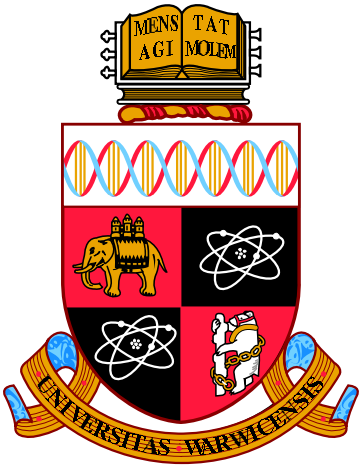


Transcriptional Regulatory Codes Underlying Arabidopsis Stress Responses

by

Richard John Hickman

Thesis

Submitted to the University of Warwick
for the degree of

Doctor of Philosophy

Department of Systems Biology

May 2012

THE UNIVERSITY OF WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

I would like to begin by thanking my supervisors, Prof. Vicky Buchanan-Wollaston and Dr Sascha Ott for their continual advice, guidance, support and motivation from the beginning until the end of the project. I would also like to express my appreciation to Dr Katherine Denby, Prof. Jim Beynon and all members of the PRESTA group for their advice and support. In addition, special thanks must go to Dr Claire Hill for being a constant source of help and guidance throughout my time in the lab.

On a personal level, I thank my family, and my parents in particular, for all the years of support that has helped me reach where I am today. Finally, I would like to thank my fiance, Francesca, for her constant help, support and love throughout these last four years.

Declarations

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. The work in this thesis has been undertaken by myself except where otherwise stated.

Abstract

Plant adaptation to stress is dependent upon the initialisation of molecular signalling networks that regulate the expression of stress-related genes. By examining high-resolution microarray datasets it has been possible to track gene expression changes over time during senescence and in response to infection by fungal pathogen *Botrytis cineria* in the model organism *Arabidopsis thaliana*. Dramatic variations in gene expression are observed at the onset of stress with different groups of genes showing different expression time-courses. This observation must, for a large part, be down to the action of different transcription factors (TFs) binding to the *cis*-regulatory DNA in the promoters of genes in each group and it is this regulatory code that underpins the gene regulatory networks that regulate stress responses. This thesis presents an interdisciplinary investigation of the regulatory codes that are responsible for controlling plant stress responses.

Computational analysis of non-coding sequences provides a powerful approach to identify patterns within DNA that may function to regulate gene expression. This thesis covers the development of Analysis of Plant Promoter-Linked Elements (APPLES), an object-orientated software framework for the analysis of non-coding DNA. Within this environment, methods were developed to probe the regulatory codes that exist within these non-coding sequences and identify regulatory motifs that may function to regulate stress responses in Arabidopsis. APPLES methods were used to identify a novel motif that is likely to play a role in regulating drought responses in Arabidopsis, with experimental approaches providing support for this view. Using known motifs that describe previously characterised TF binding sites, it was possible to identify motifs that are associated with clusters of co-regulated genes identified from the senescence and Botrytis microarray time-course datasets. This analysis revealed *cis*-regulatory elements that may contribute to generating the observed expression patterns.

In a contrasting approach to *in silico* identification of regulatory elements, the Yeast-1-Hybrid (Y1H) assay was used to experimentally identify interactions between TFs and non-coding DNA. The use of a TF library allowed the ability of approximately 1400 Arabidopsis TFs to interact with a given DNA sequence in a single assay. Using the stress-associated *ANAC092* promoter as a test case, it was possible to use this high-throughput procedure to identify TFs that can bind to the promoter of this gene. This high-throughput Y1H system was then used to perform a detailed mapping of protein-DNA interactions that can occur across the core promoters of three highly related stress inducible TF-encoding genes, *ANAC019*, *ANAC055* and *ANAC072*. Microarrays were used to assess the regulatory consequence of a subset of these interactions by perturbing the expression of interacting TFs and observing the effect on target gene expression during multiple stresses. This approach confirmed predicted regulatory relationships and therefore enhanced the current understanding of the transcriptional regulatory networks that operate during stress responses in Arabidopsis.

Abbreviations

35S	Cauliflower mosaic virus promoter
ABA	Abscisic acid
ABRE	Abscisic acid response element
AD	GAL4 activation domain
APPLES	Analysis of Plant Promoter-Linked Elements
bHLH	Basic helix-loop-helix
bp	Base pair
bZIP	Basic leucine-zipper protein
CATMA	Complete Arabidopsis Transcriptome Micro Array
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
CO_2	Carbon dioxide
Col-0	Columbia 0
Col-4	Columbia 4
DAS	Days after sowing
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate

dGTP	deoxyguanosine triphosphate
dTTP	deoxythymidine triphosphate
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
E. coli	*Escherichia coli*
ERF	Ethylene response factor
EtOH	Ethanol
g	Gram
g	Centrifugal force
GAL4	Yeast transcriptional activator
GO	Gene Ontology
GRN	Gene regulatory network
h	Hour
H_2O	Water
His	Histidine
HIS3	Yeast histidine biosynthesis gene
HSE	Heat shock element
JA	Jasmonic acid
JAZ	Jasmonate ZIM domain
kb	Kilobase
L	Litre
LB	Lysogeny broth

Leu	Leucine
LOWESS	Locally weighted sum of squares
μg	Microgram
μM	Micromolar
MEME	Multiple EM for motif elicitation
miRNA	Micro RNA
ml	Millilitre
mm	Millimetre
mM	Millimolar
min	Minute
M	Molar
OO	Object-orientated
OOP	Object-orientated programming
ORF	Open reading frame
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PRESTA	Plant Response to Environmental STress in Arabidopsis
PSSM	Position specific scoring matrix
RNA	Ribonucleic acid
RNAi	Ribonucleic acid interference
mRNA	Messenger ribonucleic acid
rRNA	Ribosomal ribonucleic acid

ROS	Reactive oxygen species
RPM	Revolutions per minute
qRT-PCR	Quantitative reverse transcriptase polymerase chain reaction
SA	Salicylic acid
SAG	Senescence associated gene
SD	Synthetic Defined
sec	Seconds
SOC	Super Optimal broth with Catabolite repression media
T-DNA	Transfer DNA
TF	Transcription factor
Trp	Tryptophan
UTR	Untranslated region
UV	Ultra Violet
VBSSM	Variational Bayesian State Space Modelling
W	Watts
WM	Weight matrix
WT	Wild-type
Y1H	Yeast-1-Hybrid
YPDA	Yeast peptone dextrose adenine

Chapter 1

Introduction

1.1 Plant stress and food security

Food insecurity and the threat of famine have historically always been at the forefront of issues affecting mankind. Today, the precarious nature of food supply is threatened by an ever-expanding population, which is expected to increase a further 50% by the year 2050. Environmental stress is a particular problem and is already responsible for reduced crop yields worldwide. Alongside this, the effects of climate change have already been noted, with more variable conditions likely to lead to increased exposure to environmental stress. An increase in global temperatures will lead to drought and the expected increase in humidity is likely to increase plant susceptibility to pathogens, which is already a major source of crop spoilage throughout the world. These factors are conspiring to greatly endanger food security, leading to social instability and increased poverty, particularly in developing countries. Clearly, however, this is not just a problem for the developing world, but is a global problem affecting the entire population.

To maintain world food supplies it is essential that we understand the mechanisms by which plants adapt to environmental stress. Plants respond to stress at both cellular and molecular level by altering the expression of many genes via complex molecular signalling networks. Understanding these pathways and identifying the regulatory code that underlies these networks will provide opportunities for enhancing the ability of crops to survive stressful conditions and increase yield through genetic manipulation.

1.2 Regulation of gene expression

1.2.1 The multiple levels of gene regulation

The genomes of higher eukaryotes typically contain tens of thousands of genes, which encode the proteins and RNAs that perform all of the structural and biochemical functions within a cell. The expression of these genes must be controlled to ensure the relevant gene products are produced at the appropriate time and place within an organism. The intricate regulation of gene expression generates complex expression patterns that ensure proper development and responses to internal and external stimuli.

Regulation of gene expression is complex and can occur at multiple levels (Figure 1.1). Within the nucleus, DNA wraps around histone proteins, which themselves assemble into higher-order structures known as nucleosomes. This tightly coiled DNA-protein complex- chromatin allows the entire genome to be compacted within a single nucleus and provides the highest level of gene regulation (reviewed in Pfluger and Wagner 2007). Chromatin controls gene usage by impeding the access of TFs and RNA polymerase to the DNA. To allow transcription to proceed, the condensed chromatin must open up, allowing the DNA to interact with all the relevant factors needed to initiate transcription. Manipulation of chromatin states requires the function of chromatin re-modelling factors, which typically function by chemically-modifying histone proteins. These modifications, such as acetylation and methylation, interfere with positive-charged histones, which disrupt the interaction with negatively charged DNA and therefore leads to more open and accessible DNA.

When free of the restrictions imposed by nucleosomes, gene expression is then regulated at the transcriptional level by regulatory proteins that bind to sites on the open, transcriptionally active DNA and act to increase or decrease the rate of mRNA production (see below).

In addition to regulation at the DNA level, a gene can be regulated post-transcriptionally at the RNA level. The mRNA transcript can be subject to alternative splicing, polyadenylation and degradation. A major form of post-transcriptional regulation is by the action of miRNAs, small (∼22 bp) non-coding RNAs which repress translation or target mRNAs for degradation (Sunkar *et al.*, 2007; Fujii *et al.*, 2005). Mature miRNAs interact with proteins to form the RNA-induced silencing complex (RISC), which guides this complex to mRNA by binding to complementary sites usually located within the 3'-untranslated region (UTR) of mRNA sequences. When targeted to a mRNA via an miRNA, RISC then uses RNase functionality to cleave the mRNA.

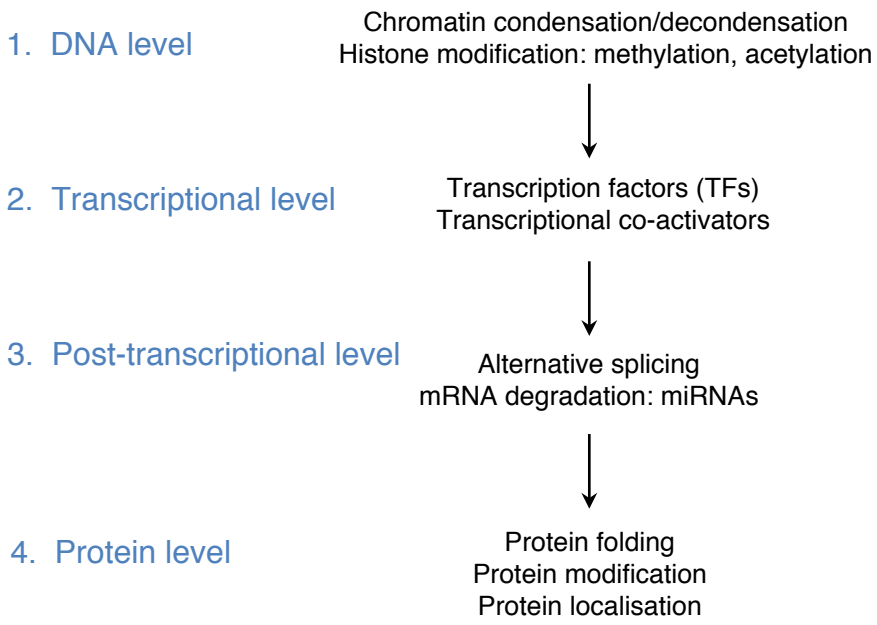


Figure 1.1: **The multiple levels of gene regulation.**

Following translation, the resulting peptide can be subject to extensive regulation that fine-tunes behaviour further. Interactions with other proteins can modify protein function (Weltmeier *et al.*, 2006; Djamei *et al.*, 2007). Chemical modification such as phosphorylation can induce conformational changes which alter functionality or result in protein re-localisation (Mao *et al.*, 2011). Phosphorylation can also increase protein stability (Lopez-Molina *et al.*, 2001), while ubiquitination targets a protein for degradation (Hardtke *et al.*, 2000).

1.2.2 Regulation of transcription

While regulation exists at multiple levels, regulation of transcription initiation is perhaps the most important. It is the level of control that is most efficient for the cell as it minimises energy wastage that may result from generating unnecessary proteins and RNAs. The initiation of transcription is controlled by DNA-binding proteins called transcription factors (TFs) that bind to short sequence elements often situated in the non-coding intergenic DNA that exists between genes. TFs interact with this *cis*-regulatory DNA in a sequence specific manner and subsequently induce or repress transcription. The regulatory effect a TF has on transcription is often influenced by interactions with other TFs and/or co-activators that act to facilitate the assembly of the core transcriptional machinery to allow gene expression.

TFs preferentially interact with specific patterns of nucleotides known as motifs. These motifs are typically short (5-15 bp long) and can consist of any pattern of the four nucleotides, often containing degenerate positions (Harbison *et al.*, 2004). TFs contact positions within the motif through specific DNA-binding domains, which interact with specific residues of closed-form DNA. Protein-DNA binding is due to non-covalent chemical interactions between the DNA-binding domain and chemical groups present on different nucleotides, and it is these specific properties of the nucleotide that are recognised by the TF. Because individual nucleotides share chemical properties, several nucleotides can be recognised by a DNA-binding domain, leading to degeneracy in the TF binding site. Moreover, not all of the positions within a motif explicitly interact with the TF, resulting in an additional source of degeneracy.

Motifs describe the binding specificities of a TF by summarising instances of TF binding sites. By observing the sequences with which a TF can interact, a more detailed description of the TF binding specificity can be made. This information can be summarised by storing it as a matrix, which describes how often each of the four nucleotides are observed at each position of the motif. These matrices are usually referred to as a position specific scoring matrix (PSSM) or a weight matrix (WM). The same information can then be described visually as a sequence logo, where the specificity of each position in the motif is measured in terms of information content (Schneider and Stephens, 1990). Approaches to modelling the binding specificity/motif for a given TF is shown in Figure 1.2.

1.2.3 Transcriptional regulatory networks

Differential expression of large sets of genes is required for the initiation of developmental programs and stress responses. These expression programs are primarily regulated by multiple TFs, where each TF can regulate multiple genes by binding to common sequence motifs present in non-coding DNA such as promoters. The interactions between TFs and regulatory DNA form complex network structures and ultimately drive the generation of complex expression patterns. While the end nodes of such networks are typically functional genes such as structural proteins or enzymes, it is primarily TFs that are responsible for network architecture. The TFs themselves are also highly regulated at the transcriptional level to ensure that the appropriate regulators are expressed at the correct time and place. The pattern of *cis*-regulatory elements that serve as target binding locations for TFs are what underpin gene regulatory networks (GRNs). The arrangement of sequence motifs within the promoters of genes form a regulatory code that is interpreted by TFs and ultimately controls gene expression during specific conditions.

(A)

Site 1	GACACGTGTC
Site 2	TCCACGTGGA
Site 3	TACACGTGTA
Site 4	GCCACGTGGC
Site 5	GTGACGTCAC
Site 6	CTGACGTCAG
Site 7	ATGACGTCAT
Site 8	NNGACGTCNN
Site 9	NNCACGTGNN

Binding sites instances

(B)

	1	2	3	4	5	6	7	8	9	10
A	1	2	0	9	0	0	0	0	3	2
C	1	2	5	0	9	0	0	4	0	3
G	3	0	4	0	0	9	0	5	2	1
T	2	3	0	0	0	0	9	0	2	1

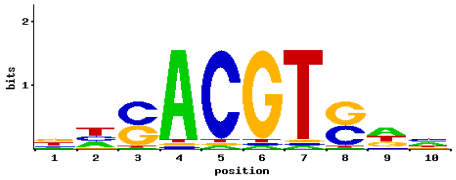
Nucleotide counts at each position in binding site

(C)

	1	2	3	4	5	6	7	8	9	10
A	0.15	0.28	0.02	0.92	0.02	0.02	0.02	0.02	0.40	0.28
C	0.15	0.28	0.52	0.02	0.92	0.02	0.02	0.42	0.03	0.40
G	0.40	0.03	0.42	0.02	0.02	0.92	0.02	0.52	0.28	0.15
T	0.28	0.40	0.02	0.02	0.02	0.02	0.92	0.02	0.28	0.15

Position Specific Scoring Matrix (PSSM)

(D)



Sequence logo

Figure 1.2: **Modelling TF binding sites.** Building models of TF binding site sequence motifs. (A) Defining the binding specificity of a TF requires the assembly of instances of TF binding. These sequences are aligned so that it is possible to observe the preferred nucleotides at each position in the motif. (B) A matrix is constructed containing the number of observed nucleotides at each position in the motif. (C) A PSSM/WM is generated using the matrix of observed nucleotide counts, by computing the ratio of the nucleotide count at a single position in the motif to the total number of nucleotides observed at that position. Pseudocounts are used to correct for small sample size (small number of binding site instances). (D) A motif can be visualised as a sequence logo, where the height of a base is proportional to its frequency at that position, while the height of the stack is proportional to the level of conservation at that position.

1.2.3.1 Methods for elucidating transcriptional networks

Transcriptional networks can be dissected experimentally using what have been termed either "TF-centred" or "gene-centred" approaches (Walhout, 2006). TF-centred approaches focus on the TF and seek to identify sites in the genome with which the TF can interact. Gene-centred methods take the opposite approach and seek to identify the

TFs that interact with a specific DNA sequence. By focusing on non-coding DNA such as promoter sequences it is possible to identify TFs that may directly regulate a gene. The two types of approach complement one another and as each has its own caveats, a combination of both is needed to comprehensively map transcriptional networks operating during complex biological processes such as stress responses.

Chromatin immunoprecipitation (ChIP) based technologies are the most common and powerful TF-centred methods and have been used to identify hundreds of target genes for certain plant TFs (Morohashi *et al.*, 2009). In this approach, the TF is chemically cross-linked to the DNA *in vivo* followed by immunoprecipitation of the TF together with the associated DNA fragment. The location of the resulting DNA fragments within the genome can then be identified using next generation sequencing (ChIP-SEQ) or microarrays (ChIP-CHIP) (Kaufmann *et al.*, 2010). The drawback of ChIP-based methods, however, is the reliance on antibodies to immunoprecipitate the TF of interest. Due to the high sequence similarity amongst many TF families, generation of a specific antibody is often difficult. Even if a suitable antibody is available, if the levels of TF within the sample are low, it may be difficult to isolate sufficient levels of chromatin.

Another approach that attempts to identify the genome wide binding profile of a TF is to first elucidate its binding specificity and then use this to scan the genome for putative binding sites and target genes (Walhout, 2006). Many experimental techniques exist that can reveal TF binding specificity, such as bacterial-1-hybrid, protein-binding microarrays and SELEX (Noyes *et al.*, 2008; Godoy *et al.*, 2011; Oliphant *et al.*, 1989). This interaction specificity can then be modelled as a PSSM and used to scan the genome for instances of the motif. Predictions of the genes targeted by the TF can be improved by scanning sequences that are likely have regulatory role, such as the core promoter of a gene.

The yeast-1-hybrid (Y1H) system is one of the most successful and popular gene-centred methods and has been used to isolate many plant TFs that physically interact with regulatory DNA sequences (Tran *et al.*, 2007; Chen *et al.*, 2010; Zhu *et al.*, 2010). This methods allows for the identification of multiple TFs from different families to interact with a piece of DNA. Subsequent analysis of interactions identified using Y1H has revealed some to have a regulatory consequence *in vivo* (Brady *et al.*, 2011), indicating that this is a powerful method for understanding transcriptional networks.

1.3 Stress responses in Arabidopsis

Plants are constantly exposed to a plethora of biotic and abiotic stresses. Abiotic stresses include those that arise as a result of the plant's environment, such as drought, cold, high salinity and heat. Biotic stresses, on the other hand, refer to living organisms such as bacteria, viruses, and fungi. In addition to environmental stress, developmental process such as senescence can also be a source of cellular stress. Such stresses pose a huge problem to plants and are the main source of reduced crop yields throughout the world. As sessile organisms, plants have been under constant evolutionary pressure to develop techniques that allow them to recognise and respond to particular stresses.

During the last decade, effort has focused on elucidating the genetic pathways responsible for stress perception, signal transduction and response in the model organism *Arabidopsis thaliana* (Arabidopsis). Post genomic era technologies, such as microarrays provide a way to analyse how the transcriptome landscape changes in response to certain stresses. Various studies have used this approach to carry out genome-wide expression profiling under a range of stress conditions in Arabidopsis (Seki *et al.*, 2001; Kreps *et al.*, 2002; Ma *et al.*, 2006; Takahashi *et al.*, 2004; Zeller *et al.*, 2009; Breeze *et al.*, 2011). These studies show substantial changes to gene expression with many genes showing strong upregulation or downregulation in response to each stress. These stress-responsive genes encode products that function to directly defend the organism from stress or those that contribute towards the regulation of the stress response. These regulatory genes can encode TFs or proteins that participate in the production of plant hormones, which play critical roles in stress-mediated signalling. These regulatory factors act together to fine-tune the expression of genes that trigger the molecular mechanisms that underpin the physiological and biochemical events that lead to stress tolerance.

1.3.1 The role of plant hormones in the stress response

Plant hormones are small signalling molecules that regulate a diverse array of biological processes such as growth, development and reproduction. These hormones also form a fundamental layer of signalling networks that control response to biotic and abiotic stress. Well-characterised hormones involved in stress responses include abscisic acid (ABA), jasmonic acid (JA), ethylene (ET), and salicylic acid (SA). Stress perception triggers changes in cellular hormone concentration, which rapidly induces signalling cascades that ultimately lead to the differential expression of genes associated with stress tolerance.

1.3.1.1 Abscisic acid (ABA)

ABA is a key phytohormone that has long been implicated in signalling pathways that regulate response to abiotic and biotic stress. In addition to stress, ABA functions in regulating plant growth and development. ABA was orignally linked to the response to water deficit, which induces the expression of genes encoding ABA biosynthesis enzymes (Zeevaart and Creelman, 1988). Further studies expanded the role of ABA as a general regulator of osmotic stresses such as drought and salt stress (reviewed in Cutler *et al.* 2010). The subsequent accumulation of ABA causes stomatal closure and induces expression of genes that encode proteins such as enzymes that synthesise osmolytes, water transporters and dehydrins that help to retain water and protect against cellular damage as well as other proteins involved in stress tolerance. ABA has also been implicated in the induction of senescence as demonstrated by the accumulation of ABA in senescing leaves (Breeze *et al.*, 2011). In addition to this observation, exogenous application of ABA also induces premature senescence (Gepstein and Thimann, 1980). Expression profiling studies have demonstrated that many ABA inducible genes are upregulated during senescence, which may reflect the induction of abiotic stress signalling pathways during this stressful process (Buchanan-Wollaston *et al.*, 2005; van der Graaff *et al.*, 2006).

In addition to abiotic stresses, ABA also has an important role in biotic stress. While the effect of ABA signalling in response to abiotic stress is generally positive, the role of ABA as a regulator of biotic stress responses is more complicated with apparent positive and negative effects (reviewed in Asselbergh *et al.* 2008). Whether ABA signalling has a positive or negative effect appears to vary according to the timing of infection (Ton *et al.*, 2009). Initially ABA induced stomatal closure can help protect the leaf against pathogen infection (Melotto *et al.*, 2006), while accumulation of ABA during infection can interfere with other plant defence responses resulting in an increase in susceptibility (de Torres-Zabala *et al.*, 2007).

In addition to functional proteins, accumulation of ABA also induces the expression of regulatory proteins such as TFs that are ultimately responsible for ABA mediated gene expression. Many ABA inducible genes contain a conserved sequence motif called the ABA-responsive element (ABRE) that has been shown to be vital for ABA induced expression (Choi *et al.*, 2000). The ABRE has the consensus (C/T)ACGTGGC and interacts with bZIP TFs known as ABRE-binding factors (ABFs). ABF2, ABF3 and ABF4 are members of the same phylogenetic clade within the bZIP family and have been shown to require ABA for full activation and to be master regulators of ABA mediated gene expression via ABREs in response to abiotic stress (Yoshida *et al.*, 2010). In the same study, a ABF knockout Arabidopsis plants displayed reduced expression of many

genes that play key roles in regulating responses to osmotic-related stresses.

In addition to ABRE motifs, many ABA-regulated genes contain binding sites for other TFs. For example, binding sites for MYC and MYB TFs, which bind MYC2 and MYB2 respectively, have been shown to have important roles in ABA mediated expression of stress related genes such as *RD22* and *ADH1* (Abe *et al.*, 2003). Another sequence motif that functions in ABA mediated expression is the dehydration response element (DRE) which has been shown to act synergistically with the ABRE to positively regulate ABA mediated expression in response to abiotic stress (Narusaka *et al.*, 2003).

1.3.1.2 Jasmonic acid (JA)

JA functions as a signalling molecule with a prominent role in mediating responses to biotic stress. Along with ET, with which it acts synergistically, JA controls defence responses to necrotrophic pathogens. The JA signalling network is primarily controlled by a collection of jasmonate ZIM-domain (JAZ) proteins, which repress transcription of JA-inducible genes through direct interaction with JA associated TFs (Katsir *et al.*, 2008). Biotic stress induces the accumulation of JA, which then induces COI1 to interact with JAZ repressors and mark them for degradation by ubiquitination (Chini *et al.*, 2007). This process relieves repression of key JA signalling TFs such as MYC2, ERF1 and ORA59, which can then induce expression of defence related genes (Pré *et al.*, 2008). TFs that regulate JA mediated expression are represented by multiple different families and therefore interact with different motifs present within the promoters of JA regulated genes. MYC2 is a bHLH TF that binds to the G-box motif, which has the consensus CACGTG (Dombrecht *et al.*, 2007). JA inducible TFs that are members of the ethylene response factor family, such as ERF1, ERF2, ERF4, ORA47 and ORA59 interact with the GCC-box, which has the consensus GCCGCC.

These JA inducible TFs regulate multiple classes of genes that encode proteins that function in plant defence. Genome-wide expression profiling studies have revealed that JA induces expression of genes involved in production of stress-associated metabolites, such as glucosinolates, phenylpropanoids, and anthocyanins (Sasaki-Sekimoto *et al.*, 2005; Pauwels *et al.*, 2009). In addition to activating the expression of defence related genes, JA also plays a role in controlling cell growth and proliferation. This regulation is through the repression of cell cycle genes, suggesting that JA regulates the reallocation of energy from growth to defence (Pauwels *et al.*, 2008).

In addition to regulating biotic stress responses, JA has also been implicated in the regulation of abiotic stress responses such as salt stress and osmotic stress (Xu *et al.*,

1994; Lehmann *et al.*, 1995). JA has also been shown to function in leaf senescence and levels of JA increase as the leaf senescences (Breeze *et al.*, 2011). Expression studies also show that genes associated with senescence are upregulated by JA treatment (Jung *et al.*, 2007). The role of JA in the senescence process may overlap with its apparent function during response to biotic stress; downregulation of genes involved in growth and photosynthesis and upregulation of stress related genes.

1.3.1.3 Ethylene (ET)

ET regulates multiple aspects of plant growth and development, and is a key signalling molecule in stress responses. In particular, ET is a prominent regulator of the defence response to necrotrophic pathogens. ET is perceived by membrane bound receptors such as ETR1, ETR2, and EIN4, which under non-stress conditions repress the activity of EIN2, a positive regulator of ET signalling, through the protein kinase CTR1 (Alonso *et al.*, 1999). Receptor interaction with ET abolishes repression of EIN2 via CTR1, which then goes on to positively regulate downstream ET signalling; EIN2 enhances the level of EIN3/EIL TFs, which in turn activate ERF family TFs such as ERF1 (Solano *et al.*, 1998). These TFs regulate the expression of genes that encode stress-related proteins. ERFs target the GCC-box (GCCGCC) present in the promoters of many stress-related genes (Okamuro *et al.*, 1997). Even though they recognise the same core motif, flanking bases may discriminate binding between different members of the ERF family (Tournier *et al.*, 2003). The importance of the ERFs in stress response signalling is noted in the observation that mutants in several ERFs result in reduced resistance to biotic stress (Oñate-Sánchez and Singh, 2002; Lorenzo *et al.*, 2004). ET is known to play a key role in senescence as indicated by *EIN2* mutants showing delayed senescence (Oh *et al.*, 1997). ET biosynthetic genes are upregulated during leaf senescence and the hormone is a major positive regulator of leaf senescence with levels of the hormone increasing during the senescence process (van der Graaff *et al.*, 2006).

1.3.1.4 Salicylic acid (SA)

SA is a key signalling molecule in the plant defence to biotic stress. In addition to this association with stress, SA has also been implicated in regulating biological processes such as growth, development and photosynthesis (Vicente and Plasencia, 2011). SA, like JA, is a key player in the response to pathogen infection, but in contrast to JA, the SA pathway is a major regulator of plant response to biotrophic pathogens. SA mediated signalling is reliant on the function of NPR1, which acts as a co-activator of SA-responsive genes. SA activates NPR1, which then translocates to the nucleus where it regulates expression of defence-related genes (Dong, 2004; Spoel *et al.*, 2009). The

SA signalling pathway also has a role in regulating gene expression during senescence (Morris *et al.*, 2000). Mutant Arabidopsis plants that are insensitive to SA display delayed developmental senescence, as highlighted by reduced chlorophyll degradation (Morris *et al.*, 2000).

1.3.1.5 Hormone crosstalk

Plants are not often exposed to a single stress, nor do they experience these as a linear sequence of events. In rapidly changing environments, plants must be able to adapt by constantly fluctuating the expression of different sets of genes. Crosstalk between hormone pathways is a phenomenon that provides a mechanism to balance responses to multiple stresses, which in turn improves the organism's viability (reviewed in Pieterse *et al.* 2009). The interactions between different hormone signalling pathways is complex with many synergistic and antagonistic interactions occurring under different contexts.

One of the best characterised examples of antagonistic crosstalk is between two biotrophic pathogen hormones, JA and SA. Biotrophic pathogens that trigger SA signalling in Arabidopsis strongly suppresses induction of JA-regulated genes activated by necrotrophic fungi (Spoel *et al.*, 2007). This antagonistic regulation is largely mediated by TFs that exist in the individual signalling cascades of each hormone. In the SA signalling pathway, TGA and WRKY TFs will repress the expression of JA inducible genes (Ndamukong *et al.*, 2007; Li *et al.*, 2004), while in the JA pathway, MYC2 can repress genes positively regulated by SA (Laurie-Berry *et al.*, 2006).

In contrast to JA-SA pathways, the JA-ET pathways are largely synergistic in terms of gene regulation with large overlap in the genes differentially expressed by JA and ET treatment (Schenk *et al.*, 2000). The Arabidopsis TFs, ERF1 and ORA59 are both key regulators of JA and ET mediated expression, indicating that they are convergence points for both pathways (Pré *et al.*, 2008; Lorenzo *et al.*, 2003). The ET and SA pathways are another example of a synergistic relationship, with ET playing an important role inducing SA defence responses and enhancing positively regulated genes.

The crosstalk that exists between hormone pathways that are classically more associated with biotic stress such as JA, SA, and ET, and the more abiotic stress associated ABA signalling pathway is complicated. It appears that ABA can attenuate JA-, ET- and SA- dependent gene expression, increasing susceptibility to biotrophic and necrotrophic pathogens (Anderson *et al.*, 2004; Mohr and Cahill, 2007; Flors *et al.*, 2008).

1.3.2 Transcriptional control of stress-associated gene expression.

Stress signalling pathways are ultimately localised to the nucleus where TFs regulate the transcription of stress-responsive genes. Initial studies predicted that approximately 1500 TFs are encoded in the Arabidopsis genome (Riechmann *et al.*, 2000). Subsequent analysis within our lab group has shown that the genome actually contains well over 2000 TFs (data not shown). These TFs fall into clearly defined families with particular members and the *cis*-regulatory elements with which they interact being linked with regulating the stress response. As described above, many TFs function within hormone signalling pathways where they function to directly regulate transcription.

Several different transcriptional regulatory systems are involved in stress-responsive gene induction, with some regulatory factors appearing to function in response to multiple conditions. Numerous studies have investigated gene expression overlap in response to different stress conditions and results demonstrate that some genes are induced by multiple stresses. This indicates that plant stress responses are mediated by overlapping pathways.

Early experiments that probed this potential cross-talk between stress-signalling pathways used microarrays to identify genes induced by ABA and other abiotic stresses such as drought, cold and osmotic stress (Seki *et al.*, 2002; Kreps *et al.*, 2002). These studies identified genes that were differentially expressed under several conditions, indicating that common sets of TFs may be functioning to regulate responses to multiple different forms of stress.

Later, the AtGenExpress project produced high-quality transcriptome data using Affymetrix ATH1 microarrays to analyse global transcriptional responses of the 24,000 genes represented on the array to many different abiotic and biotic stresses in Arabidopsis genotype Col-0. Abiotic treatments included cold, osmotic stress, salt, drought, genotoxic stress, ultraviolet light, oxidative stress, and wounding (Kilian *et al.*, 2007). Biotic stresses included infection with variety of pathogens such as virulent and avirulent *P. syringae*, *P. infestans* and pathogen related compounds such as bacterial elicitor flg22. Data were also collected for Arabidopsis treated with stress related hormones such ABA, JA, and SA (Goda *et al.*, 2008). Overall, the AtGenExpress datasets provide the most comprehensive analysis of the transcriptional changes that occur when plants are exposed to stress. All data are in the public domain and there have been a number of studies into overlapping transcriptional responses to stress using the data.

One such study analysed the changes in the Arabidopsis transcriptome in response to salt

stress in Arabidopsis and identified a strong cross-talk between stress-signalling pathways. Of 680 genes induced by high-salinity, 75% were also induced by biotic, osmotic, cold, ABA and JA, with the remaining 25% being salt-stress specific (Ma *et al.*, 2006). In a study by Swindell (2006) analysis of microarray data from the AtGenExpress project focusing on transcriptional responses to multiple abiotic stress identified extensive overlap in expression patterns and provided further evidence that there are many TFs that are functioning to regulate the responses to several different types of stress.

1.3.2.1 The NAC transcription factor family

The NAC family of TFs were initially described more than a decade ago. The first NAC genes described were *NAM* (*NO APICAL MERISTEM*) and *CUC2* (*CUP-SHAPED COTYLEDON*), which were identified in mutant screens for defective shoot apical meristem (SAM) development in Petunia and Arabidopsis respectively (Souer *et al.*, 1996; Aida *et al.*, 1997). Sequence analysis identified two Arabidopsis clones, *ARABIDOPSIS TRANSCRIPTION ACTIVATION FACTOR 1/2* (*ATAF1/2*) that were homologous to the region encoding the N-terminal part of the CUC2 protein (Aida *et al.*, 1997). The conserved N-terminal domain was named the NAC domain (for Petunia NAM and Arabidopsis ATAFI, ATAF2, and CUC2). Sequence analysis has revealed that there are at least 105 putative NAC TFs present in the Arabidopsis genome (Ooka *et al.*, 2003), together with several hundred found in the genomes of rice, soybean and tobacco (Fang *et al.*, 2008; Mochida *et al.*, 2009; Rushton *et al.*, 2008). While the NAC TF family is broadly dispersed in plants, no orthologs have been identified in other eukaryotes (Aida *et al.*, 1997).

Structurally, NAC proteins consist of two domains that exert their functionality. The highly conserved N-terminal domain contains the DNA binding ability, while the C-terminal region serves as the activation domain. The N-terminal NAC domain is approximately 150 amino acids in length and is made up of five subdomains (Ooka *et al.*, 2003). The X-ray crystallographic structure of the NAC domain of ANAC019 reveals that the binding motif does not conform to any known binding domains (Ernst *et al.*, 2004). The NAC domain adopts a twisted anti-parallel beta-sheet surrounded by two helices. The C-terminal domains of NAC TFs display significant variability amongst family members and have been characterised as having transcriptional activator or repressor functionality (Fujita *et al.*, 2004; Hu *et al.*, 2006; Tran *et al.*, 2004). The C-terminal domain has also been reported to possess protein-binding ability (Kim *et al.*, 2007).

Insights into the DNA binding specificity of the NAC TFs were first gained through studies of ANAC019, ANAC055 and ANAC072, which were identified as being able to

bind within a 63 bp fragment of the *ERD1* promoter using a Y1H system (Tran *et al.*, 2004). Gel retardation assays were used to confirm that ANAC019 and ANAC055 could specifically bind to the 63 bp fragment. ANAC019 and ANAC055 trans-activated the *ERD1* promoter demonstrating that they are functional transcriptional activators. The 63 bp promoter was mutated in order to identify a more well defined NAC recognition sequence, and identified CACG as the core NAC TF binding motif (Tran *et al.*, 2004). This limited understanding of the core NAC binding motif was expanded in a study by Olsen *et al.* (2005) that elucidated a detailed description of the NAC binding specificity. Using the SELEX approach, the authors found that ANAC019 bound to the consensus TTNCGTA while ANAC092 bound to TTGCGTGT. While not completely identical, they do share a core, CGT[GA], which is the reverse complement of the CACG motif core identified for ANAC019, ANAC055 and ANAC072 in the study by Tran *et al.* (2004). Other studies have revealed different core binding site for a Calmodulin-binding NAC TF, which bound to the core GCTT motif (Kim *et al.*, 2007). This finding suggests that other members of the NAC TF family have different binding specificities.

The NAC TFs have been implicated in regulating a diverse set of biological processes. NAC TFs were originally identified in mutant screens for defects in SAM formation and cotyledon development (Souer *et al.*, 1996; Aida *et al.*, 1997), implicating NAC TFs in the regulation of SAM and organ development. Studies have also identified regulatory roles in terms of lateral root development, growth of leaf cells and secondary wall biosynthesis (Xie *et al.*, 2000; Kato *et al.*, 2010; Zhong *et al.*, 2010).

In addition to developmental processes, NAC TFs have been heavily implicated in regulating abiotic and biotic stress responses. The first NAC TF to be functionally linked to regulating stress responses was ANAC072 (also called RESPONSE TO DESICCATION 26; RD26), which was isolated from drought-stressed plants (Fujita *et al.*, 2004). *ANAC072* expression was also found to be induced by ABA, while microarray analysis of leaves overexpressing *ANAC072* compared to wild-type identified an enrichment for ABA- and drought-associated genes.

As mentioned previously, three highly similar NAC TFs, ANAC019, ANAC055 and ANAC072 interact with the core promoter of *ERD1*, a drought response marker gene (Tran *et al.*, 2004). Expression of *ANAC019*, *ANAC055* and *ANAC072* were also induced by drought, high-salt stress, ABA and JA. Overexpression of *ANAC019*, *ANAC055* or *ANAC072* resulted in altered expression of many stress-related genes (Tran *et al.*, 2004). Interestingly, *ERD1* itself was not significantly differentially expressed in overexpressor lines, indicating that additional factors are necessary. Y1H analysis was used

to uncover another protein that bound to the *ERD1* promoter adjacent to the NAC binding site (Tran *et al.*, 2007). This DNA-binding protein, zinc finger homeodomain 1 (ZFHD1) bound to the motif AATTT, and was found to be induced by drought, high-salt and ABA (Tran *et al.*, 2007). Yeast-2-hybrid analysis revealed ZFHD1 can interact with ANAC019, ANAC055 and ANAC072, while overexpression of both *ZFHD1* and any of the three NAC TFs resulted in upregulation of *ERD1* (Tran *et al.*, 2007). These results suggest that combinatorial motifs within the *ERD1* promoter are responsible for driving gene expression.

In addition to regulating responses to abiotic stress, ANAC019 and ANAC055 have also been identified as key regulators of responses to biotic stress (Bu *et al.*, 2008). More specifically, ANAC019 and ANAC055 were shown to be responsible for regulating the JA-induced expression of defence genes. Expression of genes that are normally highly induced by JA were greatly reduced in an *anac019anac055* double mutant. Transgenic plants overexpressing *ANAC019* or *ANAC055*, however, showed enhanced JA-induced expression. Double mutants also displayed increased resistance to the necrotrophic fungus Botrytis (Bu *et al.*, 2008).

ANAC002/ATAF1 expression was strongly enhanced under drought conditions and by ABA (Lu *et al.*, 2007). Arabidopsis *ANAC002* knockout plants displayed improved drought tolerance suggesting an important role in regulating responses to water deficit. ANAC002 has also been implicated as a regulator of defence responses against necrotrophic fungal and bacterial pathogens (Wang *et al.*, 2009). In another study, work with ABA insensitive mutants found *ANAC002* expression to be induced in an ABA-independent manner, while *ANAC002* mutants further implicated the TF as a regulator of responses to necrotrophic pathogens such as Botrytis (Wu *et al.*, 2009).

