protein AINTEGUMENTA. *Nucleic acids research 28*(21), 4076–4082.

[220] Oerke, E. C. (2005, December). Crop losses to pests. *The Journal of Agricultural Science 144*(01), 31–43.

[221] Ohta, M., Y. Hayashi, A. Nakashima, A. Hamada, A. Tanaka, T. Nakamura, and T. Hayakawa (2002, December). Introduction of a Na+/H+ antiporter gene from Atriplex gmelini confers salt tolerance to rice. *FEBS letters 532*(3), 279–282.

[222] Okushima, Y., H. Fukaki, M. Onoda, A. Theologis, and M. TASAKA (2007, January). ARF7 and ARF19 Regulate Lateral Root Formation via Direct Activation of LBD/ASL Genes in Arabidopsis. *The Plant cell 19*(1), 118–130.

[223] Olsen, A. N., H. A. Ernst, L. L. Leggio, and K. Skriver (2005, October). DNA-binding specificity and molecular functions of NAC transcription factors. *Plant Science 169*(4), 785–797.

[224] Oono, Y., M. Seki, T. Nanjo, M. Narusaka, M. Fujita, R. Satoh, M. Satou, T. Sakurai, J. Ishida, K. Akiyama, K. Iida, K. Maruyama, S. Satoh, K. Yamaguchi-Shinozaki, and K. Shinozaki (2003, June). Monitoring expression profiles of Arabidopsis gene expression during rehydration process after dehydration using ca 7000 full-length cDNA microarray. *The Plant journal : for cell and molecular biology 34*(6), 868–887.

[225] Papadimitriou, S. (2007). Scientific programming with Java classes supported with a scripting interpreter. *IET Software 1*(2), 48.

[226] Park, P. J. (2009, September). ChIP–seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics 10*(10), 669–680.

[227] Park, T., S. G. Yi, S. Lee, S. Y. Lee, D. H. Yoo, J. I. Ahn, and Y. S. Lee (2003, April). Statistical tests for identifying differentially expressed genes in time-course microarray experiments. *Bioinformatics 19*(6), 694–703.

[228] Pautot, V., F. M. Holzer, J. Chaufaux, and L. L. Walling (2001). The induction of tomato leucine aminopeptidase genes (LapA) after Pseudomonas syringae pv. tomato infection is pri-

marily a wound response triggered by coronatine. *Molecular Plant-Microbe Interactions 14*(2), 214–224.

[229] Pauwels, L., G. F. Barbero, J. Geerinck, S. Tilleman, W. Grunewald, A. C. Pérez, J. M. Chico, R. V. Bossche, J. Sewell, E. Gil, G. García-Casado, E. Witters, D. Inzé, J. A. Long, G. De Jaeger, R. Solano, and A. Goossens (2010, January). NINJA connects the co-repressor TOPLESS to jasmonate signalling. *Nature 464*(7289), 788–791.

[230] Penfold, C. A., V. Buchanan-Wollaston, K. J. Denby, and D. L. Wild (2012, June). Non-parametric Bayesian inference for perturbed and orthologous gene regulatory networks. *Bioinformatics 28*(12), i233–i241.

[231] Penfold, C. A. and D. L. Wild (2011). How to infer gene networks from expression profiles, revisited. *Interface Focus 1*(6), 857–870.

[232] Pepke, S., B. Wold, and A. Mortazavi (2009, November). Computation for ChIP-seq and RNA-seq studies. *Nature methods 6*(11 Suppl), S22–32.

[233] Phillips, T. (2008). Regulation of transcription and gene expression in eukaryotes. *Nature Education 1*(1).

[234] Pounds, S. and S. W. Morris (2003, July). Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. *Bioinformatics 19*(10), 1236–1242.

[235] Powles, S. B. (1984). Photoinhibition of photosynthesis induced by visible light. *Annual Review of Plant Physiology 35*(1), 15–44.

[236] Pruneda-Paz, J. L., G. Breton, A. Para, and S. A. Kay (2009). A functional genomics approach reveals CHE as a component of the Arabidopsis circadian clock. *Science 323*(5920), 1481.

[237] Qian, J., M. Dolled-Filhart, J. Lin, H. Yu, and M. Gerstein (2001, December). Beyond synexpression relationships: local clustering of time-shifted and inverted gene expression profiles identifies new, biologically relevant interactions. *Journal of molecular biology 314*(5), 1053–1066.

[238] Rabbani, M. A. (2003, December). Monitoring Expression Profiles of Rice Genes under Cold, Drought, and High-Salinity Stresses and Abscisic Acid Application Using cDNA Microarray and RNA Gel-Blot Analyses. *PLANT PHYSIOLOGY 133*(4), 1755–1767.

[239] Ramankutty, N., A. T. Evan, and C. Monfreda (2008). Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles 22*(1).

[240] Ramonell, K. M. and S. Somerville (2002, August). The genomics parade of defense responses: to infinity and beyond. *Current opinion in plant biology 5*(4), 291–294.

[241] Reinbothe, S., B. Mollenhauer, and C. Reinbothe (1994). JIPs and RIPs: the regulation of plant gene expression by jasmonates in response to environmental cues and pathogens. *The Plant cell 6*(9), 1197.

[242] Reymond, P., H. Weber, M. Damond, and E. E. Farmer (2000, May). Differential gene expression in response to mechanical wounding and insect feeding in Arabidopsis. *The Plant cell 12*(5), 707–720.

[243] Riechmann, J. L., J. Heard, G. Martin, L. Reuber, C. Jiang, J. Keddie, L. Adam, O. Pineda, O. J. Ratcliffe, R. R. Samaha, R. Creelman, M. Pilgrim, P. Broun, J. Z. Zhang, D. Ghandehari, B. K. Sherman, and G. Yu (2000, December). Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science 290*(5499), 2105–2110.

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

[245] Rizhsky, L., H. Liang, J. Shuman, and V. Shulaev (2004). When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress. *PLANT PHYSIOLOGY 134*(4), 1683–1696.

[246] Robert-Seilaniantz, A., M. Grant, and J. D. G. Jones (2011, September). Hormone Crosstalk in Plant Disease and Defense: More Than Just JASMONATE-SALICYLATE Antagonism. *Annual Review of Phytopathology 49*(1), 317–343.

[247] Rogers, E. E. and F. M. Ausubel (1997, March). Arabidopsis enhanced disease susceptibility mutants exhibit enhanced susceptibility to several bacterial pathogens and alterations in PR-1 gene expression. *The Plant cell 9*(3), 305–316.

[248] Saijo, Y., S. Hata, J. Kyozuka, and K. Shimamoto (2001). Overexpression of a single Ca2+dependent protein kinase confers both cold and salt/drought tolerance on rice plants. *The Plant Journal 23*(3), 319–327.

[249] Sakai, H., J. Hua, Q. G. Chen, C. Chang, L. J. Medrano, A. B. Bleecker, and E. M. Meyerowitz (1998, May). ETR2 is an ETR1-like gene involved in ethylene signaling in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America 95*(10), 5812–5817.

[250] Sakamoto, H. (2004, September). Arabidopsis Cys2/His2-Type Zinc-Finger Proteins Function as Transcription Repressors under Drought, Cold, and High-Salinity Stress Conditions. *PLANT PHYSIOLOGY 136*(1), 2734–2746.

[251] Schenk, P. M., K. Kazan, I. Wilson, J. P. Anderson, T. Richmond, S. C. Somerville, and J. M. Manners (2000, October). Coordinated plant defense responses in Arabidopsis revealed by microarray analysis. *Proceedings of the National Academy of Sciences of the United States of America 97*(21), 11655–11660.

[252] Schlitt, T. and A. Brazma (2007). Current approaches to gene regulatory network modelling. *BMC bioinformatics 8*(Suppl 6), S9.

[253] Schommer, C., J. F. Palatnik, P. Aggarwal, A. Chételat, P. Cubas, E. E. Farmer, U. Nath, and D. Weigel (2008). Control of Jasmonate Biosynthesis and Senescence by miR319 Targets. *PLoS Biology 6*(9), e230.

[254] Segal, E., M. Shapira, A. Regev, D. Pe'er, D. Botstein, D. Koller, and N. Friedman (2003, June). Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature genetics 34*(2), 166–176.

[255] Seki, M., M. Narusaka, J. Ishida, T. Nanjo, M. Fujita, Y. Oono, A. Kamiya, M. Nakajima, A. Enju, and T. Sakurai (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.

[256] Shah, J., F. Tsui, and D. F. Klessig (1997). Characterization of as alicylic a cid-i nsensitive mutant (sai1) of Arabidopsis thaliana, identified in a selective screen utilizing the SA-inducible expression of the tms2 .... *Molecular Plant-Microbe ... 10*(1), 69–78.

[257] Sheen, J. (1996, December). Ca2+-dependent protein kinases and stress signal transduction in plants. *Science 274*(5294), 1900–1902.

[258] Shen-Orr, S. S., R. Milo, S. Mangan, and U. Alon (2002, April). Network motifs in the transcriptional regulation network of Escherichia coli. *Nature genetics 31*(1), 64–68.

[259] Shi, L., W. D. Jones, R. V. Jensen, S. C. Harris, R. G. Perkins, F. M. Goodsaid, L. Guo, L. J. Croner, C. Boysen, H. Fang, F. Qian, S. Amur, W. Bao, C. C. Barbacioru, V. Bertholet, X. Cao, T.-M. Chu, P. J. Collins, X.-h. Fan, F. W. Frueh, J. C. Fuscoe, X. Guo, J. Han, D. Herman, H. Hong, E. S. Kawasaki, Q.-Z. Li, Y. Luo, Y. Ma, N. Mei, R. L. Peterson, R. K. Puri, R. Shippy, Z. Su, Y. Sun, H. Sun, B. Thorn, Y. Turpaz, C. Wang, S. Wang, J. A. Warrington, J. C. Willey, J. Wu, Q. Xie, L. Zhang, L. Zhang, S. Zhong, R. D. Wolfinger, and W. Tong (2008). The balance of reproducibility, sensitivity, and specificity of lists of differentially expressed genes in microarray studies. *BMC bioinformatics 9*(Suppl 9), S10.

[260] Shinozaki, K. and K. Yamaguchi-Shinozaki (2000). Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. *Current opinion in plant biology 3*(3), 217–223.

[261] Shinozaki, K. and K. Yamaguchi-Shinozaki (2006, November). Gene networks involved in drought stress response and tolerance. *Journal of Experimental Botany 58*(2), 221–227.

[262] Shinozaki, K., K. Yamaguchi-Shinozaki, and M. Seki (2003). Regulatory network of gene expression in the drought and cold stress responses. *Current opinion in plant biology 6*(5), 410–417.

[263] Shuai, B., C. G. Reynaga-Peña, and P. S. Springer (2002, June). The lateral organ boundaries gene defines a novel, plant-specific gene family. *PLANT PHYSIOLOGY 129*(2), 747–761.

[264] Siebert, S., P. Döll, J. Hoogeveen, J. M. Faures, K. Frenken, and S. Feick (2005, August). Development and validation of the global map of irrigation areas. *Hydrology and Earth System Sciences Discussions 2*, 1299–1327.

[265] Singh, K., R. C. Foley, and L. Oñate-Sánchez (2002). Transcription factors in plant defense and stress responses. *Current opinion in plant biology 5*(5), 430–436.

[266] Singh, K. B. (1998, December). Transcriptional regulation in plants: the importance of combinatorial control. *PLANT PHYSIOLOGY 118*(4), 1111–1120.

[267] Smith, S. and I. De Smet (2012, April). Root system architecture: insights from Arabidopsis and cereal crops. *Philosophical Transactions of the Royal Society B: Biological Sciences 367*(1595), 1441–1452.

[268] Spannagl, M., K. Mayer, J. Durner, G. Haberer, and A. Fröhlich (2011, January). Exploring the genomes: From Arabidopsis to crops. *Journal of Plant Physiology 168*(1), 3–8.

[269] Sreenivasulu, N., B. Usadel, A. Winter, V. Radchuk, U. Scholz, N. Stein, W. Weschke, M. Strickert, T. J. Close, M. Stitt, A. Graner, and U. Wobus (2008, April). Barley grain maturation and germination: metabolic pathway and regulatory network commonalities and differences highlighted by new MapMan/PageMan profiling tools. *PLANT PHYSIOLOGY 146*(4), 1738–1758.

[270] Stegle, O., K. Denby, S. Mchattie, S. Meade, D. L. Wild, Z. Ghahramani, and K. Borgwardt (2009). Discovering temporal patterns of differential gene expression in microarray time series. *German Conference on Bioinformatics*.

[271] Sun, F., P. Liu, J. Xu, and H. Dong (2010, September). Mutation in RAP2.6L, a transactivator of the ERF transcription factor family, enhances Arabidopsis resistance to Pseudomonas syringae. *Physiological and Molecular Plant Pathology 74*(5-6), 295–302.

[272] Supper, J., M. Strauch, D. Wanke, K. Harter, and A. Zell (2007). EDISA: extracting biclusters from multiple time-series of gene expression profiles. *BMC bioinformatics 8*, 334.

[273] Sutheeworapong, S., O. Motonori, O. Hiroyuki, and K. Kengo (2012, September). A novel biclustering approach with iterative optimization to analyze gene expression data. *Advances and Applications in Bioinformatics and Chemistry 5*, 23–59.

[274] Tachibana, C., J. Y. Yoo, J. B. Tagne, N. Kacherovsky, T. I. Lee, and E. T. Young (2005, March). Combined Global Localization Analysis and Transcriptome Data Identify Genes That Are Directly Coregulated by Adr1 and Cat8. *Molecular and Cellular Biology 25*(6), 2138–2146.

[275] Tai, Y. C. (2010). timecourse: Statistical Analysis for Developmental Microarray Time Course Data. *R package version*.

[276] Tatematsu, K., S. Ward, O. Leyser, Y. Kamiya, and E. Nambara (2005, June). Identification of cis-elements that regulate gene expression during initiation of axillary bud outgrowth in Arabidopsis. *PLANT PHYSIOLOGY 138*(2), 757–766.

[277] Tavazoie, S., J. Hughes, M. Campbell, R. Cho, and G. Church (1999). Systematic determination of genetic network architecture. *Nature genetics 22*, 281–285.

[278] Teakle, G. R., I. W. Manfield, J. F. Graham, and P. M. Gilmartin (2002, September). Arabidopsis thaliana GATA factors: organisation, expression and DNA-binding characteristics. *Plant Molecular Biology 50*(1), 43–57.

[279] Team, R. D. C. (2007). *R: A language and environment for statistical computing*. *Vienna, Austria: R Foundation for Statistical Computing*. Nova Science Pub Inc.

[280] Thaler, J. S., P. T. Humphrey, and N. K. Whiteman (2012, May). Evolution of jasmonate and salicylate signal crosstalk. *Trends in plant science 17*(5), 260–270.

[281] Thattai, M. and A. van Oudenaarden (2001, July). Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America 98*(15), 8614–8619.

[282] Thibaud-Nissen, F., H. Wu, T. Richmond, J. C. Redman, C. Johnson, R. Green, J. Arias, and C. D. Town (2006, July). Development of Arabidopsis whole-genome microarrays and their application to the discovery of binding sites for the TGA2 transcription factor in salicylic acid-treated plants. *The Plant Journal 47*(1), 152–162.