ANAC092 has been implicated in the salt-stress response (He *et al.*, 2005). Microarray analysis of genes induced by salt stress, identified *ANAC092* as being strongly upregulated. Balazadeh *et al.* (2010b) used an inducible line to identify immediate downstream targets of ANAC092, and these were heavily enriched for known senescence associated genes. A study by Woo *et al.* (2004) examined the response to oxidative stress in an *ANAC092* mutant, by treating plants with chemicals known to induce oxidative stress. The ANAC092 loss of function mutant showed enhanced resistance to oxidative stress. These observations may indicate that ANAC092 participates in regulating a general oxidative stress resistance mechanism.

The role of ANAC092 in a regulatory network that regulates age-dependent cell death

and senescence in Arabidopsis was revealed in a study by Kim *et al.* (2009). Expression of *ANAC092* is induced by EIN2, a positive regulator of the ET signalling cascade. In young leaves, *ANAC092* expression was found to be negatively regulated by the miRNA, miR164. As the leaf ages, miRNA164 is negatively regulated in an *EIN2* dependent manner, which then allows ANAC092 levels to increase and in turn initiate senescence and cell death (Kim *et al.*, 2009).

Other members of the NAC family that have been shown to regulate stress-related processes include *ANAC029*, *ANAC059* and *ANAC078*. ANAC029 and ANAC059 positively regulate senescence (Guo and Gan, 2006; Balazadeh *et al.*, 2011), while ANAC078 regulates high-light induced flavonoid biosynthesis (Morishita *et al.*, 2009).

1.4 Leaf senescence

1.4.1 Senescence and the link to stress

A key developmental process that is linked to plant stress responses is senescence. Leaf senescence symbolises the final phase of leaf development and involves the coordinated degeneration and redistribution of the organs cellular components (reviewed in Buchanan-Wollaston 1997). Senescence is associated with large-scale physiological and biochemical changes, including the degradation of chlorophyll, proteins, lipids and nucleic acids. Chlorophyll is degraded together with chloroplast associated proteins such as Rubisco, eventually leading to the destruction of the entire chloroplast. Targeted use of proteases and lipases result in the gradual degradation of the leaf protein content and cell walls respectively. At the later stages of senescence the nucleus, together with the nucleic acids contained within it, are also degraded. This widespread degradation and subsequent remobilisation of molecules generates increased levels of reactive oxygen species (ROS), a major source of cellular damage. Oxidative stress induces stress response pathways (Neill *et al.*, 2002), and protecting cells against the effects of oxidative stress is an important aspect of the senescence process (Buchanan-Wollaston *et al.*, 2005). Mutants that display delayed senescence have been found to be more tolerant to oxidative stress (Woo *et al.*, 2004).

In addition to being an age-related process, senescence can also be induced by abiotic and biotic stress. Plants subject to drought can initiate premature senescence in order to mobilise nutrients to the reproductive organs, while leaves that undergo pathogen infection display accelerated senescence, limiting the spread of the pathogen throughout the plant. As mentioned previously, the stress associated ABA, JA, ET and SA signalling pathways have all been implicated in senescence (Buchanan-Wollaston *et al.*,

2005; Breeze *et al.*, 2011; van der Graaff *et al.*, 2006; Morris *et al.*, 2000). Consistent with this link, genes differentially expressed in response to abiotic and biotic stress overlap with those differentially expressed during senescence (Weaver *et al.*, 1998; Quirino *et al.*, 1999; Pontier *et al.*, 1999). Overlapping expression patterns highlight genes that function during multiple stress-related processes. Amongst the genes induced by stress and senescence are sets of TFs, indicating shared regulatory components operate in these processes.

1.4.2 Transcriptional control of leaf senescence

The model plant Arabidopsis has been used to study transcriptional regulation of senescence and has led to the identification of many TFs that are enhanced or repressed during this process. Microarray time-course experiments have revealed dramatic changes in gene expression at the onset of senescence, which continue to change until the leaf is fully senescent, with different sets of genes showing different expression time-courses (Buchanan-Wollaston *et al.*, 2005; Breeze *et al.*, 2011). Among the senescence enhanced genes, are hundreds of TFs from many different families. Of these TFs, however, only a minor fraction has been functionally associated with senescence.

Several members of the WRKY TF family, which generally bind to the w-box motif (TTGAC[G/T]), have been shown to play prominent roles in controlling leaf senescence in Arabidopsis. WRKY53 has been shown to be important for the progression of senescence, with *WRKY53* knockout and *WRKY53* overexpressor mutants exhibiting delayed and premature senescence respectively (Miao *et al.*, 2004). Plant defence-linked WRKY6 has been shown to positively regulate SIRK, a receptor-like kinase that is specifically enhanced during senescence (Robatzek and Somssich, 2002). WRKY70 has been implicated as a negative regulator of senescence with knockout mutants displaying premature senescence (Ulker *et al.*, 2007).

More recent work has implicated members of several other distinct TF families in the regulation of senescence. The auxin response factor (ARF) TF family, which mediates auxin dependent signalling, has also been shown to play regulatory roles in senescence. The plant hormone, auxin, is believed to suppress leaf senescence and ARF2 has been suggested to positively regulate senescence by repressing auxin signalling. *ARF2* knockout mutants exhibit delayed senescence, which might be attributed to increased auxin sensitivity, due to decreased repression of auxin signalling (Lim *et al.*, 2010). G-BOX BINDING FACTOR 1 (GBF1) was shown to promote senescence via the repression of the ROS-scavenging enzyme *CATALASE2* (*CAT2*) (Smykowski *et al.*, 2010). GBF1 knockout mutants displayed decreased ROS levels and a delayed senescence phenotype.

RELATED TO ABI3/VP1 (RAV1), a member of the RAV TF family, was also shown to regulate senescence in a positive manner (Woo *et al.*, 2010). Constitutive and inducible overexpression of *RAV1* produced an early senescence phenotype.

The NAC domain TF family is emerging as key regulators of senescence and stress responses in general, with a large proportion of this family displaying enhanced expression during senescence (Breeze *et al.*, 2011). In Arabidopsis, three NAC family members, ANAC029, ANAC059 and ANAC092, have been shown to have a regulatory function in senescence. *ANAC029* knockout and overexpression mutants exhibited delayed and premature senescence respectively (Guo and Gan, 2006). ANAC092, as has been described previously, functions in a regulatory network with EIN2 and miRNA164 to ensure that senescence occurs at the correct time in development (Kim *et al.*, 2009), while *ANAC059* has recently been shown to play an important role in regulating senescence (Balazadeh *et al.*, 2011).

1.5 Infection by *Botrytis cinerea*

1.5.1 The plant response to Botrytis infection

The necrotrophic fungal pathogen *Botrytis cinerea* (Botrytis) is a major source of crop spoilage worldwide. This is particularly prevalent in regions with humid climates. As a necrotrophic pathogen, Botrytis aims to kill host cells quickly as to obtain nutrients from the host. Botrytis is equipped with an arsenal of infection mechanisms including a variety of cell wall-degrading enzymes and toxins (reviewed in Glazebrook 2005). These pathogen-produced toxins are thought to promote host cell death by increasing cellular levels of ROS and play a major role in pathogen infection.

Infection by Botrytis induces the activation of host defence responses that employ multiple mechanisms to fight off infection. A key part of the defence mechanism includes the production of antimicrobial molecules. In Arabidopsis, the antimicrobial compound camalexin has been identified as important for resistance to some strains of Botrytis. The *PAD3* gene encodes a cytochrome P450 monoxygenase, which functions in the biosynthesis of camalexin (Böttcher *et al.*, 2009). Arabidopsis mutants lacking a functional *PAD3* gene displayed increased susceptibility to some strains of Botrytis (Ferrari *et al.*, 2003). Other defence mechanisms employed by the host are the production of molecules such as callose and lignin, which act to strengthen the plant cell wall (reviewed in Glazebrook 2005). As mentioned above, Botrytis induces host cell death in order to facilitate infection. Insensitivity to ROS also contributes to resistance to pathogen induced cell death (Govrin and Levine, 2000).

1.5.2 Transcriptional responses to Botrytis infection

The expression of defence genes is tightly controlled at the transcriptional level and many TFs, often acting through specific hormone signalling cascades, have been shown to play critical roles in regulating these defence responses. As mentioned previously, JA and ET signalling is key to induction of resistance mechanisms to necrotrophic pathogens. Mutations that attenuate the JA or ET signalling cascades result in increased susceptibility to Botrytis (Thomma *et al.*, 1998, 1999). MYC2, which is a central integrator of JA and ET signalling has been shown to play an important role in regulating host defence responses. Interestingly, MYC2 appears to regulate sets of genes that lead to enhanced resistance and genes that result in enhanced susceptibility (Glazebrook, 2005). Arabidopsis plants that overexpress *ORA59* display enhanced resistance to Botrytis, while *ORA59*-silenced plants are more susceptible (Pré *et al.*, 2008). In the same study, ERF1 was also shown to positively regulate *PDF1.2*, although levels of this key defence gene were significantly lower in *ORA59*-silenced plants, indicating the importance of ORA59 for regulating the JA-mediated expression of defence genes. Taken together, *ORA59* and *ERF1* act to integrate JA and ET signals during the response to Botrytis.

The NAC TFs ANAC019 and ANAC055 act redundantly to regulate JA mediated expression during Botrytis infection. Phenotyping analysis revealed *anac019anac055* mutants to be more resistant to Botrytis, while single knockout mutants are unaltered compared to wild type (Bu *et al.*, 2008). The increased resistance phenotype is similar to that observed in *myc2* Arabidopsis plants, while overexpression of *ANAC019* in *myc2* background rescues the miss-regulation of JA-regulated genes. These findings suggest that ANAC019 and ANAC055 function downstream of MYC2 in JA-signalling in response to Botrytis infection.

Members of the WRKY TF family play important roles in regulating defence responses to multiple different pathogens, including Botrytis (Rushton *et al.*, 2010). WRKY33 has emerged as having a particularly prominent role in regulating the response to Botrytis with *wrky33* mutants displaying increased susceptibility to the pathogen (Zheng *et al.*, 2006). The regulatory role of WRKY33 was revealed in a recent study, where it was shown to act a positive regulator of camalexin biosynthesis (Mao *et al.*, 2011). Another family member, WRKY70 has been also been linked with positively regulating responses to Botrytis infection (AbuQamar *et al.*, 2006), while WRKY50 and WRKY51 have demonstrated that they negatively regulate defence against Botrytis, possibly through the SA pathway (Gao *et al.*, 2011).

Several MYB family TFs have also been implicated in regulating responses to Botrytis.

MYB46, which is involved in regulating secondary wall biosynthesis, has recently been identified as a prominent mediator of pathogen responses (Ramírez *et al.*, 2011). Arabidopsis *myb46* lines display enhanced resistance to Botrytis, which is predicted to be through the observed misregulation cell wall proteins and enzymes (Ramírez *et al.*, 2011). MYB108 is another TF that has been implicated in regulating responses to Botrytis. MYB108 was originally isolated from Arabidopsis the BOS1 (BOTRYTIS- SUSCEPTIBLE1) gene following a T-DNA insertion mutant screen for altered Botrytis susceptibility. The T-DNA insertion was found to be at the *MYB108* locus. The *bos1* lines were shown to have increased susceptibility to Botrytis infection. Botrytis-induced expression of *BOS1* was reduced in Arabidopsis mutants that have impaired JA-signalling, indicating that *BOS1* is part of the JA cascade in response to biotic stress. Interestingly, the *bos1* plants produce increased amounts of full length *MYB108* transcript compared to wild-type (Mandaokar and Browse, 2009). In a recent thesis (Windram, 2010), different *MYB108* knockout lines were generated and assessed for altered Botrytis susceptibility. Two independent lines demonstrated increased resistance to Botrytis. Given the opposite phenotypes, this finding suggests that the T-DNA insertion present in *bos1* lines, which is present in the upstream non-coding sequence, has a positive effect on *MYB108* expression.

1.6 The PRESTA project

Plant Responses to Environmental STress in Arabidopsis (PRESTA) is a systems biology project spanning multiple lab groups at the Universities of Warwick, Essex and Exeter, that aims to elucidate transcriptional networks that regulate responses to developmental, abiotic and biotic stress. The systems approach is centred around the generation of large microarray datasets that follow gene expression changes over time in response to multiple stresses, including senescence and Botrytis.

For the senescence time-course, *Arabidopsis thaliana* plants were grown under 12 h day/12 h night conditions. Leaf 7 was tagged on emergence and biological replicates were harvested both morning and evening (7 h and 14 h into light period) every other day from 19 days after sowing (DAS) until 39 DAS. This resulted in a time-course consisting of 22 time point samples from before full expansion to fully senescent. Combining the morning and evening samples for a single day gave another version of the senescence time-course, which had 11 timepoints in total. This experiment was covered in a recent publication by Breeze *et al.* (2011).

For the Botrytis time series analysis, leaf 7 was detached from four week old *Arabidopsis*

thaliana plants and placed on a tray. Leaves were inoculated with 10μl droplets of a suspension of *Botrytis cinerea* spores at a concentration 100,000 spores per ml in grape juice. Droplets of grape juice alone were put on mock inoculated leaves. Leaves were harvested from plates every 2 h after infection up to 48 h post infection. This resulting time-course consisted of 24 timepoints comparing transcriptional differences between infected and mock infected leaves. The publication that details the Botrytis experiment is currently being prepared (Windram *et al.*, in preparation).

Both senescence and Botrytis time-course experiments were performed using Complete Arabidopsis Transcriptome Micro Array (CATMA) arrays (Sclep *et al.*, 2007). In contrast to Affymetrix arrays, which are single channel arrays, CATMA arrays are two-channel (or two-colour), and require the hybridisation of cDNA from two samples that are alternately labelled with the fluorescent dyes Cy3 and Cy5. The CATMAv3 arrays used in these studies contain probes that map to approximately 24,000 Arabidopsis coding sequences. This excellent gene coverage combined with the affordability of manufacturing the arrays made them an logical choice for such a large scale experiment.

These high-resolution datasets can be used to elucidate transcriptional networks using a variety of theoretical and experimental approaches. Network inference is a theoretical modelling approach that uses expression profiles to predict regulatory relationships between genes. The high-resolution nature of the time-course is more suited to inference methods than previous transcriptional profiling experiments consisting of fewer timepoints (Breeze *et al.*, 2011). Network inference methods such as VBSSM (Beal *et al.*, 2005) have been used to reconstruct transcriptional networks using expression data from the senescence and Botrytis time-courses.

Clustering of the genes that are statistically differentially expressed is another computational method that has been used to understand the time-course data. Grouping together sets of genes based on similarity in expression profile can reveal general insights into the behaviour of genes over the time-course and identify timepoints where there are more significant changes in expression compared to others. Sets of genes that are coexpressed over the time-course may share common biological functions and may be coregulated at the transcriptional level. Tools that analyse these clusters, such as those that use Gene Ontology, then provide a means to interpret the biological significance of expression profiles, while analysing the promoters of coexpressed genes can provide insights into the regulatory mechanisms that are responsible for generating the observed expression patterns.

A key part of elucidating transcriptional networks is the identification of TFs that bind to the promoters of differentially expressed genes, including other TFs. The PRESTA project has developed a Y1H system that uses a TF library consisting of approximately 1400 Arabidopsis TFs. Using the promoters of genes that are differentially expressed in the PRESTA time-course datasets, it is possible to use Y1H to build models of regulatory networks around these genes. Protein-DNA interactions can also be used as priors for future rounds of network inference.

1.7 Organisation of this thesis

Substantial changes to gene expression are a hallmark of the plant stress response. In part, this must be due to the action of TFs binding to regulatory DNA associated with target genes. This thesis presents an interdisciplinary investigation of the regulatory codes that are responsible for controlling transcription in response to both abiotic and biotic stress in the model plant Arabidopsis. Following on from this introductory section, chapter 2 introduces the basic concepts and methods covered in this study. Chapter 3 details the design and implementation of a software framework for the analysis of biological sequences and patterns associated with them. These computational tools are then used in chapter 4 to study regulatory codes at the sequence level, discovering a novel sequence motif associated with regulating drought tolerance, and associating known TF binding with sets of coregulated genes during senescence and the responses to Botrytis infection. Chapter 5 presents an experimental approach to discovering interactions between non-coding DNA associated with target genes; chapter 6 couples the experimental approach presented in chapter 5 with *in silico* analysis methods detailed in chapters 3 and 4 to provide a detailed mapping of the protein-DNA interactions that can occur within the promoters of key stress-associated TFs. This analysis uncovers multiple components of the regulatory code and expands our understanding of the transcriptional networks that operate during senescence and in response to infection by Botrytis.

Chapter 2

Materials and Methods

2.1 Plant material

The *myb2* (SALK_045455) and *myb108* (SALK_076395) mutants of Arabidopsis Col-0 were obtained from T-DNA insertion lines from the Salk Institute Genomic Analysis Laboratory (http://signal.salk.edu/). The SALK_076395 line was heterozygous and was selfed to generate homozygous lines by Oliver Windram at The University of Warwick. The *MYB108Ox* line is the *bos1* mutant (Mengiste *et al.*, 2003).

2.2 Plant and fungal growth

2.2.1 Plant growth

Arabidopsis seeds were stratified in 0.1% w/v agarose at 4°C for 72 h in complete darkness. Stratified seeds were sown in pre-watered Arabidopsis soil mix (6:1:1 ratio of Levington F2 compost:sand:vermiculite) in 4-cm pots (P24, Plant-pak). Pots were covered with cling film and placed in a growth chamber to germinate. The covering was removed 7 days post sowing and seedling thinned out to give one plant per pot. Plants were grown in standardised conditions under 16 h days at 20°C, 350 ppm CO_2 concentration and 120 μmol.m^2.s^{-1} light.

2.2.2 Botrytis growth

Botrytis cinerea (Botrytis) strain pepper spores (Denby *et al.*, 2004) were germinated and cultured on sterile tinned apricot halves (Tesco, UK) in Petri dishes 4 weeks prior to use. Two weeks prior to use, Botrytis was sub-cultured using the same procedure. Sub-cultures were incubated at 25°C in complete darkness. Spores were harvested in sterile water and filtered through glass wool to remove hyphae. Inoculums were prepared by

suspending spores in half strength sterile grape juice (Tesco) and concentration adjusted to 1×10^5 spores/ml. Spore concentration was measured with a hemocytometer.

2.2.3 Botrytis infected-leaf expression analysis

For Botrytis infection expression analysis, leaf 7 was tagged on emergence. Leaf 7 from 4-week old plants was detached and placed on 1% w/v agar in propagator trays. Col-0, *myb2* and *myb108* leaves were inoculated with 4-6 (depending on leaf size) evenly spaced 10μl droplets of Botrytis spores at the above concentration. An identical number of leaves from each line were mock inoculated with half-strength sterile grape juice (Tesco). Propagator lids were placed on trays and grown under 16 h days at 20°C, 350 ppm CO_2 concentration and 120 μmol.m^2.s^{-1} light and 90% humidity. Infected leaves from mutants and Col-0 were harvested at the same time post infection to provide 4 biological replicates. Samples for the comparison between *MYB108* mutant lines (*myb108* knockout and *MYB108Ox*) and Col-0 were harvested at 26 hours post infection (hpi) and 30 hpi to give two timepoints. Samples for the comparison between *myb2* knockout line and Col-0 were harvested at 24 hpi and 30 hpi to give two timepoints. Whole leaf samples were snap frozen in liquid nitrogen at the time of harvesting and stored at -80°C.

2.2.4 Dark-induced senescence expression analysis

Three week old Col-0, *myb2* and *myb108* rosettes cut from the root were transferred to Petri dishes containing water-saturated filter paper. Three rosettes were included in each Petri dish and were stored at 20°C in complete darkness. Each day, a Nikon D50 digital SLR camera was used to photograph samples illuminated with four 100 W tungsten bulbs positioned along the side of the sample. Petri dishes were placed on a background that contained a white reference point and place markers to ensure plate position and camera zoom were consistent across all Petri dish photographs (Figure 2.1). Three petri-dishes containing rosettes were photographed every day providing nine biological replicates. Images were imported into ImageJ, where a colour histogram was computed for the white reference on the background, and a lassoed area that covered leaf 5 from each rosette. This procedure was repeated for all three plates. The RGB intensities from all three replicates within a petri dish were normalised using the RGB values from the white-background reference point using an R script written by Stuart McHattie at The University of Warwick (McHattie, 2011). The same script then computes the ratio between the mean red and mean green values for leaf 5 from all nine replicates, providing a quantitative measure of leaf yellowing. Past experience in the Buchanan-Wollaston lab suggests that a red-green ratio of around 0.8 indicates the initiation of senescence. When

the average ratio of Col-0 samples was >0.8, leaf 5 from different Petri dishes for Col-0, *myb2* and *myb108* lines were harvested at the same time to give a total of 4 biological replicates for each line. Whole leaf samples were snap frozen in liquid nitrogen at the time of harvesting and stored at -80°C. The same sampling procedure was then performed on consecutive days to sample as senescence progresses.

2.3 Microarray analysis

2.3.1 RNA extraction

Snap-frozen Arabidopsis leaves were ground in 1ml Trizol reagent (Invitrogen) using a Dremel drill for 1 min until the sample was completely homogenised. The drill-bit was frozen in liquid nitrogen prior to use to prevent thawing of leaf tissue. Samples were incubated at room temperature for 5 min to allow for dissociation of nucleoprotein complex before adding 200μl chloroform. Reaction was shaken vigorously by hand for 15 sec and incubated for 3 min at room temperature. Samples were then centrifuged at 8,000 $\times$ *g* for 15 min at 4°C. The upper aqueous phase ($\sim$0.5ml) was transferred to a fresh 1.5ml Eppendorf tube followed by addition of 0.5ml of isopropanol to precipitate the RNA. Samples were mixed by inverting tubes several times and incubated at -20°C for 2 h. Samples were centrifuged at 8,000 $\times$ *g* for 20 min at 4°C. RNA pellets were washed with 1ml 75% EtOH followed by centrifugation at 8,000 $\times$ *g* for 10 min at 4°C. Supernatant was completely removed and pellet allowed to air-dry for 5 min before re-suspension in 100μl RNase free water. Total RNA was purified using Qiagen RNeasy purification kit (Qiagen) according to the manufacturers protocol except for the last step, where purified RNA was eluted from the column with 2x40μl RNase free water. Total RNA concentration was measured using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific) using 1μl sample. Total RNA quality was determined using a 2100 Bioanalyzer with the RNA 6000 Nano LabChip kit according to manufacturers instructions (Agilent). The Bioanalyzer assesses total RNA integrity by observing the 18S and 28S rRNA peaks using a high-resolution electrophoresis system. Where total RNA samples displayed no rRNA peaks or a poor 18S/28S ratio (ratio <1), total RNA was isolated from alternative single leaf samples.

2.3.2 RNA amplification

Total RNA was amplified using the MessageAmp-II aRNA Amplification Kit (Ambion Biosystems) according to the manufacturers instructions, using a single round of amplification and an *in vitro* transcription step with an incubation time of 14 h. Quality

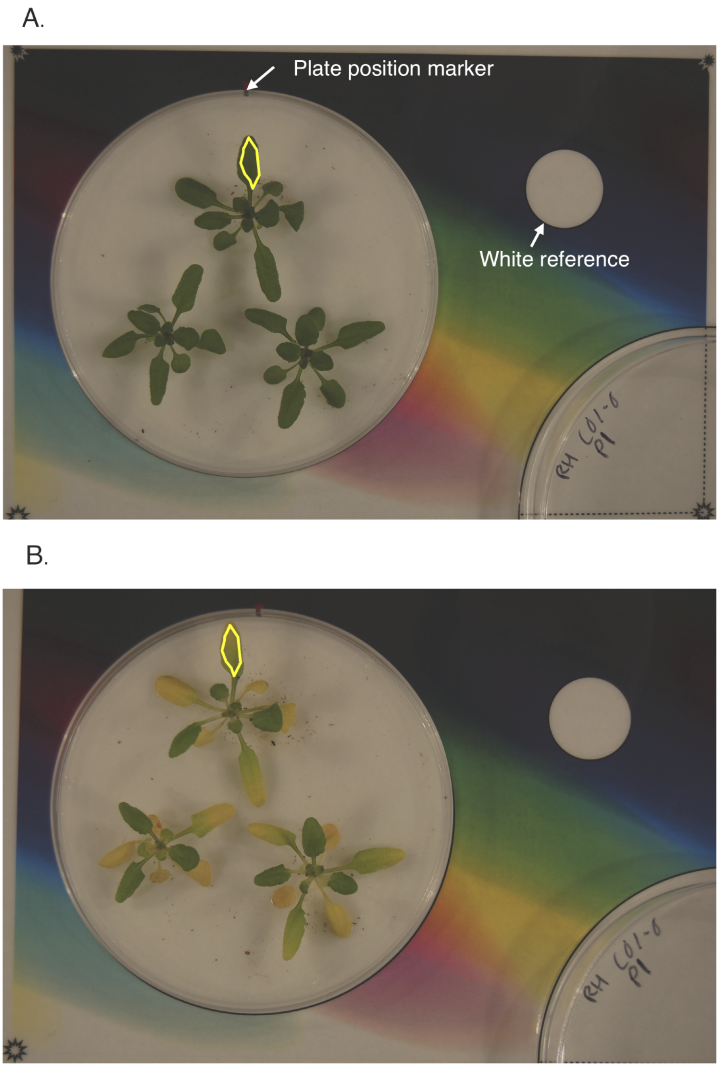


Figure 2.1: **Quantification of dark-induced senescence in single leaves.** (A) Petri dishes containing whole Arabidopsis rosettes were placed on a coloured background according to a set position marker. A camera was suspended above the background and the zoom adjusted until the corners of the background are just visible. The background contains a white reference point that serves to normalise RGB values of each leaf within a Petri dish. Photographs are imported into ImageJ where RGB values are computed for the area covering leaf 5 (yellow outline). Photographs of the same samples are taken every day until the onset of senescence. (B) The same plate as shown in (A) as it begins to visibly senesce.

of amplified RNA was determined using a 2100 Bioanalyzer with the RNA 6000 Nano LabChip kit according to manufacturers instructions (Agilent) and concentration of the purified sample was measured using a Nanodrop ND-1000 spectrophotometer. Good quality amplified RNA should display a size distribution that is approximately a normal distribution. Where the size distribution was clearly abnormal or the amplified RNA

concentration <300ng/μl, total RNA for that sample was re-amplified.

2.3.3 Microarray experimental design

For both the dark-induced senescence and Botrytis treatments, mutant samples (i.e., *myb2*, *myb108* and *MYB108Ox* lines) were compared with Col-0 samples at each given timepoint. For each sample, amplified RNA from 4 biological replicates were pooled. Each comparison consists of 4 technical replicates including two dye-swaps.

2.3.4 Direct labelling of amplified RNA

Approximately 5μg of pooled amplified RNA, generated by combining equal amounts of amplified RNA from each of the appropriate biological replicates, was combined with 0.5μl random nonamers (3μg/ul) (Invitrogen), 0.5μl of RNase inhibitor (RNase OUT; Invitrogen) and made up to a final volume of 10.5μl. Samples were incubated at 70°C in a thermocycler for 10 min. Superscript mastermix was created by combining the following reagents per reaction: 4μl 5×Superscript II First Strand Buffer (Invitrogen), 2μl 0.1M DTT (Invitrogen), 1μl dNTP mix (10mM dATP, 10mM dGTP, 10mM dTTP, 2mM dCTP) and 1 μl SuperScript II reverse transcriptase (Invitrogen). Samples were labelled by adding 8μl of mastermix with 1.5μl of either Cy3- or Cy5-dCTP (GE Healthcare) followed by incubation in the dark at 42°C for 2.5 h. 2μl of 2.5M NaOH was added to each of the labelled cDNA samples followed by incubation at 37°C for 15 min. Samples were combined with 10μl of 2M MOPS buffer and purified using a QiaQuick PCR purification kit (Qiagen) according to manufacturers instructions. At the end of the procedure, the purified cDNA was eluted with 2× 30μl of Buffer EB (Qiagen). The concentration of the purified sample was measured at 532nm (Cy3) or 635nm (Cy5) using a Nanodrop ND-1000 spectrophotometer.

2.3.5 CATMA array hybridisation

CATMAv4 array slides were prepared for hybridisation by incubating them in Coplin jars with Pre-Hybridisation buffer (1.2g Bovine Serum Albumin (Sigma Aldrich, A9418), 5× SSC, 0.1% SDS in 120ml sterile H_20) (pre-warmed to 42°C at 42°C) for 1 h. CATMA arrays slides were washed by submergence into sterile H_2O for a total of 5 washes and a final wash with isopropanol. Slides were dried by centrifugation for 1 min at 2,000× *g*. Combinations of 40pmol of the appropriate Cy3- and Cy5-labelled samples were freeze dried until nearly dry and resuspended in 50μl of hybridisation buffer (12.5μl Formamide, 12.5μl 20× SSC, 0.35μl 14% SDS, 6.25μl 4μg/μl Yeast tRNA (Invitrogen) and 18.4μll sterile $H_2O\mu$l). Resuspended sample was then incubated at 95°C for 5 min in a thermocycler followed by centrifugation at 10,000 × *g* for 1 min. The hybridisation mix

was applied to an array slide located within a hybridisation chamber (Corning) followed by the application of a coverslip (Sigma Aldrich) and chamber cover. Hybridisation chambers were placed in a humid environment at 42°C for 16 h.

Coverslips were removed by submerging array slides in 250ml Wash Solution 1 (2 × SSC, 0.07% SDS, 250ml H_20) (preheated to 42 °C) until free. Hybridised slides were then incubated in Wash Solution 1 in a hybridisation rack for 5 min on an orbital shaker. Slides were then incubated in 250ml Wash Solution 2 (0.1 × SSC, 0.07% SDS, 250ml H_20) for 10 min on an orbital shaker. Slides were then incubated in Wash Solution 3 (0.1 × SSC and 995ml H_20) for 1 min on an orbital shaker for a total of 4 washes. Slides were briefly immersed in isopropanol and dried by centrifugation for 1 min at 1,000 × *g*.

2.3.6 Microarray scanning

Arrays were scanned on a 428 Affymetrix scanner at wavelengths of 532nm and 635nm for the Cy3 and Cy5 labelled samples respectively. The gain was adjusted on a slide-to-slide basis in an attempt to give unsaturated foreground and minimal background fluorescence. Scanned images were quantified using Imagene 7.5.0 software (BioDiscovery, Inc.).

2.3.7 Expression analysis

Comparisons between mutant and wild-type samples were analysed using the R Bioconductor package limmaGUI (Wettenhall and Smyth, 2004). Raw data were normalised within arrays using a PrintTip lowess transformation and then normalised between arrays using the quantile-normalisation. The data were fitted to a linear model using a least squares method. P-values were adjusted for multiple testing using Benjamini and Hochberg method to control the false discovery rate (Benjamini and Hochberg, 1995).

2.3.8 Gene Ontology (GO) analysis

Overrepresentation within gene lists for Gene Ontology (GO) categories is analysed using BINGO (Maere *et al.*, 2005). BINGO performs hypergeometric tests to determine if a particular GO term associated with a set of genes is expected by chance when considering the number of genes associated with that term in the whole Arabidopsis genome.

2.3.9 AtGenExpress Affymetrix microarray data

Microarray data were downloaded from Weigel World (http://www.weigelworld.org/resources/microarray/AtGenExpress/). These data had been processed by gcRMA to give relative log2 gene expression values. For hormone experiments, Affymetrix CEL files

were downloaded from TAIR and processed using the gcRMA function of the Bioconductor package to generate relative log2 expression estimates. Mean expression values were computed from replicates for each treatment and respective controls, and differences used to calculate log2 fold changes. Where the effects on roots and shoots were assessed in separate experiments for a treatment, only expression values for shoots are considered.

The growth conditions for abiotic stress set are described in Kilian *et al.* (2007). Treatments were initiated in parallel 18 DAS. Following the onset of stress treatment, two biological replicates of shoot samples were taken at 0.5 h, 1 h, 3 h, 6 h, 12 h and 24 h post treatment. Arabidopsis plants of the same genotype and grown under identical conditions, were exposed to different stresses, which included heat, cold, drought, salt, osmotic, genotoxic (Bleomycin and mitomycin), oxidative, UV-B light and wounding.

The growth conditions for the hormone series are described in Goda *et al.* (2008). The hormone series includes treatment with several phytohormones. Expression data were extracted for ABA and JA where samples were taken 0.5 h, 1 h, and 3 h following treatment, and for SA where samples were taken at only 3 h post treatment.

Expression levels for biotic stresses were extracted from the AtGenExpress pathogen series. Data were extracted for bacterial pathogen genotypes DC3000, AvrRpm, HrcC, and Psph at 2 h, 6 h and 24 h post inoculation with total of 3 biological replicates. Data were extracted for bacterial elicitor Flg22 at 1 h and 4 h post treatment with 3 replicates. Data were extracted for the hemitrophic pathogen, *P. infestans* at 2 h, 6 h and 24 h post inoculation with total of 3 replicates.

2.4 Conservation analysis of promoters using EARS

Sequence conservation between promoters was analysed using the comparative genomics tool, Evolutionary Analysis of Regulatory Sequences (EARS) (Picot *et al.*, 2010). This methods compares two sequences by using a fixed length sliding window and computing the optimal alignment score for all possible window pairs at each position in both sequences. Alignment scores are converted to p-values that indicate the chance of observing that alignment score if compared to randomly chosen Arabidopsis DNA. The analysis of two sequences was performed using the online version of the tool (http://wsbc.warwick.ac.uk/ears/) with default parameters (window size of 60 bp).

2.5 Yeast-1-Hybrid

2.5.1 Yeast-1-Hybrid components

Yeast-1-hybrid (Y1H) experiments were conducted using the MATCHMAKER One-Hybrid Library Construction and Screening Kit (Clontech) with modified components generated by Dr Claire Hill at The University of Warwick. The Y1H system has three main components: 1) Bait vector that contains the reporter gene. 2) Prey vector that contains activation-domain-TF (AD-TF) fusion. 3) Yeast strain that lacks the ability to grow on minimal media. The structure of the bait and prey vectors are depicted in Figure 2.2.

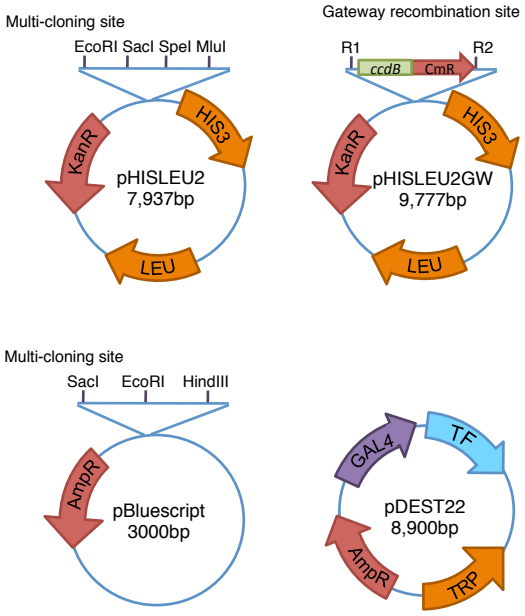


Figure 2.2: **Plasmids used in Y1H experiments.** The bait vectors pHISLEU2 and pHISLEU2GW contain a cloning site into which the bait sequence can be cloned just upstream of the *HIS3* reporter gene. *HIS3* encodes a histidine biosynthesis enzyme that allows yeast cells to grow on media lacking histidine. The two vectors differ in terms of the method required to integrate the bait sequence; pHISLEU2 contains a multi-cloning site that contains multiple different restriction enzyme recognition sites, while pHISLEU2GW is only compatible with the Gateway cloning system. pHISLEU2 and pHISLEU2 contain the *LEU* gene that allows yeast to grow on media lacking leucine. The pBluescript vector is useful for a sub-cloning applications due to its high-copy number and small size. pBluescript contains a multi-cloning site that incorporates many restriction sites; shown are restriction sites relevant for sub-cloning into bait vector pHISLEU2. The prey vector pDEST22 is a Gateway compatible vector that contains attR sites, which allows for recombination of TF ORFs flanked by attL sites. TFs are then expressed as a fusion to the GAL4-AD. pDEST22 contains the *TRP* gene that allows yeast to grow on media lacking tryptophan.

2.5.2 Amplification of promoter fragments

2.5.2.1 Restriction based primers

Promoter fragments were amplified by polymerase chain reaction (PCR) from Col-4 genomic DNA using appropriate primer pairs (Table 2.1). PCR was performed using KOD Hot Start DNA Polymerase (Novagen) using the standard kit reagents in 50μl reactions: 5μl 10x buffer, 1.5 mM $MgCl_2$, 10mM dNTPs, 0.4μM forward primer, 0.4μM reverse primer, 10ng template DNA, 1 unit of KOD polymerase enzyme. PCR program consisted of an initial denature step of 95°C for 5 min, followed by 40 cycles of 95°C for 30 sec, 60°C 30 sec and 72°C for 30 sec, corresponding to the denature, anneal and extend steps, ending with 72°C for 7 min.

2.5.2.2 Gateway cloning based primers

To generate promoter fragments suitable for Gateway cloning (Invitrogen), a two-step amplification approach was used. Promoter fragments were PCR amplified from Col-4 genomic DNA using appropriate primer pairs listed in Table 2.2. PCR reactions consisted of 5μl KOD master mix (Novagen), 10μM forward primer, 10μM reverse primer and 100ng template DNA, made up to 10μl with sterile H_2O. The following PCR thermocycler program was used: 1 cycle of 95°C for 2 min, followed by 11 cycles of 95°C for 15 s, 55°C for 15 s and 68°C for 2 min. Once this initial round of amplification had finished, a second PCR amplification was performed using generic Gateway primers: Forward primer: 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCT-3' and reverse primer: 5'-GGGGACCACTTTGTACAAGAAAGCTGGGT-3'. The second-step PCR reaction consists of 10μl of the first-step PCR reaction, 20μM forward generic primer, 20μM reverse generic primer, 25μ KOD master mix, made up to a final volume of 50μl with sterile H_2O. The PCR thermocycler program consisted of 1 cycle at 95°C for 2 min, followed by 5 cycles of 95°C for 15 sec, 45°C for 15 sec and 68°C for 2 min, then 35 cycles of 95°C for 15 sec, 55°C for 15 sec and 68°C for 2 min, ending with 68°C for 5 min.

2.5.3 Amplified product clean-up and purification

PCR products were run on analytical 1% w/v TAE agarose gel to check for clean amplification of promoter fragment. Successful PCR reaction volumes were run on a TAE 1.5% w/v agarose preparative gel, observed bands cut and purified using QiaQuick gel extraction kit (Qiagen). Concentration of purified DNA measured using a Nanodrop ND-1000 spectrophotometer.

Table 2.1: Primers used to amplify promoter fragments for use in restriction based cloning.

Gene	Fragment	Direction	Restriction	Sequence
ANAC019	F1	Fwd	SacI	5'-GGGGAGCTCTGTGAACTCCAAAT GTCTAAGCAGC-3'
ANAC019	F1	Rev	SpeI	5'-GGGTACTAGTCTACTTATATATGG ACGCAGTGTATAGGTG-3'
ANAC019	F2	Fwd	EcoRI	5'-GGGGAATTCAGTCCCCTACACGA ATATATAGACCACC-3'
ANAC019	F2	Rev	HindIII/MluI	5'-GGGAAGCTTACGCGTTGCAATCC CTTGAGTTCATTGAACCTAC-3'
ANAC019	F3	Fwd	EcoRI	5'-GGGGAATTCTCCTTCTTATTACAA CATTTGGCACAT-3'
ANAC019	F3	Rev	HindIII/MluI	5'-GGGTAAGCTTACGCGTCGACATCT AAACCACAAAACATTATTG-3'
ANAC019	F4	Fwd	EcoRI	5'-GGGGAATTCATCGTTATAACGTCT TTTTGAC-3'
ANAC019	F4	Rev	SpeI	5'-GGGAACTAGTCCATGAAATATGTG AACTGGTAAAGG-3'
ANAC019	F5	Fwd	EcoRI	5'-GGGGAATTCTGCGACCATTATATA TGGTTCTATC-3'
ANAC019	F5	Rev	MluI	5'-GGGACGCGTCGACACGTGTCCAC GTGTCTATCGTAG-3'
ANAC055	F1	Fwd	EcoRI	5'-GGGGAATTCATAAGAGGAGGTACA GTCACAC-3'
ANAC055	F1	Rev	HindIII/MluI	5'-GGGAAGCTTACGCGTCGAAGCTC TGCTACTCGTGTATGTA-3'
ANAC055	F2	Fwd	EcoRI	5'-GCCGAATTCATCCCATCATTCACT TACA-3'
ANAC055	F2	Rev	HindIII/MluI	5'-GGGAAGCTTACGCGTGATCAATTA GAGCGTCGTGATTTATG-3'
ANAC055	F3	Fwd	EcoRI	5'-GGGGAATTCGTTTGTTGTTTGTCC CTCTCTCTG-3'
ANAC055	F3	Rev	HindIII/MluI	5'-GGGAAGCTTACGCGTTGAGTTACA TAACAGTGACAATCTACG-3'
ANAC055	F4	Fwd	EcoRI	5'-GGGGAATTCGAGAAGCGTGTTTGT GTTATACGGACTT-3'
ANAC055	F4	Rev	HindIII/MluI	5'-GGGAAGCTTACGCGTTGTGTCTAT TGGTTGAGTTAGG-3'
ANAC072	F2A	Fwd	EcoRI	5'-GGGGAATTCCAGCGGACTTTTCTC ATAATCGTAGTA-3'
ANAC072	F2A	Rev	HindIII/MluI	5'-GGGAAGCTTACGCGTGTAAGTAGT TGGTTTATATAGGCTACG-3'
ANAC072	F3A	Fwd	EcoRI	5'-GGGGAATTCCCGAACGGCACCAAT ATAGTCG-3'
ANAC072	F3A	Rev	SpeI	5'-GGGACTAGTTTCCAGCGATATAAAT GTGC-3'
ANAC072	F4A	Fwd	EcoRI	5'-GGGGAATTCGACTCCAATCTATGT AGTGATC-3'
ANAC072	F4A	Rev	SpeI	5'-GGGACTAGTTCCACTGAGAAAGCA ATGAAGC-3'
ANAC072	F5A	Fwd	EcoRI	CCCCGAATTCTGAAGTCTTGCCTAAC TCTCCCAACAC-3'

Continued on next page

Table 2.1 – *Continued from previous page*

Gene	Fragment	Direction	Restriction site	Sequence
ANAC072	F5A	Rev	SpeI	5'-GGGACTAGTTGCTCCAAATAATTA GAGGCCAGTTC-3'
ANAC092	F1	Fwd	EcoRI	5'-GGGGGAATTCATACATTGTTTTCAC GAGATGGATAACATTTG-3'
ANAC092	F1	Rev	HindIII/MluI	5'-CCCCAAGCTTACGCGTCAATTTGA CCAGGAACACTTCACCTGTAAC-3'
ANAC092	F2	Fwd	EcoRI	5'-GGGGGAATTCGTCATTCTCTTCAA ACCATTCACACTG-3'
ANAC092	F2	Rev	HindIII/MluI	5'-CCCCAAGCTTACGCGTAGGTTTCG CGTTTAAATCTAATACAAC-3'
ANAC092	F3	Fwd	EcoRI	5'-GGGGGAATTCGTTACAGGTGAAGT GTTCCTGGTCAAATTG-3'
ANAC092	F3	Rev	HindIII/MluI	5'-CCCCAAGCTTACGCGTGAGCAACG AAGCTCCTCTCTATATACTAG-3'
ANAC092	F4	Fwd	EcoRI	5'-GGGGGAATTCGTTGTATTAGATTT AAACGCGAAACCTCATG-3'
ANAC092	F4	Rev	HindIII/MluI	5'-CCCCAAGCTTACGCGTTTTATCCT AATAGGGTTTCTAAAAATGATC-3'
ANAC092	F5	Fwd	EcoRI	5'-GGGGGAATTCATACATTGTTTTCA CGAGATGGATAACATTTG-3'
ANAC092	F5	Rev	HindIII/MluI	5'-CCCCAAGCTTACGCGTTTTATCCTA ATAGGGTTTCTAAAAATGATC-3'

Table 2.2: Primers used to amplify promoter fragments for use in Gateway cloning system.

Gene	Fragment	Direction	Recombination site	Sequence
ANAC072	F1B	Fwd	attB1	5'-AAAAAAGCAGGCTTCCAGCGGACT TTTCTCATAATCGTA-3'
ANAC072	F1B	Rev	attB2	5'-CAAGAAAGCTGGGTCATAGGCTACG GAAAAGCCAAAG-3'
ANAC072	F2B	Fwd	attB1	5'-AAAAAAGCAGGCTTCTAGACATTT TACATGTCACGATGG-3'
ANAC072	F2B	Rev	attB2	5'-CAAGAAAGCTGGGTCTTCCAGCGA TATAAATGTGC-3'
ANAC072	F3B	Fwd	attB1	5'-AAAAAAGCAGGCTTCACTCAATAG ATTTATCTGTC-3'
ANAC072	F3B	Rev	attB2	5'-CAAGAAAGCTGGGTCCGTCAGGAA AAGCTAAAGTTTATC-3'
ANAC072	F4B	Fwd	attB1	5'-AAAAAAGCAGGCTTCAGAGACTAC TTGCTACATTTGTGA-3'
ANAC072	F4B	Rev	attB2	5'-CAAGAAAGCTGGGTCGAGTATTAA TCAGACACGCGTGG-3'

2.5.4 Generation of bait vectors

2.5.4.1 Cloning into pHISLEU2

Amplified promoter fragments were doubly digested using appropriate restriction enzyme pairs and buffer (New England Biolabs; NEB) in 50μl reactions: 5μl buffer, 0.5μl BSA, 5μg DNA, 1μl enzyme A, μl enzyme B. Reactions were incubated for 2 h at 37°C. At the same time, pHISLEU2 plasmid was doubly digested using the same reaction scheme. Digested DNA was run on 1% w/v agarose gel and purified using QiaQuick gel extraction kit (Qiagen) according to the manufacturers protocol. Purified DNA concentrations were determined using a Nanodrop ND-1000 spectrophotometer. To ligate digested insert into cut pHISLEU2 vector a reaction volume of 10μl, containing 30fmol vector, 90fmol insert, 2μl T4 ligase Buffer (Invitrogen) and 1μl T4 ligase (Invitrogen) was incubated at 21°C overnight. Control ligations were carried out using plasmid and no insert. The ligation mixture was transformed into *E.coli* DH5α cells (Bioline) using heat shock: Ligation reaction was added to 50μl cells and incubated on ice for 10 min. The mixture was incubated at 42°C for 30 sec followed by recovery on ice for 2 min. Cells were incubated for 1 h at 37°C with shaking following addition of 200μl SOC (Super Optimal broth with Catabolite repression) media. Positive clones selected by plating on LB-agar containing kanamycin (50μg/ml).

2.5.4.2 Verifying promoter insert in pHISLEU2 vectors

Transformants were verified by colony PCR to confirm the presence of the promoter fragment using Taq polymerase (Invitrogen) and primer pairs designed to amplify the multi-cloning site: forward primer 5'-CTATCAGGGCGATGGCCCACTA-3' and reverse primer 5'-AATGCACTCAACGATTAGCG-3'. PCR thermocycler program consisted of an initial denature step of 94°C for 2 min, followed by 25 cycles of 94°C for 30 sec, 50°C 30 sec and 72°C for 2 min, corresponding to the denature, anneal and extend steps, ending with 72°C for 7 min. Positive transformants where the correct size band was observed were inoculated in 4ml LB containing kanamycin (50μg/ml) and incubated at 37°C overnight with shaking at 220 rpm. Bacterial cells were harvested by centrifugation at 10,000 $\times$ g and plasmid DNA purified using QIAprep Spin Miniprep Kit (Qiagen) according to the manufacturers protocol. Correct promoter sequence inserts were confirmed by in-house sequencing using Big Dye (Applied Biosystems) sequencing reaction mix following the manufacturers protocol.

2.5.4.3 Cloning into pHISLEU2GW

Amplified promoter fragments flanked by attB sites were sub-cloned into pDONRZeo (Invitrogen) via Gateway BP recombinant reaction between attB and attP sites. BP reaction consisted of 1μl PCR product (150ng/μl), 1μl Donor vector (150ng/μl) and 2μl TE buffer (pH8.0). 1μl of BP clonase II enzyme mix (Invitrogen) was added to the reaction and vortexed briefly. Reactions were incubated at 25°C for 2 h in thermocycler. BP reaction was transformed into Alpha select Gold efficiency competent cells (Bioline) using heat shock: 1μl of BP reaction add to 10μl Alpha select Gold efficiency competent cells and incubated on ice for 30 min. Mixture was incubated at 42°C for 30 sec followed by incubation on ice for 2 min. Cells were incubated at 37°C for 1 h with shaking following addition of 200μl SOC media. Positive clones selected by plating on LB-agar containing Zeocin (50μg/ml) and incubated overnight at 37°C.

Positive transformants were verified by colony PCR to confirm the presence of the promoter fragment using Taq polymerase as described previously, with generic Gateway primers M13F and M13R that are designed to amplify the region between recombination sites. Transformants where the correct size band was observed were inoculated in 4ml LB containing Zeocin (50μg/ml) and incubated at 37°C overnight with shaking at 220 rpm. Plasmid DNA was extracted using the QIAprep Spin Miniprep kit and sequenced in house using Big Dye as described previously.

Promoter fragments were transferred into the destination vector pHISLEU2GW via Gateway LR recombinant reaction between attL and attR sites. LR reaction consisted of 1μl pDONRZeo (150ng/μl), 1μl pHISLEU2GW (150ng/μl) and 2μl TE buffer. 1μl of LR clonase II enzyme mix (Invitrogen) was added to the reaction and vortexed briefly. Reactions were incubated at 25°C for 2 h in thermocycler. LR reaction was transformed into Alpha select Gold efficiency competent cells (Bioline) using heat shock as described above. Positive clones selected by plating on LB-agar containing Kanamycin (50μg/ml) followed by incubation overnight at 37°C. Positive transformants were verified by colony PCR and sequenced in-house to confirm the presence of the promoter fragment as described for pHISLEU2.

2.5.5 Generation of mutagenised promoter fragments

PCR products corresponding to the promoter fragments that were to be mutated and pBluescript KS vector (Stratagene) were doubly digested with appropriate restriction enzyme pairs and buffer (NEB) using the same reaction scheme as used for cloning into pHISLEU2 in section 2.5.4.1. To ligate digested insert into cut pBluescript KS vector

a reaction volume of 10μl, containing 10fmol vector, 30fmol insert, 2μl T4 ligase Buffer (Invitrogen) and 1μl T4 ligase (Invitrogen) was incubated at 16°C overnight. Control ligations were carried out using plasmid and no insert. The ligation mixture was transformed into *E.coli* DH5α cells as described in section 2.5.4.1. Positive clones selected by plating on LB-agar containing carbencillin (100μg/ml).

Mutagenised promoter constructs were created using inverse PCR with primer pairs shown in table 2.3 in 10μl reactions containing 10μM forward primer, 10μM reverse primer, 5μl KOD (Novagen) and 25ng DNA template. The following PCR thermocycler program was used: 1 cycle of 95°C for 2 min and 35 cycles of 95°C for 15 sec, 58°C for 15 sec, and 68°C for 2 min ending with 1 cycle of 68 for 5 min. After amplification, 0.5μl Dpn1 (NEB) was added to the reaction and incubated at 37°C for 1 h to digest the methylated, non-mutagenised bacterial plasmid template. Undigested DNA was extracted using a PCR purification kit (Qiagen) according to the manufacturers instructions. Fragment ends were phosphorylated using T4 polynucleotide kinase (Invitrogen) in the following reaction: 5μl 5× Forward Reaction Buffer (Invitrogen), 1μl T4 polynucleotide kinase (Invitrogen), 0.3μl ATP (100μM) and 18.7μl DNA. Reactions were incubated at 37°C for 10 min followed by 10 min at 65°C to inactivate the enzyme. Phosphorylated DNA was purified using a PCR purification kit (Qiagen) according to the manufacturers instructions. Purified DNA was ligated using T4 ligase (Invitrogen) as described in 2.5.4.1. The resulting mutagenised pBluescript KS constructs were transformed into *E.coli* DH5α cells (Bioline) using heat shock: 2 μl ligation reaction was added to 10μl cells and incubated on ice for 10 min, followed by incubation at 42°C for 30 sec and recovery on ice for 2 min. Cells were incubated for 1 h at 37°C with shaking following addition of 200μl SOC media. 100μl of cell mix were plated onto LB-agar containing kanamycin (50μg/ml). Correct mutagenised constructs were verified as in 2.5.4.2.

The mutagenised promoter construct present in pBluescript KS was sub-cloned into pHISLEU2, by treating both plasmids with relevant restriction enzymes followed by DNA purification as described in 2.5.4.1. Purified promoter fragment was then ligated into pHISLEU2 vector using the T4 ligase (Invitrogen) as described in 2.5.4.1. Ligated vector-insert constructs were transformed into *E.coli* DH5α cells (Bioline) using heat shock as described above. Correct mutagenised pHISLEU2 constructs were verified as in 2.5.4.2.

Table 2.3: Primers used for site directed mutagenesis of ANAC055 promoter fragment 4.

Mutant	Direction	Sequence
m55F4a	Fwd	5'-ATTAAACTTTTGACGAAAGAAAAA TATCCCGCCTAACTCAAC-3'
m55F4a	Rev	5'-AAGATCAGAGAGAGGGACAAACAA CAAACACGAGT-3'
m55F4b	Fwd	5'-AAAAAAAATTTGACGAAAGAAAAA TATCCCGCCTAACTCA-3'
m55F4b	Rev	5'-AAGATCAGAGAGAGGGACAAACAA CAAACACGAGT-3'
m55F4c	Fwd	5'-AAAAAAAAGACGAAAGAAAAATAT CCCGCCTAACTCAA-3'
m55F4c	Rev	5'-TTAAAGATCAGAGAGAGGGACAAA CAACAAACACGAGT-3'

2.5.6 Yeast growth media

2.5.6.1 YPDA media

Yeast extract peptone dextrose supplemented with adenine (YPDA) media contained 20g/L Glucose, 20g/L Peptone, 10/L Yeast extract and 100mg/L adenine. All media autoclaved before use.

2.5.6.2 Synthetic Dropout media

Synthetic Dropout (SD) media were prepared by combining 26.7g/L minimal SD base (Clontech, catalog number 630411) with appropriate drop-out (DO) supplements:

-leucine (-L) 0.69g/L (Clontech, catalog number 630414)

-tryptophan (-T) 0.74g/L (Clontech, catalog number 630413)

-leucine-tryptophan (-LT) 0.64g/L (Clontech, catalog number 630417)

-leucine-tryptophan-histidine (-LTH) 0.62g/L (Clontech, catalog number 630419)

Where appropriate, SD-LTH media were supplemented with 3-Amino-1,2,4-triazole (3-AT) (Sigma Aldrich).

2.5.7 Yeast transformations

Transformation of a given plasmid into yeast strain Y187 or AH109 was performed as follows: Yeast strain grown in a overnight in 10ml of appropriate selective selective media at 30°C with shaking. A 1ml aliquot of yeast culture typically containing approximately 10^8 cells was pelleted by centrifugation at 2000 rpm for 5 min. Supernatant was removed and cells resuspended in 1ml 0.1M lithium acetate, followed by centrifugation at 2000 rpm for 5 min and resuspension in 1ml fresh lithium acetate and incubation at 30°C for 1 h. A DNA mix was prepared consisting of 500ng plasmid DNA, 40μg of single-stranded carrier DNA (salmon sperm carrier DNA; Sigma Aldrich) and 290μl 50% w/v PEG 3350 (Fluka) and preheated to 30°C . 100μl yeast cell suspension was added to the DNA mix followed by gentle mixing and incubation for 50 min at 30°C followed by heat shock at 42°C for 15 min. Cells were pelleted by centrifugation at 3000 rpm for 5 min, supernatant removed and cells resuspended in 200μl sterile H_2O. Cell suspension was plated on minimal SD (synthetic drop-out)-base (Clontech) appropriate to the transformed vector and incubated until colony growth was visible.

2.5.8 TF-library construction

The normalised Y1H screening library of ORFs cloned into pDEST22 vector was generated by Dr Claire Hill (University of Warwick). The library consists of 1037 generated in the REGIA and REGULATORS projects and a further 332 cloned in-house using 312 ORFs from the ABRC and 20 in-house cloned ORFs. TF cDNAs that were not already present in pDEST22 were introduced via Gateway cloning system (Invitrogen). Individual AD-TF clones were transformed into *S. cerevisiae* AH109 strain (Clontech) followed by selection in minimal SD media lacking Tryptophan (SD-Trp). Resulting prey strains were pooled into sets of 12 AD-TF in order to generate workable libraries. Two alternate pooling strategies were used to generate two uniquely pooled versions of the library. Pooled TF libraries were stored as glycerol stocks in 96 well plates.

2.5.9 Screening a promoter fragment against AD-TD library

2.5.9.1 Mating bait and prey strains to generate co-transformants

Cultures of each AD-TF library in AH109 yeast strain were grown in 2.2ml deep well 96 well plates (Thermo scientific) containing 500μl SD-T media. Plates were covered with Gas-permeable seal and incubated for 72 h at 30°C with moderate shaking. The day prior to screening, 10ml SD-L media were inoculated with appropriate bait strains and grown overnight at 30°C and at 200 rpm. Using an 8 channel pipette, 3μl of each cultured pool of TFs from a given 96 well plate were spotted onto a YPDA plate, allowed

to dry and overlaid with 3μl of bait strain and incubated at 28°C for 24 h. This process was repeated for each 96-well plate that contains a cultured AD-TF library.

Colonies were then replicated onto plates containing SD-LT, SD-LTH and SD-LTH supplemented with 3-amino-triazole (3-AT, a competitive inhibitor of the *HIS3* gene) using "velvets", a material to which yeast cells can adhere. A velvet is placed on a replica-plating block. Contact with the YPDA plate transfers yeast colonies to the velvet. These colonies can then be transferred to the selective plates. The amount of 3-AT used to supplement SD-LTH plates varied between promoter fragments, depending on the amount of autoactivation observed. Replicated plates were incubated at 30°C for 24 h. Velvets and the replica-plating block were then used to remove dead yeast from plates, leaving living yeast cells to be fully exposed to selective media. Plates were incubated at 30°C for 72 h.

2.5.9.2 Selection and characterisation of positive interactions

Following incubation of the diploid yeast cells, plates were studied for colony growth compared to the background. Colonies can only grow if the TF present in the pDEST22 plasmid present in the diploid cells can interact with the bait sequence present in the pHISLEU2 plasmid. Positive colonies were picked from the mating screen and re-streaked onto SD-LT agar plates and incubated at 30°C for 24 h. Identity of TF present in positive colonies was confirmed by colony PCR. Cells were prepared by boiling a small amount of colony in 10μl of 20mM NaOH for 10 min. Amplifications were performed in 21μl reactions containing 1.2μl prepared yeast cells, 2μl 10× buffer (Invitrogen), 2μl Orange-G dye, 0.6μl $MgCl_2$ (50mM, Invitrogen), 0.4μl dNTPs (10mM, Invitrogen), 10μM forward primer, 10μM reverse primer, 0.1μl Taq polymerase (Invitrogen). A PCR program of 30 cycles of 94°C for 1 min, 50°C for 1 min and 72°C for 2 min, and 1 cycle at 72°C for 7 min. Amplification was assessed by running 3μl PCR reaction on a 1% w/v agarose gel. Amplified DNA was purified using Multiscreen-HTS filter plates (Millipore). Amplified DNA was sequenced in-house using Big Dye (Applied Biosystems) sequencing reaction mix following the manufacturers protocol with the primer 5'-CTAACGTTCATGATAACTTCATG-3'.

2.5.9.3 Re-testing positive interactions by direct co-transformation

Bait vectors containing the promoter fragment to be tested were transformed into AH109 prey strains carrying the AD-TF of interest as described in section using SD-LT selection plates. As a control, bait vectors were transformed into AH109 prey strains expressing AD alone (empty pDEST22 vector). Colonies were used to inoculate 10ml SD-LT liquid

media and grown overnight at 30°C with shaking. Yeast culture concentrations were calculated using a spectrometer and 10^8 cells added to a 2ml Eppendorf tube. The culture was centrifuged for 5 min at 1000 $\times$ g at room temperature. Supernatant was removed and pellet resuspended in 1ml sterile H_20, and centrifuged for 5 min at 1000 $\times$ g at room temperature. Supernatant was removed and pellet resuspended in 1ml sterile H_20. Serial dilutions were performed to give 10^8 cells/ml, 10^7 cells/ml, 10^6 cells/ml, 10^5 cells/ml and 10^4 cells/ml. Diluted yeast cultures were spotted onto the following selective plates: SD-LT, SD-LTH, SD-LTH+5mM 3-AT, SD-LTH+10mM 3-AT, and SD-LTH+20mM 3-AT. Positive AD-TF-bait interactions were identified by observing positive colony growth on media that is more selective than that on which control transformants exhibit growth.

Chapter 3

Software and Method Development

3.1 Introduction

A significant proportion of any eukaryotic genome consists of non-coding DNA, which is vital for controlling gene transcription in space and time. In essence, the non-coding DNA contains the regulatory code that is responsible for manipulating transcription. Identification and characterisation of these sequence features, and how they interact with one another in order to regulate transcription, is key to developing a systems-level understanding of gene expression. The availability of an ever increasing number of fully sequenced genomes, combined with equally large amount of gene expression data, has driven the development of computational techniques that allow scientists to probe the function of these non-coding sequences.

Since the sequencing of the first genomes, a recurring problem has plagued the analysis of the resulting sequence data. Bioinformaticians often want to ask questions of the sequence data using different tools, but even though these tools often complement each other, with the output from one serving as input to another, the two are usually not directly compatible. The end result is that the user must spend a significant amount of time performing repetitive tasks such as parsing files and converting between different file formats, which otherwise could be spent examining the real biological question. This problem is not helped by individual coding efforts which result in software that is written to be used there and then and not designed for easy further development, resulting in what is sometimes termed "PhD-ware" or "Postdoc-ware". In this scenario, the likely outcome is that further down the line researchers repeat what has already been implemented.

Sequence analysis software that is truly re-usable and can be readily developed by others requires a common representation of the biological entities relevant to the research, along with the tools that are used to study them. Such a strategy is suited to an object-orientated programming (OOP) approach to software design. OOP is a method of programming which models the problem space as a set of real world objects. This results in limited code that describes a finite set of entities relevant to the problem, which is easier to organise, has a more intuitive structure, and is easier to develop.

OO-approaches to software development for biological research do already exist with the BioPerl toolkit probably the best example of this (Stajich *et al.*, 2002). BioPerl was the first published software framework to be based on a collection of objects which represent biological entities and contain the methods required to manipulate them. This means that users can execute and view the results of common bioinformatics tools such as BLAST and ClustalW with relatively little effort. However, BioPerl does not explicitly deal with the creation, manipulation and application of sequence features that may be important in regulating transcription. Approaching this more specific requirement, the Transcription Factor Binding Site (TFBS) framework (Lenhard and Wasserman, 2002) is a set of OO Perl modules that are designed for the identification and analysis of TF-binding sites. TFBS contains objects that model common elements related to transcriptional regulation, including weight matrices (WMs) and external tools that generate a new WM from sets of sequences. The functionality offered by TFBS is limited, however, and is written in old OO Perl, which is particularly verbose.

This chapter presents the Analysis of Plant Promoter Linked-Elements (APPLES) toolkit. APPLES is a collection of OOP modules that provide a framework in which sequence analysis can be performed, and is geared towards exploiting transcriptional datasets such as microarray time series; namely identifying non-coding sequences that may be important for regulation of transcription. APPLES is written in MooseX, an extension of the Perl OO system, which offers a much more concise syntax compared with the older OO-Perl.

The APPLES project has numerous contributors, but I personally wrote approximately 10,000 lines of code and together with Dr Laura Baxter (University of Warwick) developed the core of the APPLES software platform. The vast majority of the APPLES content described in this chapter was developed by myself. The resulting code from this chapter has contributed towards the following publication:

Baxter, L., Jironkin, A., Moore, J., Hickman, R., Barrington, C., Krusche, P., Tiskin, A., Beynon, J., Denby, K., Ott, S. **(2012)** Conserved Noncoding Sequences Shed Light on the Core Regulatory Network in Dicotyledonous Plants. (*Submitted*)

3.2 Results

3.2.1 Object-orientated design

There are three main concepts in OOP which are harnessed to guide development: modularity, inheritance and interoperability. APPLES is conceptualised as a set of objects that represent biological and theoretical entities, which work together in order to answer questions related to biological sequences. APPLES objects can be divided into two broad groups, depending on whether they represent biological features related to biological sequences or approaches to manipulate and analyse these sequences. This OO-approach allows us to break down the sequence analysis problem into smaller components, which ultimately reduces the complexity of the program. Each instance of an object is generated via a set of instructions which specify its attributes and behaviour, and is known as a class. The primary APPLES classes are listed and described in Table 3.1 and graphically depicted in Figure 3.1. The biological entities that need to be modelled are sequences and the patterns that occur within them that contribute towards regulation of gene expression.

The class design is such that the objects have the key attributes and methods required to describe themselves and offer the expected functionality, while they can be expanded through the key OO principle of inheritance. For example, a biological sequence is represented as a genomic interval, through the class *Genomic_Interval*. A genomic interval has no context and so inheritance is used to extend this basic object to model a sequence believed to have a regulatory function - a *ReMo* (Figure 3.2A). The same approach can be used to model sequence patterns. The base class is the *Generic_Sequence_Pattern* and from this more specific patterns are described, such as *k-mer*, *Generic_WM* and *Generic_Pair_WM*- which represent a short DNA sequence, a weight matrix model of a TF binding site, and a pair of weight matrices that model two TF binding sites that co-occur with certain spacing rules respectively (Figure 3.2B).

The net result of this design approach is a common representation for each biological sequence feature. This means that the objects containing the methods that use these sequence features know what to expect as input format and know what format to output: for example, a *Generic_Pattern_Matching_Model* object is defined by a specific *Generic_Sequence_Pattern* and the method of matching that pattern. It takes a set of

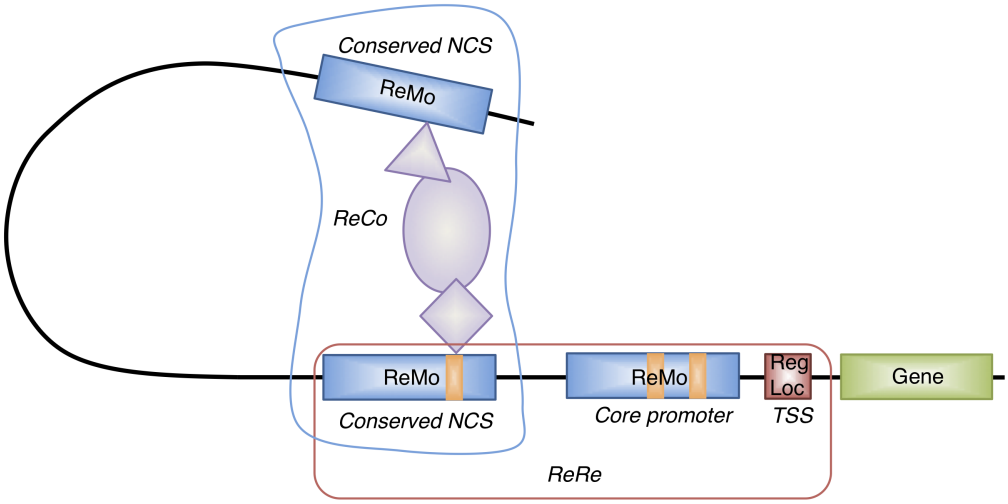


Figure 3.1: **Biological entities incorporated into the APPLES object structure.** APPLES objects model the biological features relevant to the analysis of non-coding sequences. A gene (green box) can have several regulatory features. The transcription start site (TSS) is modelled as a regulated locus (Reg_Loc) and represents a genomic locus. Within the sequence surrounding the TSS there may be stretches of DNA that are required for binding factors important for transcriptional regulation, a regulatory module (ReMo), such as the core promoter or an evolutionarily conserved non-coding sequence (NCS). Within a ReMo there may be binding sites (BiSi; orange boxes) that interact with regulatory proteins. All of the ReMos that may be regulating a gene are termed a regulatory region (ReRe). ReMos may interact with one or more proteins which aggregate to form a regulatory complex (ReCo) that is important for regulating transcription.

ReMos as input, and outputs a set of binding sites (BiSis) based on matches to the *Generic_Sequence_Pattern*.

3.2.2 APPLES functionality

The analysis tools offered by the APPLES software are designed to meet the requirements of typical sequence analysis tasks with a particular emphasis on the identification of motifs present in non-coding DNA sequences. In the simplest case, given a single promoter sequence, there are methods to identify significant instances of a known motif. This is useful for gaining a general idea of which TFs may be important for regulating a specific gene and guiding the design of promoter reporter constructs or bait sequences for Y1H experiments. This simple approach can be extended to assess whether a motif is statistically enriched within the sequence as a whole, compared to the background. Given a set of coregulated genes, there is a method to identify if their promoters are enriched for a known motif, which then may be responsible for coregulation. An alterna-

Table 3.1: List of major APPLES classes.

Class name	Description
Reg_Loc	A regulated locus. Describes a single point within a genome that is regulated at the transcriptional level. E.g., the transcription start site for a specific gene.
Genomic_Interval	A single contiguous piece of biological sequence. Can be any one of 3 defined sequence types: DNA, RNA or protein. E.g., an intergenic sequence.
Genomic_Interval_Set	A set of genomic intervals, where set members can be from the same or different species. E.g., a set of intergenic sequences.
ReCo	A regulatory complex. Describes the binding of a set of proteins to a set of genomic intervals to exert some transcriptional regulation effect on a specified regulated locus.
ReMo	A regulatory module. Derived from a Genomic_Interval. Its purpose is to distinguish regions to be considered as important for binding of a regulatory complex from any other genomic sequence. E.g., a core promoter.
ReMo_Set	A set of ReMos. E.g., a set of promoter sequences.
BiSi	A binding site. BiSi represents the location of TF binding to a DNA sequence.
ReRe	A regulatory region. A collection of ReMos for one Reg_Loc. E.g., a set of phylogenetically conserved non-coding regions present in the vicinity of a gene.
Generic_Sequence_Pattern	The description of a sequence pattern that may be functional. E.g., a PSSM or k-mer.
Generic_Pattern_Matching_Model	The model that describes how to identify occurrences of a theoretical sequence pattern in a biological sequence. E.g., scanning a PSSM across a promoter sequence.

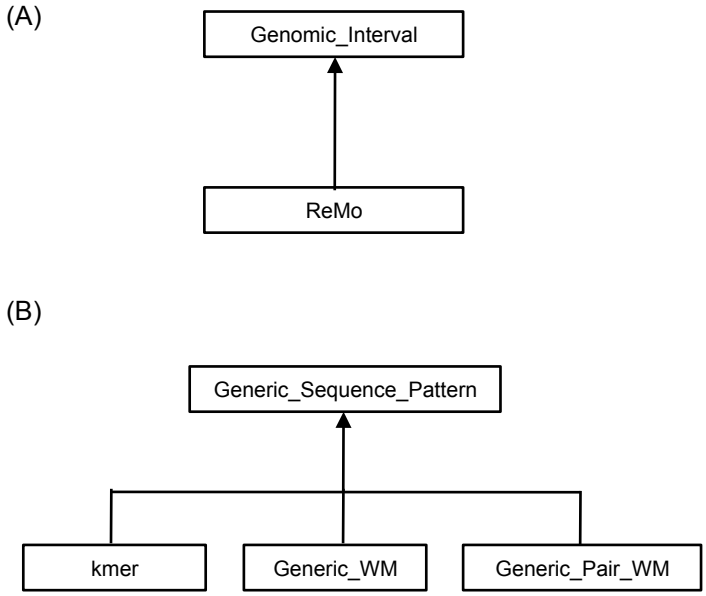


Figure 3.2: **Example APPLES object models.** Class hierarchy for objects that represent sequences (A) and sequence patterns (B).

tive, yet complementary approach, is not to search for instances of known motifs but to discover overrepresented sequence patterns within a set of coregulated gene promoters using *de novo* methods. This has the advantage of identifying novel motifs that may not have been experimentally elucidated. The methods used for each of the supported analysis tools are detailed below.

3.2.2.1 Genome sequence retrieval

Sequences can be retrieved via either local installations or remote connection to the publicly available Ensembl database (Hubbard *et al.*, 2009). Sequences can also be provided by users if specified as FASTA format. Orthologous sequences can be retrieved from local or remote installations of Ensembl Compara and Genbank.

3.2.2.2 Motif scanning

The most abundant representation for sequence patterns (motifs) is a PSSM/WM. Within APPLES motifs are represented in this matrix format. For a given PSSM, we scan a sequence for instances of the motif by computing the matrix similarity score (MSS) (Kel *et al.*, 2003) at each position on both strands. In order to define the MSS, we first need to introduce the basic similarity score. The basic similarity score, S, of a given PSSM on a sequence, T, of the same length, n, is given by:

$$S = \sum_{i=1}^{n} C_i f_{i,T(i)} \tag{3.1}$$

where $f_{i,T(i)}$ is the frequency of nucleotide $T(i)$ occurring at position i of the PSSM. C_i is an entropy parameter that describes the degree of nucleotide conservation at position i:

$$C_i = \sum_{B \in \{A,C,G,T\}} f_{i,B} ln(4 \cdot f_{i,B}) \tag{3.2}$$

The basic score, S, is then normalised to the $[0, 1]$ range to give the MSS:

$$MSS = \frac{S - S_{min}}{S_{max} - S_{min}} \tag{3.3}$$

where S_{min} is the lowest possible score obtained using the PSSM,

$$S_{min} = \sum_{i=1}^{n} C_i \min_{B \in \{A,C,G,T\}} f_{i,B} \tag{3.4}$$

and S_{max} is the highest score:

$$S_{max} = \sum_{i=1}^{n} C_i \max_{B \in \{A,C,G,T\}} f_{i,B} \tag{3.5}$$

The significance of the MSS is hard to interpret as certain PSSMs will naturally obtain higher scores than others. In order to assign significance to each MSS, p-values are computed by calculating the frequency of observing a MSS as high or higher based on a score distribution of that PSSM on random sequence (Figure 3.3A). The random sequence used to build this PSSM hit profile is automatically generated when the scoring procedure is invoked, and is built from a sequence model that is learned from the genome the sequences currently being analysed are derived from. If the genome changes then p-values are computed using a hit profile generated using random sequence derived from this alternate genome, improving the accuracy of the resulting p-values. As opposed to a score between zero and one, the p-value better reflects the probability of observing a match as good or better by chance alone and serves as a more appropriate score.

3.2.2.3 Motif-sequence overrepresentation

The ability to scan genomic sequence and highlight the strongest hits is adequate for some applications. Often, however, it is more appropriate to obtain a score that describes the overall statistical enrichment for a motif in a given sequence. Such a score is required in order to distinguish sequences that are overrepresented for a given motif compared to the background. In order to do this, scores for multiple instances of a motif within a sequence are combined to give a score that accounts for the strength and multiplicity of hits, and provides an estimate to the number of occurrences of the motif in the sequence. Given a set of MSS p-values for a motif m that cover a sequence of length n, we take the top k non-overlapping hits and perform the binomial test for the occurrence of k or more sites:

$$P(occ\{m\} \geq k) = \sum_{i=k}^{n} \binom{n}{k} p^k (1-p)^{n-k} \tag{3.6}$$

where n is the number of positions within the sequence that can be scored with the motif and p is the MSS p-value associated with the kth best match. This procedure is illustrated in Figure 3.3B.

By positively integrating multiple motif scores, the more matches to a motif a sequence exhibits the more significant the binomial enrichment score will be. Due to short length and degeneracy, a good match for many motifs will occur in a sequence of 500bp or more by chance alone. The binomial score accounts for the length of the query sequence resulting in a more statistically rigorous score that should lead to a reduced false positive rate. The binomial score will give more significance to sequences that have more motifs present and will also reveal sequences that exhibit clustering of multiple instances of weaker motifs, a feature that could otherwise be missed.

3.2.2.4 Motif-sequence set overrepresentation

Using the approach described in section 3.2.2.3, the binomial test indicates whether the occurrences of a motif in a sequence is unexpected compared to its distribution in randomly generated sequence. However, this test does not assess the significance relative to genomic sequence and so a score is required that takes the genomic background into account.

A natural question is then to ask whether a set of sequences is enriched for the motif. This is an important task in transcriptional data analysis. A common example is the identification of TF-binding sites within a set of promoter sequences, where the corresponding genes are believed to be coregulated. Essentially, the task is to assess

whether a motif is discriminatively enriched within this set of promoters compared to genomic promoters. Using a threshold, t, binomial probabilities for each sequence are binarised to mark whether or not it is significantly enriched for a motif. The number of sequences that are enriched for this motif are then compared to the number of occurrences of the motif in a background set. In order to assess the statistical significance of these occurrences within a cluster of sequences, a p-value is computed using the cumulative hypergeometric test:

$$Pval = \sum_{i=s}^{S} \frac{\binom{r}{i}\binom{R-r}{S-i}}{\binom{R}{S}} \tag{3.7}$$

where S is the number of sequences in a cluster, s is the number of sequences that have the motif, R is the number of sequences in the universe and r is the number of sequences within the universe that are enriched for the motif. Typically, the universe will be all of the genes in the genome or the set of differentially expressed genes. This procedure is illustrated in Figure 3.3C.

3.2.2.5 Motif finding

Characterisation of all sequence patterns that function in regulating gene expression is far from complete. Due to this gap in knowledge, it is often desirable to try and identify novel motifs that exist within a set of sequences. In order to perform *de novo* motif finding APPLES invokes the popular motif finder, MEME (Bailey *et al.*, 2006), which uses the expectation maximisation algorithm to discover sequence motifs. All motifs discovered by such a tool are output as a *Generic_WM* objects. These sequence patterns can then be used by any pattern matching model such as those that score sequences for motifs as described in the preceding sections.

3.2.2.6 Motif similarity and clustering

Databases that contain a large number of PSSMs that represent known TF-binding sites often contain multiple PSSMs that are slight variations of the same core sequence. As PSSMs can be considered as probability distributions, the similarity between two motifs of equal length can be computed as the relative entropy of PSSM_1 with respect to PSSM_2, also known as the Kullback-Leibler (KL) distance:

$$d = \sum_{i=1}^{L} \sum_{B \in \{A,C,G,T\}} f^1_{i,B} log_2 \frac{f^1_{i,B}}{f^2_{i,B}} \tag{3.8}$$

where L is the PSSM length and $f^1_{i,B}$ and $f^2_{i,B}$ are the probability of finding base B at posi-

tion i in PSSM_1 and PSSM_2 respectively. Since $d(PSSM_1, PSSM_2) = d(PSSM_2, PSSM_1)$ does not hold in general, we compute both relative distances and take the average. The KL distance can only be computed between an alignment of two PSSMs if they both have the same length. Where the two aligned PSSMs do not meet this requirement, the ends of each PSSM are expanded with background nucleotide probabilities until they are of the same length. A visual depiction of this procedure is given in Appendix A.

Clustering similar motifs together results in a more manageable number of PSSMs to be used in a study. Starting from a set of singleton clusters, PSSMs are subject to hierarchical clustering. In this process, each PSSM is assigned to a cluster where it is the only member. The KL distance is computed for all pairwise comparisons and the two PSSMs that display the highest degree of similarity are placed together in the same cluster. This process continues, where the inter-cluster distance is the average of pairwise KL distances. The iterative merging continues until the minimal inter-cluster distance is above some threshold value. Clustering more broadly or finely adjusts the dependence amongst results.