[283] Thilmony, R., W. Underwood, and S.-Y. He (2006, April). Genome-wide transcriptional analysis of the Arabidopsis thalianainteraction with the plant pathogen Pseudomonas syringaepv. tomatoDC3000 and the human pathogen Escherichia coliO157:H7. *The Plant Journal 46*(1), 34–53.

[284] Thomma, B. P., K. Eggermont, I. A. Penninckx, B. Mauch-Mani, R. Vogelsang, B. P. Cammue, and W. F. Broekaert (1998, December). 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 of the United States of America 95*(25), 15107–15111.

[285] Thomma, B. P. H. J., I. A. M. A. Penninckx, B. Cammue, and W. F. Broekaert (2001). The complexity of disease signaling in Arabidopsis. *Current Opinion in Immunology 13*(1), 63–68.

[286] Tilstone, C. (2003). DNA microarrays: vital statistics. *NATURE-LONDON- 424*(6949), 610–612.

[287] Tkačik, G. and A. M. Walczak (2011, April). Information transmission in genetic regulatory networks: a review. *Journal of physics. Condensed matter : an Institute of Physics journal 23*(15), 153102.

[288] Toledo-Ortiz, G. (2003, July). The Arabidopsis Basic/Helix-Loop-Helix Transcription Factor Family. *The Plant cell 15*(8), 1749–1770.

[289] Ton, J., V. Flors, and B. Mauch-Mani (2009, June). The multifaceted role of ABA in disease resistance. *Trends in plant science 14*(6), 310–317.

[290] Torres, M. A., J. D. G. Jones, and J. L. Dangl (2005, September). Pathogen-induced, NADPH oxidase–derived reactive oxygen intermediates suppress spread of cell death in Arabidopsis thaliana. *Nature genetics 37*(10), 1130–1134.

[291] Tran, L.-S. P. (2004, September). 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.

[292] Trémousaygue, D., L. Garnier, C. Bardet, P. Dabos, C. Hervé, and B. Lescure (2003, March). Internal telomeric repeats and 'TCP domain' protein-binding sites co-operate to regulate gene expression in Arabidopsis thaliana cycling cells. *The Plant journal : for cell and molecular biology 33*(6), 957–966.

[293] Tremousaygue, D., A. Manevski, C. Bardet, N. Lescure, and B. Lescure (1999, December). Plant interstitial telomere motifs participate in the control of gene expression in root meristems. *The Plant journal : for cell and molecular biology 20*(5), 553–561.

[294] Uno, Y., T. Furihata, H. Abe, R. Yoshida, K. Shinozaki, and K. Yamaguchi-Shinozaki (2000, October). 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 of the United States of America 97*(21), 11632–11637.

[295] Urano, K., Y. Kurihara, M. Seki, and K. Shinozaki (2010, April). 'Omics' analyses of regulatory networks in plant abiotic stress responses. *Current opinion in plant biology 13*(2), 132–138.

[296] Urao, T., T. Katagiri, and T. Mizoguchi (1994). Two genes that encode Ca 2+-dependent protein kinases are induced by drought and high-salt stresses in Arabidopsis thaliana. *Molecular and General ... 244*, 331–340.

[297] Usadel, B. (2005, July). Extension of the Visualization Tool MapMan to Allow Statistical Analysis of Arrays, Display of Coresponding Genes, and Comparison with Known Responses. *PLANT PHYSIOLOGY 138*(3), 1195–1204.

[298] Usadel, B., A. Nagel, D. Steinhauser, Y. Gibon, O. E. Bläsing, H. Redestig, N. Sreenivasulu, L. Krall, M. A. Hannah, F. Poree, A. R. Fernie, and M. Stitt (2006). PageMan: an interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. *BMC bioinformatics 7*, 535.

[299] van Helden, J., A. F. Rios, and J. Collado-Vides (2000, April). Discovering regulatory elements in non-coding sequences by analysis of spaced dyads. *Nucleic acids research 28*(8), 1808–1818.

[300] van Kan, J. (2006). Licensed to kill: the lifestyle of a necrotrophic plant pathogen. *Trends in plant science 11*(5), 247–253.

[301] Van Tunen, A. J., R. E. Koes, C. E. Spelt, A. R. Van der Krol, A. R. Stuitje, and J. N. Mol (1988). Cloning of the two chalcone flavanone isomerase genes from Petunia hybrida: coordi-

nate, light-regulated and differential expression of flavonoid genes. *The EMBO Journal 7*(5), 1257.

[302] Vandepoele, K., M. Quimbaya, T. Casneuf, L. De Veylder, and Y. Van de Peer (2009, June). Unraveling Transcriptional Control in Arabidopsis Using cis-Regulatory Elements and Coexpression Networks. *PLANT PHYSIOLOGY 150*(2), 535–546.

[303] Vandepoele, K., K. Vlieghe, K. Florquin, L. Hennig, G. T. S. Beemster, W. Gruissem, Y. Van de Peer, D. Inzé, and L. De Veylder (2005, September). Genome-wide identification of potential plant E2F target genes. *PLANT PHYSIOLOGY 139*(1), 316–328.

[304] Velculescu, V. E., L. Zhang, B. Vogelstein, and K. W. Kinzler (1995, October). Serial analysis of gene expression. *Science 270*(5235), 484–487.

[305] Vermeirssen, V., B. Deplancke, M. I. Barrasa, J. S. Reece-Hoyes, H. E. Arda, C. A. Grove, N. J. Martinez, R. Sequerra, L. Doucette-Stamm, M. R. Brent, and A. J. M. Walhout (2007, August). Matrix and Steiner-triple-system smart pooling assays for high-performance transcription regulatory network mapping. *Nature methods 4*(8), 659–664.

[306] Vogelstein, B., D. Lane, and A. J. Levine (2000, November). Surfing the p53 network. *NATURE-LONDON- 408*(6810), 307–310.

[307] Walhout, A., G. F. Temple, M. A. Brasch, and J. L. Hartley (2000). GATEWAY recombinational cloning: Application to the cloning of large numbers of open reading frames or ORFeomes. *Methods in ... 328*, 575–592.

[308] Walhout, A. J. M. and M. Vidal (2001, July). High-Throughput Yeast Two-Hybrid Assays for Large-Scale Protein Interaction Mapping. *Methods (San Diego, Calif.) 24*(3), 297–306.

[309] Walley, J. W., S. Coughlan, M. E. Hudson, M. F. Covington, R. Kaspi, G. Banu, S. L. Harmer, and K. Dehesh (2007, October). Mechanical stress induces biotic and abiotic stress responses via a novel cis-element. *PLoS Genetics 3*(10), 1800–1812.

[310] Wang, H., M. E. Heinz, S. D. Crosby, M. Johnston, and R. D. Mitra (2008, September). 'Calling Cards' method for high-throughput identification of targets of yeast DNA-binding proteins. *Nature Protocols 3*(10), 1569–1577.

[311] Wang, K. L. C., H. Li, and J. R. Ecker (2002). Ethylene biosynthesis and signaling networks. *The Plant cell 14*(suppl 1), S131–S151.

[312] Wang, X., H. Wang, J. Wang, R. Sun, J. Wu, S. Liu, Y. Bai, J.-H. Mun, I. Bancroft, F. Cheng, S. Huang, X. Li, W. Hua, J. Wang, X. Wang, M. Freeling, J. C. Pires, A. H. Paterson, B. Chalhoub, B. Wang, A. Hayward, A. G. Sharpe, B.-S. Park, B. Weisshaar, B. Liu, B. Li, B. Liu, C. Tong, C. Song, C. Duran, C. Peng, C. Geng, C. Koh, C. Lin, D. Edwards, D. Mu, D. Shen, E. Soumpourou, F. Li, F. Fraser, G. Conant, G. Lassalle, G. J. King, G. Bonnema, H. Tang, H. Wang, H. Belcram, H. Zhou, H. Hirakawa, H. Abe, H. Guo, H. Wang, H. Jin, I. A. P. Parkin, J. Batley, J.-S. Kim, J. Just, J. Li, J. Xu, J. Deng, J. A. Kim, J. Li, J. Yu, J. Meng, J. Wang, J. Min, J. Poulain, J. Wang, K. Hatakeyama, K. Wu, L. Wang, L. Fang, M. Trick, M. G. Links, M. Zhao, M. Jin, N. Ramchiary, N. Drou, P. J. Berkman, Q. Cai, Q. Huang, R. Li, S. Tabata, S. Cheng, S. Zhang, S. Zhang, S. Huang, S. Sato, S. Sun, S.-J. Kwon, S.-R. Choi, T.-H. Lee, W. Fan, X. Zhao, X. Tan, X. Xu, Y. Wang, Y. Qiu, Y. Yin, Y. Li, Y. Du, Y. Liao, Y. Lim, Y. Narusaka, Y. Wang, Z. Wang, Z. Li, Z. Wang, Z. Xiong, and Z. Zhang (2011, August). The genome of the mesopolyploid crop species Brassica rapa. *Nature genetics 43*(10), 1035–1039.

[313] Wang, X.-J., J. L. Reyes, N.-H. Chua, and T. Gaasterland (2004). Prediction and identification of Arabidopsis thaliana microRNAs and their mRNA targets. *Genome Biology 5*(9), R65.

[314] Wang, Z., M. Gerstein, and M. Snyder (2009, January). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics 10*(1), 57–63.

[315] Wei, H., S. Persson, T. Mehta, V. Srinivasasainagendra, L. Chen, G. P. Page, C. Somerville, and A. Loraine (2006, August). Transcriptional Coordination of the Metabolic Network in Arabidopsis. *PLANT PHYSIOLOGY 142*(2), 762–774.

[316] Weinmann, A. S. and P. J. Farnham (2002, January). Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. *Methods (San Diego, Calif.) 26*(1), 37–47.

[317] Welchen, E. and D. H. Gonzalez (2005). Differential expression of the Arabidopsis cytochrome c genes Cytc-1 and Cytc-2. Evidence for the involvement of TCP-domain protein-

binding elements in anther-and meristem-specific expression of the Cytc-1 gene. *PLANT PHYSIOLOGY 139*(1), 88–100.

[318] Wettenhall, J. M. and G. K. Smyth (2004, December). limmaGUI: A graphical user interface for linear modeling of microarray data. *Bioinformatics 20*(18), 3705–3706.

[319] White, J. V., I. Muchnik, and T. F. Smith (1994, December). Modeling protein cores with Markov random fields. *Mathematical biosciences 124*(2), 149–179.

[320] Wilkinson, J. Q., M. B. Lanahan, H. C. Yen, J. J. Giovannoni, and H. J. Klee (1995, December). An ethylene-inducible component of signal transduction encoded by never-ripe. *Science 270*(5243), 1807–1809.

[321] WILLIAMSON, B., B. Tudzynski, P. Tudzynski, and J. A. L. Van Kan (2007, September). Botrytis cinerea: the cause of grey mould disease. *Molecular Plant Pathology 8*(5), 561–580.

[322] Windram, O., P. Madhou, S. Mchattie, C. Hill, R. Hickman, E. Cooke, D. J. Jenkins, C. A. Penfold, L. Baxter, E. Breeze, S. J. Kiddle, J. Rhodes, S. Atwell, D. Kliebenstein, Y.-S. Kim, O. Stegle, K. Borgwardt, C. Zhang, A. Tabrett, R. Legaie, J. Moore, B. Finkenstädt, D. L. Wild, A. Mead, D. Rand, J. Beynon, S. Ott, V. Buchanan-Wollaston, and K. J. Denby (2012, October). Arabidopsis Defense against Botrytis cinerea: Chronology and Regulation Deciphered by High-Resolution Temporal Transcriptomic Analysis. *The Plant cell 24*(9), 3530–3557.

[323] Windram, O. P. F. *Using a systems biology approach to elucidate transcriptional networks regulating plant defence*. Ph. D. thesis.

[324] Winkel-Shirley, B. (2002). Biosynthesis of flavonoids and effects of stress. *Current opinion in plant biology 5*(3), 218–223.

[325] Witkowski, A., S. Bellamkonda, and T. Bozkaya (2005). Query by excel. In *Proceedings of the 31st ....*

[326] Wu, H., M. K. Kerr, X. Cui, and G. A. Churchill (2003). *MAANOVA: A Software Package for the Analysis of Spotted cDNA Microarray Experiments*. Statistics for Biology and Health. New York: Springer-Verlag.

[327] Wu, J. (2006, July). ChIP-chip Comes of Age for Genome-wide Functional Analysis. *Cancer Research 66*(14), 6899–6902.

[328] Xie, Z. (2005, August). Expression of Arabidopsis MIRNA Genes. *PLANT PHYSIOLOGY 138*(4), 2145–2154.

[329] Xue, G.-P. and C. W. Loveridge (2003, December). HvDRF1 is involved in abscisic acid-mediated gene regulation in barley and produces two forms of AP2 transcriptional activators, interacting preferably with a CT-rich element. *The Plant Journal 37*(3), 326–339.

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

[331] Yan, J., C. Zhang, M. Gu, Z. Bai, W. Zhang, T. Qi, Z. Cheng, W. Peng, H. Luo, F. Nan, Z. Wang, and D. Xie (2009, August). The Arabidopsis CORONATINE INSENSITIVE1 protein is a jasmonate receptor. *The Plant cell 21*(8), 2220–2236.

[332] Yang, S., S. Wang, X. Liu, Y. Yu, L. Yue, X. Wang, and D. Hao (2009, October). Four divergent Arabidopsis ethylene-responsive element-binding factor domains bind to a target DNA motif with a universal CG step core recognition and different flanking bases preference. *FEBS Journal 276*(23), 7177–7186.

[333] Yeung, K. Y., M. Medvedovic, and R. E. Bumgarner (2004). From co-expression to co-regulation: how many microarray experiments do we need? *Genome Biology 5*(7), R48.

[334] Yona, G., W. Dirks, S. Rahman, and D. M. Lin (2006, July). Effective similarity measures for expression profiles. *Bioinformatics 22*(13), 1616–1622.

[335] Yonekura-Sakakibara, K., T. Tohge, R. Niida, and K. Saito (2007). Identification of a flavonol 7-O-rhamnosyltransferase gene determining flavonoid pattern in Arabidopsis by transcriptome coexpression analysis and reverse genetics. *The Journal of biological chemistry 282*(20), 14932–14941.

[336] Young, E. T., K. M. Dombek, C. Tachibana, and T. Ideker (2003, July). Multiple pathways are co-regulated by the protein kinase Snf1 and the transcription factors Adr1 and Cat8. *The Journal of biological chemistry 278*(28), 26146–26158.

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

[338] Yu, H., X. Chen, Y. Y. Hong, Y. Wang, P. Xu, S. D. Ke, H. Y. Liu, J. K. Zhu, D. J. Oliver, and C. B. Xiang (2008, April). Activated Expression of an Arabidopsis HD-START Protein Confers Drought Tolerance with Improved Root System and Reduced Stomatal Density. *THE PLANT CELL ONLINE 20*(4), 1134–1151.