3.2.3 APPLES workflow

All of the biological features relevant to transcriptional regulation, along with the methods used to manipulate and interrogate them, are encapsulated within objects giving the toolkit a modular structure. These interconnected objects can then participate collectively as a system allowing for a generalised workflow to be described (Figure 3.4). A typical scenario would involve trying to identify sequence patterns that are enriched within a set of promoters from coexpressed genes. Starting with just a set of gene identifiers, the upstream sequences can be retrieved from the Ensembl database. These sequences are stored as a *Genomic_Interval_Set* object, which can then be used by several different approaches to identify overrepresented sequence patterns. Starting with known motifs for a set of given TFs, the corresponding PSSMs can be retrieved directly from the PSSM database. These PSSMs can be clustered in order to produce a partially non-redundant set of sequence patterns. A PSSM is then used to define a *Pattern_Matching_Model* object. This object can subsequently compute the binomial score for each sequence and each PSSM, indicating which patterns are enriched within each promoter. This analysis can be extended by using a *Pattern_Matching_Model* that can assess whether a PSSM is overrepresented within the set of promoter sequences compared to the occurrences of that motif in promoters at a genome-wide level. The resulting score indicates whether the set of promoters as a whole is enriched for a given PSSM. An alternative approach to using PSSMs describing known TF binding sites is to search for motifs in a *de novo* manner. This is achieved by using the same *Genomic_Interval_Set*

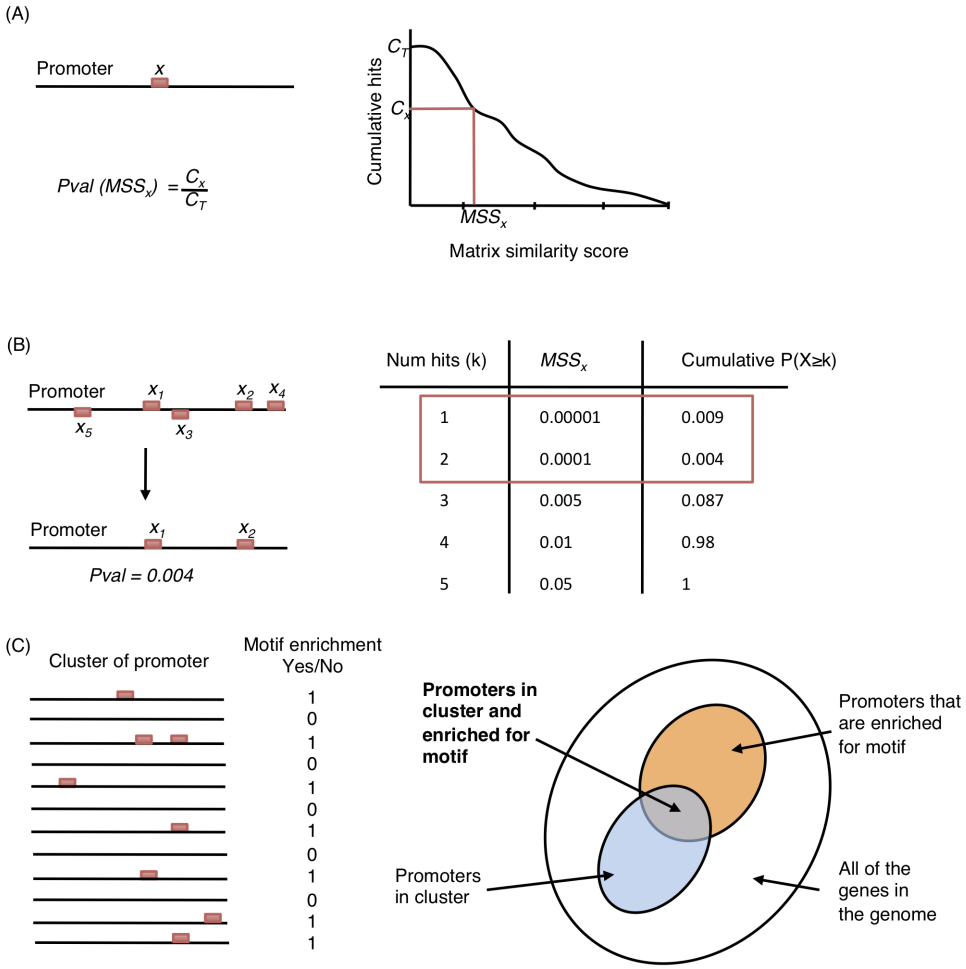


Figure 3.3: **Scoring sequences within APPLES.** (A) Scoring an instance of a motif x using a PSSM. The motif is scored at a single position by computing matrix similarity score (MSS_x), which is then converted to a p-value that represents the probability of observing a score of MSS_x or better by using a score distribution obtained by applying the PSSM to a random sequence produced by a background model. The y-axis indicates the number of observed scores as high or higher than MSS_x. (B) Scoring a sequence for a motif. A score that describes the overall abundance of motif matches in a sequence using the binomial formula. For a given sequence, the n best non-overlapping MSS are identified and converted to p-values and used to determine an overall enrichment score for motif within the sequence. The p-values are used as input to the binomial formula, which computes the cumulative probability of observing k or more hits with a pval($MSS_{x,i}$), given that the motif was scored at each position on forward and reverse strands. In the example shown here, the cumulative binomial probability was computed for observing $1, .., 5$ matches, with 2 matches giving the strongest p-value. (C) Computing overrepresentation of a motif within a set of sequences. Given a set of sequences, the motif-enrichment score is computed for each sequence and labelled as significant by applying a threshold. The set of sequences of the same type within the entire genome are marked using the same criteria and the cumulative hypergeometric statistic is computed to assess the significance of the overlap between the number of sequences in the set (blue oval) and the total number of sequences that are enriched for the motif (orange oval) given the total number of sequences in the genome.

object as input to a *Motif_Finding* object and invoking the motif finding method within it. The revealed motifs are output as *Generic_WM* objects, which can then be used with the same methods described above to assess the statistical significance of these motifs within the set of promoter sequences.

As the output from one method can be used as input to another, tailored analyses that would otherwise take many lines of code to perform, can be written using a collection of method calls that result in more concise scripts. A script that exhibits ease of use is shown in Figure 3.5. The code underlying such analyses would take a newcomer thousands of lines of code to write from scratch. Making use of the APPLES objects allows users to write short scripts that still provide maximal control over parameters and allows users to ask different questions of the data by tweaking small sections of the script in a short space of time.

3.3 Discussion

Understanding the mechanisms that control gene expression requires tools that can interrogate non-coding sequences containing the transcriptional regulatory code. APPLES provides a programming environment that allows users to perform multiple analyses using these sequences. The functionality described is particularly pertinent to the data analysis requirements of systems biology projects, such as the PRESTA project, which generate a wealth of expression data that provide a rich source of information for guiding sequence analysis. The pattern matching methods in APPLES facilitate these studies by providing a set of techniques that can be used to identify statistically significant sequence patterns that may contribute to regulating gene expression. Many excellent sequence analysis tools exist, but are often released as stand-alone applications that cannot directly interface with other relevant utilities. This lack of interconnectivity makes it hard to combine several methods into a pipeline that meets all the requirements of a typical sequence analysis. For instance, the MEME suite (Bailey *et al.*, 2006) offers a collection of some of the most popular sequence analysis tools, but combining these into a workflow is not straightforward. In contrast to sets of tools that are difficult to fit together in a workflow, the OO-approach allows the use of multiple tools that are connected together with common sets of objects. OOP developments that are specifically geared towards sequence analysis are not common and often lack some of the methods needed to analyse clusters of sequences. For example, the TFBS (Lenhard and Wasserman, 2002) toolkit is an OOP approach for the analysis of non-coding sequence yet it lacks some methods covered by APPLES. Also, unlike APPLES, TFBS is written using the old style Perl OO system, which is more low level than MooseX, and therefore more difficult for novice

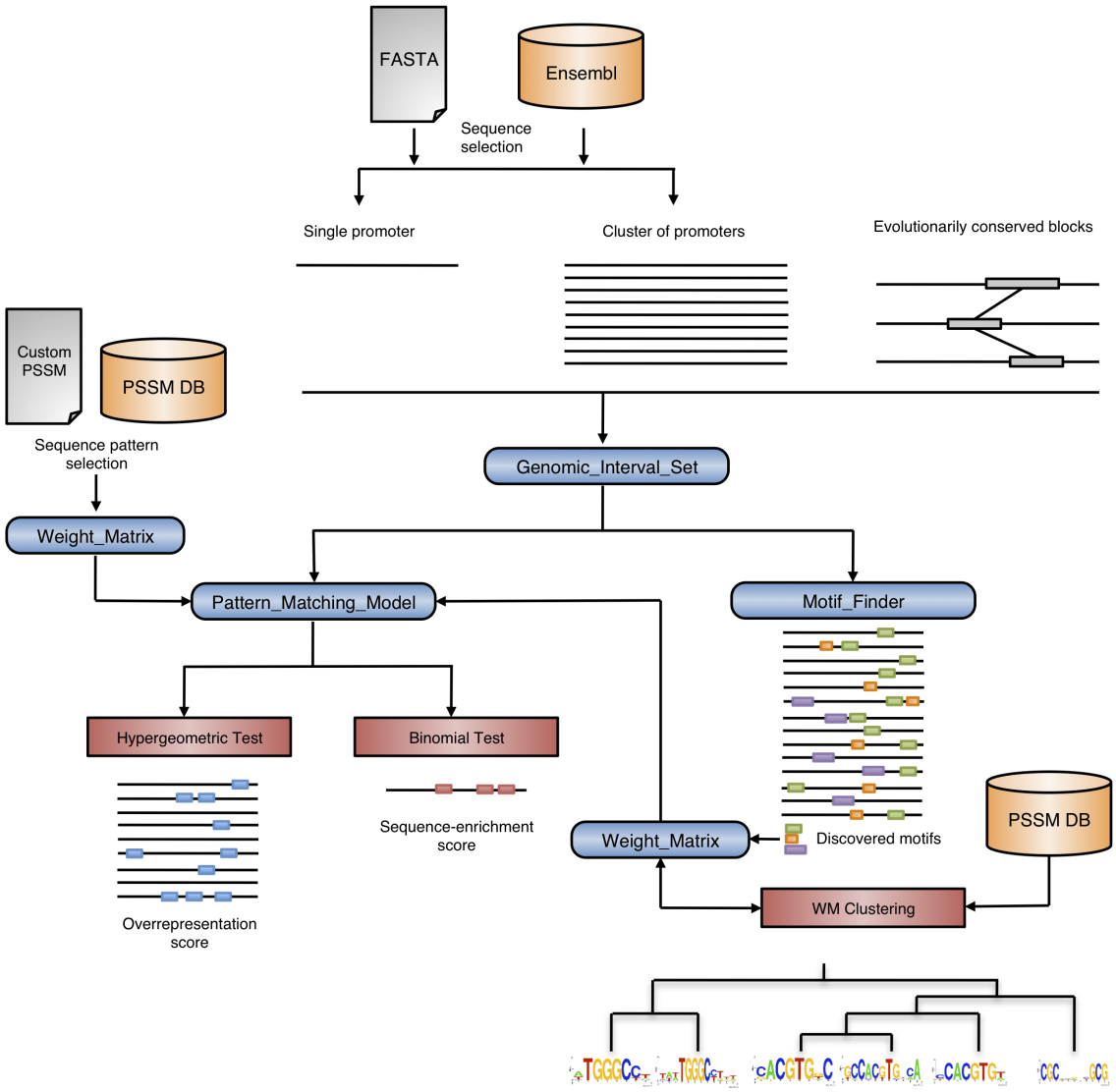


Figure 3.4: **APPLES workflow for promoter analysis.** Object orientated design allows users to link APPLES objects to generate workflows. The figure demonstrates the relationships between objects (blue rounded boxes) and methods (red boxes) that could link together in order to perform analysis of non-coding sequences. Arrows indicate links between methods and objects. The output from a method is sometimes an object that can be used as input to another method. A typical analysis of non-coding sequences will begin by assembling the sequences corresponding to the genes of interest. These sequences can be defined in a fasta file or retrieved from a genome database (DB) such as Ensembl. For a given set of orthologous sequences, evolutionarily conserved regions can be identified within APPLES. All sequences are defined as a Genomic_Interval_Set object, irrespective of origin or context, providing a common representation that can be used by any APPLES method that probes the sequence architecture. A set of sequences can then be used by a pattern matching model as input for *de novo* motif finding. Pattern_matching_model objects assess the statistical enrichment of a weight matrix within a sequence or set of sequences. The weight matrix that defines a pattern matching model is either user-defined, retrieved from the PSSM DB or derived from motif discovery programs. Weight matrix clustering can be used to cluster a group of matrices discovered using motif finding programs or any other set of weight matrix objects.

```
# -----------
# parameters
# -----------
my $core_promoter_max_length = 500;
my $core_promoter_min_length = 50;
my $stop_at_neighbouring_gene_for_core_promoter = TRUE;
my $max_number_of_sites = 5; # up to how many times do you expect your weight matrix to occur in a single ReRe?
my $threshold = 0.05; # to decide for single ReRe if pattern is present or absent
my $random_seed = 13;

# ------------------
# sequence database
# ------------------
my $genome_sequence_database_parameters == Ensembl_Database_Parameters->new(alias => 'arabidopsis', location =>
'local', dbname => 'arabidopsis_thaliana_core_30_55_9');

# -------------
# get ReRe_Set
# -------------
my $parameters_maker = Parameters_Maker->new();
my $rere_set_cons_params = $parameters_maker->default_rere_set_constructor_parameters_arabidopsis(
    $core_promoter_max_length,
    $core_promoter_min_length,
    $stop_at_neighbouring_gene_for_core_promoter
    );
my $rere_set_maker = ReRe_Set_Maker->new();
my $rere_set = $rere_set_maker->make_rere_set_for_all_reg_locs_in_a_genome_through_job_handler($rere_set_cons_params,
$genome_sequence_database_parameters);

# --------------------------------
# set up sequence pattern to test
# --------------------------------
my $pssms = ['M00434']; # PIF3
my $wm_util = WM_Utilities->new();
my $tvp = $wm_util->get_transfac_version_parameter();
my @wm_objects = @{$wm_util->get_bifa_server_wm_objects($pssms, $tvp, TRUE, TRUE)};
my $weight_matrix = $wm_objects[0];
my $empirical_scoring_parameters = $parameters_maker->default_model_based_empirical_wms_parameters($random_seed);
my $pattern_model = Multi_Occurrence_WM_Pattern_Model->new(
    pattern => $weight_matrix,
    scoring_model => $empirical_scoring_parameters,
    maximum_number_of_sites => $max_number_of_sites,
);

# ----------------------------------------
# hypergeometric tests against some clusters
# ----------------------------------------
my @geneIDlista = qw (AT1G34070-TAIR-G AT1G54050-TAIR-G AT1G53540-TAIR-G);
my @geneIDlistb = qw (AT5G12110-TAIR-G AT5G52640-TAIR-G AT5G25450-TAIR-G);
push( my @clusters, \@geneIDlista, \@geneIDlistb);

my @p_values = $pattern_model->hypergeometric_overrepresentation_test(TRUE,$rere_set,\@clusters,$threshold,TRUE);
```

Figure 3.5: **Sample script that uses APPLES objects.** The script assesses the enrichment for the PIF3 binding site (Transfac record, M00434) within a set of gene clusters using the hypergeometric test. Parameters include the length of promoters to test and the value used to threshold the binomial scores for each promoter sequence. Any parameters such as promoter length and PSSM to test can be changed to assess the impact of this on results.

programmers to use.

3.3.1 Functional decomposition

A key feature of the APPLES software concept is the principled design of objects to represent and describe biological features that are eternal in nature. There will always be biological sequences, regulatory sequences, and sequence patterns, and because the methods are built around these entities the system is more robust and extendable. Modelling biological entities in a generalised manner will facilitate reuse of code and methods in the future, as the generality allows other users to extend current object models, if needed, to meet new requirements. The high degree of re-usability allows other scientists to start asking questions immediately, and as code need not be built from scratch, productivity is enhanced. Because care was taken with the initial design, the software will not only serve to answer questions associated with the current project, but also future challenges- allowing users to build on past developments. This flexibility is important in order to exploit data from emerging technologies such as high-throughput sequencing, which will result in more sequenced genomes and expression data. For example, the rapid rate at which plant genomes are sequenced will get to the point where, for a given gene, the ortholog can be found in enough genomes for motif finding to be performed using the promoters from these sequences.

3.3.2 Future development

As with any software development there are always improvements to be made to functionality. The biggest area of further development is the incorporation of more published pattern matching/finding tools in the APPLES framework. For example, the ability to invoke more *de novo* motif finding algorithms would be to great advantage, as the use of multiple methods can improve the identification of relevant motifs (Harbison *et al.*, 2004). The bioinformatics field is littered with motif discovery programs; the more popular of these, which are not currently supported, include AlignAce (Hughes *et al.*, 2000), Consensus (Hertz and Stormo, 1999) and Weeder (Pavesi *et al.*, 2004).

Elaborate gene expression programs are likely to be orchestrated by the concerted action of multiple different motifs, rather than just one class of sequence patterns. A natural extension to current functionality would be to develop methods that can identify combinations of sequence motifs, associated within a set of sequences. This problem might be approached by identifying patterns in a *de novo* manner, or by examining combinatorial patterns of known TF binding site motifs within a regulatory sequence.

The key reason behind the design of APPLES was that it should be easy to use and develop, and it was because of this that it was extended using Perl MooseX. Perl is one of the easiest languages for a novice to learn and is frequently used in biological research. The OO behaviour is implemented through MooseX, which is high-level compared with other OO-languages. This ease of use does come at a cost, however, because it is much slower than other languages; this performance deficit could be reduced by writing CPU intensive tasks in faster languages.

Chapter 4

Transcriptional Regulatory Codes in Stress Responses

4.1 Introduction

Large-scale changes in gene expression are a hallmark of the stress response in Arabidopsis and are essential for triggering the appropriate set of events that lead to stress tolerance. Gene expression is partly controlled by TFs binding to *cis*-regulatory DNA in order to manipulate transcription of neighbouring genes. TFs exhibit sequence-specific binding behaviour, which limits the pattern of nucleotides with which they can interact. Hence, the presence/absence of sequence motifs in the non-coding sequence surrounding genes partly defines the regulatory scope for a TF. Conversely, the presence of the same motif in multiple promoters of different genes allows the same TF to regulate many different targets and underpins the concept of coregulation. Genes that are coregulated can often have a similar expression pattern and biological function.

Technologies that can measure gene expression on a genome-wide level, such as microarrays, allow the identification of genes whose expression pattern changes under certain conditions. Such an approach has been used within the PRESTA project to track changes in gene expression over time in response to stress-related treatments and assess the effects of overexpression of stress related TFs. In either case, the end product is a set of genes that are differently expressed in response to the specific treatment or mutation. This analysis can be expanded further by identifying genes with a similar expression pattern, suggesting that these genes could be regulated by an overlapping set of TFs. If this is true, the promoter regions of these coregulated genes should contain a similar set of sequence motifs that allow the common TFs to bind and subsequently influence transcription.

This chapter uses the aforementioned APPLES toolkit to analyse sets of coexpressed genes identified within the PRESTA project and to identify promoter motifs that may form part of the regulatory code that contributes towards generating observed expression patterns. First, a set of genes differentially expressed in transgenic plants overexpressing the key thermotolerance related TF *HEAT SHOCK FACTOR 3* (*HSF3*) (Lohmann *et al.*, 2004) is analysed in order to delineate its downstream targets.

Drought imposes major limits on crop productivity worldwide. Dehydration is often accompanied by an increase in temperature, and overlap has been observed in genes induced by both stresses in isolation and in concert (Rizhsky *et al.*, 2004). This finding suggests that TFs implicated in the response to heat stress may be involved in regulating drought tolerance. The heat shock factor (HSF) TFs are major determinants of thermotolerance in Arabidopsis (Baniwal *et al.*, 2004), but members of this family are also induced by drought stress (Swindell *et al.*, 2007), suggesting some HSFs may also regulate responses to dehydration.

Collaborators at The University of Essex have shown through genetic analysis with transgenic Arabidopsis plants that overexpressing one family member, *HSF3*, leads to increased drought resistance (Figure 4.1A-B). Studies with HSF3 have already demonstrated that it regulates the immediate activation of heat shock gene transcription (Lohmann *et al.*, 2004). In addition to enhanced drought tolerance, the *HSF3Ox* line (overexpressing *HSF3* under the control of the 35S promoter) also displayed the desirable trait of improved productivity under both wild-type and drought conditions (Figure 4.1C). This observed effect on drought tolerance with little impact on productivity suggests a key role for HSF3 in mediating the response to water deficit in Arabidopsis, leading to the question: how does HSF3 confer these phenotypes?

As a TF, it is likely that HSF3-mediated drought tolerance is accomplished by manipulating the expression of functional proteins that combat drought stress at the biochemical and physiological level. In order to delineate the network of genes operating downstream of HSF3, microarrays were used to compare the transcriptome of *HSF3Ox* plants with wild-type plants under non-stressed conditions. This study revealed 295 genes upregulated as a result of the *HSF3* overexpression. As a TF, HSF3 can influence gene transcription by binding directly to the promoters of target genes, or indirectly via TFs that are themselves directly induced by HSF3. Thus, the list of differentially expressed genes will contain sets of both direct and indirect targets. In order to delineate this local network structure, the promoter sequences of the differentially expressed genes were

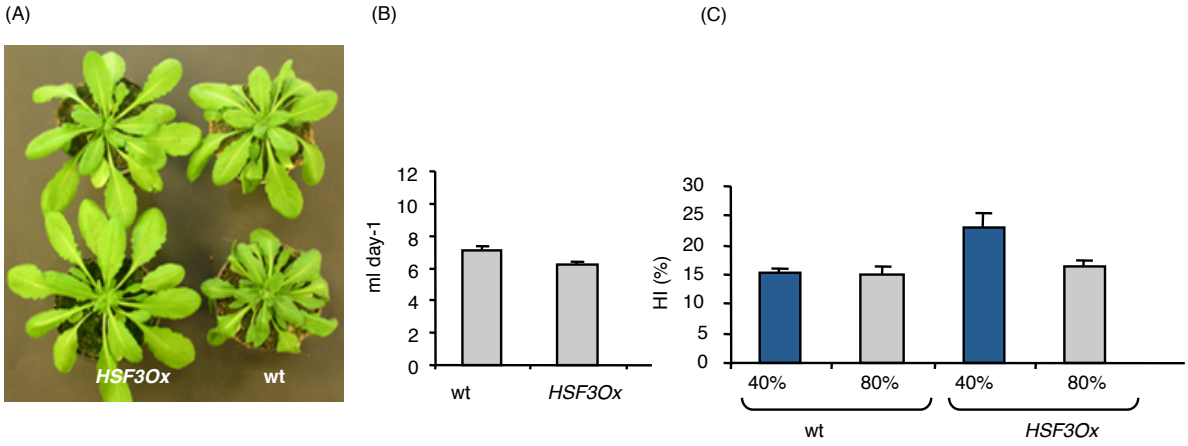


Figure 4.1: **HSF3 regulates drought tolerance and productivity.** (A) Typical phenotype of transgenic Arabidopsis plants overexpressing *HSF3* (*HSF3Ox*) compared to wild-type (wt) control 14 days since last watering. (B) Rate of water loss in *HSF3Ox* plants compared to wt control averaged over 13 days with no watering. The difference is significant; $p = 0.017$. (C) Harvest Index (HI) in *HSF3Ox* and wt plants grown with different relative soil water content (rSWC). Plants were subject to water-replete (80% rSWC) and water-limited (40% rSWC) growth regimes. The difference between *HSF3OX* and wt at 40% rSWC is significant; $p = 0.005$. All experiments and analyses performed by Bechtold and colleagues at The University of Essex.

analysed in order to identify motifs that may serve as binding sites for HSF3. Genes whose promoters are enriched for these motifs are possible direct targets for this key drought response TF.

This analysis, of a fairly limited transcriptional response driven by just one TF, is then expanded to reveal mechanisms driving more complex networks of genes controlled by some of the hundreds of TFs differentially expressed throughout senescence and in response to infection by Botrytis. Sequence analysis, using clusters of genes that are coexpressed during both conditions is used to identify promoter motifs enriched within each group and reveal sequence features that may be responsible for driving the pattern of coexpression.

Several parts of this chapter have been included in the following papers, which have either been published, submitted or are currently in preparation:

Breeze, E., Harrison, E., McHattie, S., Hughes, L., Hickman, R., Hill, C., Kiddle, S., Kim, Y.-S., Penfold, C. A., Jenkins, D., Zhang, C., Morris, K., Jenner, C., Jackson, S., Thomas, B., Tabrett, A., Legaie, R., Moore, J. D., Wild, D. L., Ott, S., Rand, D., Beynon, J., Denby, K., Mead, A., and Buchanan-Wollaston, V. **(2011)**. High-resolution

temporal profiling of transcripts during arabidopsis leaf senescence reveals a distinct chronology of processes and regulation. *The Plant Cell*

Bechtold, U., Fryer, M., Lawson, T., Richard, F., Sparrow, P., Kim, Y.-S., Hickman, R., Ott, S., Beynon, J., Buchanan-Wollaston, V., Baker, N., Morison, J., Schffl, F., Denby, K., Mullineaux, P. **(2012)** Arabidopsis HEAT SHOCK TRANSCRIPTION FACTOR A1b is a major determinant of seed yield and constitutively regulates basal resistance to diverse abiotic and biotic stresses. *Submitted*

Windram, O., Madhou, P., Kiddle, S., Hill, C., Hickman, R., Cooke, E., McHattie, S., Jenkins, D., Penfold, C., Kim, Y.-S., Baxter, L., Zhang, C., Tabrett, A., Moore, J., Wild, D., Mead, A., Rand, D., Beynon, J., Ott, S., Buchanan-Wollaston, V., Denby, K. Defence against Botrytis cinerea: chronology and regulation deciphered by high-resolution temporal transcriptomic analysis. *(In preparation)*

4.2 Results

4.2.1 Elucidating the regulatory role of HSF3

The popular *de novo* motif finder, MEME (Bailey *et al.*, 2006), was used to discover putative HSF3 binding sites in the promoters of the genes upregulated by *HSF3* overexpression. Promoter sequences were defined as the 500 bp immediately upstream of the predicted transcription start site (retrieved from Ensembl Plants sequence database; release 50). MEME takes a set of DNA sequences along with numerous optional arguments that fine-tune its behaviour, and outputs any potentially novel motifs overrepresented within a set of sequences. MEME was invoked within the APPLES software framework allowing several runs of the program with differing parameters to be performed simultaneously. Genes were ranked based on the level of increased expression, and sets of promoter sequences were created from the top 10, 20, 30, 50, 75, 100, 200 and 300 genes. Using sets of promoters corresponding to the top 10, 20, 30 and 50 ranked genes, MEME identified potential overrepresented motifs in all of these datasets. Figure 4.2 shows the motif identified by MEME using the top 50 up-regulated gene promoter sequences. This dataset was optimal in terms of the conservation of the hits and the total number of sequences in which the motif was identified. The motif could not be learned from lists with more than 50 sequences. This is likely to be because MEME finds it increasingly difficult to identify motifs with high levels of nucleotide conservation as the number of sequences in the dataset increases and the increasing noise dilutes any signals present. This putative HSF3 binding element (pHSF3E) is similar to experimentally elucidated vertebrate HSF binding site motifs V$HSF1_01 and V$HSF1_02 present in

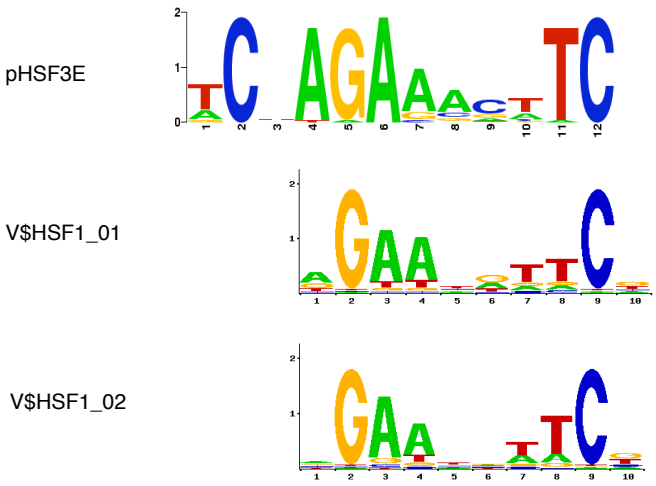


Figure 4.2: **Similarity of putative HSF3 binding site to known HSF motifs.** MEME searched for overrepresented motifs in the 500 bp sequences upstream of the predicted TSS of genes upregulated in *HSF3Ox* lines. The putative HSF3 binding element (pHSF3E) was identified by MEME using the top 50 upregulated gene promoter sequences. The pHSF3E is similar to experimentally determined binding sites for vertebrate HSFs, V$HSF1_01 and V$HSF1_02 retrieved from the TRANSFAC database.

the TRANSFAC database suggesting that it may have a functional role *in vivo*. The pHSF3E consists of three domains that are palindromic with respect to each other. This observation fits in with the requirement for plant HSFs to function as trimers (Schöffl *et al.*, 1998).

As the pHSF3E was only determined from the top 50 upregulated genes, there was no information regarding motif enrichment in the other upregulated gene promoters. Because the MEME motif had been identified using the APPLES tools, the pHSF3E was stored using standard WM pattern representation and could be used to test for enrichment within the promoter of each upregulated gene. Promoter sequences were scanned with the pHSF3E PSSM and the motif-sequence overrepresentation score (section 3.2.2.3) computed. This score was used because it accounts for a potential multiplicity of hits. Out of the 295 upregulated gene promoter sequences 73 had a high multi-score ($p \leq 0.05$). To further assess the significance of the pHSF3E within this promoter set, the occurrences of the motif were compared against ten sets of 295 promoters from randomly selected genes. The number of promoters enriched in random sets was notably lower than in the *HSF3Ox* upregulated set (Table 4.1). The hypergeometric test was used in order to calculate the significance of overlap between genes that were upregulated in *HSF3Ox* lines and the set of genes whose promoters are enriched for the pHSF3E in the entire genome. The overlap was significant ($p < 10^{-10}$) and further indi-

Table 4.1: Frequency of pHSF3E in promoters of genes upregulated in *HSF3Ox* plants and in randomly selected promoter sets.

Sequence set	Binomial p-val ≤ 0.05
HSF3 upreg	73
Random A	23
Random B	23
Random C	27
Random D	28
Random E	33
Random F	31
Random G	21
Random H	14
Random I	21
Random J	28

cates that the pHSF3E is statistically overrepresented within the set of genes responsive to HSF3. A list of the genes enriched for the pHSF3E is shown in Appendix B.

Functional *cis*-regulatory elements can exhibit positional bias with respect to the TSS (Cooper *et al.*, 2006), and strengthen the case for pHSF3E having a genuine functional role. Figure 4.3A shows the occurrences of pHSF3E within a representative subset of the 73 promoters enriched for the motif, as identified by MEME. Where a promoter was enriched for pHSF3E (binomial p-value $\leq$0.05), the positions of all occurrences of the motif were plotted as a histogram (Figure 4.3B). As a control, using the same enrichment criteria, the position of matches to the motif within the sets of 295 promoters picked at random from the Arabidopsis genome were plotted (Figure 4.3C). Comparison of the two histograms indicates that there is a clear preference for positions -150 to -50 relative to the TSS within the promoters of the genes upregulated by *HSF3* overexpression, and suggests that HSF3 has a bias in terms of where it binds to activate transcription. Studies with eukaryotic genomes have demonstrated that functional *cis*-regulatory elements are often located within 200 bp of the TSS (Tabach *et al.*, 2007). It is possible that this location-sensitive behaviour may be driven by the necessity for TFs to interact with the core transcriptional machinery, imposing minimal and maximal positional constraints on motif location. The preference for the motif to locate 50 bp from the TSS could be a result of a steric constraint imposed by the need for RNA polymerase docking, forcing the TF binding site to locate a minimal distance downstream of the TSS. The apparent preferred maximal distance of 150 bp from the TSS may simply be due to the need for the TF to be close enough to interact with the transcriptional machinery. Regardless of the underlying biophysical rationale, such positional bias provides further support to

the argument that the instances of pHSF3E identified within the set of differentially expressed genes are authentic binding locations for HSF3.

In order to test the hypothesis that HSF3 is directly targeting some genes through the pHSF3E, collaborators at The University of Essex performed experiments where *HSF3Ox* and wild-type plants were treated with cycloheximide, a chemical that inhibits protein synthesis. Following treatment, the expression levels of HSF3 direct targets should be unchanged as the extremely high levels of HSF3 should be sufficient to continue to drive transcription. Conversely, the transcription of indirect targets should be reduced, as cycloheximide will inhibit the synthesis of TFs required for activation. Expression levels of each gene tested were determined as *HSF3Ox* relative to wild-type (Figure 4.4). The majority of putative HSF3 direct targets tested were not significantly affected by the cycloheximide treatment when compared to those genes that were not enriched for the pHSF3E, which displayed reduced expression levels. Such an accurate prediction rate provides confidence that the predicted direct targets not tested are correct as well. This confirms the hypothesis that HSF3 can directly regulates genes through the pHSF3E.

In order to understand the biological processes and molecular functions associated with predicted HSF3 direct targets, Gene Ontology (GO) analysis was performed using BINGO (Maere *et al.*, 2005)(Figure 4.5). Putative direct targets of HSF3 are highly enriched for genes involved in the response to heat or temperature stimulus, suggesting that HSF3 directly regulates the activity of a large set of heat shock response genes. A large proportion (20/25) of these heat responsive genes are heat shock proteins (HSPs). HSPs function to regulate the folding, localisation and degradation of proteins in stressed and un-stressed cells (Feder and Hofmann, 1999). Drought stress can cause denaturation and damage to proteins, suggesting a requirement for HSP function during these conditions. HSF3 mediated drought tolerance may then be attributed to the direct activation of HSPs that go on to protect the cell by mediating protein stability. The predicted direct target list is also overrepresented for genes involved in the responses to oxidative stress and high light. Again, HSPs are primarily responsible for this functional enrichment suggesting that HSF3 may regulate the response to multiple stresses through activation of HSPs that combat the general negative effects of stress on protein stability. Indeed, the lack of GO classes linked to water deprivation response suggests that HSF3 mediated drought tolerance is activated via a set of proteins that have not yet been functionally linked to dehydration yet appear to confer tolerance to other abiotic stresses.

In addition to HSPs, the heat responsive TF *MULTIPROTEIN BRIDGING FACTOR C* (*MBF1c*) is a target of HSF3. MBF1c is heavily linked with regulating expression

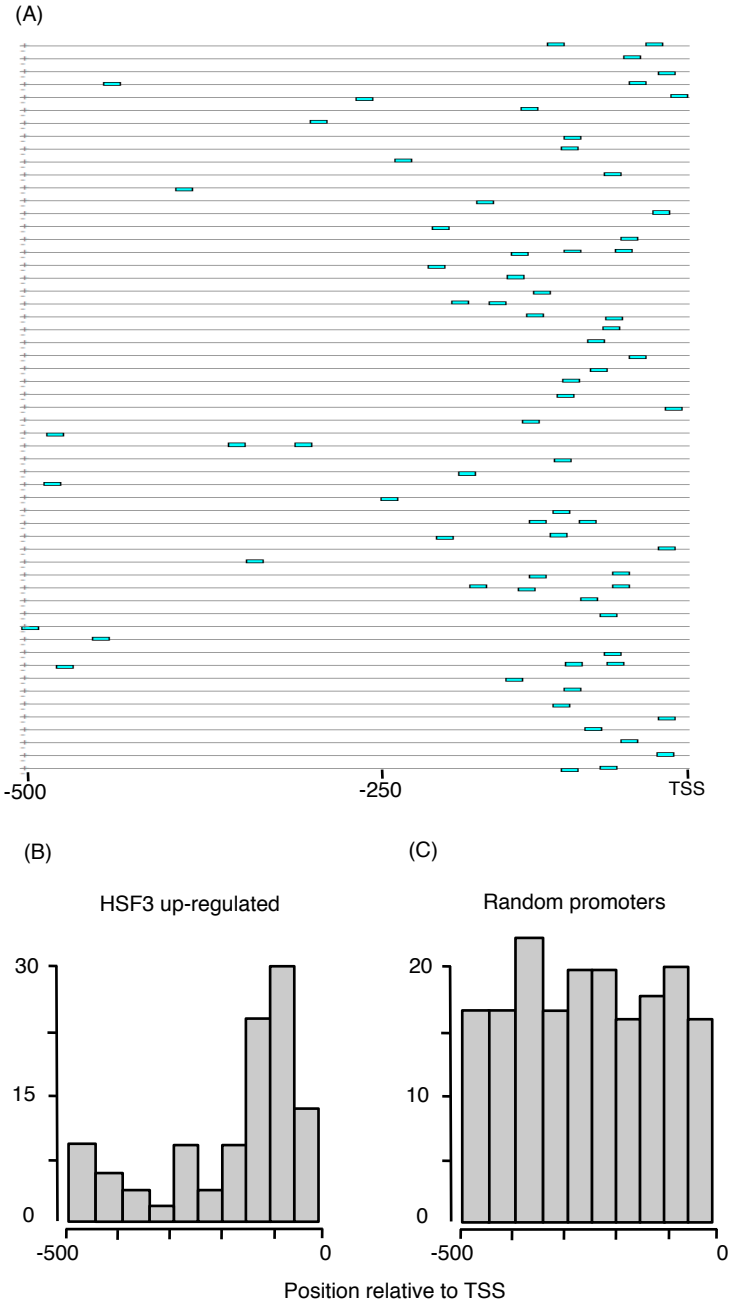


Figure 4.3: **Positional bias of the pHSF3E.** (A) Visualisation of the occurrences of the pHSF3E in a representative subset of the 73 promoters that are enriched for the motif. Image is taken directly from MEME html output. (B) The number of times that the pHSF3E is present in 50bp bins within the promoters of genes upregulated in *HSF3Ox*. All occurrences of the pHSF3E were defined as the instances of the motif that contribute to a binomial score $\leq$0.05. (C) The positional distribution of occurrences of the pHSF3E in a set of promoters from randomly selected genes.

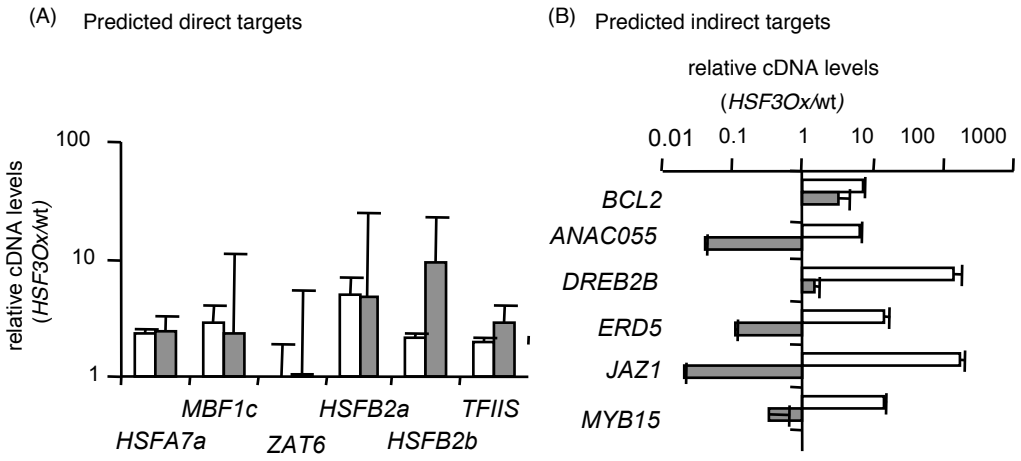


Figure 4.4: **Experimental validation of HSF3 direct/indirect targets.** Ten day old *HSF3Ox* and wild-type (wt) seedlings were treated with cycloheximide or water. Expression levels were determined using qRT-PCR using whole seedlings harvested 4 h post treatment. Expression levels of predicted direct targets in *HSF3Ox* relative to wt for cycloheximide and water treatment shown as grey bars and white bars respectively. All experiments and analyses performed by Bechtold and colleagues at The University of Essex.

of stress responsive genes and overexpression of *MBF1c* enhances tolerance to osmotic and heat stress (Suzuki *et al.*, 2005). *MBF1c* is induced by drought, heat stress and a combination of both suggesting that MBF1c contributes towards the response to dehydration by functioning downstream of HSF3 (Rizhsky *et al.*, 2004). It is likely that MBF1c, together with other TFs go on to regulate the other genes that are not direct targets of HSF3.