[339] Yu, H., N. M. Luscombe, J. Qian, and M. Gerstein (2003, August). Genomic analysis of gene expression relationships in transcriptional regulatory networks. *Trends in genetics : TIG 19*(8), 422–427.

[340] Zhang, H. X. and E. BLUMWALD (2001). Transgenic salt-tolerant tomato plants accumulate salt in foliage but not in fruit. *Nature Biotechnology 19*(8), 765–768.

[341] Zhang, J. Z. (2004, June). From Laboratory to Field. Using Information from Arabidopsis to Engineer Salt, Cold, and Drought Tolerance in Crops. *PLANT PHYSIOLOGY 135*(2), 615–621.

[342] Zheng, Y., N. Ren, H. Wang, A. J. Stromberg, and S. E. Perry (2009, October). Global Identification of Targets of the Arabidopsis MADS Domain Protein AGAMOUS-Like15. *The Plant cell 21*(9), 2563–2577.

[343] Zheng, Z., S. A. Qamar, Z. Chen, and T. Mengiste (2006, November). Arabidopsis WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens. *The Plant Journal 48*(4), 592–605.

[344] Zhou, N., T. L. Tootle, F. Tsui, D. F. Klessig, and J. Glazebrook (1998, June). PAD4 functions upstream from salicylic acid to control defense responses in Arabidopsis. *The Plant cell 10*(6), 1021–1030.

[345] Zhu, J. K. (2000, November). Genetic analysis of plant salt tolerance using Arabidopsis. *PLANT PHYSIOLOGY 124*(3), 941–948.

[346] Zipfel, C. (2008, February). Pattern-recognition receptors in plant innate immunity. *Current Opinion in Immunology 20*(1), 10–16.

[347] Zipfel, C., S. Robatzek, L. Navarro, E. J. Oakeley, J. D. G. Jones, G. Felix, and T. Boller (2004, April). Bacterial disease resistance in Arabidopsis through flagellin perception. *Nature 428*(6984), 764–767.

[348] Zou, C., K. Sun, J. D. Mackaluso, A. E. Seddon, R. Jin, M. F. Thomashow, and S.-H. Shiu (2011, September). Cis-regulatory code of stress-responsive transcription in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences 108*(36), 14992–14997.

# Appendix A

# Oligonucleotides for cloning promoter fragments

Table A.1: Oligonucleotides for Y1H promoter fragments with Gateway cloning

*a)*

| | *MYB102* promoter |
|---|---|
| | Fragment 1 |
| Forward oligo | 5' - aaa-aaa-gca-ggc-ttc-TAT-CAT-CTT-GGA-AAT-ATA-AAA-TGT-AAA-CAC-G - 3' |
| Reverse oligo | 5' - caa-gaa-agc-tgg-gtc-AGG-GTC-CGT-AAG-GGG-AAG-TAC-AAA-GTA-TTT-ATA-GGG - 3' |
| | Fragment 2 |
| Forward oligo | 5' - aaa-aaa-gca-ggc-ttc-TTA-ACA-AAA-CTG-CAC-TTT-TTT-CAA-CGT-CAC-AGC - 3' |
| Reverse oligo | 5' - caa-gaa-agc-tgg-gtc-GAT-AGT-GTT-TTA-GTC-TTT-GGA-AAT-ATA-GAA-ATA-TAG-ATG-C - 3' |
| | Fragment 3 |
| Forward oligo | 5' - aaa-aaa-gca-ggc-ttc-TTG-AAT-GTA-CAA-TGA-AAC-TAC-ATA-TTT-CTA-C - 3' |
| Reverse oligo | 5' - caa-gaa-agc-tgg-gtc-TAC-GAA-ACC-TTG-AAA-CAA-ATG- |

| | |
|---|---|
| | TC - 3’ |
| | *b)* |
| | *TCP13* promoter |
| | Fragment 1 |
| Forward oligo | 5’ - aaa-aaa-gca-ggc-ttc-TAT-TAT-ATA-TAA-CTT-CAC-GTC-AAT-GTA-TGT-TTG-ATT-TTG-GC - 3’ |
| Reverse oligo | 5’ - caa-gaa-agc-tgg-gtc-AGA-AAG-AAG-AAG-ATG-CTT-TTG-GAA-GTG-AAT-ATG-AGA-ACC-C - 3’ |
| | Fragment 2 |
| Forward oligo | 5’ - aaa-aaa-gca-ggc-ttc-TAG-ATC-CAA-CTT-CAT-TTC-AAT-ATA-TCA-CGA-GTG-C - 3’ |
| Reverse oligo | 5’ - caa-gaa-agc-tgg-gtc-AAG-AAG-AAA-ATC-CCA-AAA-AAA-GTA-TAC-GCG - 3’ |
| | Fragment 3 |
| Forward oligo | 5’ - aaa-aaa-gca-ggc-ttc-CCA-TAT-ACT-GTG-TGT-ATA-TTA-TAT-ATT-ACC - 3’ |
| Reverse oligo | 5’ - caa-gaa-agc-tgg-gtc-TAT-ATT-ATT-TAT-GAT-CCT-TGA-TTT-TTT-TTT-TTG-CTC - 3’ |
| | *c)* |
| | *BANQUO2* promoter |
| | Fragment 1 |
| Forward oligo | 5’ - aaa-aaa-gca-ggc-ttc-TCT-GAA-TAT-AAG-AAA-ATA-GAA-GAT-ATA-TAT-CTA-TCT-TCG-G - 3’ |
| Reverse oligo | 5’ - caa-gaa-agc-tgg-gtc-AAG-AAA-GAA-GGA-GAA-GAA-ATG-GTG - 3’ |
| | Fragment 2 |
| Forward oligo | 5’ - aaa-aaa-gca-ggc-ttc-TAA-CAG-ATA-TTA-TAA-TGT-TAC-TAA-TTA-AC - 3’ |
| Reverse oligo | 5’ - caa-gaa-agc-tgg-gtc-TGT-CTA-ACT-TTT-ATC-TAT- |

| | |
|---|---|
| | TTA-CTT-TAA-AAA-AAG-ACA-GC - 3’ |
| | Fragment 3 |
| Forward oligo | 5’ - aaa-aaa-gca-ggc-ttc-TCT-AAC-AAG-ACA-TAT-TAG-TCG-AGC-TTT-TGG - 3’ |
| Reverse oligo | 5’ - caa-gaa-agc-tgg-gtc-ATT-ATA-ATG-CTA-TTA-TAT-TTT-TAA-TAT-TGT-AG - 3’ |
| | *d)* *At1g71520* promoter |
| | Fragment 1 |
| Forward oligo | 5’ - aaa-aaa-gca-ggc-ttc-GTT-TTC-TTT-TCA-AGT-TTC-AGC - 3’ |
| Reverse oligo | 5’ - caa-gaa-agc-tgg-gtc-TCA-AGT-CCG-AAT-AAA-TTG-AAT-GAA-TG - 3’ |
| | Fragment 2 |
| Forward oligo | 5’ - aaa-aaa-gca-ggc-ttc-CAT-TGT-AGA-GAC-TAG-AGA-GTA-CAT-TGC - 3’ |
| Reverse oligo | 5’ - caa-gaa-agc-tgg-gtc-ACG-ACT-TGA-TTC-AAA-AAG-TCT-G - 3’ |
| | Fragment 3 |
| Forward oligo | 5’ - aaa-aaa-gca-ggc-ttc-TGT-ATT-TGA-CAT-ACA-ACA-CCT-C - 3’ |
| Reverse oligo | 5’ - caa-gaa-agc-tgg-gtc-GTG-GTA-CTC-AAA-TTT-AAC-AC - 3’ |
| | *d)* *At2g33710* promoter |
| | Fragment 1 |
| Forward oligo | 5’ - aaa-aaa-gca-ggc-ttc-TGT-TCA-GGA-TGA-AAA-TAA-TTA-TTT-AAA-TCC-ACC - 3’ |
| Reverse oligo | 5’ - caa-gaa-agc-tgg-gtc-GTC-TAT-TCT-CTG-CAT-CTA-TGT- |

| | |
|---|---|
| | TAT-ATA-TAG-ACG-C - 3’ |
| | Fragment 2 |
| Forward oligo | 5’ - aaa-aaa-gca-ggc-ttc-TGC-TAT-TTT-ATT-TAT-GTG-ATT-TAT-ACA-GTA-CG - 3’ |
| Reverse oligo | 5’ - caa-gaa-agc-tgg-gtc-ATG-ATC-CAA-GTC-TTG-TGT-TAA-TTA-CGG - 3’ |
| | Fragment 3 |
| Forward oligo | 5’ - aaa-aaa-gca-ggc-ttc-AAA-TGG-CAC-GAA-TCG-TCC-AAA-TTG-C - 3’ |
| Reverse oligo | 5’ - caa-gaa-agc-tgg-gtc-ATA-TAT-ATA-TAT-ATA-TAT-ATA-TAT-AGA-TAC-ATA-TGT-GTT-TTG-C - 3’ |
| | *e)* |
| | *At3g53600* promoter |
| | Fragment 1 |
| Forward oligo | 5’ - aaa-aaa-gca-ggc-ttc-ATA-ATA-TTA-AAC-ATG-TTA-GAC-AGA-CG - 3’ |
| Reverse oligo | 5’ -caa-gaa-agc-tgg-gtc-GTT-TTT-ACG-GTT-GTA-TTT-CGA-AAA-GAG-AAA-AGA-GC - 3’ |
| | Fragment 2 |
| Forward oligo | 5’ - aaa-aaa-gca-ggc-ttc-GAC-TTT-CAA-CTC-TTT-AAT-CTA-AAG-TCT-AAA-CCA-CAA-TCT-AAA-TCC-GG - 3’ |
| Reverse oligo | 5’ - caa-gaa-agc-tgg-gtc-CTT-GCC-ATA-TTT-TAT-TTT-GTT-TGG-TAA-CCG - 3’ |
| | Fragment 3 |
| Forward oligo | 5’ - aaa-aaa-gca-ggc-ttc-GAC-GGC-GTC-GCG-CCT-CCT-TTG-TCG-G - 3’ |
| Reverse oligo | 5’ - caa-gaa-agc-tgg-gtc-CTA-TGT-GTG-TGA-TAT-AAG-ACA-TCA-TCA-ATG-ATG-G - 3’ |
| | Generic oligos |

| | |
|---|---|
| Forward oligo | 5 - GGG-GAC-AAG-TTT-GTA-CAA-AAA-AGC-AGG-CT - 3 |
| Reverse oligo | 5 - GGG-GAC-CAC-TTT-GTA-CAA-GAA-AGC-TGG-GT - 3 |

Table A.2: Oligonucleotides for Y1H promoter fragments of *MYB2* with Gateway cloning, where the fragments produced will be 100bp in length, with the exception of Fragment 5, which is of length 120bp

| *MYB2* promoter | |
|---|---|
| Fragment 1 | |
| Forward oligo | 5' - aaa-aaa-gca-ggc-ttc-ATA-TCC-TTT-TTA-TAA-AAT-ACT-AC - 3' |
| Reverse oligo | 5' - caa-gaa-agc-tgg-gtc-AGA-TTT-GAA-GTG-ATT-AAG-CAA-TGT-GCG - 3' |
| Fragment 2 | |
| Forward oligo | 5' - aaa-aaa-gca-ggc-ttc-AAA-AAA-TAA-AAA-TTG-AAC - 3' |
| Reverse oligo | 5' - caa-gaa-agc-tgg-gtc-TTT-AAA-AGT-AGT-ATT-TTA-TAA-AAA-GG - 3' |
| Fragment 3 | |
| Forward oligo | 5' - aaa-aaa-gca-ggc-ttc-AAA-TGG-AGA-GCT-AAT-TAT-GTT-TAG-C - 3' |
| Reverse oligo | 5' - caa-gaa-agc-tgg-gtc-CAT-AAG-TGT-TAT-GTT-C - 3' |
| Fragment 4 | |
| Forward oligo | 5' - aaa-aaa-gca-ggc-ttc-GTG-TTA-CTA-TAC-ATC-TGA-AC - 3' |
| Reverse oligo | 5' - caa-gaa-agc-tgg-gtc-ATT-ATG-CTA-AAC-ATA-ATT-AGC - 3' |
| Fragment 5 | |
| Forward oligo | 5' - aaa-aaa-gca-ggc-ttc-ATA-ACT-TAC-GTC-TGC-GAT-AC - 3' |
| Reverse oligo | 5' - caa-gaa-agc-tgg-gtc-CCT-TTG-ACT-TGT-TCA-GAT-GTA-TAG-TAA-CAC-GC - 3' |

Table A.3: Oligonucleotides for sequencing entry clones into pDONRZeo

| | |
|---|---|
| Forward oligo | 5' - GTAAAACGACGGCCAG - 3' |
| Reverse oligo | 5' - CAGGAAACAGCTATGAC - 3' |

Table A.4: Oligonucleotides for performing colony PCR on growing colonies from Y1H screens

| | |
|---|---|
| Forward oligo | 5' - CTA-ACG-TTC-ATG-ATA-ACT-TCA-TG - 3' |
| Reverse oligo | 5' - GAA-GTG-TCA-ACA-AVG-TAT-CTA-CC - 3' |

# Appendix B

# Gene members and expression profiles of potential regulons presented in Table 4.9

All ATG identifiers and gene names presented in the subsequent tables are correct according to TAIR9 release of the CATMA probe annotation mapping version 8.

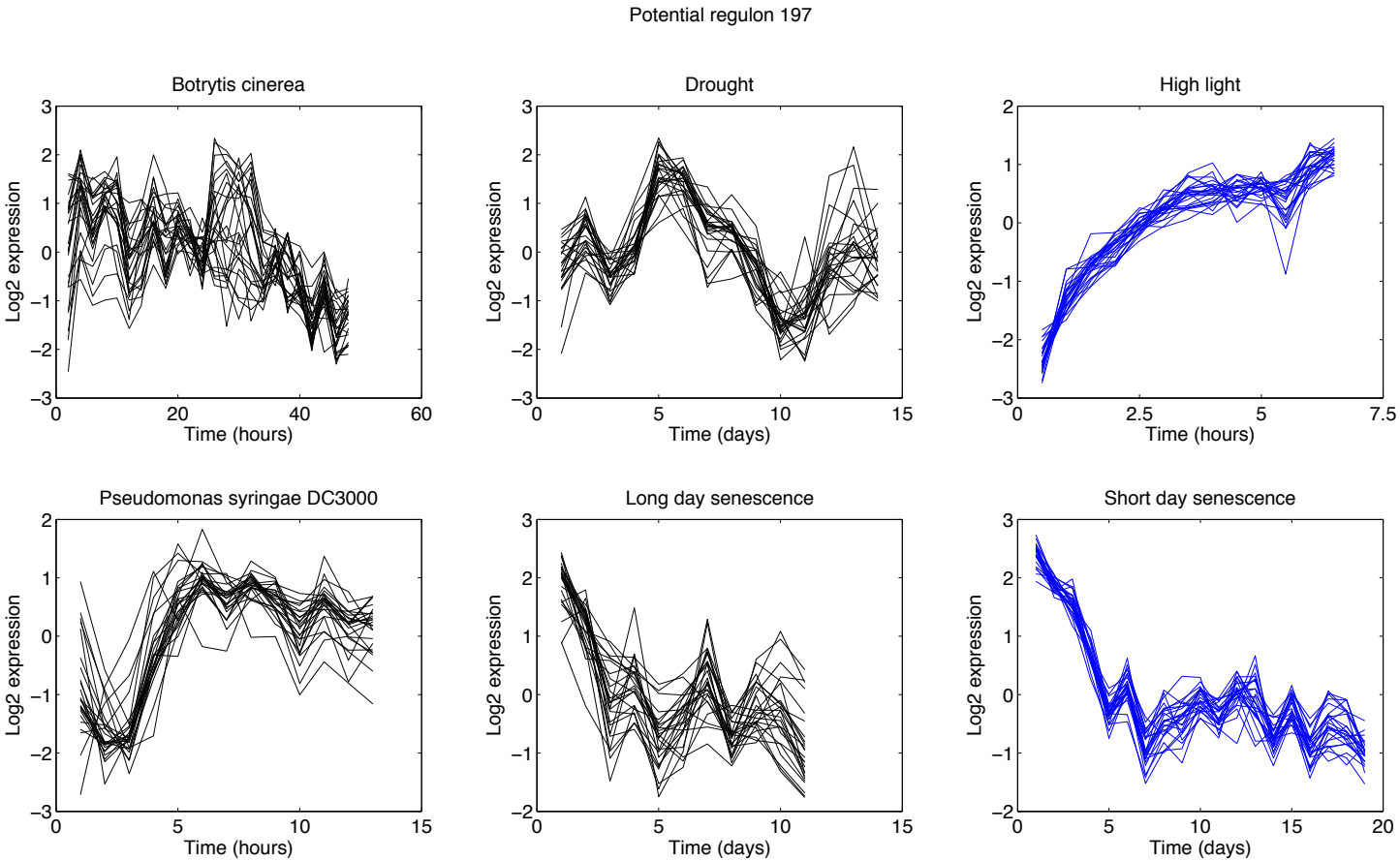



Figure B.1: Potential regulon 197 presented in Table 4.9. Expression profiles of genes significantly co-expressed in high light and short day senescence stress conditions. Gene membership is presented in Table B.1

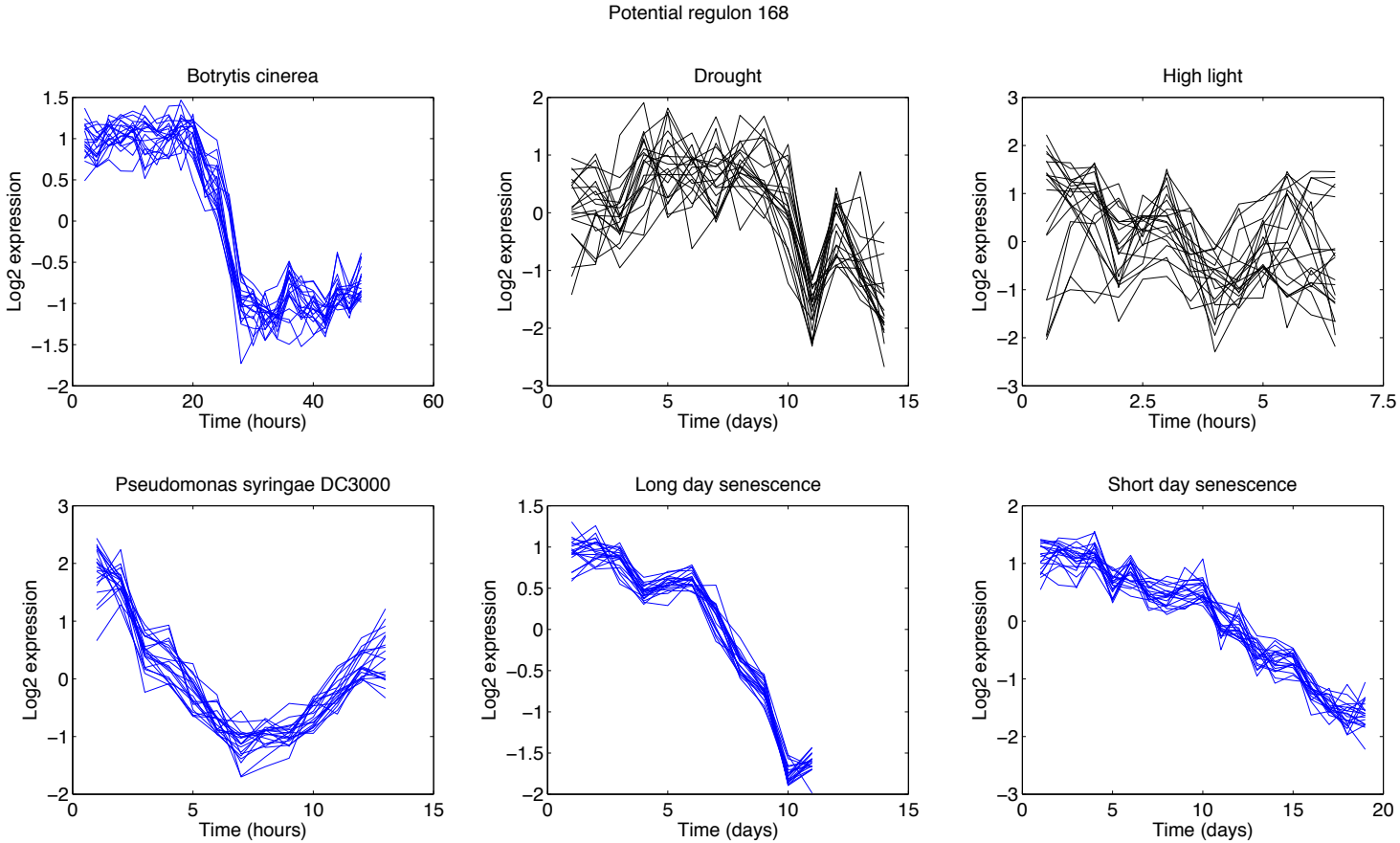



Figure B.2: Potential regulon 168 presented in Table 4.9. Expression profiles of genes significantly co-expressed in *B. cinerea*, *P. syringae* DC3000, long day and short day senescence stress conditions. Gene membership is presented in Table B.2

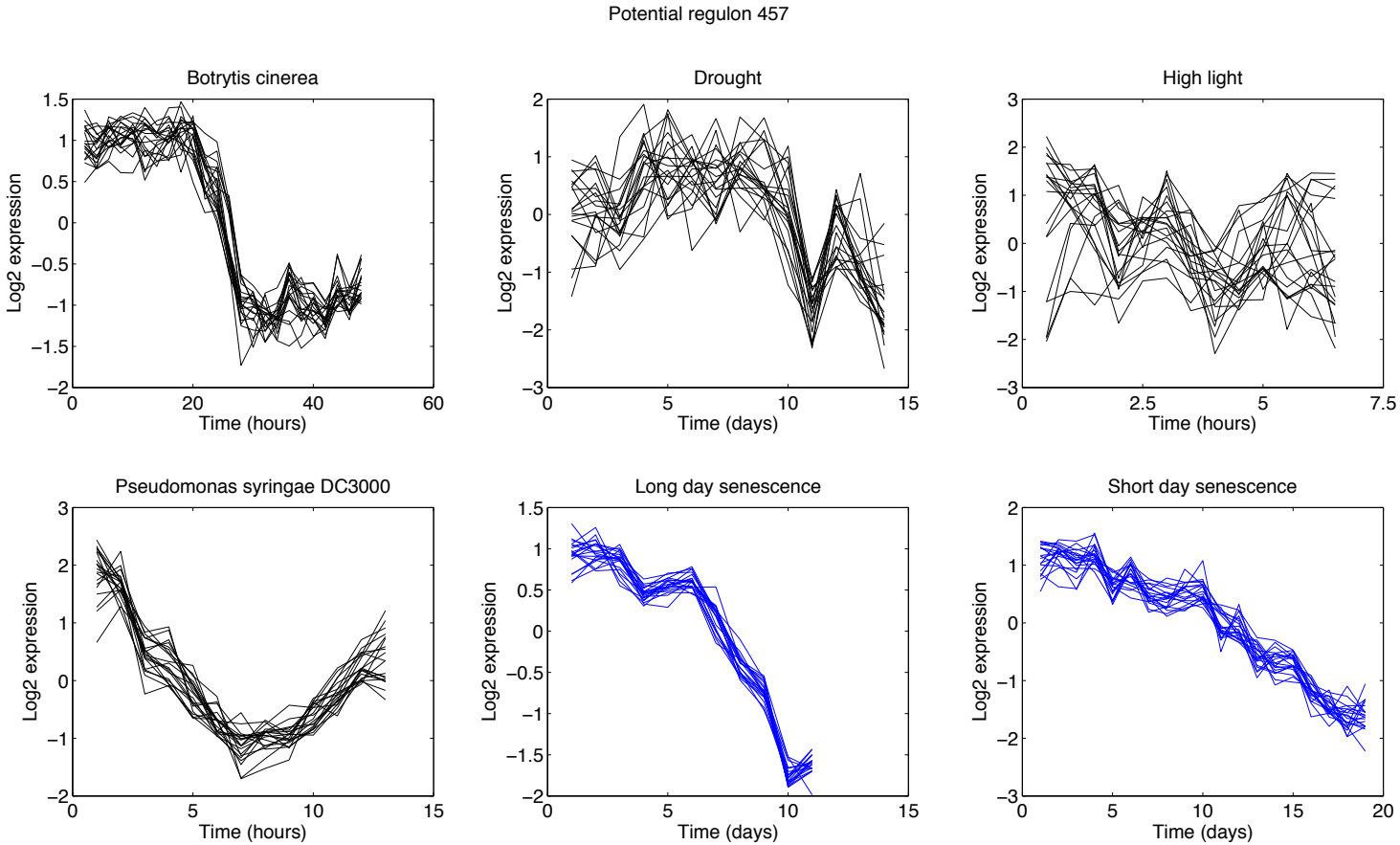



Figure B.3: Potential regulon 457 presented in Table 4.9. Expression profiles of genes significantly co-expressed in long day and short day senescence stress conditions. Gene membership is presented in Table B.3

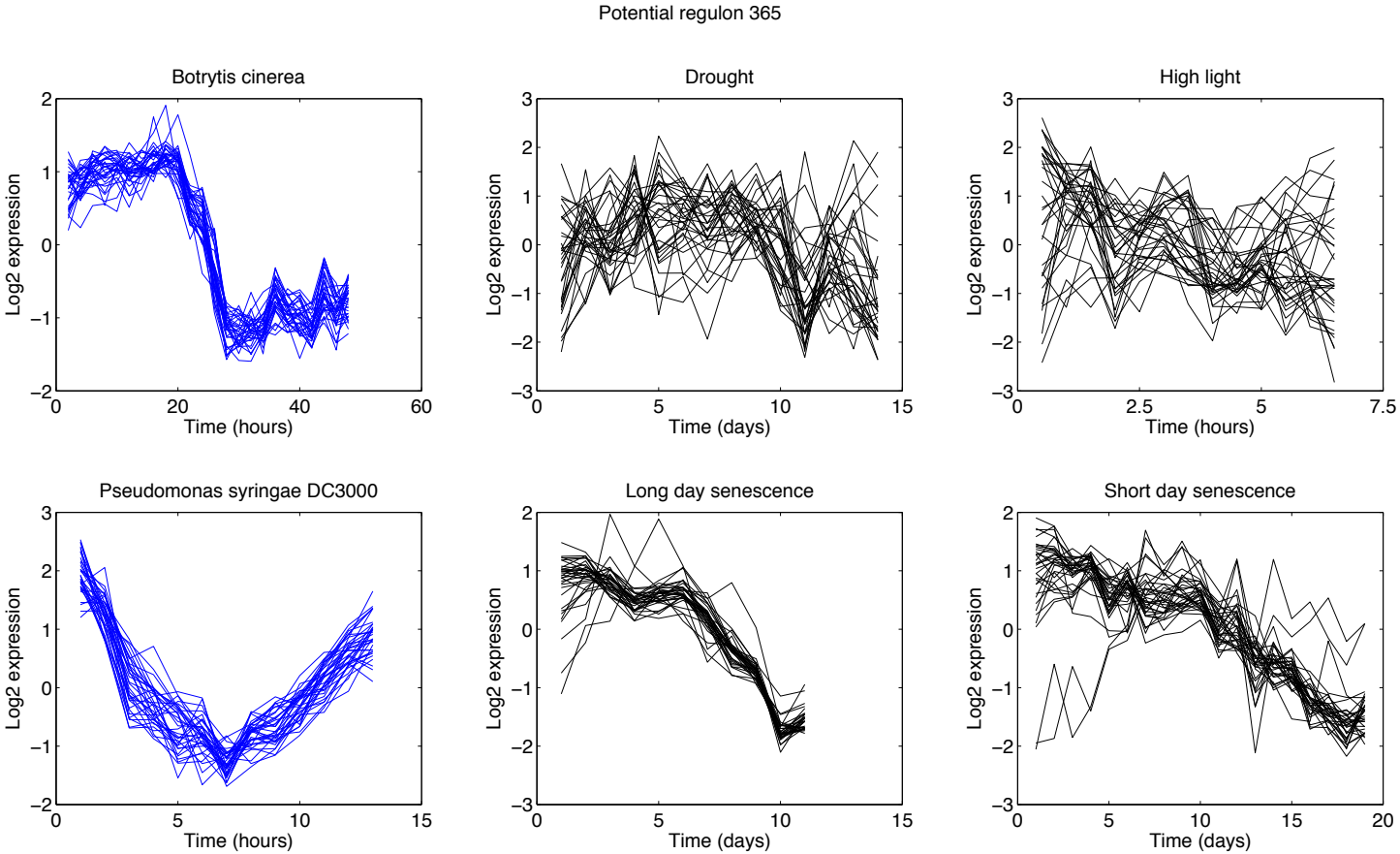



Figure B.4: Potential regulon 365 presented in Table 4.9. Expression profiles of genes significantly co-expressed in *B. cinerea* and *P. syringae* DC3000 stress conditions. Gene membership is presented in Table B.4

Table B.1: Gene members of potential regulons presented in Table 4.9, showing ATG identifiers and common gene name, where applicable

| Potential regulon 197, where genes are co-expressed over high light and short day senescence stress conditions | |
|---|---|
| ATG identifier | Gene name |
| AT1G77940 | |
| AT1G04480 | |
| AT1G12960 | |
| AT1G70600 | |
| AT2G32220 | |
| AT2G35240 | |
| AT2G41650 | |
| AT2G25210 | |
| AT2G31140 | |
| AT2G37270 | ATRPS5B |
| AT3G23940 | |
| AT3G44590 | |
| AT3G61100 | |
| AT3G06680 | |
| AT3G07110 | |
| AT4G15000 | |
| AT4G25740 | |
| AT4G30800 | |
| AT4G38100 | |
| AT4G18100 | |
| AT5G02870 | |
| AT5G23900 | |
| AT5G53070 | |
| AT5G60670 | |
| AT5G67510 | |