4.2.2 Identification of promoter motifs regulating leaf senescence

Leaf senescence is a developmental process that attempts to reclaim the organ's cellular components and distribute them to other parts of the plant. In addition to being a key developmental process senescence can also be induced prematurely by biotic and abiotic stress, while the process itself creates a highly stressful environment within the cell, with stress response pathways induced as a consequence. In order to optimise the reallocation of vital nutrients, the plant must balance this process, limiting the detrimental impact of large scale degradation, oxidation and remobilisation of molecules by tightly regulating the onset, progression and completion of senescence through complex GRNs.

In order to understand the GRNs that operate during senescence, microarray analysis was used to generate a high-resolution gene expression time-course that tracked the activity of the Arabidopsis transcriptome in a single leaf from near full expansion to

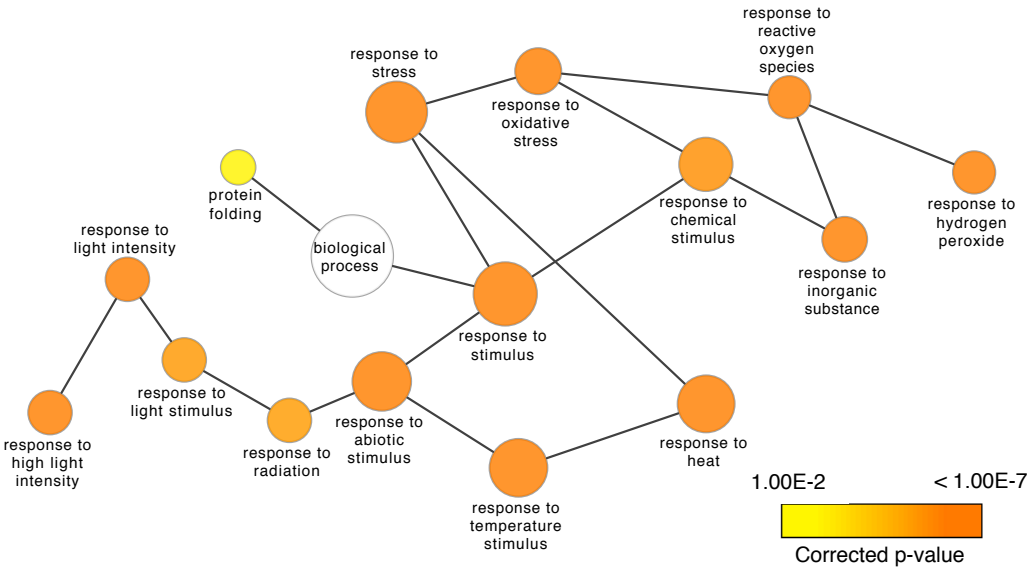


Figure 4.5: **Gene Ontology (GO) categories overrepresented in the set of putative HSF3 direct targets.** The set of HSF3 direct targets predicted based on motif enrichment were analysed using BINGO. Shown are GO categories that are significantly overrepresented. The significance of GO enrichment is indicated by colour. P-values are subject to Benjamini and Hochberg correction. Edges indicate parent-child relationships between GO terms.

fully senescent (Breeze *et al.*, 2011). Samples were taken in the morning and afternoon every 2 days from 19 days after sowing (DAS) to 39 DAS, to obtain 22 timepoints in total. Differential expression analysis identified 6323 genes that underwent significant changes in expression over the entire time-course. The set of differentially expressed genes were clustered using SplineCluster (Heard and Holmes, 2006) in order to identify sets of coexpressed genes resulting in 48 clusters (Figure 4.6). The variety in expression profiles demonstrates the complexity of the senescence process and suggests that it is elaborately regulated at the transcriptional level. A large proportion of the genes differentially expressed throughout the process encode TFs (Breeze *et al.*, 2011), which will be partly responsible for generating the observed coexpression patterns by binding to the promoters of genes in each group. Therefore, identification of the sequence motifs that interact with senescence enhanced TFs is an important step in identifying coregulated genes and expanding the GRNs that control this complex process.

To gain an initial understanding of the regulatory mechanisms of genes underlying differential gene expression during senescence, promoters corresponding to 500 bp upstream of the predicted transcription start site of genes in each cluster were screened for overrepresentation of known TF binding motifs. This analysis was performed within APPLES

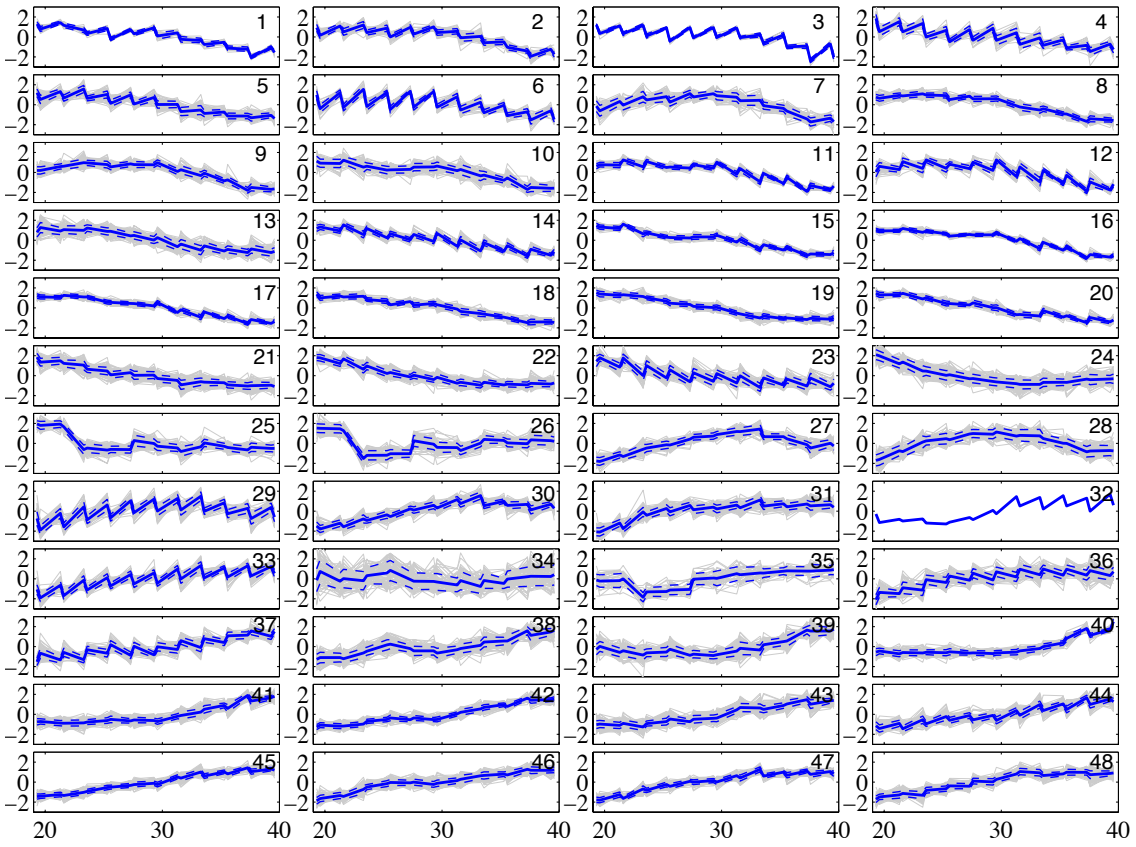


Figure 4.6: **Clustering analysis of genes differentially expressed during senescence.** SplineCluster was used to group the 6323 genes differentially expressed during senescence based on similarity in expression profile using mean normalised expression values. Clusters were taken from Breeze *et al.* (2011). The blue line on each plot represents the mean expression values at each time point. The majority of clusters follow one of two general trends, clusters 1-24 are down-regulated and 27-48 are up-regulated.

using the motif-sequence set enrichment procedure (section 3.2.2.4) which computes the cumulative hypergeometric statistic to assess whether a motif is significantly enriched within a set of sequences compared to the background. In this instance, for each known motif, it assesses the significance of the overlap in a given cluster, and the set of all promoter sequences in the genome that are enriched for that motif. Plant PSSMs were retrieved from the TRANSFAC database (Matys *et al.*, 2006) and the PLACE database (Higo *et al.*, 1999). This set was supplemented with PSSMs for a heat shock element (TRANSFAC matrix record M00146) and two NAC TF binding sites (Olsen et al., 2005) since these stress-associated motifs were absent from these databases. The set of PSSMs were clustered in order to reduce redundancy, and a representative of each cluster was chosen for screening.

Clusters were defined as overrepresented for a given motif if the hypergeometric p-value was below 0.05 following correction for the number of clusters tested using Bonferroni correction. By focusing on TF binding site motifs that were overrepresented within at least one cluster, it is possible to identify motifs that are associated several different TF families (Figure 4.7). By examining the clusters that show significant enrichment for each motif, we observe that specific sequences are associated with particular expression patterns. Figure 4.8 displays the statistical enrichment for all of the selected motifs in all of the senescence clusters. This figure demonstrates that for many of these motifs, the enrichment signal is primarily associated with clusters that have similar expression profiles. Full list of motifs enriched in senescence clusters shown in Appendix C Table 1.

GO term analysis of the clusters revealed several of the downregulated clusters (clusters 1-24) are enriched for biological processes associated with photosynthesis and cell growth (Breeze *et al.*, 2011). Consistent with these results, several of these clusters are significantly enriched for sequence motifs associated with these processes. The G-box motif is known to function as a positive and negative mediator of transcription in photosynthetic genes by interacting with members of both the bZIP and bHLH superfamilies (Waters and Langdale, 2009). The bZIP TF HY5 is a major positive regulator of light signalling and acts through binding to G-box motifs within the promoters of photosynthetic genes. The bHLH PHYTOCHROME INTERACTING FACTOR (PIF) proteins, which also function to regulate gene expression in response to light, repress transcription of photosynthetic genes by binding to the G-box motif (Martínez-García *et al.*, 2000). Thus, it is possible that the G-box may be acting to mediate repression of photosynthetic genes during senescence. Downregulated clusters are also heavily overrepresented for the site-II element, which bind TCP family TFs. TCPs are implicated in regulating cell growth and proliferation (Cubas *et al.*, 1999) and can function as activators or repressors (Kosugi and Ohashi, 2002; Li *et al.*, 2005). This potential for negative regulation may explain why the site-II motif is mostly overrepresented in clusters where genes appear to follow a similar pattern of reduced expression as senescence progresses. Binding sites for E2F TFs, key regulators of cell proliferation (Ramirez-Parra *et al.*, 2003), are enriched in cluster 22, which is consistent with this cluster being enriched with genes annotated with the GO term 'Cell Cycle' as described in Breeze *et al.* (2011). It would be expected that genes in this cluster are downregulated over the time-course, as cell division will have stopped in the mature leaf. E2F TFs can exhibit repressive functionality and therefore may be responsible for repressing the expression of cell cycle genes through the E2F binding motif as the leaf ages.

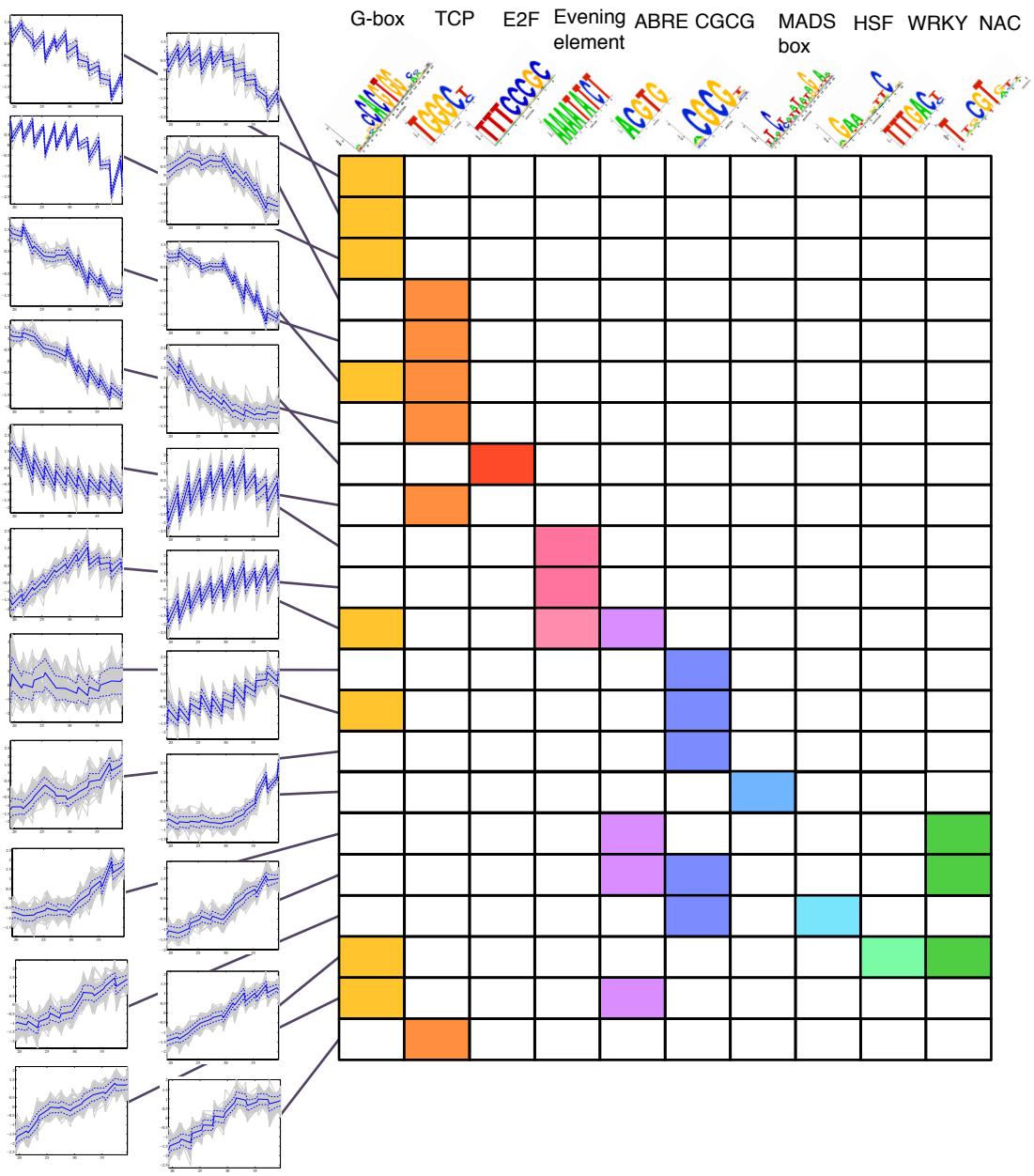


Figure 4.7: **Overrepresentation of known TF binding sites is associated with distinct expression profiles during the senescence time-course.** By focusing on TF binding sites that were enriched within at least one cluster it was possible to identify motifs linked with different TF families. Coloured boxes indicate overrepresentation of a motif (p-val <0.001) in a given cluster (left hand side). Chosen motifs serve as binding sites for TFs that have been previously implicated in senescence/stress response or important signalling pathways. By examining the clusters that show significant enrichment for each motif we observe that particular sequences are associated with particular expression patterns.

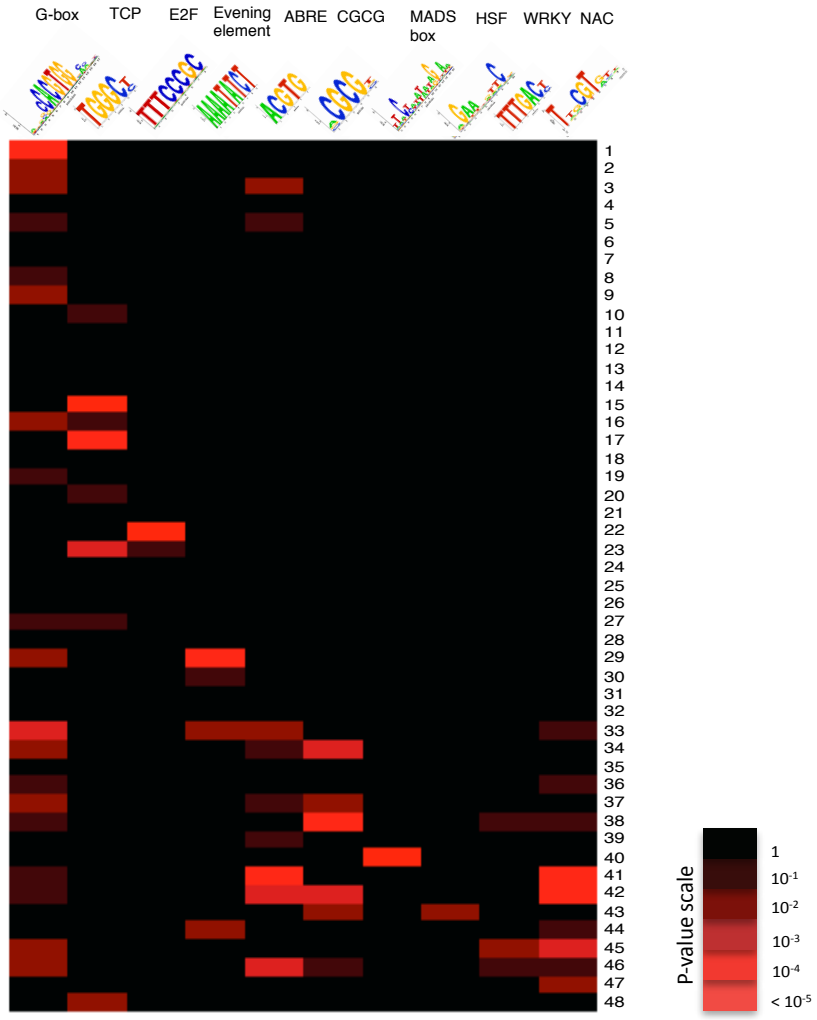


Figure 4.8: **Overrepresentation of known TF binding motifs in promoters of coexpressed genes.** Horizontal axis displays logo representation of known TF binding motifs. Vertical axis indicates senescence cluster numbers. Coloured boxes indicate statistical enrichment (hypergeometric p-value) of each motif in each cluster.

The promoters of coexpressed genes that are upregulated during senescence (clusters 27-48) contain sequence motifs that are known to interact with TFs that are themselves enhanced during senescence. The heat shock element (HSE) is overrepresented in a single cluster, and several HSFs are upregulated during senescence. As described previously, HSFs are induced by a wide range of stresses where they activate the expression of HSPs, which facilitate protein folding, localisation and degradation (Swindell *et al.*, 2007). During senescence large scale destruction of macromolecules will generate reactive oxygen species which damage and denature proteins. HSPs may combat this by

assisting in the re-folding of proteins, which allows senescence to progress at maximum rate. The CGCG motif, which is enriched in several upregulated clusters has been shown to bind CAMTA TFs (Yang and Poovaiah, 2002) and has been implicated in signalling responses to wounding, cold, and other stresses (Walley et al., 2007; Doherty et al., 2009). As senescence progresses the stress imposed at the cellular level increases and may induce stress response pathways that overlap with those that act via the CGCG motif. MADS box TFs which bind to the CArG box are the major regulators of flowering in Arabidopsis but have recently been shown to be capable of delaying senescence (Chen *et al.*, 2011) suggesting that the genes containing these sequence motifs may be enhanced during senescence by de-repression.

The ABA-responsive element (ABRE) is overrepresented in multiple upregulated clusters, correlating with the observed increase in ABA levels as senescence progresses (Breeze *et al.*, 2011). ABRE-BINDING FACTORS (ABFs) are known to activate target genes in an ABA-dependent manner (Nakashima et al., 2006). ABF3 and ABF4 are master regulators of ABA signalling (Yoshida *et al.*, 2010) and are enhanced during senescence and are possible candidates for mediating ABA-dependent transcription during the process. Recently it was shown that motifs that are variations of the core ABRE sequence (ACGTG) are enriched within the promoters of genes responsive to abiotic stress but not biotic stress (Zou *et al.*, 2011). This core motif forms a super-family of regulatory elements with specificity driven by flanking nucleotides. Genes that are enhanced during senescence overlap with those induced by abiotic stress and therefore may be regulated by ABRE-like sequences with TF-binding specificity driven by subtle variation in flanking bases. Future work should focus on analysing these potential source of selectivity.

The pattern of enrichment for the NAC motif is particularly striking, both in terms of statistical significance and the similarity in expression profile (Figure 4.8). The NAC recognition sequence is overrepresented in several upregulated clusters that share similar expression profiles. The NAC domain TFs constitute a large proportion of the senescence-enhanced TFs and are known to play significant roles in regulating leaf senescence in Arabidopsis (Miao et al., 2004; Guo and Gan, 2006; Kim et al., 2009). Even though the enrichment patterns displayed in figures 4.7 and 4.8 are for a sequence motif that was specifically derived from the binding specificities of ANAC092 (Olsen *et al.*, 2005), it is likely that other members of the NAC family can interact with this sequence and subtle variations of it. The NAC motif registers some of the most significant p-values observed during the motif analysis, implying that this sequence pattern plays a fundamental role in contributing towards the generation of the observed expression

profiles (Figure 4.8). In addition to this, the distribution of the enrichment is heavily biased towards the upregulated gene clusters, with virtually no enrichment in the set of downregulated clusters, even when considering weaker p-values. This pattern confirms the role of the NAC motif and NAC TFs in regulating genes important for the progression of senescence.

Recently, Balazadeh *et al.* (2010a) used inducible *ANAC092* overexpressor lines and microarrays to predict downstream target genes. The use of an inducible promoter coupled with expression profiling soon after induction means that we can be more confident that a large proportion of these genes are direct targets of ANAC092. The study identified 170 genes that were upregulated following artificial induction of *ANAC092* expression. Of these genes, 102 are upregulated in the senescence time-course and 75%, including ANAC092 itself, are members of clusters that are significantly enriched for NAC motifs (clusters 41, 42, 44 and 45). This finding is further evidence to suggest that ANAC092, and other NAC TFs, are key regulators of senescence associated genes. The observation that such a significant proportion of ANAC092 target genes are present in clusters that are overrepresented for NAC motifs suggests that the motif analysis is identifying biologically relevant regulatory elements that are likely to function in the GRNs that operate during senescence. This finding provides confidence that the other motifs that are significantly enriched in specific clusters are also likely to represent functional elements that are important for coordinating the senescence process.

The elaborate manipulation of expression patterns during senescence is likely to be the result of not just one motif and corresponding TF family, but through the combinatorial action of many different TFs binding to multiple different sequence motifs. In order to gain a preliminary understanding of any potential combinatorial regulation, a simple analysis was performed in which promoters that are enriched for a given known motif are input to MEME in order to discover any additional motifs present in these sequences. As expected MEME could always re-discover the known motif that was enriched in all of the input sequences, but in general additional novel motifs could not be identified. However, one slight exception was observed using promoters that were enriched for the Barley NAC69 motif (Transfac record, M01055). Using the 15 promoter sequences from cluster 42 that were enriched for the NAC69 motif, MEME identified an additional motif present in 10 of these sequences (Figure 4.9A). The TOMTOM (Gupta *et al.*, 2007) search tool was then used to identify and visualise similar known TF binding sites (Figure 4.9B). The motifs that exhibit the most similarity to the discovered motif were binding sites for Nkx2-5 and TTF1 (Transfac records M00240 and M00432 respectively). Both Nkx2-5 and TTF1 are homeodomain TFs (Chen and Schwartz, 1995; Leon *et al.*, 2009). This

result is interesting because homeodomain TFs have been shown to function together with NAC TFs in Arabidopsis in order to drive expression of a gene involved in abiotic stress (Tran *et al.*, 2007). Multiple members of the NAC and homeodomain TF families display enhanced expression during senescence and so taken together with the observed combinatorial motif patterns, this hints at a possible mechanism of combinatorial control.

4.2.3 Identification of promoter motifs regulating the response to infection by *Botrytis cineria*

Botrytis cineria (Botrytis) is a common necrotrophic fungal pathogen that can infect the tissue of a wide range of plants. The pathogen infects the host producing a myriad of chemicals and proteins in an attempt to induce host cell death, leaving it free to extract nutrients from dead tissue (reviewed in Glazebrook 2005). The plant attempts to counter this infection through the concerted action of many genes that induce the various components of the plants's defence system. Previous studies have identified TFs that are essential for the response to infection indicating that these regulatory proteins play important roles in orchestrating the induction of defence related programs. In order to investigate the genetic mechanisms that control the response to infection by Botrytis, whole transcriptome analysis was performed on Arabidopsis leaves sampled at 2 hour intervals until 48 hours post infection (hpi). Over the 48 hour period, large scale changes in gene expression are observed and differential expression analysis identifies 9838 genes that change significantly compared to mock infection. The extent of the change in the Arabidopsis transcriptome following infection is striking and indicates the complexity of the response at the genetic level. These results are currently being prepared for publication (Windram et al., in preparation).

Clustering analysis performed using the SplineCluster tool generated 44 clusters of genes that are coexpressed over time (Figure 4.10). A major transcriptional transition is observed approximately 24 hpi, with many genes exhibiting sharp up- or down-regulation at this timepoint. Clusters also reveal genes that undergo differential expression earlier and later than this timepoint. The expression of some sets of genes is gradually enhanced or reduced, while others have variable expression patterns. The broad range of expression patterns brought about by the infection suggests that many different processes are highly regulated at the transcriptional level. Genes that are coexpressed as the leaf responds to infection may be coregulated by overlapping sets of TFs. Thus, the promoters of genes that share a similar expression pattern are likely to be overrepresented for sequence motifs that interact with these TFs. In order to investigate this hypothesis, a similar analysis was performed on the sets of coexpressed genes identified

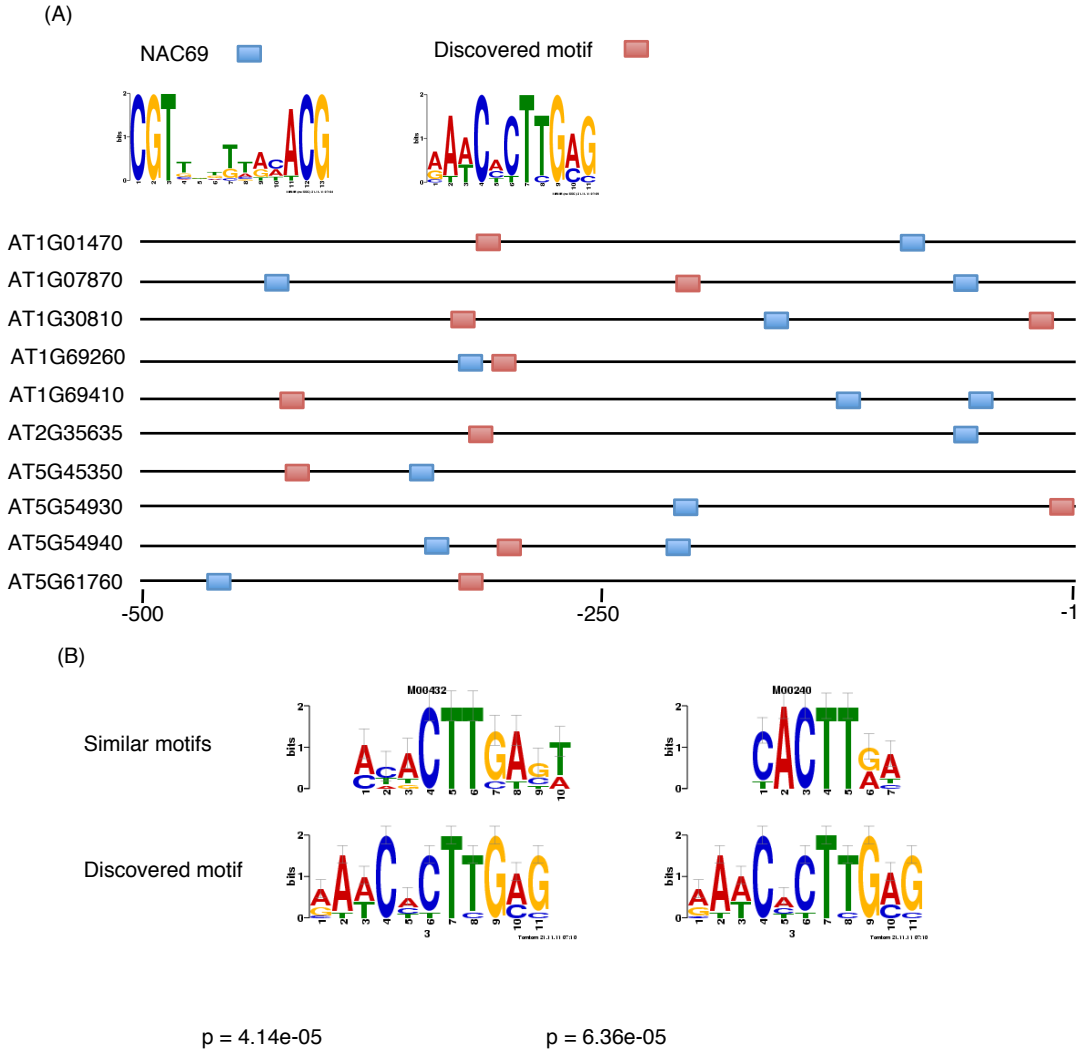


Figure 4.9: **Hints at combinatorial motif patterns.** (A) MEME analysis using promoter sequences from the cluster that was overrepresented for the NAC69 motif. The NAC69 motif (blue boxes) was re-discovered along with another motif (red boxes). Shown are the positions of the co-occurences of the two motifs. (B) Comparison of the discovered motif with known DNA motifs. A comparison of the motif identified by MEME and known motifs from multiple databases was performed using the TOMTOM tool. The most significant similarity is shown in the alignments between the PSSMs for NAC69 and Nkx2-5 (left) and TTF1 (right). Both Nkx2-5 and TTF1 are homeodomain TFs.

in this study as used previously with clusters of genes from the senescence time-course data (Section 4.2.2).

The analysis (Figure 4.11 and 4.12) shows that clusters that have similar expression profiles are specifically overrepresented for similar motifs. Full list of motifs enriched

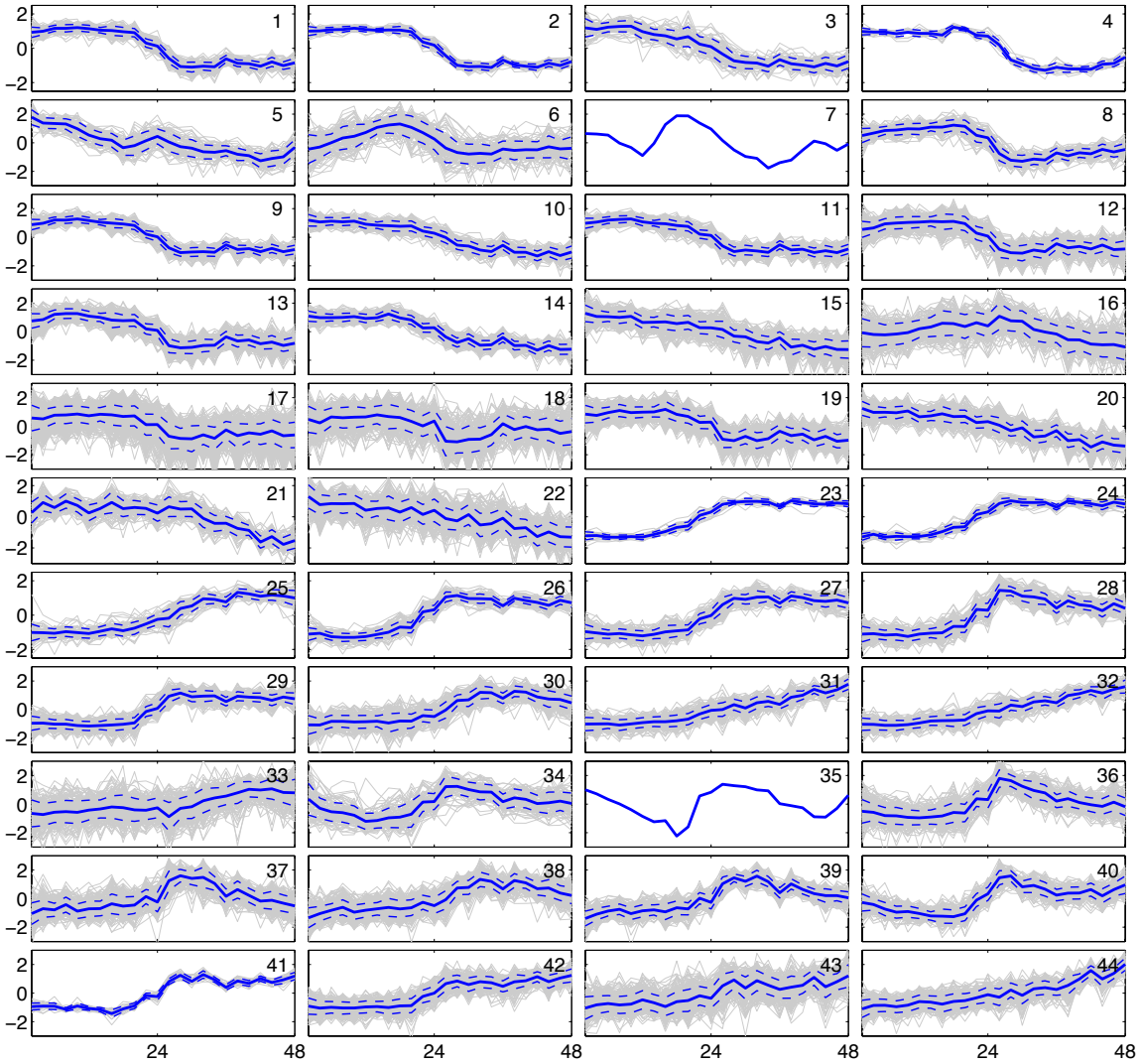


Figure 4.10: **Clustering analysis of genes differentially expressed following infection by Botrytis.** SplineCluster was used to group the 9838 genes differentially expressed following Botrytis infection. Clustering was performed using mean normalised expression profiles of infected samples. The blue line on each plot represents the mean expression values at each time point. Clusters were taken from Windram et al., (in preparation).

in Botrytis clusters shown in Appendix C Table 2. Clusters of genes that are suddenly downregulated approximately 20 hpi are enriched for the variants of the G-box and I-box motifs, which have been shown to function in light-regulated gene expression (Cashmore, 1990; Menkens *et al.*, 1995). These clusters are enriched for GO terms related to photosynthesis and associated processes. The global downregulation of photosynthetic genes during biotic stress has been suggested to be a mechanism to reallocate resources in order to support the defence response (Bilgin *et al.*, 2010). The concerted downregulation of

such a large number of genes suggests that the downregulation of photosynthetic genes is highly coordinated. The G-box motif has been shown to interact with TFs that exert repressive functionality, suggesting that repression of photosynthetic genes can be mediated through these motifs. As previously described PIFs, for example bind to the the G-box motif present in the promoters of some photosynthetic genes (Martínez-García *et al.*, 2000).

Site-II motifs are overrepresented within the promoters of genes that are downregulated post infection. As described previously, TCP proteins have been associated with regulating a wide range of biological processes including development (Kieffer *et al.*, 2011; Li *et al.*, 2005; Hervé *et al.*, 2009), cell proliferation (Aggarwal *et al.*, 2011), the circadian clock (Pruneda-Paz *et al.*, 2009; Giraud *et al.*, 2010) and JA biosynthesis (Schommer *et al.*, 2008). Consistent with this, some of the clusters that are enriched for the site-II motif are enriched for GO terms related to developmental processes. Developmental processes may be downregulated as part of the defence response to reallocate resources to fight infection.

Binding sites for TFs that have been implicated in pathogen related defence responses are also overrepresented in clusters of differentially expressed genes. Motifs that interact with Dof family TFs are enriched within the promoters of genes that fall into clusters that have different trends in expression profile. Dof TFs, of which several are differentially expressed following infection, have been implicated in regulating a wide number of processes including glucosinolate biosynthesis (Skirycz *et al.*, 2006). It has been suggested that glucosinolates play key roles in pathogen infection (Bednarek *et al.*, 2009). The GCC-box, which interacts with AP2/ERF TFs is overrepresented in one cluster. Several members of this family known to bind to this element are differentially expressed in this time-course such as ORA59 and ERF1 (Zarei *et al.*, 2011). A MYB TF binding motif is enriched within the promoters of genes that are upregulated in response to infection. Several members of the MYB family have regulatory roles in response to biotic stress (Mengiste *et al.*, 2003; Clay *et al.*, 2009; Ramírez *et al.*, 2011).

A dramatic shift in gene expression is observed approximately 24 hpi with clusters enriched for GO categories related to defence and stress. Examining the motifs enriched in these clusters identifies motifs associated with TF families that have key roles in regulating the responses to pathogen infection. Figure 4.12 depicts the same data as shown in Figure 4.11, except that it shows the enrichment score for each motif in all clusters. This representation better demonstrates the distribution of motif enrichment over clusters that have similar expression patterns. This is particularly evident in the

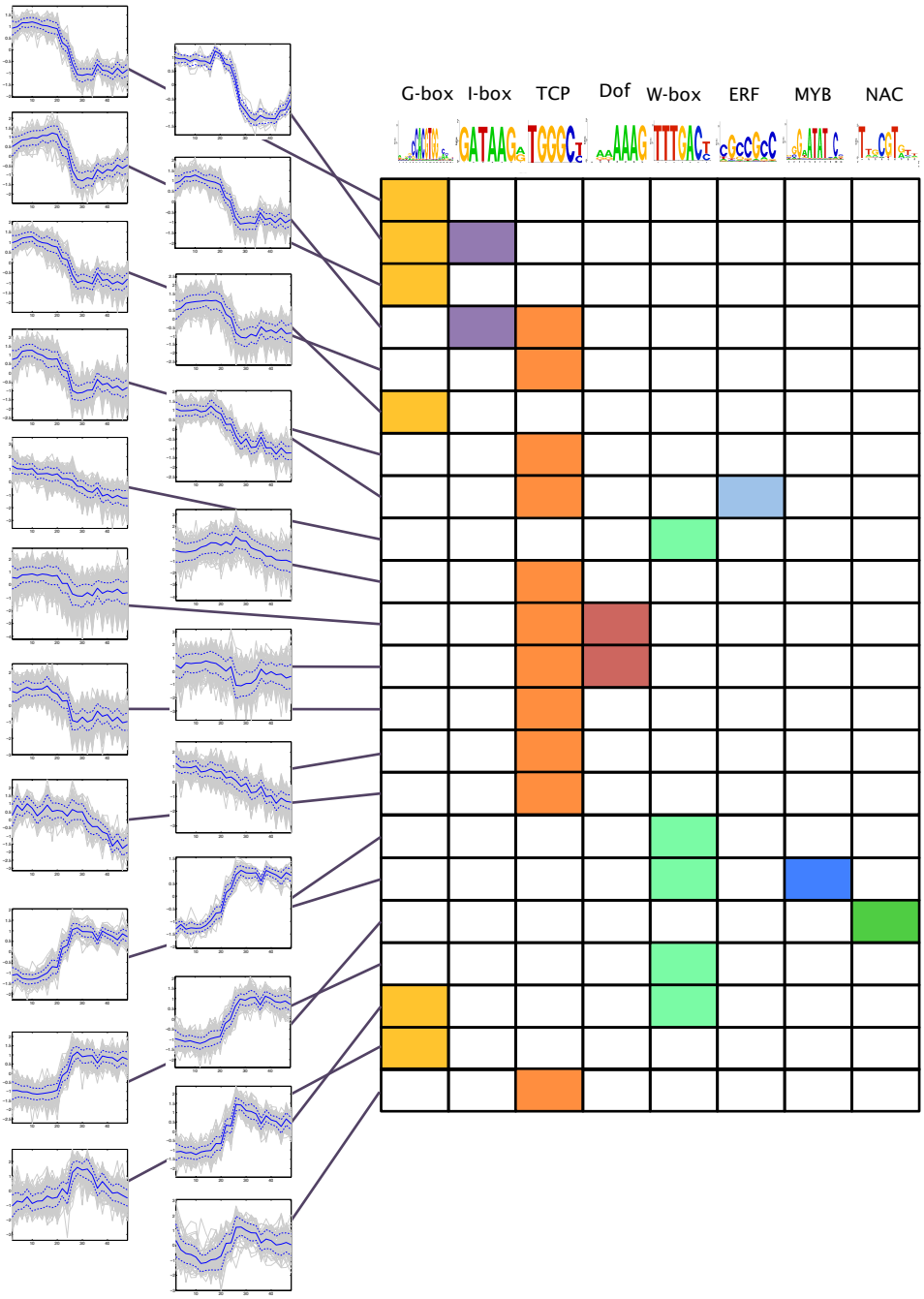


Figure 4.11: **Overrepresentation of known TF binding sites is associated with distinct expression profiles during the Botrytis time-course.** Figure displays selected TF binding sites that are enriched within at least one cluster. Binding motifs (horizontal axis) represent several different TF families that are associated with response to pathogens and/or stress related pathways. Coloured boxes indicate overrepresentation of motif (p-val <0.001) in a given cluster (left hand side). TF binding sites are associated with particular expression profiles.

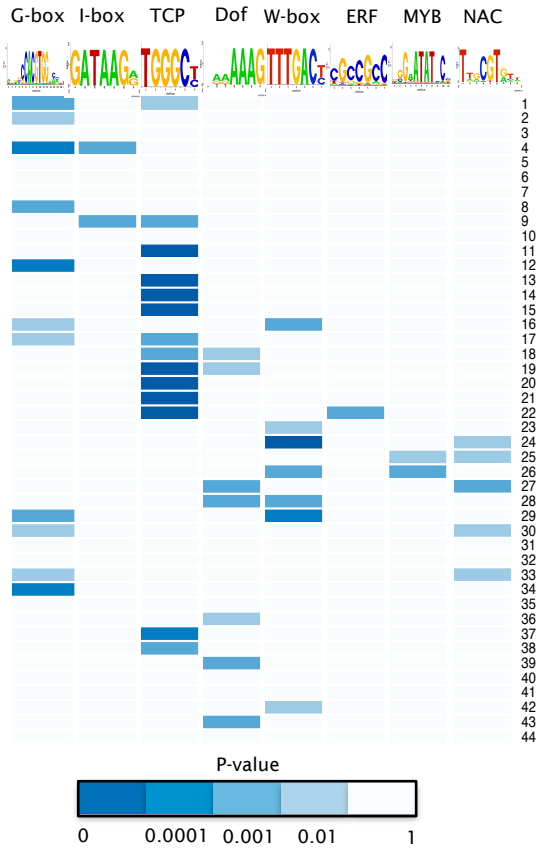


Figure 4.12: **Overrepresentation of known TF binding motifs in promoters of co-expressed genes during response to Botrytis infection.** Horizontal axis displays logo representation of known TF binding motifs. Vertical axis indicates Botrytis cluster numbers. Coloured boxes indicate statistical enrichment (hypergeometric p-value) of each motif in each cluster.

clusters that are sharply induced approximately 20-24 hpi.

The w-box motif, which selectively binds members of the WRKY TF family, is over-represented in clusters where genes are rapidly induced around 24 hpi. W-box motifs are known to be enriched within the promoters of genes induced by biotic stress and many WRKY TFs have demonstrated that they are functionally important for the response to Botrytis and other pathogens (Zheng *et al.*, 2006; Rushton *et al.*, 2010). W-boxes are enriched in clusters that are overrepresented for GO terms such as 'Biotic Stimulus' and stress related processes. The pattern of the w-box enrichment is also consistent with the overrepresentation of WRKY TFs upregulated at specific times during the time-course. The pattern of enrichment over all clusters is striking and specifically enriched in clusters that are rapidly activated in response to Botrytis infection. Such discriminative enrich-

ment in this set of highly induced genes suggests that interactions between motif and TF are important for generating the expression patterns observed and play an important role in plant defence.

Binding sites for NAC TFs are overrepresented in a cluster that is sharply induced 24 hpi with additional enrichment in clusters that have similar expression profiles. The pattern of enrichment is consistent with the upregulation of many members of the NAC family around this time. Although NAC TFs are heavily linked with regulating the responses to abiotic stimuli, there is mounting evidence that they function in the response to biotic stress and Botrytis in particular (Nakashima *et al.*, 2011). Indeed, recent studies have implicated the NAC family members, ANAC002, ANAC019, ANAC055, and ANAC092, as having important roles in regulating the response to biotic stress (Carviel *et al.*, 2009; Bu *et al.*, 2008; Wang *et al.*, 2009; Wu *et al.*, 2009). All of these aforementioned NAC TFs are differentially expressed over the time-course suggesting that they, probably together with other NAC TFs may be interacting with the NAC binding site motifs overrepresented in promoters of differentially expressed genes.

The patterns of enrichment for both the WRKY and NAC TFs within the promoters of genes that are rapidly induced in response to infection, highlights the importance of these TF families in regulating the response to infection.

4.3 Discussion

The pattern of *cis*-regulatory elements within the Arabidopsis genome describes a regulatory code that is interpreted by TFs that bind to this DNA in a sequence-specific manner to activate or repress transcription. In this chapter, a combination of both *de novo* motif finding and analysis using known TF binding sites was used to investigate the transcriptional mechanisms that regulate stress responses in Arabidopsis. Using these techniques, direct targets of the key stress induced TF, HSF3, were predicted and validated. This success was followed by an initial understanding of the sequence motifs that contribute towards the GRN programs that underpin leaf senescence and the response to Botrytis infection.

The analysis of the HSF3 regulon demonstrates the power of *in silico* analysis by identifying a regulatory motif that has genuine predictive potential. A motif that may serve as a binding site for HSF3 was discovered in the promoters of genes that were upregulated in Arabidopsis plants that overexpress HSF3. It was then possible to predict direct targets based on enrichment for this motif in the promoters of all genes downstream of HSF3.

Experimental analysis confirmed this hypothesis for a subset of genes allowing for the elucidation of the parts of the HSF3 network structure. Future work with ChIP-based methods such as ChIP-SEQ would allow for additional confirmation of HSF3 binding and could be used to reveal direct targets of HSF3 under a range of environmental conditions.

In order to understand HSF3 behaviour and function further it would be interesting to assess the effect the pHSF3E has on transcription through the use of synthetic promoters; the pHSF3E could be inserted into randomly chosen promoters and the activity of adjacent genes compared to wild-type promoters during stress treatments or HSF3 overexpression. This analysis could be extended to examine the effect of HSF3 binding site location with respect to the TSS. From the work performed in this chapter, it appears that the binding of HSF3 to promoters is location-sensitive. Synthetic promoter-reporter constructs with the pHSF3E positioned at differing distances upstream of the TSS would be a simple way to probe the effect of position on activation of transcription.

A limitation of the analysis was the definition of promoters as 500 bp upstream of the predicted TSS, which resulted in a constrained search space. The decision to use this length of sequence was made in order to reduce the signal to noise ratio in the input sequences, improving the chances of the MEME algorithm discovering a relevant motif. Intergenic regions are often far greater than 500 bp, and TFs can bind at extended distances, upstream in order to manipulate transcription.

The ability to predict direct targets of HSF3 based on motif enrichment means that it could potentially be used to predict HSF3 targets in other plant species where an *HSF3* ortholog is known. This would be particularly useful for crop species where manipulation of *HSF3* ortholog expression results in the beneficial phenotypes observed in Arabidopsis. Elucidation of the direct targets of any HSF3 ortholog would contribute to a clearly defined understanding of the regulatory network that confers stress tolerance and would allow the fine tuning of gene behaviour to enhance the efficiency of the stress response and maximise yield.

Developmental processes and responses to external stimuli require extensive changes in gene expression. Microarray time-course experiments allow changes in the transcriptome to be tracked over time and for the identification of genes that are coexpressed and may therefore be coregulated. In this chapter, promoter analysis was performed to identify TF binding site motifs that may contribute to the regulation of sets of genes that have similar expression profiles. Using known TF binding motifs, it was possible to identify clusters of genes that may be regulated by a particular TF family with

enrichment often correlating with the biological processes associated with those genes. The coherence between expression patterns, motif enrichment and biological processes provides confidence that these microarray time-course experiments accurately capture the gene expression dynamics that underly both senecence and the response to Botrytis infection.

The importance of both the NAC and WRKY TF families in regulating the biological processes that promote senescence and pathogen defence was highlighted through the significant enrichment of potential binding sites for these regulators in the promoters of genes upregulated in both conditions. This enrichment suggests that both NAC and WRKY TF families function in regulatory programs operating in different contexts and adds to existing studies that implicate these TFs as having important roles in regulating abiotic and biotic stresses (Nakashima *et al.*, 2011; Rushton *et al.*, 2010). As with all of the known motifs used in this study, they only relate to specific TFs and it is likely that different members of the same family can bind to these motifs. Full characterisation of each motif instance will require the use of experimental methods such as Y1H that can identify which individual TFs interact with each motif within each promoter. Once candidate TFs have been identified their effect on gene expression can be assessed using mutant Arabidopsis lines in which the candidate TF expression is perturbed. Assaying the effect of this TF using microarrays or qRT-PCR will test predictions from the motif analysis and reveal the genes that lie downstream of this factor.

The advantage of using known TF binding sites is that they have already been shown as functional elements and this provides confidence that any observed enrichment may be relevant *in vivo*. The disadvantage of this approach is that the pool of TFs that have experimentally determined binding sites is small, and motifs are often derived in species other than Arabidopsis. An additional problem is that many of the PSSMs that model these known binding sites are limited in how accurately they describe the binding specificity of a TF. This usually occurs when the PSSMs are derived from a limited number of known binding sites. Many motifs in the PLACE database for example, are derived from a single binding event between a TF and one genomic locus such as a promoter deletion experiment. PSSMs that do not accurately describe the binding specificity of a TF will have poor predictive power. The use of *in vitro* methods for determination of TF-binding specificities such as protein-binding microarrays (Godoy *et al.*, 2011), bacterial-one-hybrid (Noyes *et al.*, 2008), Y1H (Grove *et al.*, 2009), SELEX and *in vivo* methods such as ChIP-chip and ChIP-SEQ (Kaufmann *et al.*, 2010) should begin to alleviate this problem in the future.

As with the HSF3 network analysis the study of known motifs within the promoters of clusters of coexpressed time-course genes was constrained by the use of sequences only 500 bp upstream of the TSS. Again, this reduces the noise that may interfere with statistical analyses, but may result in a false negative rate as there will be bias against any motifs that preferentially occur >500 bp upstream of the TSS. Future investigations that use alternative definitions of promoter length and position may achieve different results and reveal other motifs as having a likely functional role in regulating either process.

The regulation of transcription in response to exposure to stress and developmental processes is complex and likely to be controlled by multiple different *cis*-regulatory elements and the TFs that interact with them. While there are many TFs in the Arabidopsis genome there are certainly not enough for each TF to regulate a single gene in a binary manner. Studies in Arabidopsis as well as other higher organisms have shown that gene expression is the result of the combinatorial action of many different TFs and the synergistic and antagonistic interactions that occur between them. Most of the analysis presented in this chapter studies motifs in isolation; therefore much information regarding combinations of motif pairs or higher order patterns remains undiscovered. The analysis of promoters that are enriched for the palindromic NAC motif hints at possible coregulation and suggests that there are more complex patterns within these sequences that are responsible for regulating transcription (Figure 4.9). Future work will require a move away from single motif analysis towards the development of methods to identify sets of motifs. Experimental elucidation of multiple functional *cis*-regulatory elements is a difficult and costly task, making computational prediction of regulatory modules an important task in bioinformatics. Experimenting with these tools may lead to identification of novel combinations of TF binding sites that function to regulate stress responses in Arabidopsis. Following identification of statistically significant combinatorial patterns it would be interesting to test whether these sites are required for activation of transcription, and if the presence of combinations of motifs had synergistic or antagonistic effects on the levels of transcription. Y1H analysis could be used to identify TFs that bind to each element, while Y2H could be used to reveal protein-protein interactions that can occur between these TFs, providing insights into the regulatory complexes that could potentially occur on a promoter.

The current approach to TF binding site detection explicitly de-couples sequence and expression analysis; differentially expressed genes are clustered in order to identify potentially coregulated genes, and then the promoters of these genes are analysed for overrepresented sequence motifs. The assumption is that coregulated genes will have similar expression profiles. This is not always the case as complex responses are likely

to involve different sets of TFs acting independently to generate similar expression patterns. Experimental noise arising from microarray technology itself contributes to this problem further still and leads to inaccuracies in expression measurements. This leads to clustering results where coregulated genes may not fall in the same cluster but are instead distributed over multiple clusters. This situation appears to be evident in the clustering of both PRESTA time series datasets, as the pattern of enrichment for particular motifs is distributed across multiple clusters that subtly differ in expression profile (see figures 4.8 and 4.12). Since the clustering and sequence analysis is uncoupled, there is no option to re-evaluate the cluster assignments to better reflect coregulation. Future work should focus on developing methods that optimise both similarity in expression and sequence patterns. In an attempt to solve this problem, methods have been developed that combine both regulatory sequence and expression data for motif discovery. (Segal *et al.*, 2003; Kundaje *et al.*, 2005). Both of these methods perform clustering and motif discovery simultaneously, iteratively re-allocating genes to different clusters in an attempt to optimise similarity in expression profile and motif content. Method development along these lines will help refine predictions of key elements of the regulatory code that are responsible for generating complex expression patterns observed during stress responses.

Chapter 5

High-throughput Yeast-1-Hybrid

5.1 Introduction

Physical interactions between TFs and *cis*-regulatory DNA are essential for the proper control of target gene expression. The interactions between these *trans*- and *cis*-acting factors form the basis of the transcriptional networks that tightly coordinate proper spatial and temporal gene expression. A key challenge in order to elucidate regulatory networks is to identify the protein-DNA interactions that can occur within upstream non-coding DNA. One of the most popular methods used to identify proteins that can interact with a given DNA sequence is the Yeast-1-Hybrid (Y1H) system (Li and Herskowitz, 1993). The Y1H system has been used to isolate many plant TFs that physically interact with regulatory DNA sequences (Tran *et al.*, 2007; Chen *et al.*, 2010; Zhu *et al.*, 2010). Briefly, the Y1H consists of three key components: 1. The 'Prey' vector library, which consists of TFs expressed as fusions to the yeast GAL4 activation domain (AD). 2. The 'Bait' DNA sequence cloned in front of a histidine biosynthesis reporter gene. 3. The host yeast strain, which cannot synthesise histidine. Interaction between a TF and the target sequence induces expression of the histidine biosynthesis reporter gene, which allows the host yeast strain to grow on minimal media lacking histidine (Figure 5.1). An attractive feature of this method is the ability to use it in a high-throughput setting via the use of extensive AD-TF libraries. Using this version of the Y1H system it is possible to identify interactions between genes and multiple TFs in a relatively short amount of time.

The PRESTA project has assembled a normalised Y1H screening library of ORFs consisting of 1369 TFs, individually cloned into prey vector pDEST22. This provides excellent, although not full coverage of the predicted Arabidopsis TF complement. In general, different lab groups have implemented subtly different versions of the Y1H

system using slightly differing components and operating distinct screening protocols. Differences include the design and combination of the bait and prey vectors, mating vs co-transformation based screening methods and the length of the DNA fragments to be used as bait. The approach taken in the PRESTA project operates a distinct mating-based high-throughput screening protocol with unique combinations of vectors. Therefore, this new approach needed to be tested with varying parameters and quality checks in order to identify an appropriate screening strategy to be used in future with more DNA bait sequences.

The creation of the AD-TF library, pooled AD-TF library and development of the Y1H mating based screening system was performed by Dr Claire Hill (PRESTA) at The University of Warwick.

In this chapter the suitability of the PRESTA Y1H system is assessed by examining the effectiveness of the high-throughput mating system, how well the results from this kind of screen can be corroborated via co-transformation experiments and also the most appropriate length of bait fragments is explored. In order to do this the promoter of the stress induced *ANAC092* was used as a test case.

5.1.1 *ANAC092* as a candidate for testing the Y1H procedure

The NAC family of TFs has been implicated in regulating the responses to a number of stress stimuli in Arabidopsis. As described in the introduction, ANAC092 is one of the best studied family members and has been shown to play important regulatory roles in response to multiple stresses including salt stress, senescence and pathogens (He *et al.*, 2005; Kim *et al.*, 2009; Carviel *et al.*, 2009). Studying the behaviour of the *ANAC092* transcript during the senescence and Botrytis time-course experiments shows expression is clearly enhanced during these treatments, further supporting the importance of this TF in stress responses (Figure. 5.2). Taken together, all of the experimental evidence points towards ANAC092 as a central regulator of stress responses allowing us to speculate that the gene itself is likely to be highly regulated at the transcriptional level. A substantial amount of regulation will arise from interactions between TFs and the promoter region of *ANAC092* making the upstream sequence an interesting candidate for testing the Y1H procedure. Any interactions identified could represent relevant *in vivo* interactions, identifying TFs that may function in GRNs with ANAC092.

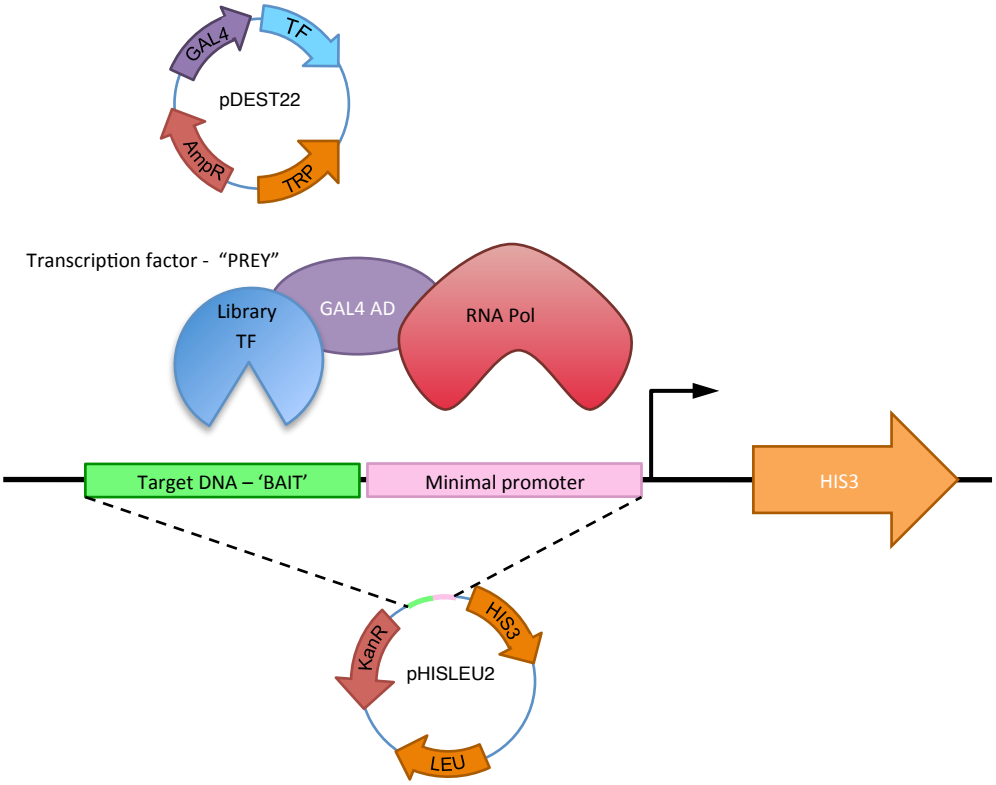


Figure 5.1: **The Yeast-1-Hybrid (Y1H) system.** The Y1H system is a method for characterising interactions between a DNA 'bait' sequence and a TF 'prey'. Bait sequences are cloned upstream of a minimal promoter adjacent to the *HIS3* reporter gene in the pHISLEU2 (or pHISLEU2GW) vector. Prey proteins, usually a TF, are cloned upstream of the GAL4-AD where they are expressed as fusion protein in the vector pDEST22. Both bait and prey vectors are transformed in yeast that lacks the ability to synthesise histidine. Yeast growth occurs if the prey TF can interact with the DNA bait sequence, bringing the GAL4-AD into close proximity of the minimal promoter and recruiting the core transcriptional machinery to transcribe *HIS3*. Expression of *HIS3* allows yeast to grow on media lacking histidine.

5.2 Results

5.2.1 Establishing a strategy for high-throughput Y1H screens using a TF library

Classical Y1H experiments typically involve screening a DNA sequence of interest against a cDNA library generated from a specific transcriptome (Tran *et al.*, 2004, 2007). Such libraries are typically unnormalised, contain a high proportion of non DNA-binding proteins, and often lack cDNAs corresponding to TFs that are expressed for short temporal intervals. Ideally, we want to assess the ability of many individual Arabidopsis TFs to interact with the promoter of *ANAC092* in an unbiased manner. This was achieved by using the TF library generated within the PRESTA project. The prey library consists

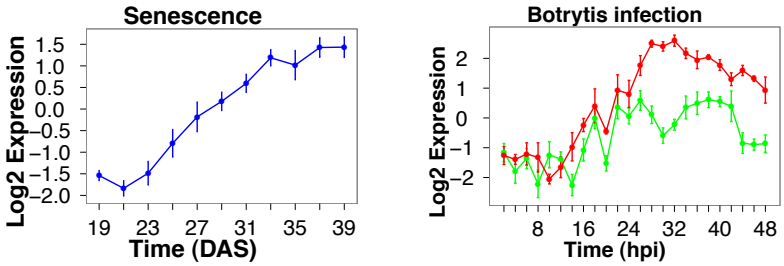


Figure 5.2: **Expression time-course profiles of *ANAC092* during stress treatments.** Lines represent the mean of the biological replicates and error bars indicate standard error. Genes have a mean expression of zero. Left: Expression changes over time of *ANAC092* from 19 days after sowing (DAS) until 39 DAS in a single leaf. Right: Expression changes of *ANAC092* in following infection by Botrytis from 2 hours post infection (hpi) until 48 hpi. Red line: Infection sample, Green line: mock infected sample.

of a collection of Arabidopsis TF cDNAs cloned into a tryptophan selectable vector (pDEST22) as a C-terminal translational fusion to GAL4-activation domain (AD).

The length of the bait sequence used in Y1H assays fluctuates from experiment to experiment and between lab groups with some favouring short promoter fragments and others using significantly longer sequences (Pruneda-Paz *et al.*, 2009; Martinez *et al.*, 2008). As the PRESTA Y1H system had not been used before it was decided to screen multiple bait sequences of varying sizes covering 1000 bp upstream of the *ANAC092* translation start site. Ultimately, the test was to investigate whether the same interactions can be identified with a sequence spanning the complete 1000 bp upstream of the ATG as can be achieved using multiple partially overlapping sequences of smaller size that cover the same region. Obviously, to increase throughput it would be preferable to only screen one bait sequence per gene. Therefore, in addition to a sequence spanning 1000 bp upstream of *ANAC092*, four partially overlapping sequences spanning the same distance were cloned upstream of the *HIS3* reporter gene in the novel leucine selectable vector, pHISLEU2 (created by Dr Claire Hill, The University of Warwick), and transformed into the yeast strain Y187 in order to create Y1H bait strains.

Prey vectors were transformed into AH109 strains in 96-well arrays and pooled so that a workable prey library for subsequent screening could be assembled. Pooling prey strains from each array by row, and then by column, created two alternate versions of the pooled library. This approach allows for the initial validation of interactions through a comparison of the results of both pooled libraries, while also reducing the potential negative effects of competition arising from strains in the same pool having different growth rates.

Following growth in the appropriate selective medium, the prey library and bait strains were mated on solid media. Diploid strains were then replicated onto solid selective media with increasing stringency. Positive growth on media lacking histidine plus 3-amino-triozol (3-AT) indicates interaction between bait and prey. The high-throughput strategy allowed the screening of several promoter fragments per screen. The Y1H workflow is shown in Figure 5.3.

5.2.2 Protein-DNA interactions revealed by Y1H screen

Y1H assays revealed multiple interactions between the promoter fragments of *ANAC092* and members of the AD-TF library (Figure 5.4). One of the key questions was the potential discrepancy in protein-DNA interactions identified using a bait sequence spanning the entire region 1000 bp upstream of the coding sequence (fragment 5) compared to the same coverage with smaller partially overlapping bait fragments (fragments 1-4). The results clearly show that the larger 1000 bp fragment only identifies a subset of the potential protein-DNA interactions that can occur across the upstream sequence. The omission of positive interactions between fragment 5 and TFs isolated with other fragments suggest that there are possibly spatial constraints regarding the location of a TF-binding site with respect to the *HIS3* TSS. This may be a universal feature of yeast transcription initiation, or specific to transcription from the pHISLEU2 plasmid or plasmids in general.

Interactions between members of the AD-TF library and the promoter sequences included members from multiple different protein families such as the TCP, bHLH, bZIP, NAC and homeodomain. This observation suggests that the Y1H system has no particular preference for specific TF families. Looking more closely at the interaction results there are several TFs that can be linked to processes in which ANAC092 is known to function. For example multiple interactions between the *ANAC092* promoter and TFs implicated in the response to salt stress and others were uncovered. For example, *ETHYLENE AND SALT INDUCIBLE 1* (*ESE1*) is highly induced by salt stress. In addition to this, expression analysis demonstrated that *ESE1* expression was suppressed in an *ein2* mutant suggesting that this might be involved in the regulatory link between EIN2 and ANAC092 (Zhang *et al.*, 2011). *ABF3* and *ABF4*, which encode ABA-related TFs, are also induced by high-salinity (youn Kang *et al.*, 2002; Kim *et al.*, 2004). Indeed the link between ABF3 and *ANAC092* promoter may be particularly relevant as *ANAC092* is upregulated in transgenic plants overexpressing *ABF3* under drought conditions suggesting ABF3 is a positive regulator of ANAC092 (Abdeen *et al.*, 2010).

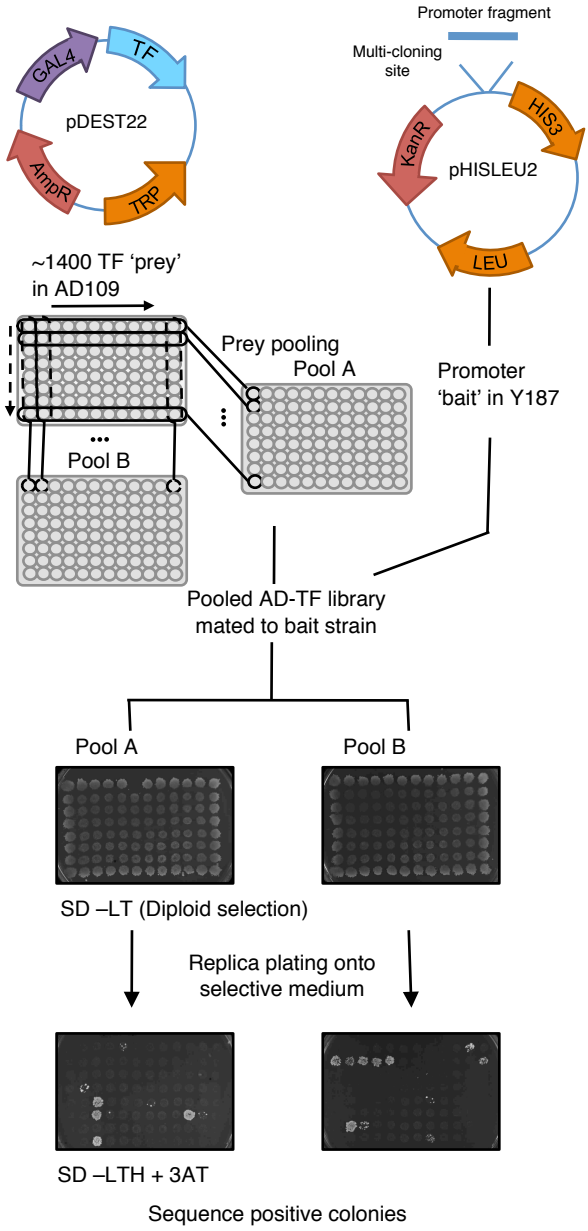


Figure 5.3: **High-throughput Y1H screening assay workflow.** Prey plasmid (pDEST22) consists of an Arabidopsis TF cloned in frame with the GAL4-activation domain (AD) allowing expression of the TF as a C-terminal translational fusion to the AD. Prey library consists of 1400 Arabidopsis AD-TF plasmids individually transformed into AD109 yeast strains. Workable screening libraries in different configurations are generated by pooling prey strains by rows and columns. Bait plasmids (pHISLEU2) are generated by cloning a promoter fragment upstream and proximal to the *HIS3* reporter gene. Bait strains are created by transformation into Y187. Screening of a bait sequence requires mating of the prey and bait strains on solid YPDA medium. Co-transformed strains are selected by replica plating of colonies onto dropout media lacking tryptophan and leucine (-LT) to select for prey and bait plasmids respectively. Protein-DNA interactions are assessed by activation of the *HIS3* gene following replica plating onto media lacking tryptophan, leucine and histidine (-LTH). Autoactivation via the bait sequence is quenched by using -LTH media supplemented with 3-amino-triazol (3-AT), which inhibits histidine biosynthesis. Positive growth indicates interactions between an AD-TF and the bait sequence. Identity of TF is confirmed via sequencing and BLAST.

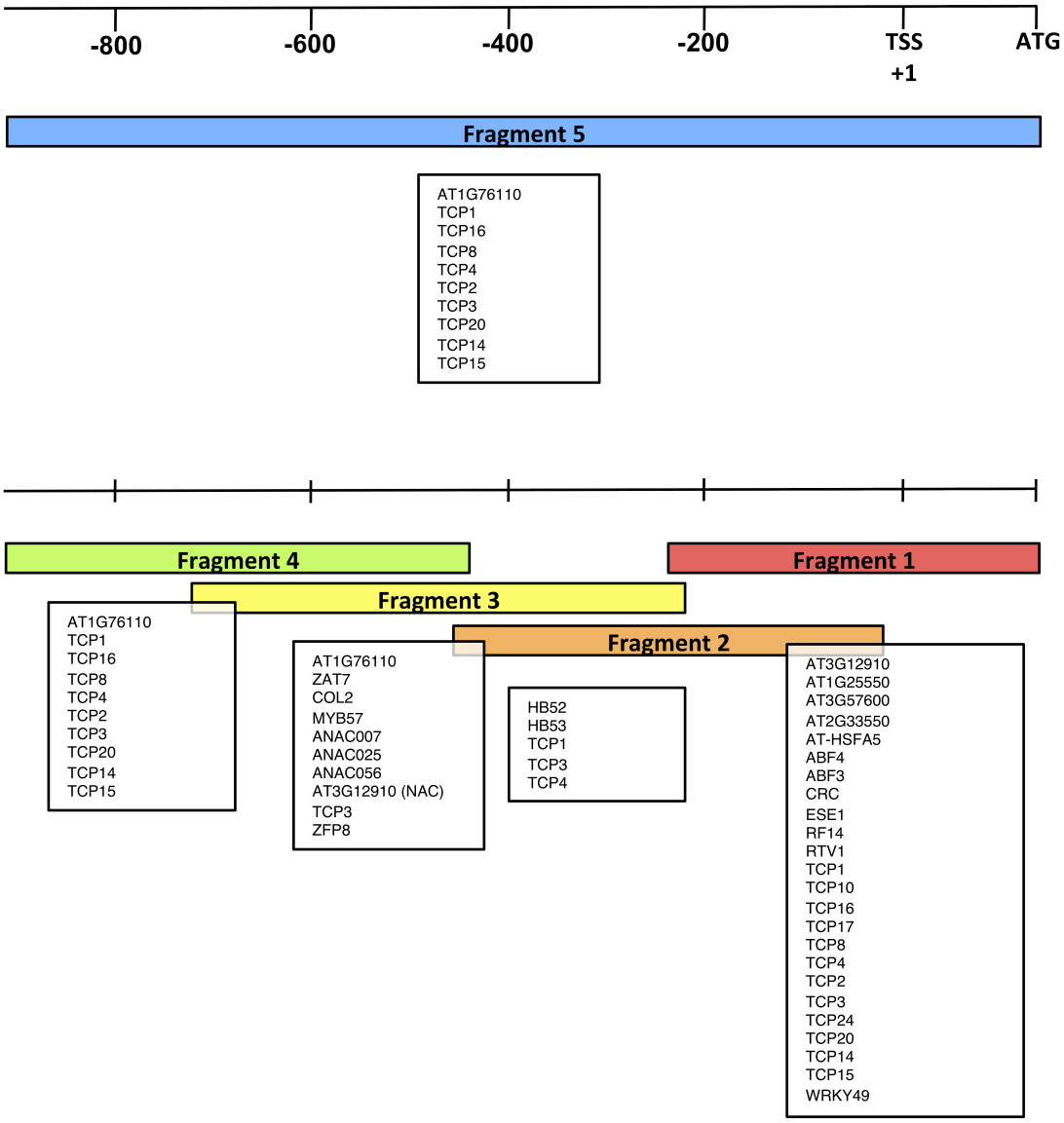


Figure 5.4: **Interactions between TFs and *ANAC092* promoter fragments identified in Y1H screen.** DNA fragments of differing lengths covering approximately 1000 bp upstream of the *ANAC092* coding sequence (ATG) were cloned in front of the *HIS3* reporter gene in the pHISLEU2 bait vector. Promoter fragments tested (coloured boxes) included partially overlapping sequences approximately 500 bp in length (Fragments 1-4) and a fragment spanning the entire 1000 bp upstream region (Fragment 5). Interactions between bait sequences and members of the prey library were revealed by identifying colonies that grow on selective media lacking histidine. Bait strains exhibiting widespread growth across the entire prey library indicated auto-activation via the bait sequence. In these instances, colony growth on selective media supplemented with 3-AT were indicative of protein-DNA interactions. Individual TFs present in positive colonies were identified by PCR of the prey insert followed by sequencing. TFs that interacted with bait sequences are listed below the appropriate fragments.

A particularly interesting set of interactions were identified between fragment 3 and several NAC TFs. Members of this family that bound to the promoter of *ANAC092* were ANAC007, ANAC025, ANAC056 and AT3G12910, an un-annotated TF that contains the conserved NAM domain. None of these NAC TFs have been explicitly linked to function in regulating stress responses although all bar *ANAC007* are differentially expressed during either senescence or following Botrytis infection indicating that they are contributing towards regulation of these processes. A large proportion of the NAC TF family display differential expression during natural senescence (Breeze *et al.*, 2011) and in response to Botrytis infection (Windram *et al.*, in preparation) suggesting that this family plays a key role in orchestrating the transcriptional changes that control these processes.

Members of the NAC TF family share similar types of DNA-binding domain and so it is possible that additional TFs can bind to this promoter fragment, through the same *cis*-regulatory motif. In order to test this hypothesis, all senescence and Botrytis enhanced NAC TFs were directly tested for the ability to interact with fragment 3 of *ANAC092* promoter. Enhanced TFs are defined by being upregulated in the differentially expressed gene lists collated for the publications of the senescence (Breeze *et al.*, 2011) and Botrytis time-course experiments (Windram *et al.*, in preparation). A list of the NAC TFs tested is shown in Appendix D. This test also served as a practical way to assess the reproducibility of the interactions identified in the original mating-based screen. Direct testing was performed by transforming Y187 fragment3-pHISLEU2 bait strains with each of the senescence/Botrytis enhanced AD-TF-pDEST22 prey vectors followed by spotting onto selective media lacking histidine plus 3-AT at varying concentrations. The co-transformation Y1H screen reproduced the protein-DNA interactions with ANAC025, ANAC056 and AT3G12910 and identified an additional member of the NAC family, ANAC018, which can bind to fragment 3 (Figure. 5.5). Also shown is colony growth resulting from yeast bait strains transformed with the NAC TFs ANAC029, ANAC092, and ANAC059, which all have well understood regulatory roles in various stresses-related processes (Guo and Gan, 2006; Kim *et al.*, 2009; Balazadeh *et al.*, 2011). These members of the NAC family did not exhibit strong growth, if any, on SD-LTH + 50 mM 3-AT plates when compared to ANAC025, ANAC056, AT3G12910 and ANAC018 prey vectors. This suggests that the specific interaction between ANAC025, ANAC056, AT3G12910, ANAC018 and the promoter of *ANAC092* is more significant than other stress related NAC TFs.

Y1H identifies protein-DNA interactions that can occur *in vitro*, but does not indicate the regulatory consequences of these interactions *in planta*. In order for a TF

to directly regulate another gene it must be expressed at the same time and place as the target. In the simplest scenario, expression of a TF gene would be induced at approximately the same time as the gene it is directly regulating. We would also assume that the expression of the TF and its target would be either positively or negatively correlated, depending on whether the TF is activating or repressing the target gene. In order to assess which of the NAC TFs that bound to the *ANAC092* promoter are realistic targets under a given condition, genes that are differentially expressed during either senescence or following Botrytis infection were plotted against *ANAC092* transcript levels (Figure 5.6). *ANAC056* and *ANAC018* are differentially expressed during senescence but not Botrytis infection. *ANAC025* is differentially expressed during the response to Botrytis infection but not during senescence. *AT3G12910* is differentially expressed during senescence and the response to Botrytis infection. Interestingly, both *ANAC018*, *ANAC056* and *AT3G12910* expression is enhanced in senescence in concert with that of *ANAC092*. Similarly, *ANAC025* and *AT3G12910* expression appears to be activated at the same time post Botrytis infection. This observation suggests that ANAC018 and ANAC056 may regulate *ANAC092* during senescence, while ANAC025 may regulate *ANAC092* during the response to Botrytis infection and AT3G12910 may have the capacity to regulate *ANAC092* in both contexts.

ANAC092 has been linked to function under stresses other than Botrytis and senescence (Balazadeh *et al.*, 2010b), meaning that interactions identified in the Y1H screen may be relevant during different conditions. The AtGenExpress project profiled Arabidopsis under a comprehensive set of stresses. Figure 5.7 shows the expression patterns of *ANAC092* and the putative regulators identified from Y1H during multiple abiotic, hormone and biotic treatments. Clearly, *ANAC092* is induced by multiple stress such as osmotic, salt and *P. syringae*. Interestingly, the TF that was shown to bind to the promoter of *ANAC092* that displays the clearest positive correlation in terms of stress induced expression is *ABF3*. This TFs expression is enhanced under osmotic, salt and biotic stress. In addition to this, *ABF3* along with *ANAC092* is induced by treatment with ABA alone, suggesting that *ANAC092* may be regulated in an ABA dependent manner. One of the interacting NAC TFs, AT3G12910 also displays similar expression pattern when compared to *ANAC092* in response to osmotic stress. Taken together with coexpression in response to Botrytis and senescence, it is further evidence that there is a regulatory association between the two TFs.

Studying the expression patterns of the interacting TFs, many are downregulated following exposure to stress. For example, many TCPs are downregulated during abiotic and biotic stress (Figure 5.7) and also during the senescence and Botrytis time-course

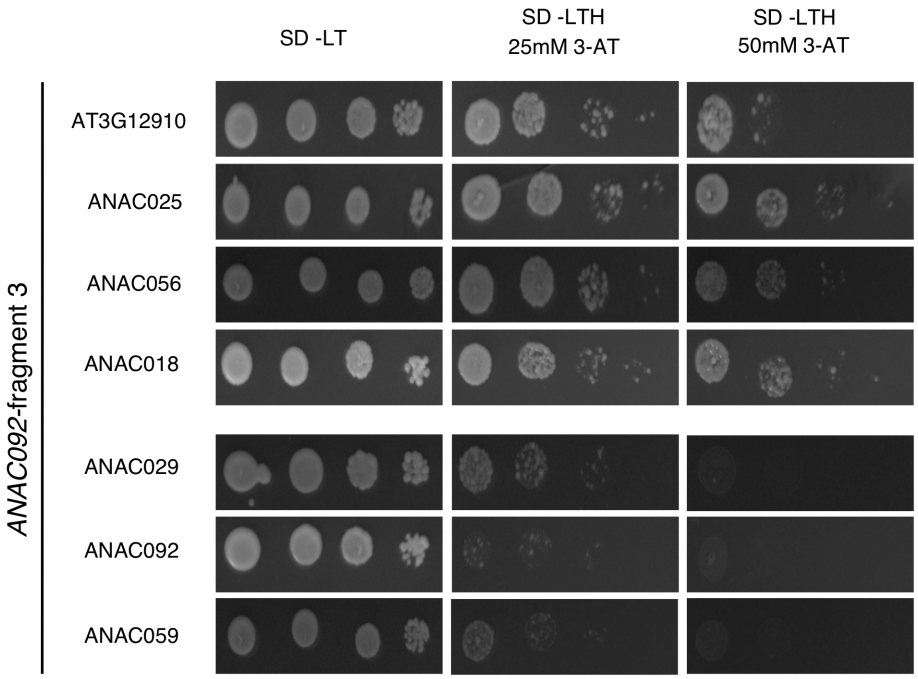


Figure 5.5: **Co-transformation of bait and prey plasmids to re-test interactions and identify novel ones.** Yeast cells were transformed with fragment3-pHISLEU2 and a collection of AD-TF-pDEST22 vectors consisting of senescence and Botrytis infection enhanced NAC TFs. Transformants were spotted onto SD-LTH+25mM 3-AT and SD-LTH+50mM 3-AT in serial dilutions of 10^8, 10^7, 10^6 and 10^5 cells. Positive growth on selective media containing 50mM 3-AT confirmed interaction of fragment 3 with AT3G12910, ANAC025 and ANAC056 and identified an additional interaction with ANAC018. Colony growth of co-transformants with other key stress related NAC TFs are shown for comparison. Co-transformants containing ANAC029, ANAC092, and ANAC059 have growth levels much lower than the other TFs.