Table B.2: Gene members of potential shared regulons presented in Table 4.9, showing ATG identifiers and common gene name, where applicable

| Potential shared regulon 168, where genes are co-expressed over *B. cinerea*, *P. syringae* DC3000, long day and short day senescence stress conditions | |
|---|---|
| ATG identifier | Gene name |
| AT1G03600 | |
| AT1G06680 | PSBP-1 |
| AT1G15820 | LHCB6 |
| AT1G21500 | |
| AT1G30380 | PSAK |
| AT1G74730 | |
| AT1G75690 | |
| AT2G30570 | PSBW |
| AT2G30790 | PSBP-2 |
| AT2G39470 | PPL2 |
| AT3G08920 | |
| AT3G54050 | |
| AT4G17560 | |
| AT4G24750 | |
| AT4G28750 | PSAE-1 |
| AT5G36700 | PGLP1 |
| AT5G38410 | |
| AT5G51110 | |
| AT5G53490 | |
| AT5G64040 | PSAN |

Table B.3: Gene members of potential shared regulons presented in Table 4.9, showing ATG identifiers and common gene name, where applicable

| Potential shared regulon 457, where genes are co-expressed over long day and short day senescence stress conditions | |
|---|---|
| ATG identifier | Gene name |
| AT1G07070 | |
| AT1G14410 | WHY1 |
| AT1G18440 | |
| AT1G26880 | |
| AT1G31660 | |
| AT1G48830 | |
| AT1G72370 | P40 |
| AT1G77750 | |
| AT1G77940 | |
| AT2G02450 | ANAC034/ANAC035 |
| AT3G09630 | |
| AT2G31610 | |
| AT2G31725 | |
| AT2G32060 | |
| AT2G36620 | RPL24A |
| AT2G38300 | |
| AT2G19670 | PRMT1A |
| AT2G27710 | |
| AT2G37270 | ATRPS5B |
| AT3G04840 | |
| AT3G05560 | |
| AT3G14390 | |
| AT3G21300 | |
| AT3G23940 | |
| AT3G23990 | HSP60 |
| AT3G47370 | |
| AT3G51190 | |
| AT3G56340 | |
| AT3G60245 | |
| AT3G06680 | |
| AT3G07110 | |
| AT3G16780 | |
| AT3G25520 | ATL5 |
| AT3G28900 | |
| AT4G10480 | |
| AT4G12600 | |
| AT4G13170 | |
| AT4G16141 | |
| AT4G17390 | |
| AT4G25890 | |
| AT4G31700 | RPS6 |
| AT4G31710 | ATGLR2.4 |
| AT4G16720 | |
| AT5G02870 | |
| AT5G09510 | |
| AT5G10920 | |
| AT5G16130 | |
| AT5G20720 | CPN20 |
| AT5G22440 | |

Table B.4: Gene members of potential shared regulons presented in Table 4.9, showing ATG identifiers and common gene name, where applicable

| Potential shared regulon 365, where genes are co-expressed over *B. cinerea* and *P. syringae* DC3000 stress conditions | |
|---|---|
| ATG identifier | Gene name |
| AT1G03600 | |
| AT1G06680 | PSBP-1 |
| AT1G12900 | GAPA-2 |
| AT1G18460 | |
| AT1G30380 | PSAK |
| AT1G32060 | PRK |
| AT1G50900 | |
| AT1G52230 | PSAH2 |
| AT1G74470 | |
| AT1G75690 | |
| AT1G16460 | ATRDH2 |
| AT2G21330 | |
| AT2G30790 | PSBP-2 |
| AT3G08940 | LHCB4.2 |
| AT3G12345 | |
| AT3G50685 | |
| AT3G55800 | SBPASE |
| AT3G56910 | PSRP5 |
| AT3G56940 | CRD1 |
| AT3G63520 | CCD1 |
| AT4G01800 | |
| AT4G02920 | |
| AT4G25050 | ACP4 |
| AT4G25080 | CHLM |
| AT5G13630 | GUN5 |
| AT5G14910 | |
| AT5G16400 | ATF2 |
| AT5G17870 | PSRP6 |
| AT5G19940 | |
| AT5G23060 | CaS |
| AT5G51110 | |
| AT5G57345 | |
| AT5G58250 | |

Table B.5: Gene members of potential shared regulons presented in Table 4.9, showing ATG identifiers and common gene name, where applicable

| Potential shared regulon 199, where genes are co-expressed over *B. cinerea* and short day senescence stress conditions | |
|---|---|
| ATG identifier | Gene name |
| AT1G07070 | |
| AT1G22780 | PFL |
| AT1G70600 | |
| AT2G02450 | ANAC034/ANAC035 |
| AT2G05120 | |
| AT2G32060 | |
| AT2G33370 | |
| AT2G35240 | |
| AT2G38300 | |
| AT2G41650 | |
| AT2G44120 | |
| AT2G37190 | |
| AT2G39390 | |
| AT3G05560 | |
| AT3G05590 | RPL18 |
| AT3G09500 | |
| AT3G47370 | |
| AT3G57490 | |
| AT4G25740 | |
| AT4G34555 | |
| AT4G36130 | |
| AT5G02450 | |
| AT5G09510 | |
| AT5G23535 | |
| AT5G23740 | RPS11-BETA |
| AT5G45775 | |
| AT5G56710 | |
| AT5G58420 | |

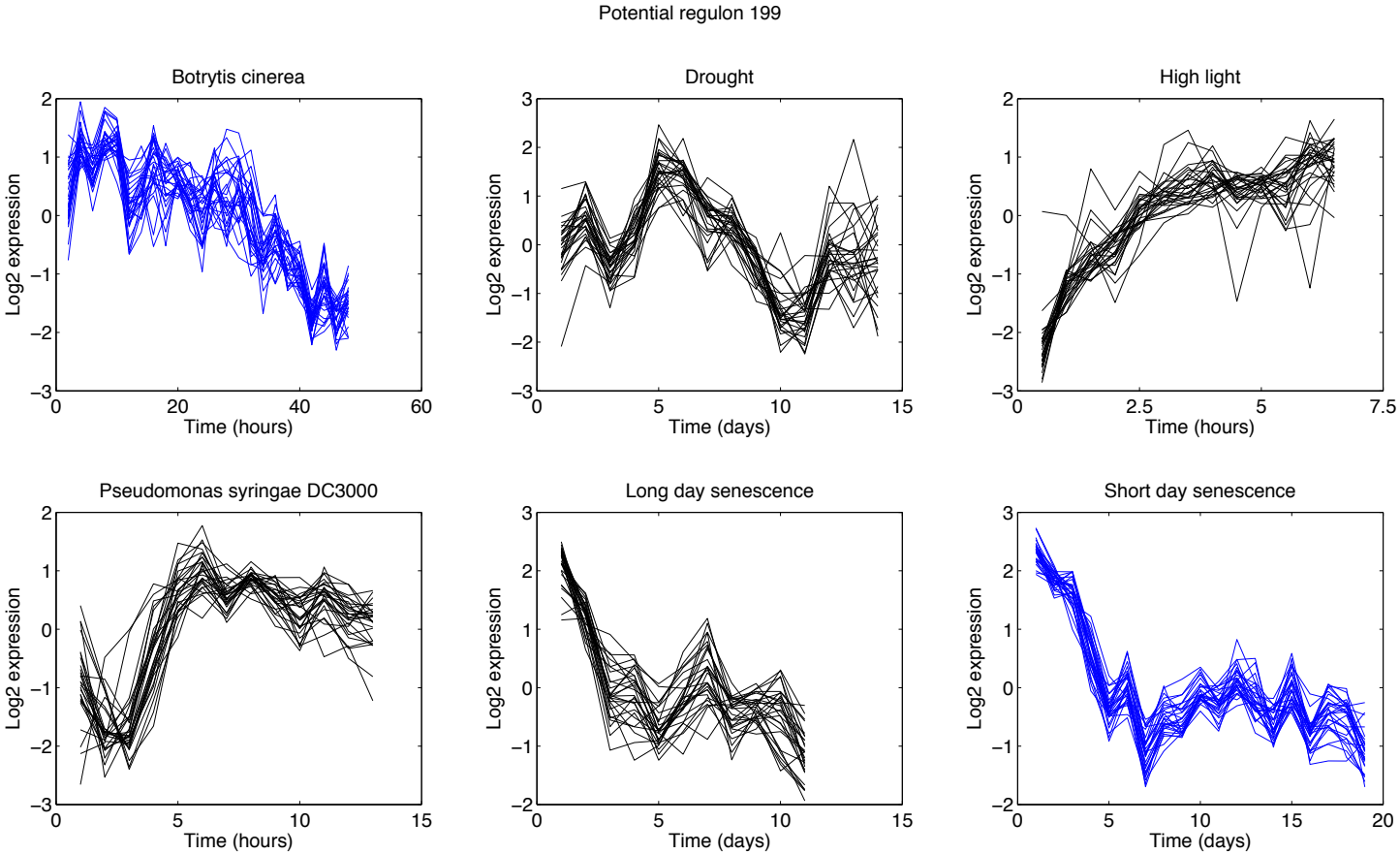



Figure B.5: Potential regulon 199 presented in Table 4.9. Expression profiles of genes significantly co-expressed in *B. cinerea* and short day senescence stress conditions. Gene membership is presented in Table B.5

Table B.6: Gene members of potential shared regulons presented in Table 4.9, showing ATG identifiers and common gene name, where applicable

| Potential shared regulon 320, where genes are co-expressed over long and short day senescence stress conditions | |
|---|---|
| ATG identifier | Gene name |
| AT1G02770 | |
| AT1G14410 | WHY1 |
| AT1G17560 | HLL |
| AT1G26880 | |
| AT1G44960 | |
| AT1G48570 | |
| AT1G48830 | |
| AT1G54690 | GAMMA-H2AX |
| AT1G60770 | |
| AT1G77750 | |
| AT2G02450 | ANAC034/ANAC035 |
| AT2G02740 | WHY3 |
| AT2G31610 | |
| AT2G31725 | |
| AT2G32060 | |
| AT2G33210 | HSP60-2 |
| AT2G19670 | PRMT1A |
| AT2G27710 | |
| AT3G04840 | |
| AT3G23830 | GRP4 |
| AT3G23940 | |
| AT3G47370 | |
| AT3G51190 | |
| AT3G54090 | |
| AT3G60245 | |
| AT3G06680 | |
| AT3G07110 | |
| AT3G28900 | |
| AT4G10480 | |
| AT4G12600 | |
| AT4G13170 | |
| AT4G16141 | |
| AT4G25630 | FIB2 |
| AT4G25890 | |
| AT4G31710 | ATGLR2.4 |
| AT5G07090 | |
| AT5G11340 | |
| AT5G15520 | |
| AT5G16130 | |
| AT5G20720 | CPN20 |
| AT5G22440 | |
| AT5G23535 | |
| AT5G27820 | |
| AT5G47700 | |
| AT5G52650 | |
| AT5G60390 | |
| AT5G60670 | |
| AT5G64670 | |

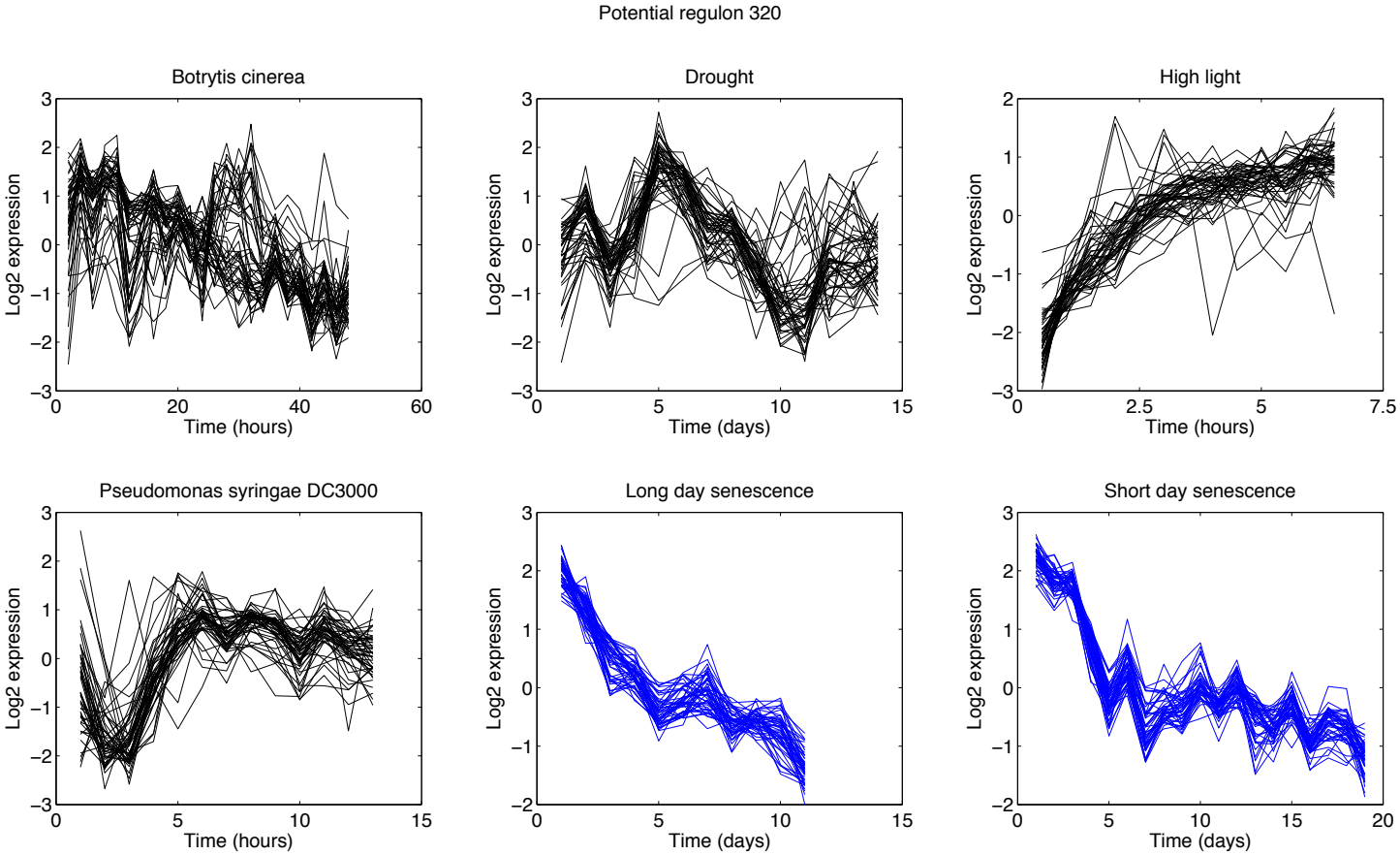



Figure B.6: Potential regulon 320 presented in Table 4.9. Expression profiles of genes significantly co-expressed in long and short day senescence stress conditions. Gene membership is presented in Table B.6

Table B.7: Gene members of potential shared regulons presented in Table 4.9, showing ATG identifiers and common gene name, where applicable

| Potential shared regulon 408, where genes are co-expressed over high light and short day senescence stress conditions | |
|---|---|
| ATG identifier | Gene name |
| AT1G09690 | |
| AT1G26910 | |
| AT1G04480 | |
| AT1G70600 | |
| AT2G35240 | |
| AT2G41650 | |
| AT2G19670 | PRMT1A |
| AT2G27720 | |
| AT3G23940 | |
| AT3G44590 | |
| AT3G53740 | |
| AT3G61100 | |
| AT3G04920 | |
| AT3G07110 | |
| AT4G00620 | |
| AT4G30800 | |
| AT4G38100 | |
| AT4G18100 | |
| AT5G02870 | |
| AT5G23900 | |
| AT5G60670 | |
| AT5G61170 | |
| AT5G67510 | |

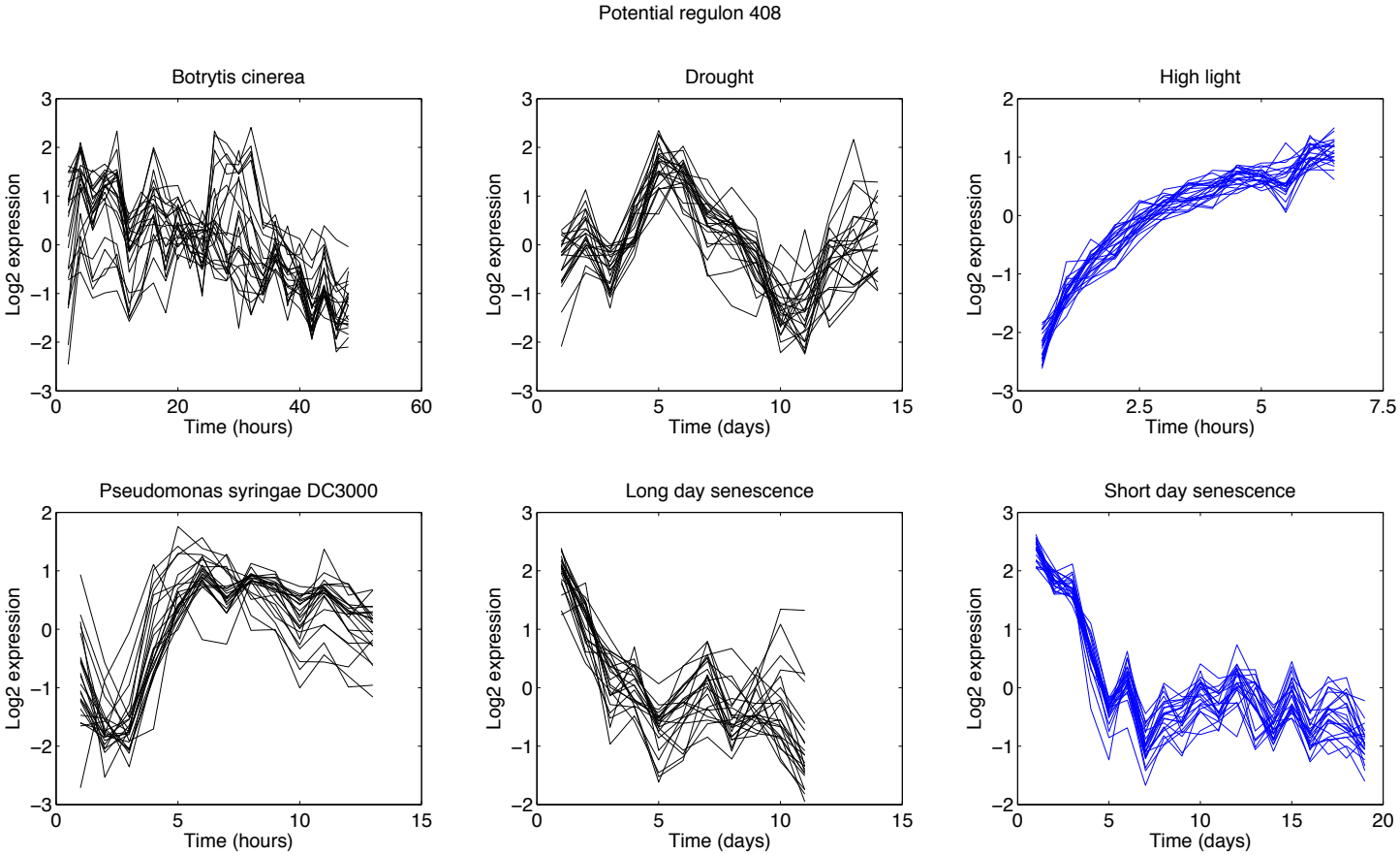



Figure B.7: Potential regulon 408 presented in Table 4.9. Expression profiles of genes significantly co-expressed in high light and short day senescence stress conditions. Gene membership is presented in Table B.7

Table B.8: Gene members of potential shared regulons presented in Table 4.9, showing ATG identifiers and common gene name, where applicable

| Potential shared regulon 117, where genes are co-expressed over long and short day senescence stress conditions | |
|---|---|
| ATG identifier | Gene name |
| AT1G07070 | |
| AT1G13960 | WRKY4 |
| AT1G18440 | |
| AT1G26880 | |
| AT1G35550 | |
| AT1G48830 | |
| AT1G72370 | P40 |
| AT1G77940 | |
| AT2G02450 | ANAC034/ANAC035 |
| AT2G33370 | |
| AT2G36620 | RPL24A |
| AT2G47570 | |
| AT2G37190 | |
| AT2G37270 | ATRPS5B |
| AT3G04350 | |
| AT3G04840 | |
| AT3G05560 | |
| AT3G44750 | HDA3 |
| AT3G47370 | |
| AT3G51190 | |
| AT3G56340 | |
| AT3G60245 | |
| AT3G06680 | |
| AT3G07110 | |
| AT3G16780 | |
| AT3G28900 | |
| AT3G49910 | |
| AT4G13170 | |
| AT4G15000 | |
| AT4G31700 | RPS6 |
| AT5G02870 | |
| AT5G09510 | |
| AT5G52650 | |
| AT5G58420 | |
| AT5G59850 | |
| AT5G60390 | |
| AT5G60670 | |
| AT5G61170 | |
| AT5G63510 | GAMMA-CAL1 |

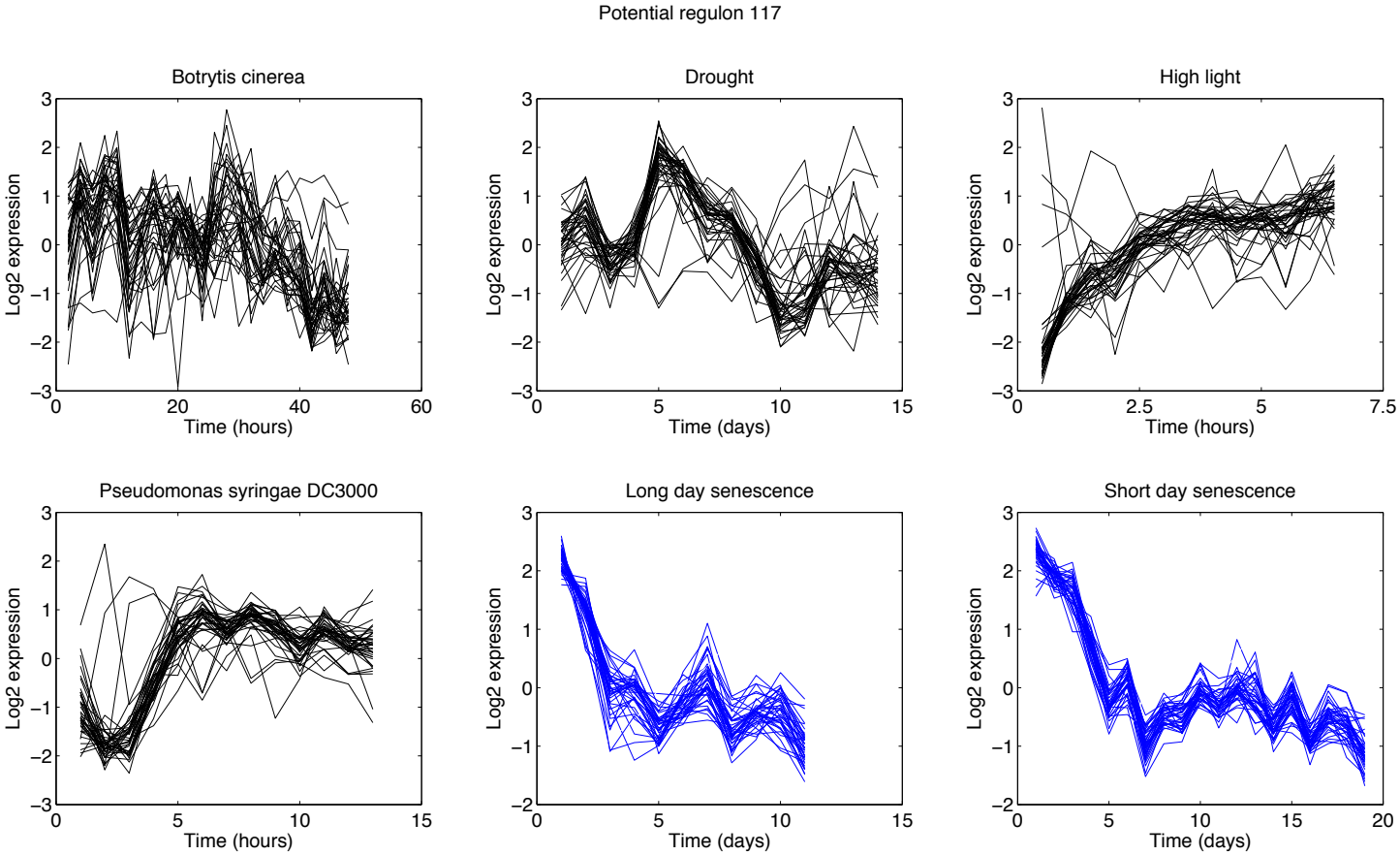



Figure B.8: Potential regulon 117 presented in Table 4.9. Expression profiles of genes significantly co-expressed in long and short day senescence stress conditions. Gene membership is presented in Table B.8

Table B.9: Gene members of potential shared regulons presented in Table 4.9, showing ATG identifiers and common gene name, where applicable

| Potential shared regulon 344, where genes are co-expressed over *B. cinerea* and short day senescence stress conditions | |
|---|---|
| ATG identifier | Gene name |
| AT1G22780 | PFL |
| AT1G24793 | |
| AT1G48830 | |
| AT1G70600 | |
| AT2G32060 | |
| AT2G33370 | |
| AT2G35240 | |
| AT2G41650 | |
| AT3G05560 | |
| AT3G05590 | RPL18 |
| AT3G09500 | |
| AT3G11250 | |
| AT3G53870 | |
| AT3G57490 | |
| AT3G24830 | |
| AT3G58700 | |
| AT4G34555 | |
| AT4G16720 | |
| AT5G02450 | |
| AT5G09510 | |
| AT5G23740 | RPS11-BETA |
| AT5G23900 | |
| AT5G27850 | |
| AT5G45775 | |
| AT5G56710 | |
| AT5G58420 | |

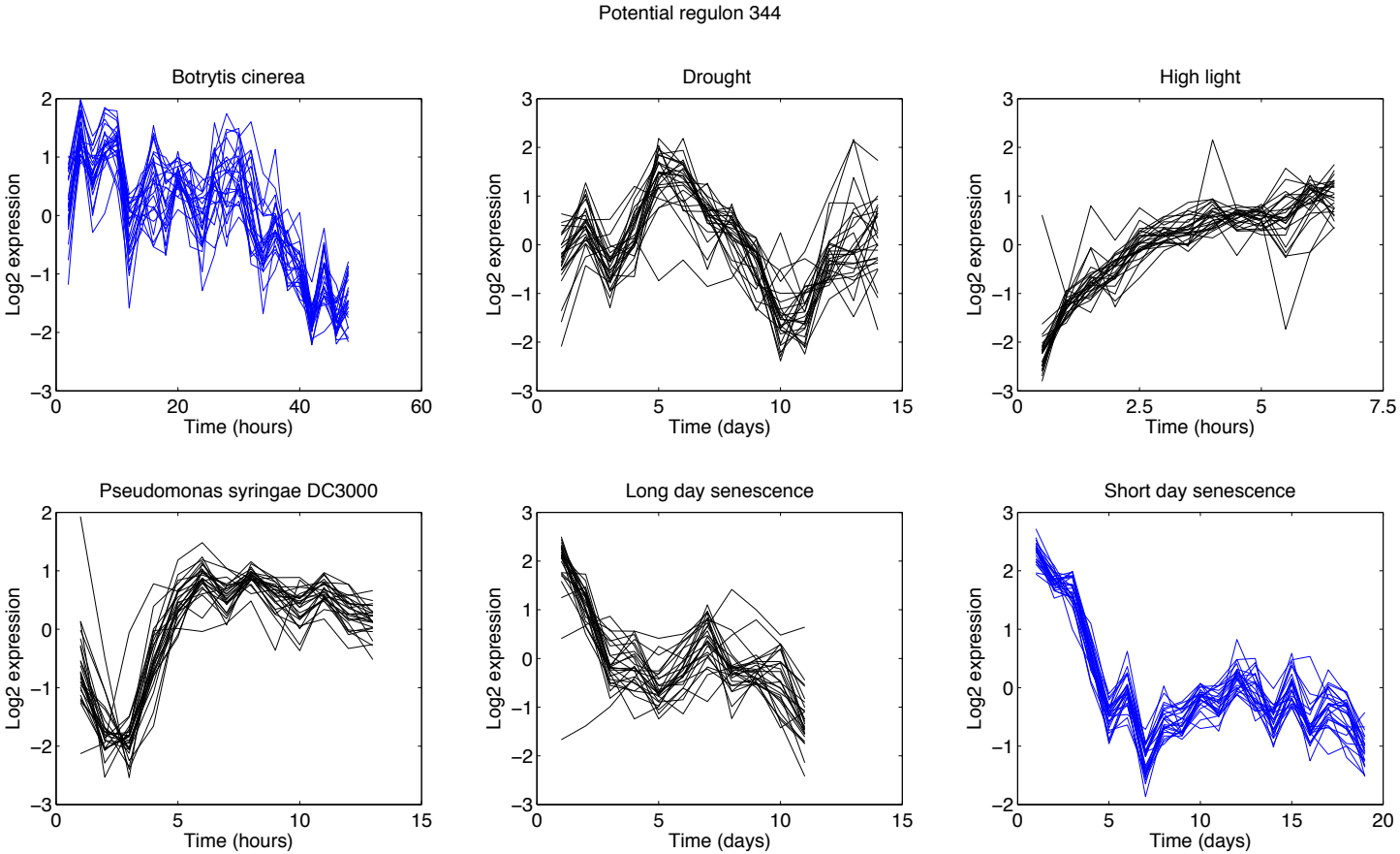



Figure B.9: Potential regulon 344 presented in Table 4.9. Expression profiles of genes significantly co-expressed in *B. cinerea* and short day senescence stress conditions. Gene membership is presented in Table B.9