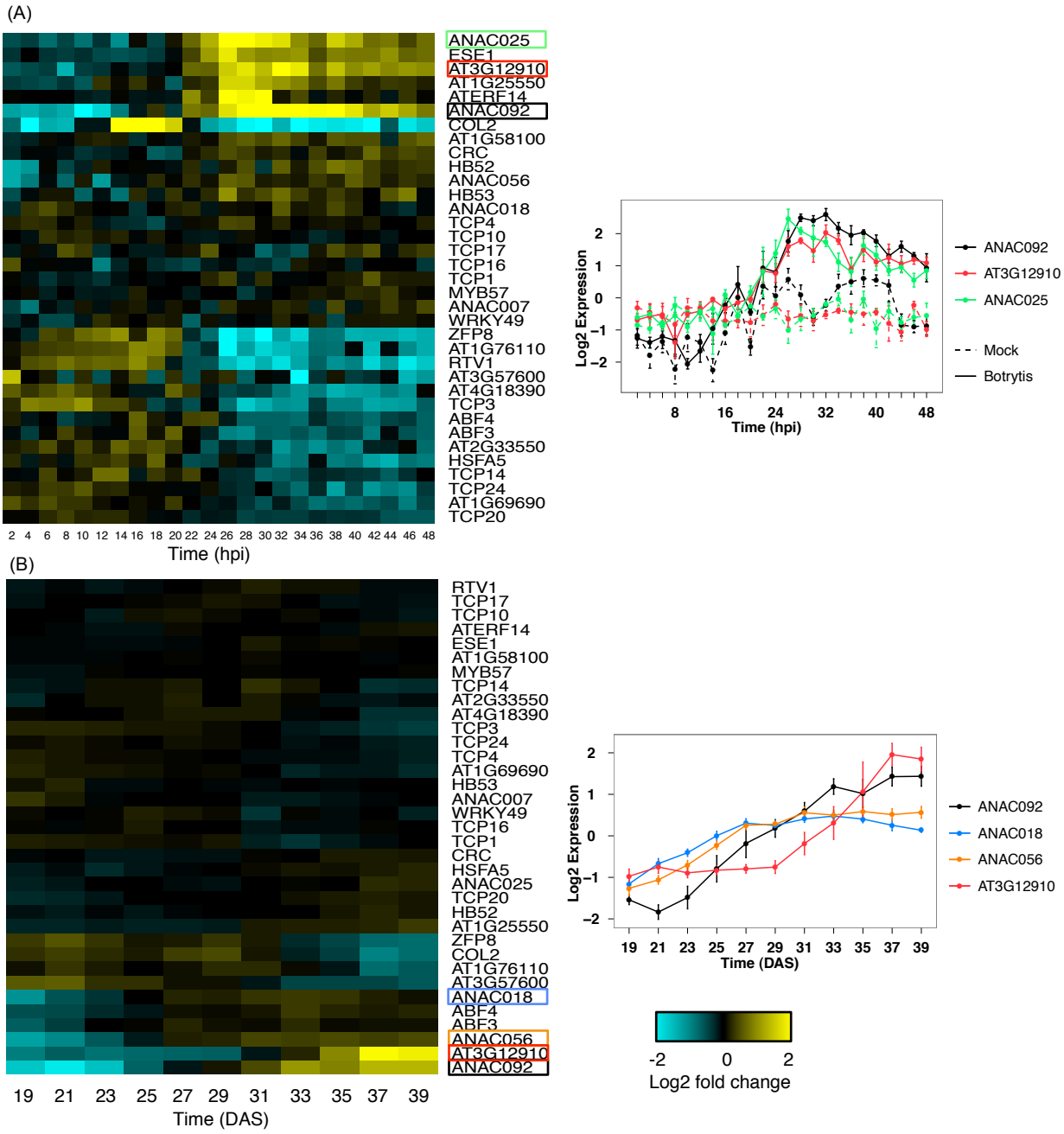


Figure 5.6: **Expression profiles of TFs that bind upstream of *ANAC092* during senescence and Botrytis time-courses.** Shown is a heatmap representation of the interacting TFs in the senescence or Botrytis datasets. Genes have a mean expression of zero. Lines represent the mean of the biological replicates and error bars indicate standard error. *ANAC018* and *ANAC056* are differentially expressed during senescence, while *AT3G12910* and *ANAC025* are differentially expressed during Botrytis infection. (A) Left: Expression patterns of interacting TFs in response to Botrytis infection. Right: Expression pattern of *ANAC092*, *AT3G12910* and *ANAC025* following Botrytis infection. Gene expression profiles from infected leaves and mock infected leaves are shown in solid and dashed lines respectively. Expression for both treatments measured every 2 hours from 2 hours until 48 hpi (B) Left: Expression patterns of interacting TFs during senescence. Changes in expression tracked over time from 19 DAS to 39 DAS. Right: Expression profiles of *ANAC092*, *ANAC018*, *ANAC056* and *AT3G12910* during senescence.

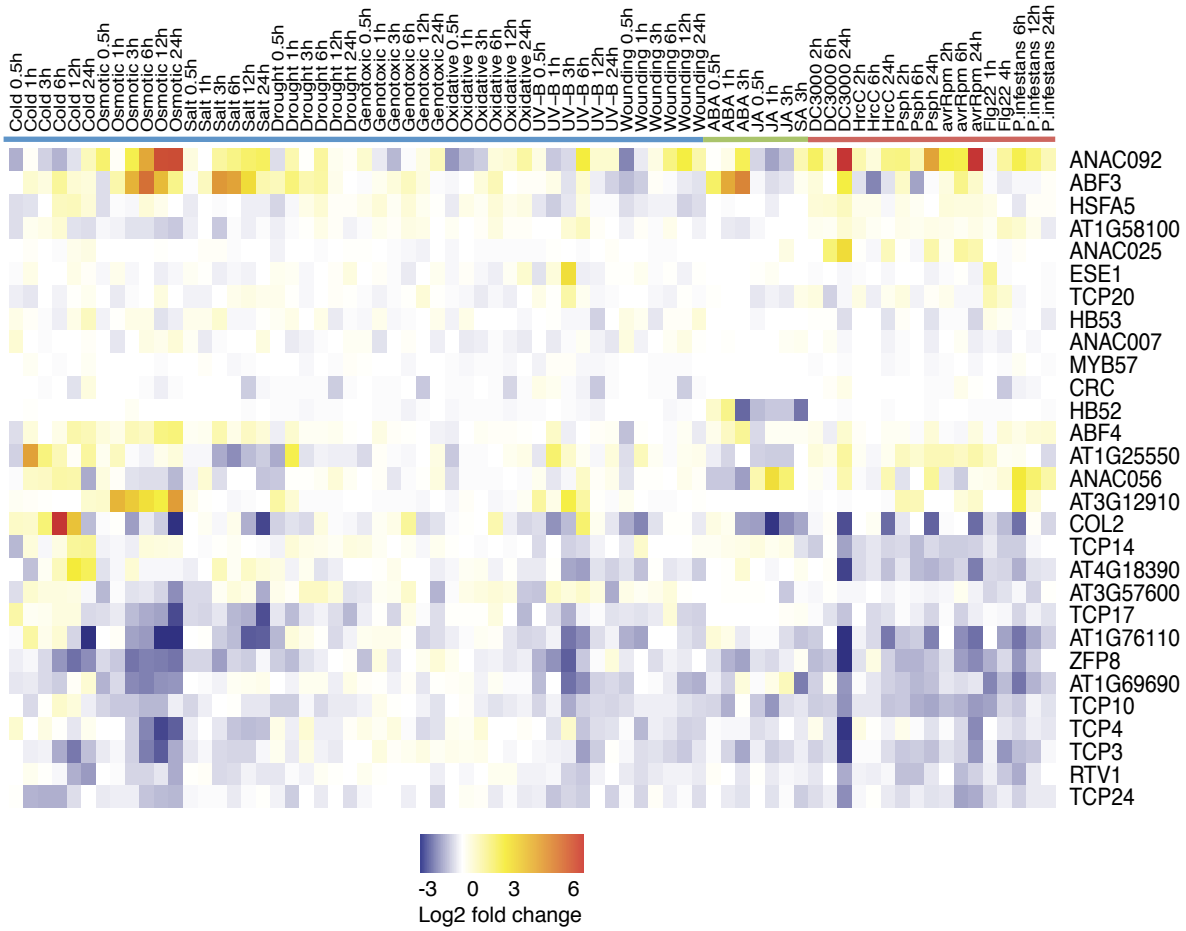


Figure 5.7: **Expression profiles of TFs that bind upstream of *ANAC092* during abiotic, hormone and biotic stress treatments.** Shown is heatmap representation of expression patterns based on normalised AtGenExpress Affymetrix microarray data. Abiotic stress, hormone and biotic stress treatments are indicated by the blue, green and red bars respectively. Expression ratios (treatment/control) indicated by the colour scale bar. Blue indicates downregulation and yellow-red upregulation.

(Figure 5.6). It is possible that such TFs may repress *ANAC092* expression during non-stressed conditions, inhibiting the stress response and allowing the plant to grow and develop with maximum efficiency. Stress responses impose demands on the plant's energy resources and so it is critical to strictly control the activity of these genes. Comparative analysis of TF and target gene expression under multiple contexts is a simple way of identifying the potentially relevant interactions operating during different processes.

5.3 Discussion

High-throughput Y1H assays are becoming an increasingly popular technique for the identification of protein-DNA interactions. This gene-centred approach has been exploited to map interactions between TFs and regulatory DNA, helping to characterise GRNs operating in *C. elegans* (Deplancke *et al.*, 2006; Martinez *et al.*, 2008; Arda *et al.*, 2010) and Arabidopsis (Pruneda-Paz *et al.*, 2009; Brady *et al.*, 2011). Using the promoter of the highly stress-inducible TF, ANAC092, we assessed the ability of the Y1H procedure generated within the PRESTA project to be used in a high-throughput setting. Using promoter fragments of different lengths as bait sequences, it was possible to observe distinct colony growth on selective media compared to the background. Protein-DNA interactions could be confirmed by direct co-transformation of yeast cells with bait and prey vectors, indicating that the high-throughput mating based system identify good quality reproducible interactions. The co-transformation based assay did reveal interactions not identified in the previous screen, therefore highlighting a limitation of the mating based approach.

The use of two independently pooled versions of the AD-TF library was vindicated as colony growth corresponding to certain TFs were sometimes only observed in one version of the library. This could be due to growth competition between TFs in a particular pool prior to mating and/or after mating. In either scenario, certain yeast transformants may outgrow others, masking or preventing the growth of other colonies even if an interaction is occurring. Future Y1H studies should maintain the use of two alternate AD-TF libraries.

A greater number of protein-DNA interactions were identified when fragmented promoter sequences were used as bait compared to the longer 1000 bp sequence that covered the same upstream region. This finding suggests interactions between AD-TF fusions and DNA-elements are less able to activate reporter gene expression at certain distances from the TSS. A study in *S. cerevisia* demonstrated transcriptional activation by the GAL4 TF diminished the further the GAL4 binding site is positioned upstream relative

to the TATA-box (Dobi and Winston, 2007). This finding suggests that future attempts to characterise protein-DNA interactions along large promoter regions with Y1H experiments require the use of a collection of small (approximately 500 bp) overlapping promoter fragments. Some Y1H protocols, however, do use promoter sequences up to and over 1000 bp in length and achieve good results. For example the method developed by Deplancke *et al.* (2004) uses such promoter fragments. This may be because this particular approach uses vectors that integrate the DNA bait-reporter construct directly into the yeast genome leaving it fully chromatinised. These conditions presumably better reflect the *in vivo* context of the promoter where the constraints imposed by distal affects are less significant.

The stress inducible TF ANAC092 is known to be extensively regulated at the post-transcriptional level through interaction with a miRNA Kim *et al.* (2009), but there is little knowledge with respect to how *ANAC092* is regulated at the transcriptional level. The interactions between TFs and the promoter region of *ANAC092* identified by the Y1H screens reveal novel insights into possible mechanisms of transcriptional regulation. Many of the TFs that bind upstream of *ANAC092* have been linked with regulating stress responses. The link between *ESE1* and *ANAC092* is particularly interesting as it provides a possible link between *EIN2* and *ANAC092* in a salt induced GRN. Interactions between the *ANAC092* promoter and the key mediators of ABA signalling, ABF3 and ABF4, provide a possible link with ANAC092 and the ABA signalling pathway, while the former TF may also directly regulate ANAC092 under drought conditions (Abdeen *et al.*, 2010). Several NAC TFs bound upstream of *ANAC092* providing a possible mechanism of intra-family regulation. Even though all of the positive interactions presumably occur through the same DNA-element, not all of the potential regulators appear to be able to bind and exert regulatory affect in any condition. The NAC homolog (*AT3G12910*) displays overlapping expression with *ANAC092* under both senescent and Botrytis infected tissue, suggesting that it may function to regulate the target gene under multiple conditions. The observation that *ANAC025* are differentially expressed following Botrytis infection, while *ANAC018* and *ANAC056* are differentially expressed during senescence and not vice-versa suggests possible context specific regulation by these factors.

The context free nature of the Y1H system means that it is possible to identify TFs that have the capacity to bind to regulatory DNA and not just under a specific condition. This provides us with many candidates that may operate only under specific conditions. In addition, the Y1H system will only provide information on physical protein-DNA interactions and provides no insight into what the regulatory consequence of these in-

teractions are *in vivo*. Confirmation of the regulatory interaction and the conditions in which they operate can be probed by perturbing the expression of the predicted regulator and observing the effect this change has on the expression of the target gene. In Arabidopsis this can be achieved using regulator specific T-DNA knock-out or RNA interference lines in conjunction with microarrays or qRT-PCR. Such an approach would be particularly interesting in order to confirm the regulatory effects of the NAC TFs that bound to *ANAC092* promoter during different treatments.

As is the case with any technique, there are limits as to what information the Y1H system can capture. Y1H will not be able to identify interactions with certain TFs present in the library even if these TFs can bind to a specific promoter element *in vivo*. This may be because the TF may not be able to establish appropriate DNA binding specificity due to expression as a fusion protein and not its native form. Interactions will not be picked up if a particular TF needs to be post-translational chemical modification, by phosphorylation or acetylation for example, in order to be functionally active. Y1H is not compatible with TFs that need to interact with other proteins in order to bind to the DNA as heterodimers or higher order protein complexes. In spite of these limitations, the Y1H system is a powerful and invaluable tool for the elucidation of protein-DNA interactions. Used in concert with other molecular biology techniques this system had the potential to greatly develop our systems level understanding of gene regulation during stress responses.

Chapter 6

Local Network Reconstruction

6.1 Introduction

A significant proportion of the Arabidopsis genome encodes TFs, which occur in large families that share similar types of DNA-binding domains. The expansion of TF families is owed primarily to gene duplication events and the subsequent increased potential for complexity is thought to have driven the evolution of increasingly complex GRNs (Teichmann and Babu, 2004). Global expression profiling experiments have identified many individual members of various TF families that are induced in response to different stress conditions. In spite of this, information on the position of these individual TFs in the GRNs that operate in response to different stresses and their relative contribution towards stress tolerance is limited. Many highly-related TFs are induced by multiple stresses and appear to have both overlapping and distinct functions, yet how individual regulatory roles have evolved for each duplicated gene is unclear.

As described in the introduction, genetic studies have highlighted the important role of the NAC family of TFs in regulating stress responses in Arabidopsis. NAC proteins share a conserved N-terminal NAC domain and form one of the largest plant specific TF families with over 100 members encoded in the Arabidopsis genome (Ooka *et al.*, 2003). NAC TFs have been functionally implicated in a variety of stress related programs such as the response to drought, high-salinity, bacterial pathogens, fungal pathogens and senescence (Fujita *et al.*, 2004; Hu *et al.*, 2006; Tran *et al.*, 2004). Many members of NAC family have overlapping expression patterns that are induced by several stresses suggesting potential functional redundancy and crosstalk between members is operating during these stress responses (Nakashima *et al.*, 2011).

Some of the most well studied members of the NAC family are three highly-related TFs,

ANAC019, ANAC055 and ANAC072. These TFs were originally identified as drought inducible proteins that could bind to the *ERD1* promoter (Tran *et al.*, 2004). Further analysis demonstrated that expression was induced by salt stress, and the stress-associated signalling hormones JA and ABA. Overexpression of either *ANAC019*, *ANAC055* and *ANAC072* led to enhanced drought tolerance in transgenic Arabidopsis plants, highlighting their potential importance in response to abiotic stress (Tran *et al.*, 2004). Studies with the necrotrophic pathogen Botrytis indicate that ANAC019 and ANAC055 have an important role in regulating JA-mediated defence responses (Bu *et al.*, 2008).

TFs that function upstream of other TFs are likely to constitute an important part of the stress response regulatory network. Thus, an important step in dissecting and characterising stress GRNs is the identification of upstream regulators of TFs. As discussed previously, high-throughput Y1H assays that screen for interactions between DNA-sequences and members of a TF library are becoming an increasingly popular approach for elucidation of GRNs in Arabidopsis (Mitsuda *et al.*, 2010; Ou *et al.*, 2011; Castrillo *et al.*, 2011; Brady *et al.*, 2011).

In this chapter, an approach to identifying the factors that lie upstream of ANAC019, ANAC055 and ANAC072 in stress-response transcriptional networks is presented. The high-throughput Y1H system described in the previous chapter is used to identify TFs that can interact with the promoters of the three NAC genes, providing excellent candidates for direct regulation. This analysis is complemented with microarray analysis of mutants of the potential regulators, in an attempt to reveal the potential regulatory consequence of the interactions *in planta*.

Nearly, all of the experimental and theoretical work performed in this chapter was done so by the author. An exception to this is the Botrytis infected Arabidopsis leaf time series with the *myb108* mutant and *MYB108* overexpressor, which was performed by Emma Cooke at The University of Warwick. All analysis and interpretations using these data were performed by the author.

6.2 Results

6.2.1 *ANAC019*, *ANAC055* and *ANAC072* are differentially expressed in response to many stresses.

In order to acquire insight into transcriptional networks that operate during plant stress responses, three genes that are have been implicated in regulatory roles in a wide number of stresses were selected for further study. The NAC TFs ANAC019, ANAC055 and

ANAC072 have been associated with multiple forms of biotic and abiotic stress and are therefore excellent candidates from which to build stress related gene regulatory networks.

To confirm that expression is enhanced in multiple different contexts, expression patterns of *ANAC019*, *ANAC055* and *ANAC072* were compared under multiple stress conditions (Figure 6.1). Visualisation of Affymetrix microarray data from the AtGenExpress repository indicates that expression of all three NACs are induced by multiple different abiotic, biotic and hormone treatments (Figure 6.1A). The expression of all NAC TFs is induced by multiple abiotic stresses; all three NAC TFs are induced by cold, osmotic, salt and drought treatment. *ANAC019* and *ANAC072* expression are more strongly activated under osmotic, salt and drought conditions compared to *ANAC055*. Expression of *ANAC072* expression is heavily induced by cold stress compared to the other two TFs. Expression of *ANAC019* and *ANAC072* is also induced by wounding, while *ANAC055* remains relatively unchanged. Studying the expression behaviour of the NACs during the senescence time-course, it is clear that all three display enhanced expression (Figure 6.1B).

Examining expression patterns in response to hormone treatment shows that all three NACs are induced to different extents by ABA and JA at least at one time point, while none are induced by SA. ABA slightly stimulates expression of *ANAC055*, while the effect on *ANAC072* and *ANAC019*, in particular, is more prominent. Treatment with JA strongly induces expression of *ANAC055* and *ANAC019*, while the effect on *ANAC072* appears less significant. These results suggest that all three of these NAC TFs may play a role in ABA and JA mediated signalling, while the difference in hormone responsiveness suggests that ANAC055 functions in a JA dependent manner, ANAC072 is more associated with ABA signalling, while ANAC019 may have a prominent role in both ABA and JA mediated signalling.

The expression of all three genes, *ANAC019*, *ANAC055* and *ANAC072*, is induced by at least one form of biotic stress. All three are heavily induced by both virulent (DC3000) and avirulent (avrRpm) strains of *P. syringae* 24 hpi, while only *ANAC019* and *ANAC055* appear to be activated in response to a non-host bacterial pathogen (PshP). *ANAC019* is the only TF out of the three whose expression is induced by a type-III secretion system deficient bacterial strain (HrcC). *ANAC019* and *ANAC055* expression is induced by the bacterial flagellin-derived peptide, Flg22, while *ANAC072* remains unchanged. Only *ANAC055* expression is induced by the hemibiotrophic pathogen *P. infestans*. Examining the expression profiles of the *ANAC019*, *ANAC055* and *ANAC072*

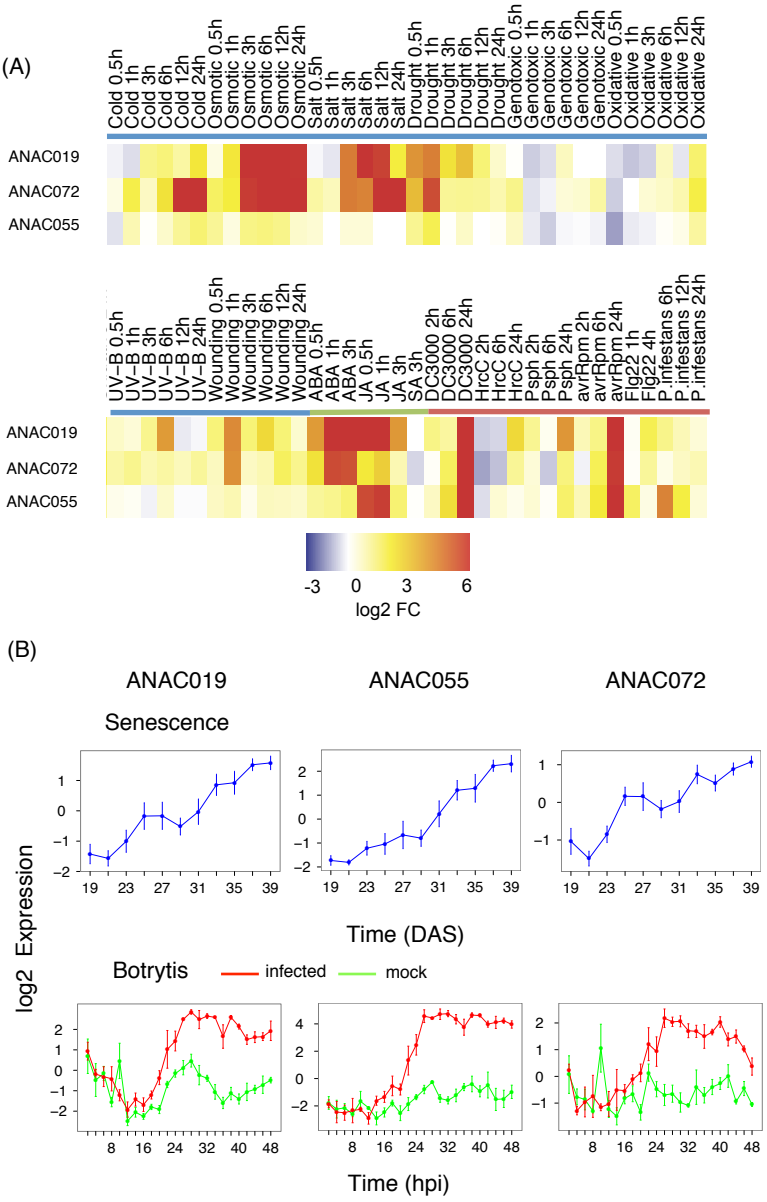


Figure 6.1: ***ANAC019*, *ANAC055* and *ANAC072* expression is induced by multiple stresses.** (A) Abiotic, hormone and biotic stress specific expression of NAC TFs. Shown is heatmap representation of expression patterns based on normalised AtGenExpress Affymetrix microarray data. Abiotic stress, hormone and biotic stress treatments are indicated by the blue, green and red bars respectively. Expression ratios (treatment/control) indicated by the colour scale bar. Blue indicates downregulation and yellow-red upregulation. (B) time-course expression profiles of NAC TFs during senescence and Botrytis time-courses. Lines represent the mean of the biological replicates and error bars indicate standard error. Genes have a mean expression of zero. (Top) Changes in expression during senescence tracked over time from 19 DAS to 39 DAS. (Bottom) time-course expression profiles of NAC TFs in response to infection by Botrytis (red) and mock treatment (green). Expression for both treatments measured every 2 hours from 2 hpi until 48 hpi.

during the Botrytis infection time-course clearly shows that all three TFs are significantly upregulated following infection (Figure 6.1C).

In summary, expression of *ANAC019*, *ANAC055* and *ANAC072* is clearly induced in response to a wide range of both abiotic and biotic stresses. Under several conditions all three display the same general trend in expression pattern, which suggests that they may be coregulated. This is evident in the responses to bacterial pathogens and some abiotic stress such as drought. A particularly convincing case can be observed during the response to Botrytis infection: all three NACs undergo differential expression compared to mock treatment at approximately the same time (20 hpi) suggesting that they may share common regulatory inputs. Following exposure to some stresses however, the expression patterns of the three NACs do not correlate- as, for example, observed during response to cold, osmotic and salt stress. These differing expression patterns strongly suggest that these NAC TFs may also be regulated in distinct ways.

6.2.2 Core promoter conservation of *ANAC019*, *ANAC055* and *ANAC072*

Diversity in regulation between homologous NAC TFs can be partially attributed to important similarities and differences in promoter architecture. To investigate the extent of promoter conservation between these highly related TFs the comparative genomics tool, Evolutionary Analysis of Regulatory Sequences (EARS) (Picot *et al.*, 2010), was used to identify shared and unique regions between the promoters of *ANAC019*, *ANAC055* and *ANAC072*. The analysis of two sequences was performed using the online version of the tool (http://wsbc.warwick.ac.uk/ears/) with default parameters. The resulting conservation profiles of each pair-wise comparison of promoter sequences are shown in Figure 6.2. The promoters of *ANAC019* and *ANAC055* show the greatest levels of conservation (Figure 6.2A). This may be expected as phylogenetic analysis indicates that *ANAC019* and *ANAC055* are extremely similar at the sequence level (Ooka *et al.*, 2003; Tran *et al.*, 2004). Two clear peaks of conservation are observed, suggesting that these regions are important for coregulation of the two genes. Interestingly, large sections of the promoters bare no significant similarity indicating that these stretches of sequence may harbour elements that contribute to differences in expression pattern. The conservation profiles comparing the promoters of *ANAC072* with *ANAC019* and *ANAC055* suggest no large-scale conservation of regulatory sequence between either pair (Figure 6.2B). This finding indicates that the ANAC072 promoter structure has diverged considerably with minimal stretches of non-coding sequence retained.

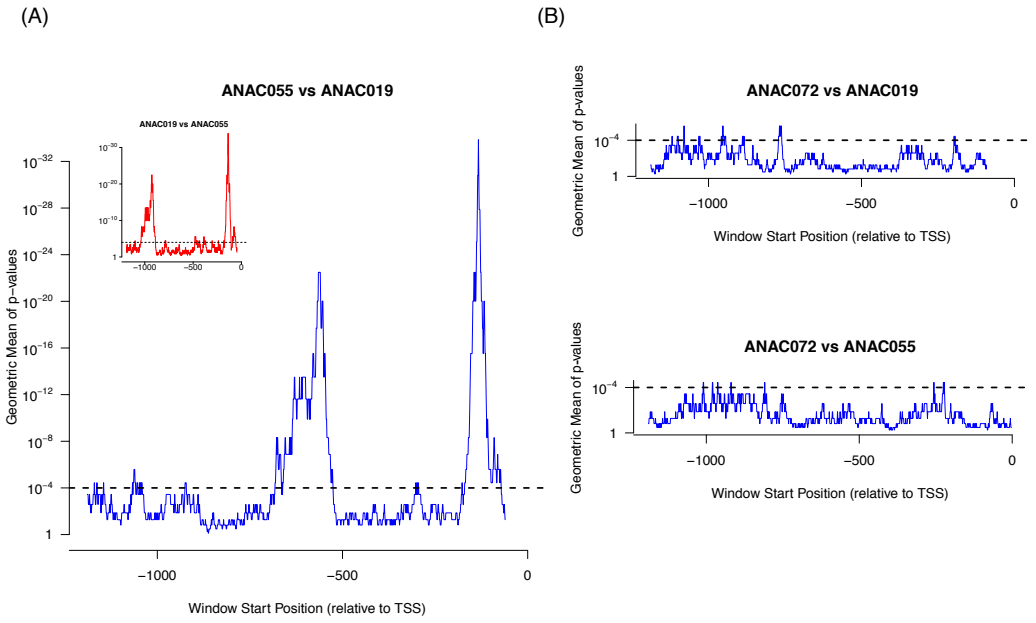


Figure 6.2: **NAC paralog core promoter sequence conservation.** Sequence conservation profiles were created for the pairwise comparisons of 1.25 kb promoters of *ANAC019*, *ANAC055* and *ANAC072* using the EARS tool (Picot *et al.*, 2010). (A) Conservation profile of *ANAC055* vs *ANAC019*. Blue line indicates *ANAC055* vs *ANAC019*. Inset shows conservation profile of *ANAC019* vs *ANAC055*. Two clear peaks of conservation are observed corresponding to two regions of significant promoter conservation. (B) Conservation profiles of *ANAC072* vs *ANAC019* (top) and *ANAC055* (bottom) show no large significant regions of similarity exist between these promoters.

6.2.3 Y1H assays identify TFs that interact with *ANAC019*, *ANAC055* and *ANAC072* promoters

To identify TFs that can physically interact with the promoter regions of *ANAC019*, *ANAC055* and *ANAC072*, partially overlapping promoter sequences, spanning approximately 800 – 1200 bp upstream of the predicted transcription start site of each gene, were used as bait to test against an Arabidopsis AD-TF library. A collection of shorter overlapping fragments, or 'promoter-hiking' (Pruneda-Paz et al. 2009), was used in order to reduce distal affects on transcription in yeast. To assess the ability of many individual Arabidopsis TFs to interact with promoter sequences in an unbiased manner, screening was performed with the Y1H 'prey' library, consisting of approximately 1400 Arabidopsis TFs as described in Chapter 5. Protein-DNA interactions were identified by mating bait and prey strains and selecting diploid cells for growth on selective media compared to background levels of growth.

An overview of the Y1H screening strategy and the promoter fragment design for each target gene is shown in Figure 6.3. Promoter fragments of *ANAC019* and *ANAC055* were designed to extend upstream far enough to cover regions that were heavily conserved between two promoters, based on the EARS analysis. *ANAC072* promoter fragments were designed to span 1000 bp upstream of the TSS. The original design of *ANAC072* promoter fragments 3 and 4 gave much higher background growth compared to all other tested fragments, with autoactivation of the *HIS3* gene apparent even on selective media supplemented with up to 150mM 3-AT. The inability to identify colony growth distinct from background levels for these fragments, led to the bait sequences being re-designed. High-levels of autoactivation were observed with two partially overlapping fragments, suggesting that the overlapping sequence may be responsible for autoactivation. Bait sequences were re-designed to omit the overlapping region between the two fragments, and when there were used in the Y1H screen, autoactivation could be quenched by addition of 3-AT.

This high-throughput screen generated an initial set of protein-DNA interactions, which were re-tested by co-transforming yeast with positive bait-prey pairs and comparing colony growth with that of cells co-transformed with bait and empty prey vector (pDEST22 alone). The co-transformation approach was also used to identify additional interactions that were missed in the mating screen. A representative set of results that indicate protein-DNA interactions via positive colony growth on selective media compared to their respective controls is shown in Figure 6.4. Figures showing Y1H growth results for those re-tested via co-transformation procedure are shown in Appendix E. The nature of the colony growth between co-transformants containing different promoter fragments was variable, with some colonies able to grow on selective media even though they only contain an empty prey vector (pDEST22 alone). These bait sequences, such as *ANAC055* fragment 3, for example, presumably have the ability to autoactivate transcription of the reporter gene. Such behaviour is not a problem as it is still possible to identify protein-DNA interactions by observing colony growth on more stringent selective media compared to the control. The end result was a set of high-confidence protein-DNA interactions, which included 27, 30 and 20 TFs that could bind to the promoters of *ANAC019*, *ANAC055* and *ANAC072*, respectively and a total of 116 interactions between TFs and all bait fragments (Table 6.1). Analysing the set of interactions with all three revealed a broad range of binding proteins, which included members of multiple TF families such as TCP, bZIP, bHLH, MYB, AP2/ERF, and homeodomain families, evenly distributed among all three NAC promoters. Large overlap was observed between the sets of positive interactors with some TFs able to bind to all NAC promoters, while some interactions were unique to single genes.

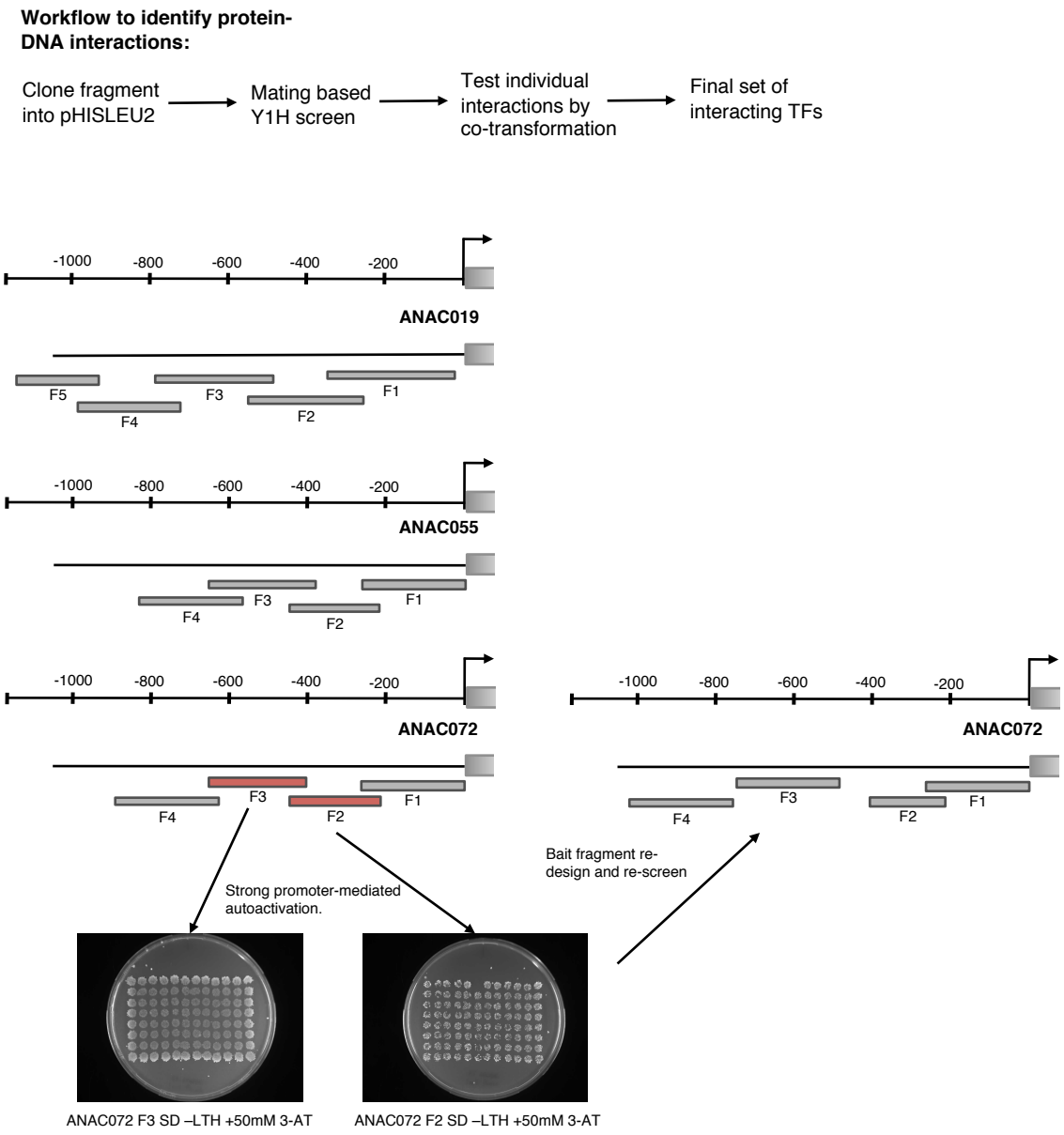


Figure 6.3: **Overview of Y1H screening strategy for identification of TFs that interact with *ANAC019*, *ANAC055* and *ANAC072* promoters.** Partially overlapping promoter fragments for *ANAC019*, *ANAC055* and *ANAC072* were cloned into bait vector pHISLEU2. Constructs were transformed into yeast strain Y187 in order to generate bait strains and mated with pooled versions of the TF library. Protein-DNA interactions are identified by colony growth of diploid cells on selective media compared to background levels of growth. Interactions were re-tested by co-transforming yeast strain AH109 with bait and prey vectors followed by plating on selective media. Co-transformation Y1H assays were also used to identify additional protein-DNA interactions. When tested in the mating based assay, the originally designed *ANAC072* promoter fragments 3 and 4 exhibited extremely high levels of autoactivation, that could not be quenched, even by very high levels of 3-AT. *ANAC072* promoter fragments were subsequently re-designed to yield fragments where background growth could be quenched by addition of 3-AT.