Table B.10: Gene members of potential shared regulons presented in Table 4.9, showing ATG identifiers and common gene name, where applicable

| Potential shared regulon 280, where genes are co-expressed over *B. cinerea* and high light stress conditions | |
|---|---|
| ATG identifier | Gene name |
| AT1G06680 | PSBP-1 |
| AT1G29930 | CAB1 |
| AT2G30790 | PSBP-2 |
| AT3G16140 | PSAH-1 |
| AT3G46780 | PTAC16 |
| AT3G61470 | LHCA2 |
| AT4G00400 | GPAT8 |
| AT4G15560 | CLA1 |
| AT4G28750 | PSAE-1 |
| AT4G33220 | PME44 |
| AT5G46110 | APE2 |
| AT5G64040 | PSAN |
| AT5G66570 | PSBO1 |

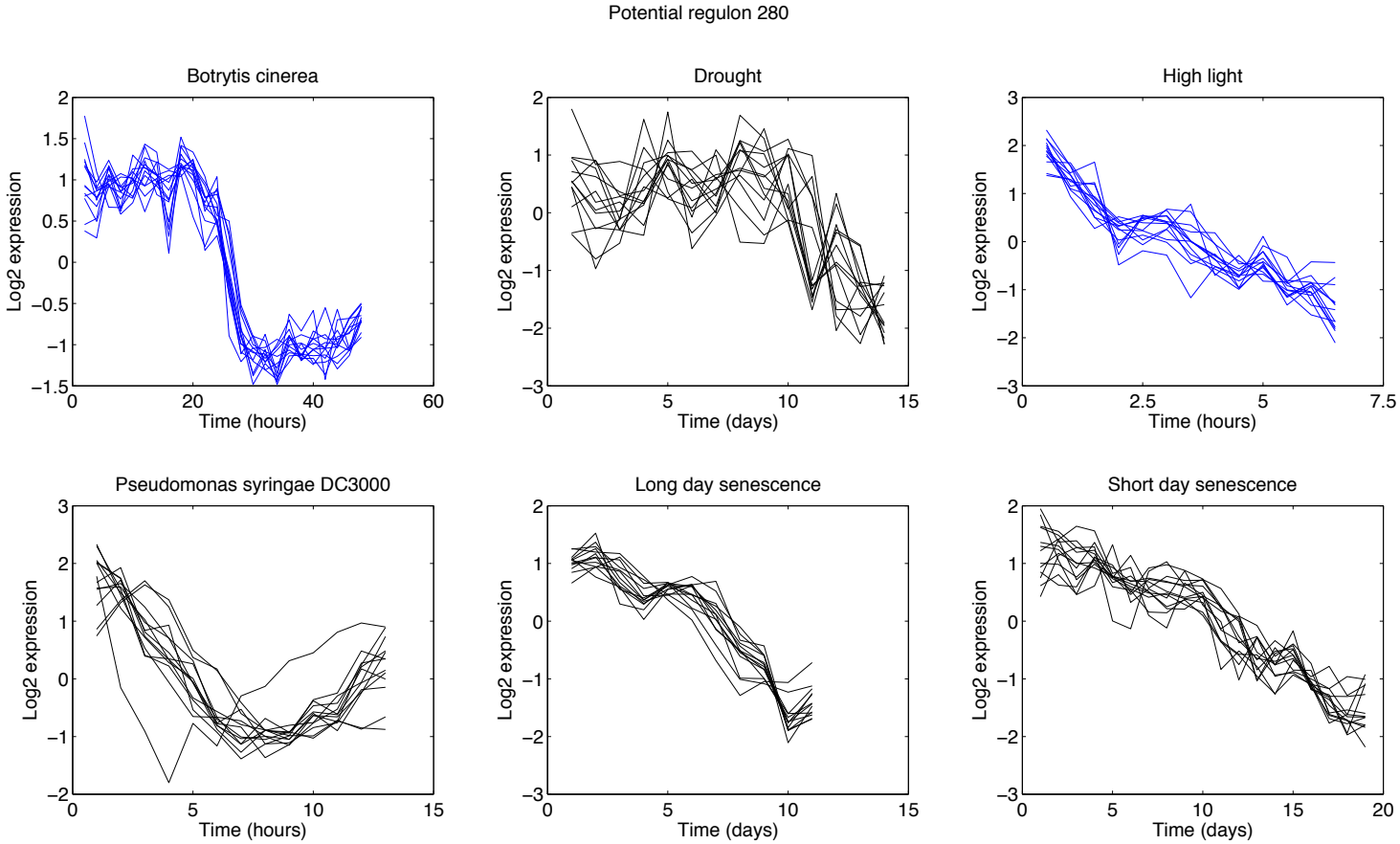



Figure B.10: Potential regulon 280 presented in Table 4.9. Expression profiles of genes significantly co-expressed in *B. cinerea* and high light stress conditions. Gene membership is presented in Table B.10

Table B.11: Gene members of potential shared regulons presented in Table 4.9, showing ATG identifiers and common gene name, where applicable

| Potential shared regulon 416, where genes are co-expressed over high light and short day senescence stress conditions | |
|---|---|
| ATG identifier | Gene name |
| AT1G02830 | |
| AT1G26910 | |
| AT1G48830 | |
| AT1G77940 | |
| AT1G12960 | |
| AT1G74060 | |
| AT2G05220 | |
| AT2G19730 | |
| AT2G21580 | |
| AT2G32220 | |
| AT2G25210 | |
| AT2G37270 | ATRPS5B |
| AT2G39460 | RPL23AA |
| AT2G40590 | |
| AT3G44590 | |
| AT3G06680 | |
| AT3G07110 | |
| AT3G55280 | RPL23AB |
| AT4G00100 | ATRPS13A |
| AT4G15000 | |
| AT4G33070 | |
| AT4G34555 | |
| AT4G18100 | |
| AT4G31985 | |
| AT5G02870 | |
| AT5G50810 | TIM8 |
| AT5G60670 | |
| AT5G25757 | |
| AT5G62300 | |

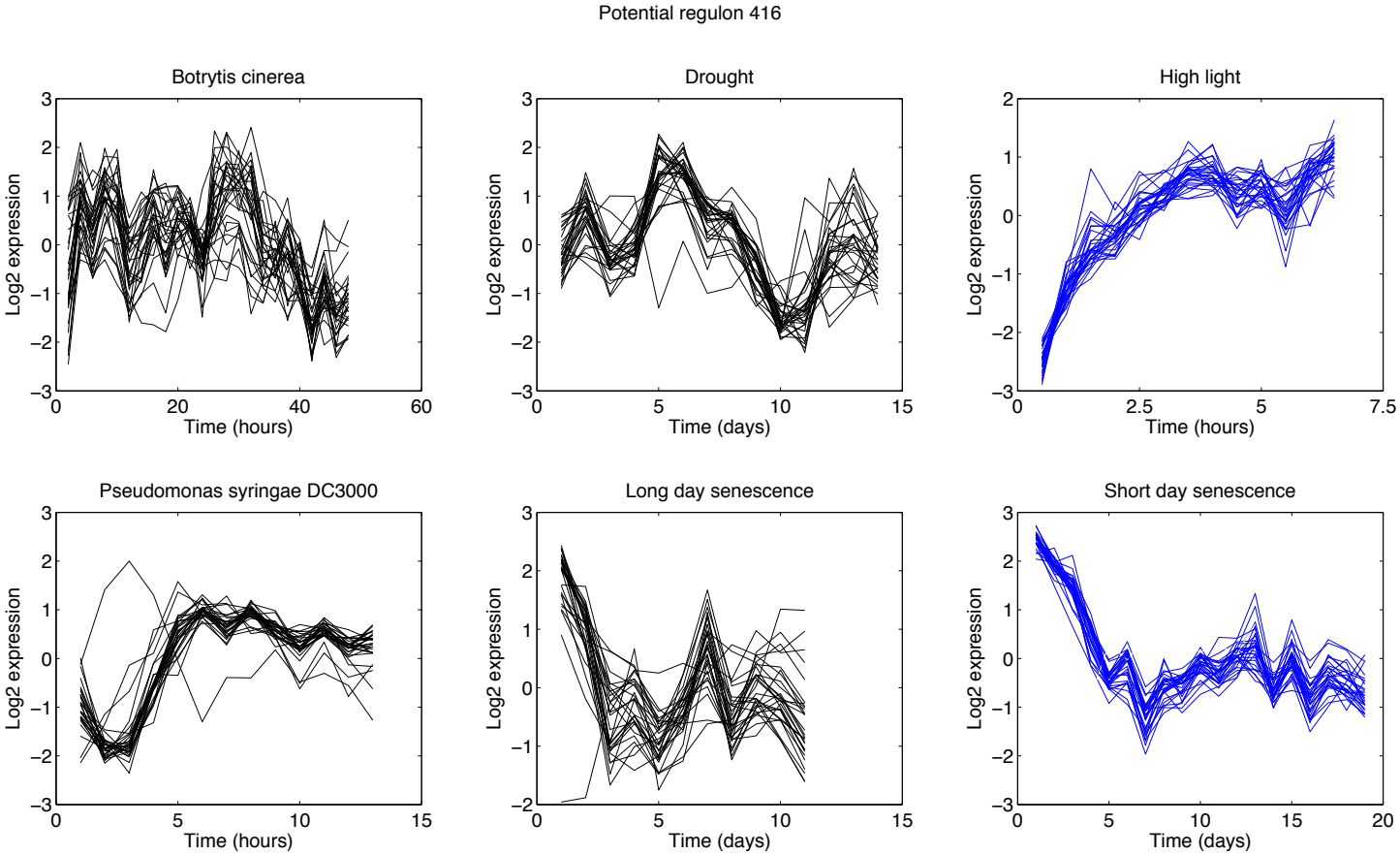



Figure B.11: Potential regulon 416 presented in Table 4.9. Expression profiles of genes significantly co-expressed in high light and short day senescence stress conditions. Gene membership is presented in Table B.11

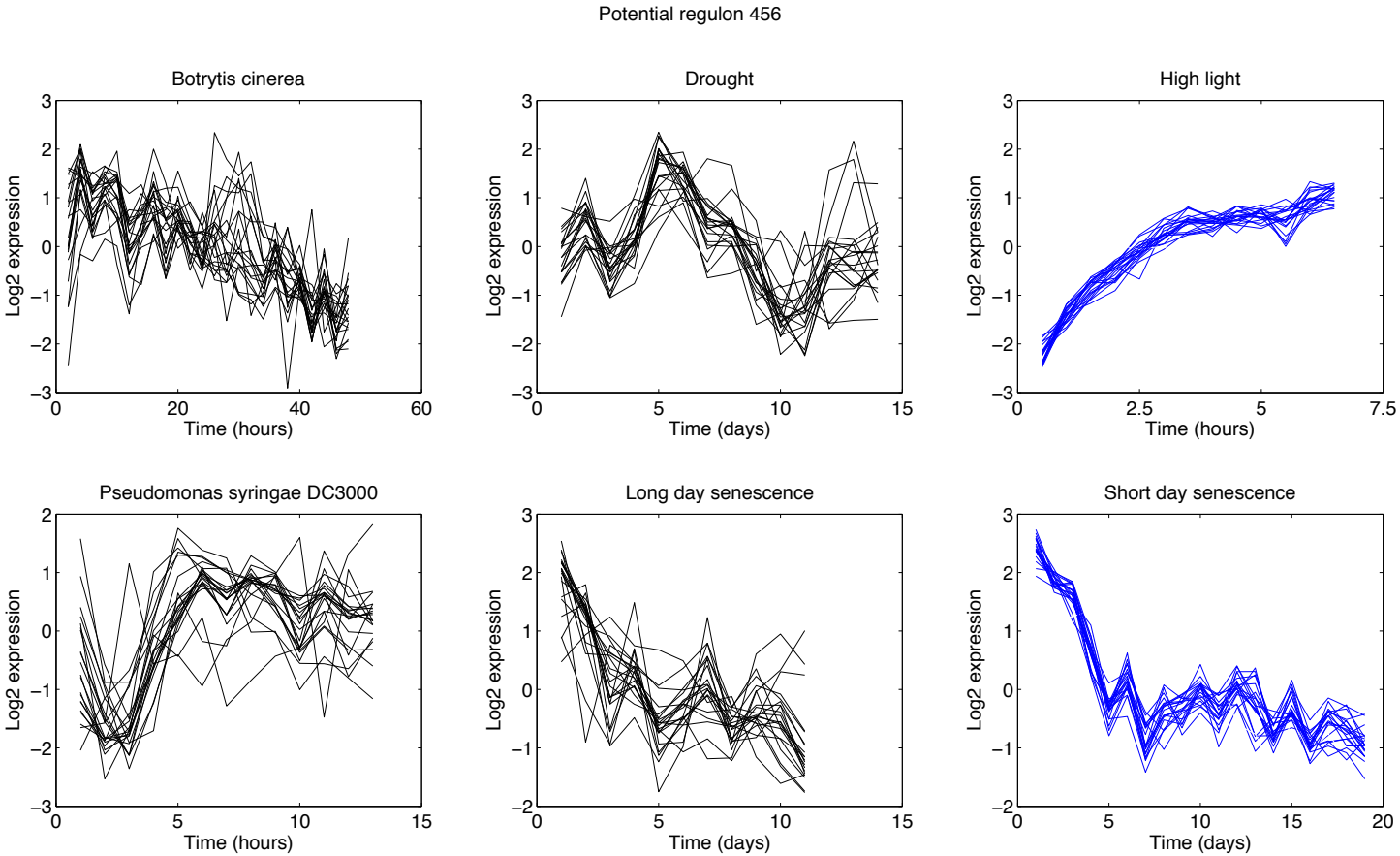



Figure B.12: Potential regulon 456 presented in Table 4.9. Expression profiles of genes significantly co-expressed in high light and short day senescence stress conditions. Gene membership is presented in Table B.12

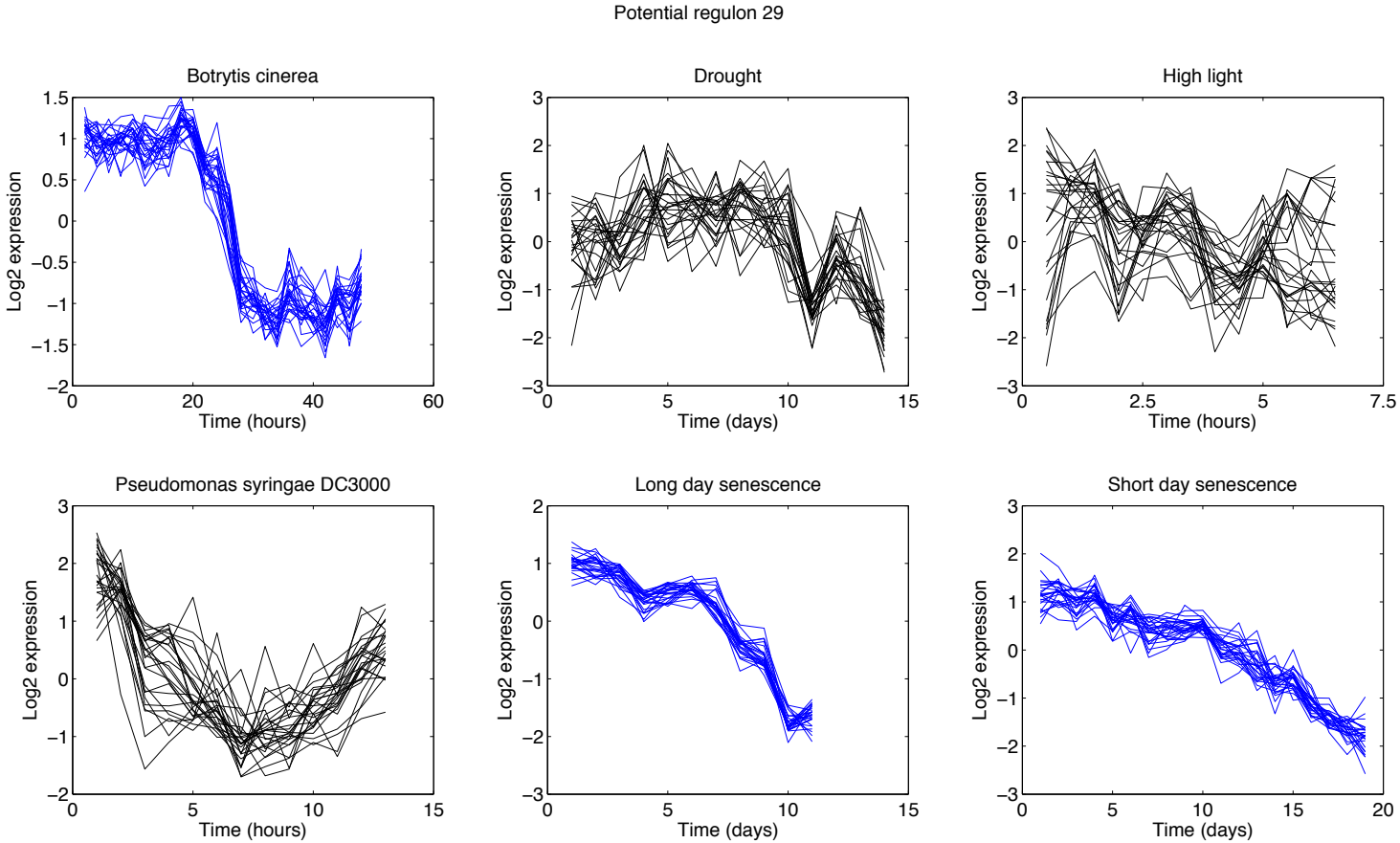



Figure B.13: Potential regulon 29 presented in Table 4.9. Expression profiles of genes significantly co-expressed in *B. cinerea*, long day and short day senescence stress conditions. Gene membership is presented in Table B.13

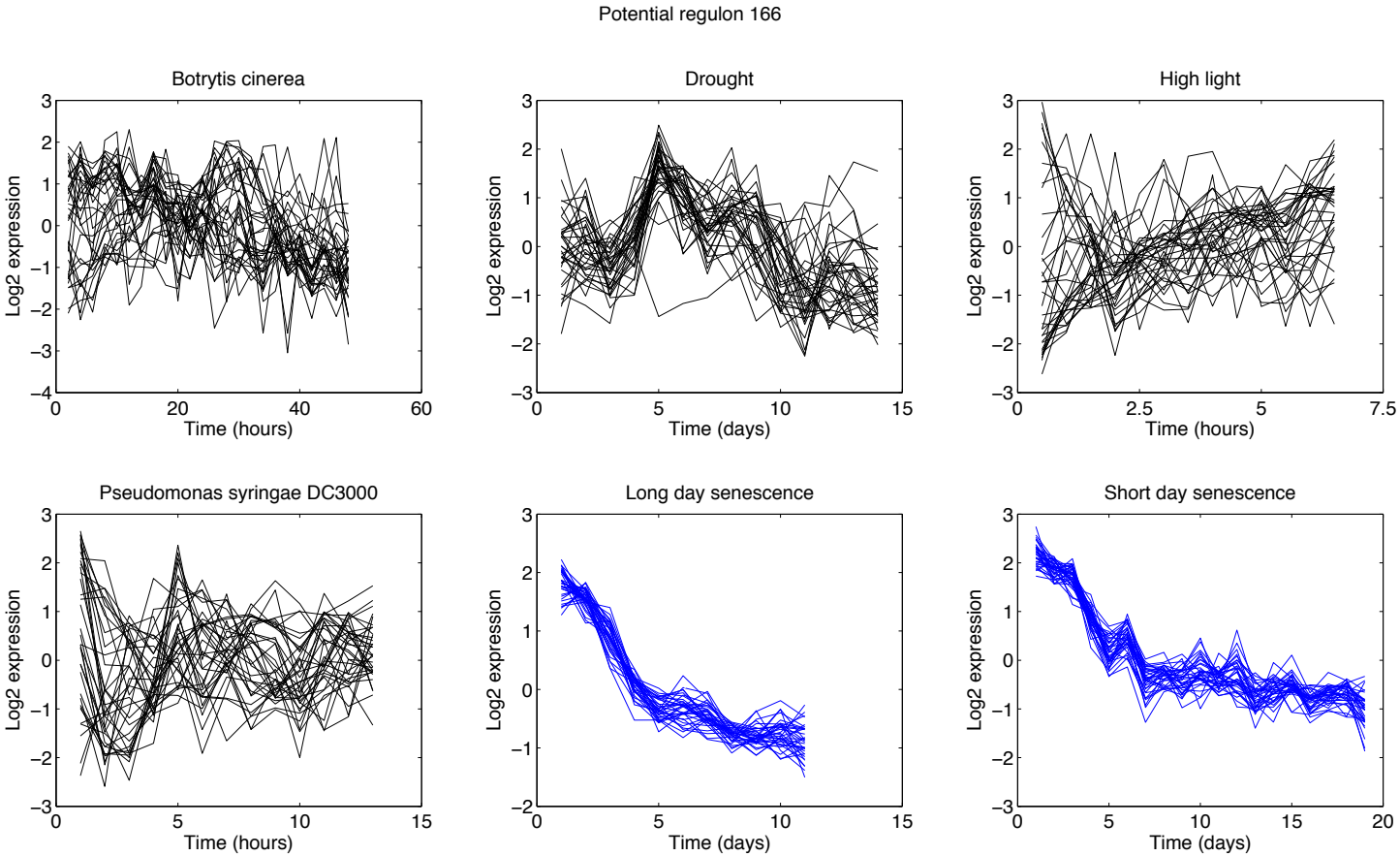



Figure B.14: Potential regulon 166 presented in Table 4.9. Expression profiles of genes significantly co-expressed in long day and short day senescence stress conditions. Gene membership is presented in Table B.14

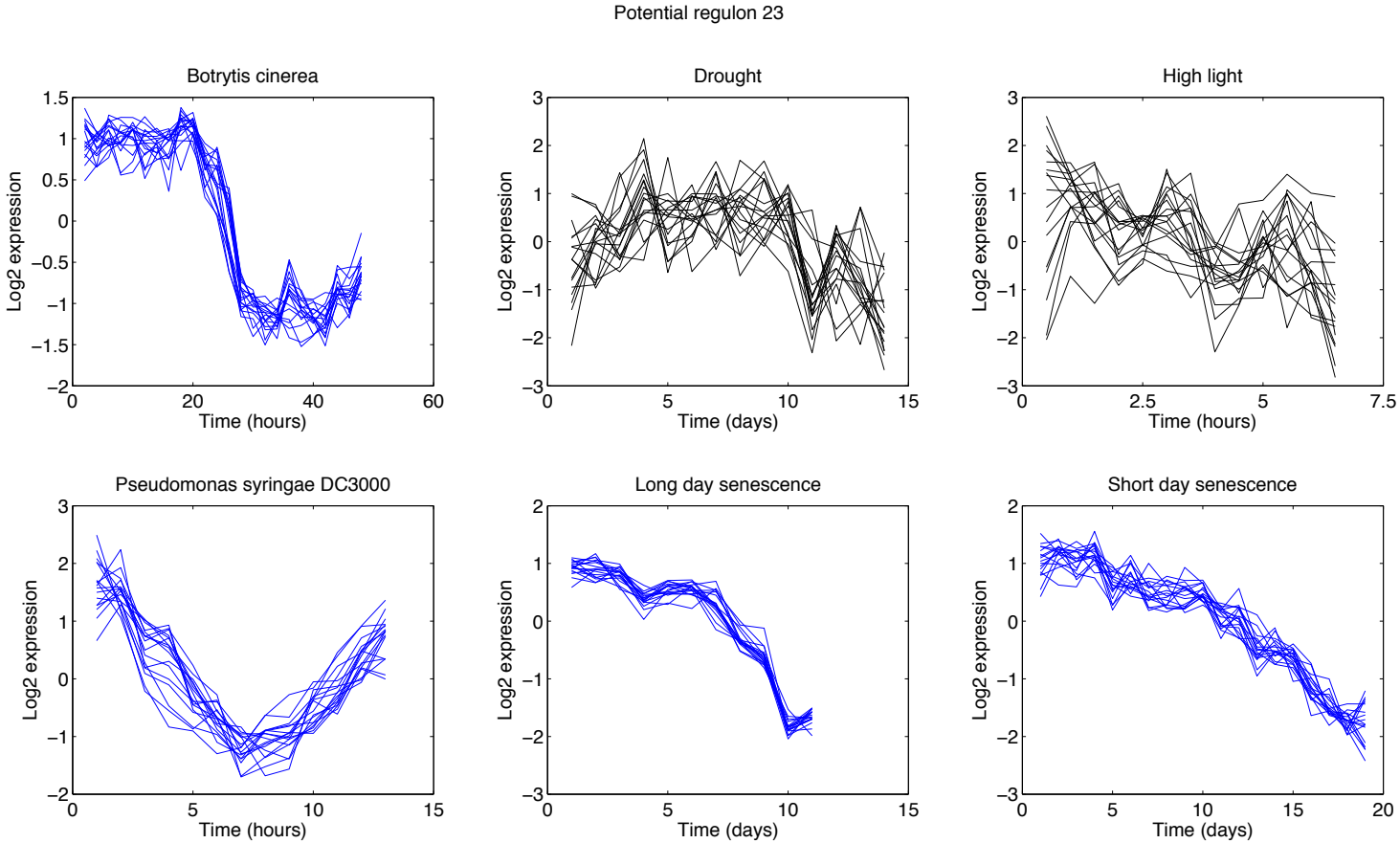



Figure B.15: Potential regulon 23 presented in Table 4.9. Expression profiles of genes significantly co-expressed in *B. cinerea*, *P. syringae* DC3000, long day and short day senescence stress conditions. Gene membership is presented in Table B.15

Table B.12: Gene members of potential shared regulons presented in Table 4.9, showing ATG identifiers and common gene name, where applicable

| Potential shared regulon 456, where genes are co-expressed over high light and short day senescence stress conditions | |
|---|---|
| ATG identifier | Gene name |
| AT1G70600 | |
| AT2G32220 | |
| AT2G35240 | |
| AT2G19670 | PRMT1A |
| AT2G31140 | |
| AT2G37270 | ATRPS5B |
| AT3G23940 | |
| AT3G44590 | |
| AT3G44750 | HDA3 |
| AT3G61100 | |
| AT3G07110 | |
| AT5G49590 | |
| AT4G13170 | |
| AT4G24780 | |
| AT4G25740 | |
| AT4G30800 | |
| AT4G38100 | |
| AT4G18100 | |
| AT5G02870 | |
| AT5G23900 | |
| AT5G60670 | |

Table B.13: Gene members of potential shared regulons presented in Table 4.9, showing ATG identifiers and common gene name, where applicable

| Potential shared regulon 29, where genes are co-expressed over *B. cinerea*, long day and short day senescence stress conditions | |
|---|---|
| ATG identifier | Gene name |
| AT1G03600 | |
| AT1G06680 | PSBP-1 |
| AT1G08380 | PSAO |
| AT1G15820 | LHCB6 |
| AT1G20340 | DRT112 |
| AT1G30380 | PSAK |
| AT1G50900 | |
| AT1G52230 | PSAH2 |
| AT1G55670 | PSAG |
| AT1G67090 | RBCS1A |
| AT1G74970 | RPS9 |
| AT1G29930 | CAB1 |
| AT2G06520 | PSBX |
| AT2G30570 | PSBW |
| AT2G30790 | PSBP-2 |
| AT3G08030 | |
| AT3G08920 | |
| AT3G08940 | LHCB4.2 |
| AT3G56910 | PSRP5 |
| AT4G03470 | |
| AT4G28750 | PSAE-1 |
| AT5G01530 | |
| AT5G38410 | |
| AT5G53490 | |
| AT5G64040 | PSAN |
| AT5G38430 | |

Table B.14: Gene members of potential shared regulons presented in Table 4.9, showing ATG identifiers and common gene name, where applicable

| Potential shared regulon 166, where genes are co-expressed over long and short day senescence stress conditions | |
|---|---|
| ATG identifier | Gene name |
| AT1G14410 | WHY1 |
| AT1G28580 | |
| AT1G47210 | CYCA3;2 |
| AT1G66620 | |
| AT1G69770 | CMT3 |
| AT1G77750 | |
| AT1G23030 | |
| AT2G02450 | ANAC034/ANAC035 |
| AT2G10940 | |
| AT2G21790 | RNR1 |
| AT2G24490 | RPA2 |
| AT2G31725 | |
| AT2G36620 | RPL24A |
| AT2G24170 | |
| AT3G06880 | |
| AT3G14740 | |
| AT3G14900 | |
| AT3G18730 | TSK |
| AT3G23740 | |
| AT3G23940 | |
| AT3G24495 | MSH7 |
| AT3G25100 | CDC45 |
| AT3G54560 | HTA11 |
| AT3G25520 | ATL5 |
| AT3G27360 | |
| AT3G28900 | |
| AT3G53580 | |
| AT4G12970 | |
| AT4G16141 | |
| AT4G25890 | |
| AT4G28310 | |
| AT4G28780 | |
| AT4G31710 | ATGLR2.4 |
| AT5G08020 | RPA70B |
| AT5G66005 | |

Table B.15: Gene members of potential shared regulons presented in Table 4.9, showing ATG identifiers and common gene name, where applicable

| Potential shared regulon 23, where genes are co-expressed over *B. cinerea*, *P. syringae* DC3000, long day and short day senescence stress conditions | |
|---|---|
| ATG identifier | Gene name |
| AT1G03600 | |
| AT1G06680 | PSBP-1 |
| AT1G08380 | PSAO |
| AT1G15820 | LHCB6 |
| AT1G30380 | PSAK |
| AT1G29930 | CAB1 |
| AT2G30570 | PSBW |
| AT3G54050 | |
| AT3G55800 | SBPASE |
| AT3G56940 | CRD1 |
| AT3G47470 | LHCA4 |
| AT4G10340 | LHCB5 |
| AT4G25080 | CHLM |
| AT4G28750 | PSAE-1 |
| AT5G38410 | |
| AT5G51110 | |
| AT5G64040 | PSAN |