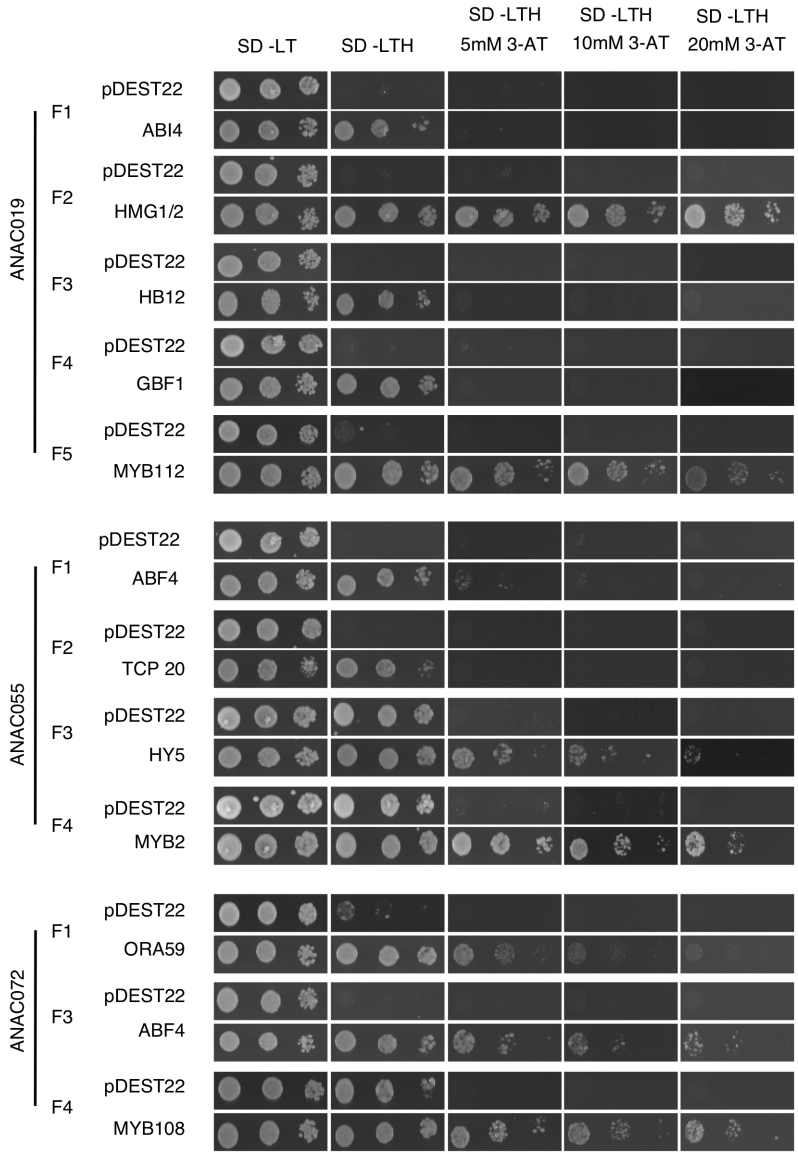


Figure 6.4: **Identification of TFs that interact with the promoters of *ANAC019*, *ANAC055* and *ANAC072*.** Shown are representative examples of protein-DNA interactions revealed by Y1H assays. Interactions between promoter fragments and TFs originally identified in a mating based Y1H screen were re-tested by co-transforming yeast strain AH109 with appropriate pairs of bait and prey vectors, followed by growth on media lacking histidine and increasing concentrations of 3-AT. Autoactivation is assessed with control co-transformants containing promoter fragments plus empty pDEST22 vector alone. Protein-DNA interactions are confirmed by growth of promoter-TF co-transformants on more stringent selective media compared to controls (pDEST22). Examples are shown for each promoter fragment that can interact with at least one TF.

Transcription Factor		ANAC019					ANAC055				ANAC072			
AGI identifier	Name	F1	F2	F3	F4	F5	F1	F2	F3	F4	F1	F2	F3	F4
AT4G34000	ABF3				●		●●		●●		●●			
AT3G19290	ABF4	●●			●●		●●		●●		●		●	●
AT2G40220	ABI4	●●					●●							
AT1G19350	BES1													●●
AT1G59640	BPEP				●●									
AT4G25490	CBF1										●●			
AT4G25470	CBF2										●			
AT4G25480	CBF3										●●			
AT5G51990	CBF4										●			
AT4G36730	GBF1				●		●●		●●					
AT4G01120	GBF2				⋆		●		●					
AT2G44910	HB04	●												
AT5G65310	HB05			●										
AT3G61890	HB12			●										
AT5G66700	HB53	⋆												
AT1G76110	HMG1/2		●●											
AT5G11260	HY5								●●					
AT3G17609	HYH								●					
AT5G54680	ILR3				●●				●●					
AT3G06490	MYB108					⋆			●	●●				⋆
AT1G48000	MYB112					⋆			●	●				●
AT1G25340	MYB116					⋆			●●	●●				●●
AT2G47190	MYB2					⋆			●●	●				●
AT3G27810	MYB21					⋆			●	●				⋆
AT3G01530	MYB57								●	●				
AT1G06160	ORA59	⋆					●●				●			
AT5G61270	PIF7	●●			●●		●●		●●		●●		●	
AT1G67260	TCP1	●●	●●	●			●●				●		●●	
AT3G47620	TCP14	●●	●	●●			●●	●●			●			
AT1G69690	TCP15	●●	●●	●●			●●	●●			●			
AT3G45150	TCP16	●	●	●●			●	●			●●			
AT3G27010	TCP20	●●	●	●			●●	●			●			
AT1G53230	TCP3	●●	●●		●●		●	●●			●●		●	
AT3G15030	TCP4	●	●●	●			●●	●						
AT2G22750	bHLH				●●				●●					
AT4G14410	bHLH104								●●					
AT4G34590	bZIP11								●●					
AT2G35530	bZIP16								●●					
AT1G75390	bZIP44						●●		●●					
AT1G06850	bZIP52								●					
AT3G44460	bZIP67								●●					

Table 6.1: **Positive interactions between TFs and promoters of *ANAC019*, *ANAC055* and *ANAC072*.** Filled circles indicate protein-DNA interaction as identified with the mating based screen. Single and double circles represent interactions observed via colony growth in a single and both versions of the pooled library respectively. Black bullets indicate that interactions were re-tested and confirmed by co-transformation. Grey bullets indicate interaction not subject to re-testing. Stars indicate protein-DNA interaction was identified via colony growth of co-transformed bait and prey vectors only.

6.2.3.1 *In silico* analysis can predict TF binding locations.

In silico analysis of each promoter sequence used in the Y1H screens was performed to identify occurrences of motifs that could serve as binding sites for TFs that interact with each promoter. Binding sites for each TF were retrieved from the TRANSFAC (Matys *et al.*, 2006) and PLACE (Higo *et al.*, 1999) databases and used to scan relevant fragments. The binding sites specific to most of the interacting TFs have not been experimentally characterised and in these cases, a binding site for a homologous TF was retrieved. The matrix similarity score (Section 3.2.2.3) was computed at each position and motif instances that achieved a score <0.001 were judged candidate binding sites. This analysis revealed the presence of putative binding sites for all interacting TFs and provides likely binding positions. The positions and instances of the putative TF binding sites in the promoters of *ANAC019*, *ANAC055* and *ANAC072* are shown is Figures 6.5, 6.6 and 6.7 respectively.

Instances of binding sites were identified for most TFs even though only a minority were specifically characterised. The similarity between characterised binding sites and instances of these motifs within promoters was particularly strong for TFs where their specific binding specificities have been elucidated. For example, the top hit for the ABI4 binding site (Transfac record; M00958) within promoter fragment 1 of both *ANAC019* and *ANAC055* has very high similarity to the PSSM that describes the ABI4 binding site (Figures 6.5 and 6.6). Strong matches were also observed for the ERF, ABF, homeodomain, PIF, HY5 and CBF binding sites. The consistency between TF-promoter interactions and the *in silico* motif matching results suggests that these are genuine interactions.

The only interactors for which a strong match to the experimentally described binding site was not always identified in relevant promoter fragments were the TCP TFs, which exhibit a broad range of binding to all NAC promoters. Binding motifs for the TCP family are currently described as GGNCCCAC and TGGGC(C/T). A strong match to the GGNCCCAC motif is found in promoter fragment 1 of *ANAC019*, *ANAC055* and *ANAC072*, but no other significant instances can be identified in fragments that interact with TCPs. The site II motif, which has the consensus TGGGC(C/T) has been shown to interact with TCP TFs (Trémousaygue *et al.*, 2003) but exact instances of this motif are not present in fragments that bind TCPs. This observation may reflect a lack of detailed descriptions of TCP binding sites forcing the use of a consensus sequence that lacks information regarding binding site degeneracy. The widespread nature of the binding suggests these TCP proteins may be able to recognise sequences that are suboptimal instances of current PSSM descriptions.

Examining the relative positions of each motif along each promoter reveals interesting combinatorial patterns. The most striking of these is the co-occurence and order of the G-box like element (core ACGT), TCP binding site and ERF/ABI4 binding sites near the TSS of *ANAC019*, *ANAC055* and *ANAC072* (Figures 6.5 – 6.7). The conservation of these binding sites and their relative positions between all three NAC promoters suggests that this combinatorial pattern may be functionally important. Interestingly, these elements fall within the first highly-conserved non-coding region between *ANAC019* and *ANAC055* (Figure 6.8). A recent publication (Zou *et al.*, 2011) used wild-type and mutagenised versions of the *ANAC019* promoter to demonstrate that several of these motifs contribute towards stress induced expression of *ANAC019*. Truncated promoter-β-glucuronidase fusion constructs were used to analyse the contributions of three motifs towards regulating salt-induced expression of *ANAC019*. Two of these motifs overlap with predicted binding locations for TFs that interact with the promoter fragment covering these sites (Figure 6.8A).

Another combinatorial motif pattern present within the fragments that bind multiple different TFs is observed between the predicted MYB binding site and motifs that share the G-box core. Again, this combinatorial pattern is conserved across all NAC promoters and is present in the second region conserved between *ANAC019* and *ANAC055* indicating that it may be important for regulation of NAC gene expression (Figure 6.8B).

6.2.3.2 Functional analysis of interacting TFs

The Y1H system identifies protein-DNA interactions in a context free manner, indicating which TFs have the capacity to interact with a stretch of DNA, but lacks information regarding firstly, whether such a binding event takes place *in planta* and if it does under what conditions the interaction can occur. This binary protein-DNA interaction data also lacks information regarding the potential regulatory consequence of a TF binding to the promoter of a gene. To expand the understanding of GRNs that operate during stress responses, it is necessary to pin down TFs that bind to promoters and contribute towards regulating expression of the neighbouring gene. In terms of gene expression, the relationship between a TF and its target is often complicated, but in general, if a TF undergoes differential expression, then the expression of target genes is expected to change accordingly. Using this hypothesis, microarray data that measure the expression of genes under a wide range of stress conditions was used to identify possible regulatory relationships between the three NAC genes and the TFs that bind to their promoters.

As described earlier in the chapter, NAC genes are induced by a broad range of stresses.

Using the AtGenExpress data it is possible to compare their expression with that of putative regulators in various conditions (Figure 6.9). A major observation is that many interactors are not differentially expressed in response to stress treatments covered by the AtGenExpress project, suggesting that under these conditions at least, these TFs do not exert a regulatory influence over the NAC TFs. Some TFs are induced by certain stresses, however, and it is possible to start identifying coexpressed TF-target pairs. *HB12* and *ABF3*, for example, display similar expression patterns to all three NAC TFs over all treatments, while *MYB2* and *MYB112* have similar expression patterns in response to most stresses but not hormone treatments.

While the AtGenExpress dataset covers a wide-range of contexts, it suffers from minimal time-resolution. The expression time-courses for each individual experiment consists of a maximum of 6 timepoints with the response to many stresses covered by fewer. Potential relationships between TFs and targets are easier to predict if more information is available regarding the temporal changes in gene expression. The high-resolution senescence and Botrytis time-course datasets provide such detail, and can be used to identify interactors that are upregulated or downregulated around the same time as the predicted targets (Figure 6.10). Again, many interactors do not appear to be differentially expressed in response to infection by Botrytis or during senescence, while some TFs do display similar expression patterns to their respective predicted targets.

Analysis of the biological processes associated with the interactors based on Gene Ontology (GO) classifications revealed enrichment for TFs associated with Response to Hormone Stimulus (hypergeometric corrected p-value <0.05). Given that *ANAC019*, *ANAC055* and *ANAC072* are all induced by various hormone treatments, interactions between promoters and this group of TFs may be relevant *in vivo*. More specifically, the hormone responsive TFs isolated in the Y1H screen are associated with mediating ABA and/or JA dependent signalling. The set of proteins that bind to the NAC promoters are enriched for TFs linked to the response to ABA (corrected p-value <0.05). All three NAC TFs are induced by ABA and are likely to be targeted by TFs known to mediate ABA signalling. Key ABA related TFs ABF3 and ABF4, were found to bind at multiple locations along the promoter of *ANAC019*, *ANAC055* and *ANAC072*. ABF TFs recognise the ABA-responsive element (ABRE) motif in order to promote ABA-induced gene expression (Yoshida *et al.*, 2010). The ABRE occurs multiple times within each NAC promoter suggesting that these elements may be responsible for ABA induced expression via ABFs *in vivo*. *ABF3* expression correlates well with the expression of *ANAC019* and *ANAC072* in response to osmotic stress, salt stress and in response to ABA treatment, suggesting a role in regulating these TFs during some abiotic stresses. The observable

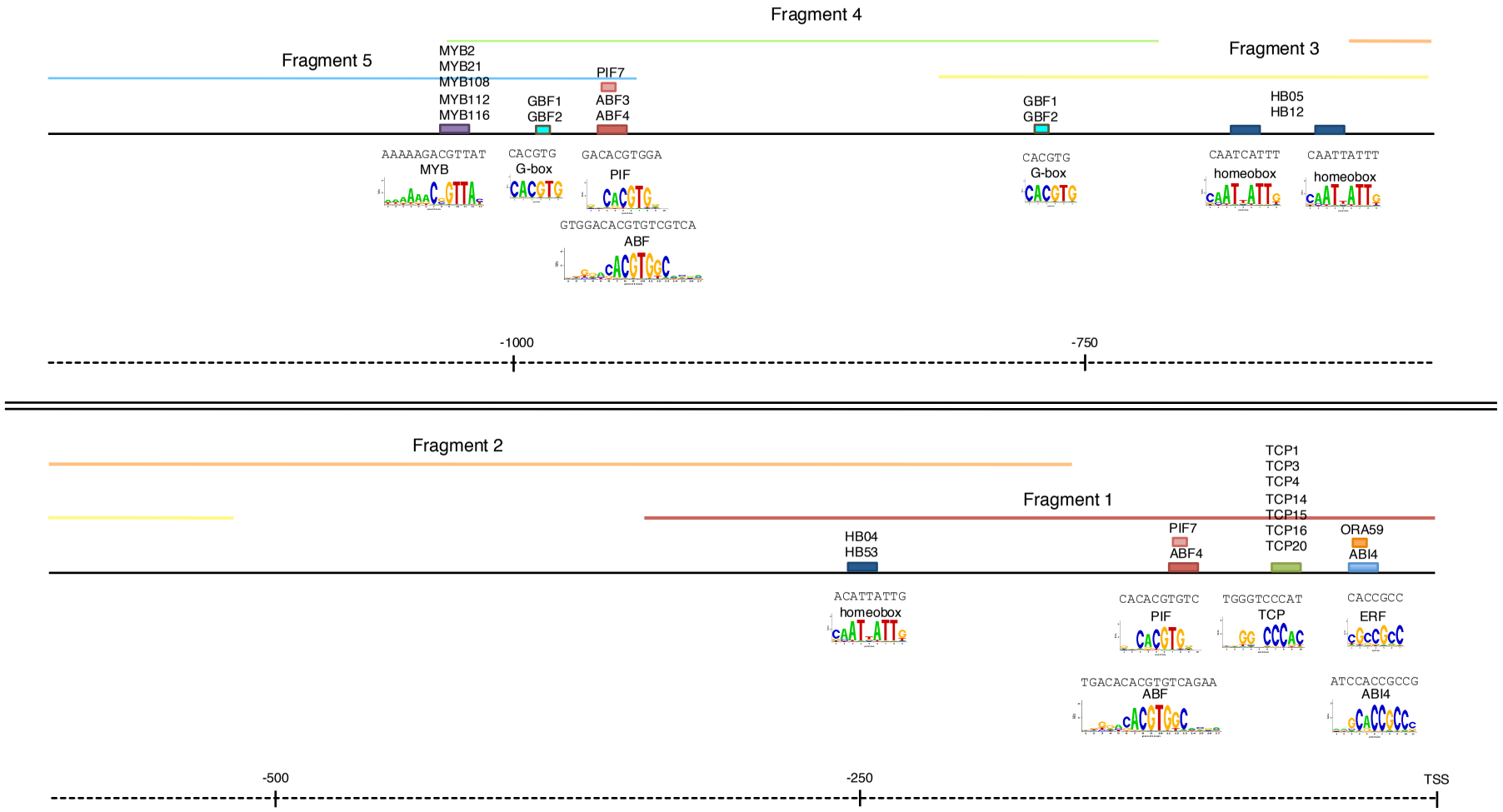


Figure 6.5: **Identification of putative binding locations for TFs that interact with the promoter of *ANAC019* in Y1H assays.** Shown is a scaled schematic of the *ANAC019* promoter (black line) with the positions of candidate binding sites for TFs that interact with specific fragments in Y1H assays. Position relative to TSS is shown on the dashed line. Putative binding sites (coloured boxes) were identified by scanning promoter fragments (coloured lines) with PSSMs that describe known binding specificities for interacting TFs or family members. TFs predicted to bind to a motif are displayed above putative binding sites. The sequence logo for each TF binding site is shown below the instance of each motif. PSSM name is given above logo. Predicted *cis*-regulatory elements serve as candidate binding locations highly related TFs.

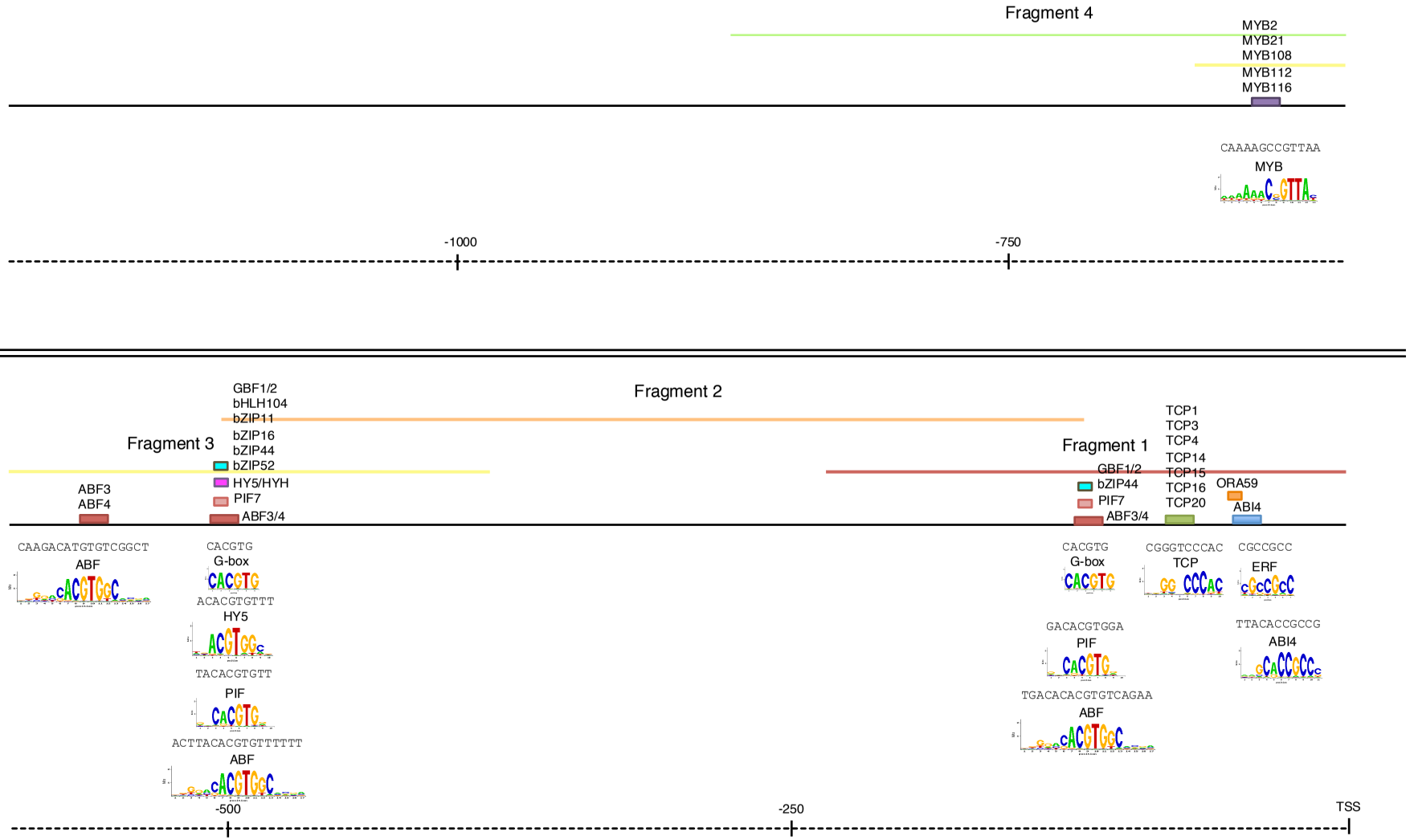


Figure 6.6: **Identification of putative binding locations for TFs that interact with the promoter of *ANAC055* in Y1H assays.** Shown is a scaled schematic of the *ANAC055* promoter (black line) with the positions of candidate binding sites for TFs that interact with specific fragments in Y1H assays. Position relative to TSS is shown on the dashed line. Putative binding sites (coloured boxes) were identified by scanning promoter fragments (coloured lines) with PSSMs that describe known binding specificities for interacting TFs or family members. TFs predicted to bind to a motif are displayed above putative binding sites. The sequence logo for each TF binding site is shown below the instance of each motif. PSSM name is given above logo. Predicted *cis*-regulatory elements serve as candidate binding locations highly related TFs.

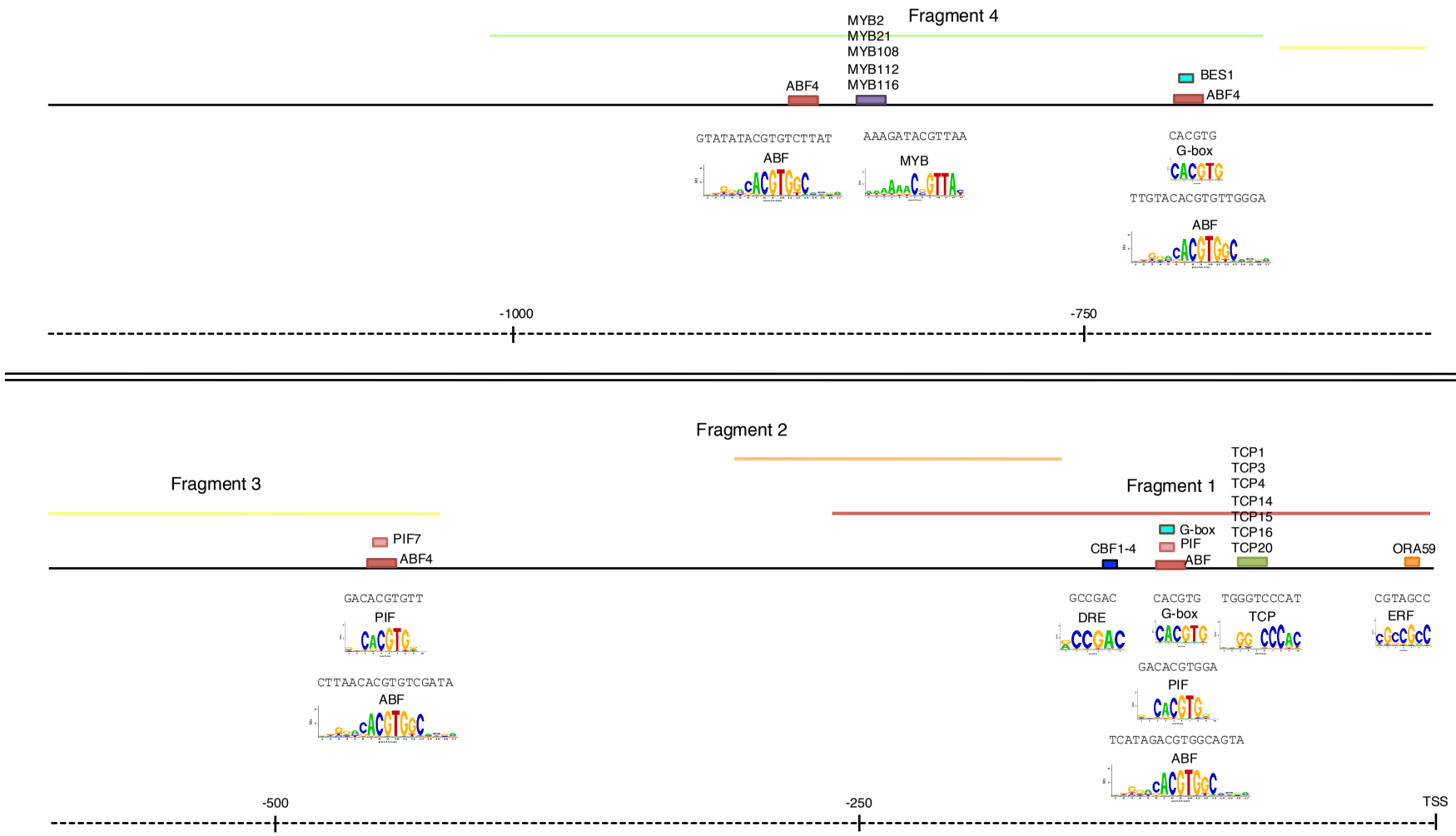


Figure 6.7: **Identification of putative binding locations for TFs that interact with the promoter of *ANAC072* in Y1H assays.** Shown is a scaled schematic of the *ANAC072* promoter (black line) with the positions of candidate binding sites for TFs that interact with specific fragments in Y1H assays. Position relative to TSS is shown on the dashed line. Putative binding sites (coloured boxes) were identified by scanning promoter fragments (coloured lines) with PSSMs that describe known binding specificities for interacting TFs or family members. TFs predicted to bind to a motif are displayed above putative binding sites. The sequence logo for each TF binding site is shown below the instance of each motif. PSSM name is given above logo. Predicted *cis*-regulatory elements serve as candidate binding locations highly related TFs.

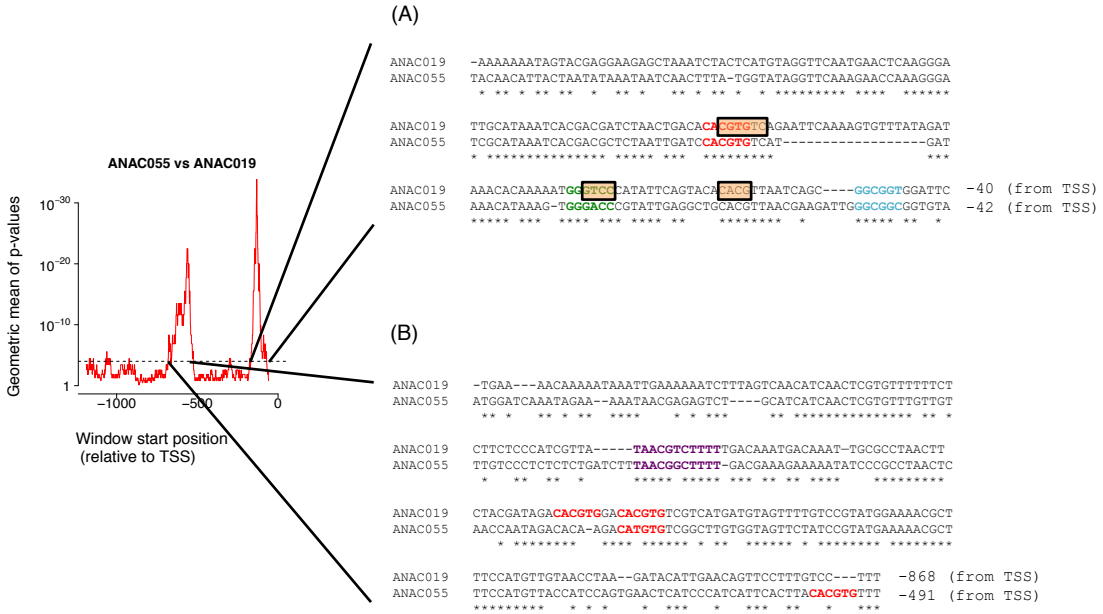


Figure 6.8: **Conserved regions contain combinatorial motif patterns.** Conserved regions originally identified in the EARS analysis are covered by promoter fragments that demonstrate TF-DNA interactions in the Y1H screen. Sequence alignments of the promoter regions conserved between *ANAC019* and *ANAC055* were generated using ClustalW (http://www.ebi.ac.uk/clustalw/). Coloured text indicates putative TF binding sites identified by *in silico* sequence analysis. (A) Alignment of conserved region between *ANAC019* and *ANAC055* nearest the TSS. The G-box variant motifs, TCP binding site, and ERF/ABI4 binding site are shown in red, green and blue respectively. Orange boxes indicate sequences that were identified by Zou *et al.* (2011) as functionally important for salt-induced expression of *ANAC019*. (B) Sequence alignment of the second conserved block identified in the EARS analysis indicates the combinatorial pattern of putative MYB binding site (purple letters) and G-box like elements (red letters) is conserved between *ANAC019* and *ANAC055*.

correlation between *ABF3* upregulation and that of *ANAC055* is not as strong as for the other two NAC TFs, although *ANAC055* is responsive to ABA treatment and therefore could be regulated by ABF3. Interestingly, *ABF4* expression only appears to be slightly induced following ABA treatment in the AtGenExpress experiment, but has been functionally linked to mediating ABA signalling (Yoshida *et al.*, 2010).

The upstream regulatory sequences of *ANAC019* and *ANAC055* also interact with another key mediator of hormone signalling, ABI4, an AP2 family TF that binds to the consensus sequence GCACCGCC, which also occurs in the promoters of these genes. Interestingly the occurrence of the ABI4 motif in both promoter fragments is within 100 bp of the TSS, suggesting that the relative position may be important for these *cis*-regulatory elements. ABI4 is an important mediator of ABA signalling during seed

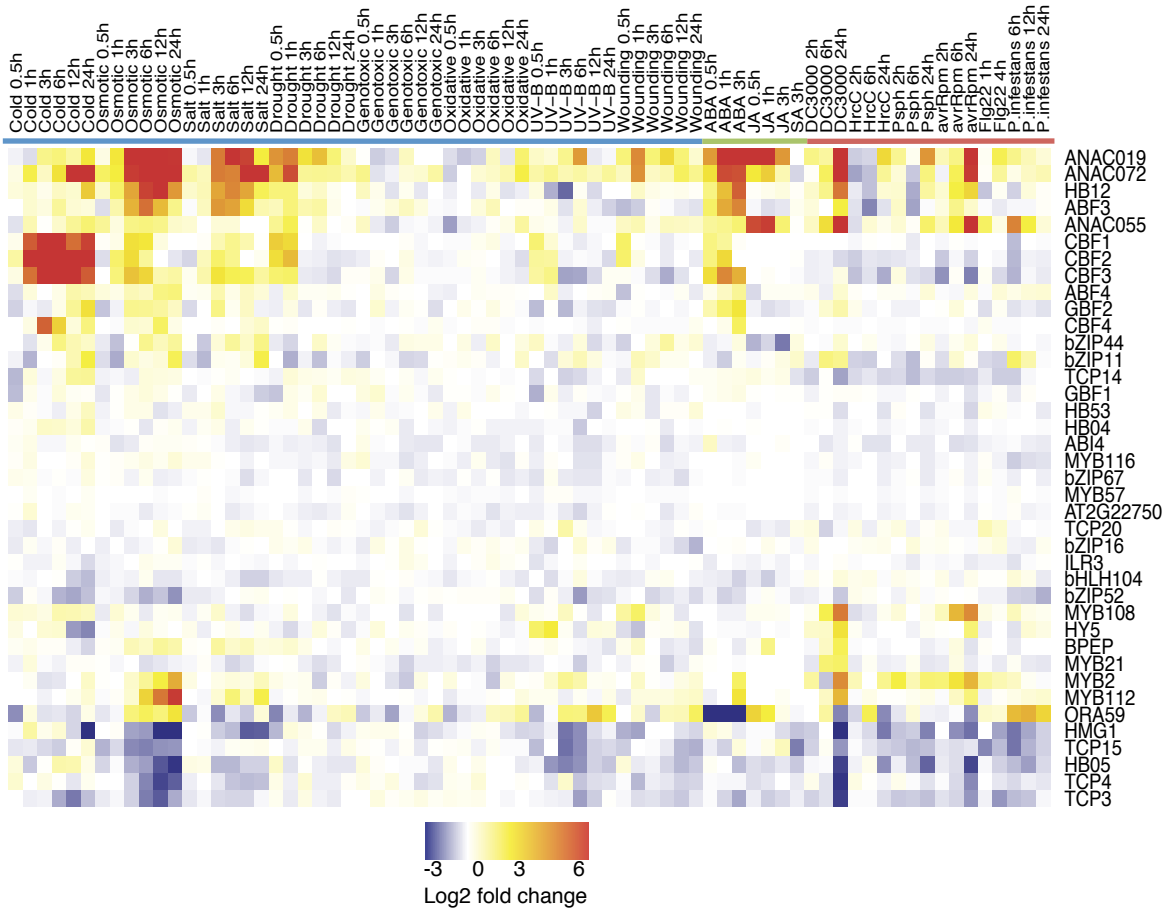


Figure 6.9: **Temporal expression patterns of interacting TFs in response to various environmental stimuli.** Expression ratios (treatment/control) for all interacting TFs and *ANAC019*, *ANAC055* and *ANAC072* were computed from the AtGenExpress datasets and displayed as a heatmap. Blue indicates downregulation and yellow-red indicates upregulation. Clustering genes on expression reveals interactor gene expression patterns that correlate with those of *ANAC019*, *ANAC055* and *ANAC072*.

development and germination (Finkelstein *et al.*, 1998), and it has recently been linked to regulating defence responses in a JA dependent manner (Kerchev *et al.*, 2011).

The key integrator of JA and ET signalling, ORA59, was found to bind to the promoters of *ANAC019*, *ANAC055* and *ANAC072*. Consistent with this, *ORA59* expression is induced early after JA treatment. *ORA59* is coexpressed with all three NAC TFs during the response to Botrytis infection and expression is induced along with *ANAC055* following infection by *P. infestans*. Artificially enhancing or repressing the expression of *ORA59* makes Arabidopsis plants less and more susceptible to fungal pathogens respectively (Pré *et al.*, 2008). The key importance of ORA59 in regulating defence responses

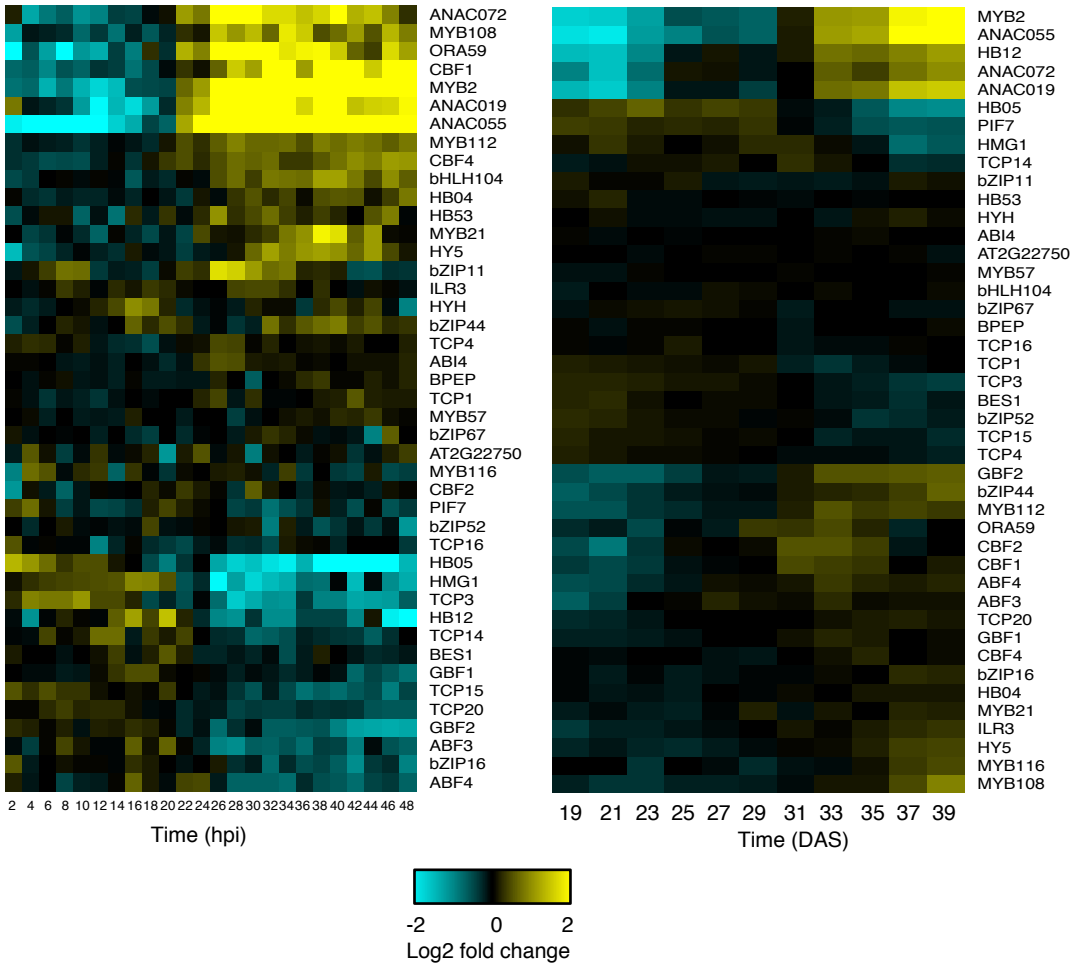


Figure 6.10: **Expression profiles of interacting TFs in senescence and Botrytis-infection time-courses.** The heatmaps display expression levels of *ANAC019*, *ANAC055* and *ANAC072*, and TFs that interact with these NAC gene promoters over time in response to infection by Botrytis (left) and during senescence (right). Cyan indicates downregulation and yellow indicates upregulation. Clustering on expression identifies genes that are coexpressed with the NAC TFs in either condition.

makes it a plausible regulator of the NAC TFs during some stresses. ORA59 mediates JA and ET signalling by interacting with subtle variants of the GCCGCC consensus sequence. Significant occurrences of this motif were found less than 100 bp upstream of the TSS of all three NACs, suggesting that this interaction may be conserved amongst these promoters.

Another interaction of possible functional significance is that of PIF7, which interacts with all NAC promoter sequences. Binding sites for PIF TFs are found in the vicinity

of all PIF7 binding events. PIF7, under circadian control, has been shown to negatively regulate stress-related TFs (Kidokoro *et al.*, 2009). The expression of *PIF7* is downregulated in response to Botrytis infection and during senescence, an opposite trend to that of *ANAC019*, *ANAC055* and *ANAC072* in these datasets. Inverted expression profiles fit the hypothesis that PIF7 is repressing expression of the three NAC genes under non-stress conditions, which is then relieved following exposure to stress. Inputs from the circadian clock are also observed though the binding of HY5 and HYH to the promoter of *ANAC055*. These interactions suggest possible mechanisms of NAC TF regulation by the circadian clock.

Four closely related proteins; CBF1, CBF2, CBF3 and CBF4 interacted with the promoter of *ANAC072* via the bait fragment positioned closest to the TSS (Figure 6.11A). *CBF1-4* are heavily induced by drought and/or cold, and have been shown to bind to the promoters of stress related genes via the dehydration response element (DRE), which has the consensus (A/G)CCGAC (Sakuma *et al.*, 2002). Interestingly, the only instance of the DRE in the entire intergenic region upstream of *ANAC072* is present in the fragment that displayed CBF protein binding, pinpointing the likely location of interaction. An instance of the DRE motif occurs within the promoter of *ANAC055*, indicating a potential binding location for the CBF TFs; however, when the fragment that carries this sequence was tested in Y1H assays, no interaction was observed (Figure 6.11B). *ANAC072* expression is induced more strongly by drought and cold stress compared to *ANAC019* and *ANAC055* (see Figure 6.1), suggesting a more specific regulatory role for this TF. The absence of interactions between CBF TFs and either promoter of *ANAC019* or *ANAC055* implies the existence of a possible mechanism to delineate the regulatory function of ANAC072 in response to drought and cold stress.

6.2.4 MYB TFs bind to the promoters of *ANAC019*, *ANAC055* and *ANAC072*

A particularly interesting set of protein-promoter interactions was observed between members of the R2R3 MYB TF superfamily and the promoters of all three NAC TFs tested in the study. MYB2, MYB21, MYB108, MYB112 and MYB116 all interact with bait fragments corresponding to the promoters of *ANAC019*, *ANAC055* and *ANAC072* (Figure 6.12A). A descriptive binding site motif model, assembled from multiple instances of the binding site for any MYB TFs that interact with the NAC promoters is not available, making localisation of potential binding site for the MYB TFs difficult. The R2R3 MYB TF superfamily is characterised by a shared R2R3 DNA binding domain. An alignment of all Arabidopsis MYB TFs, followed by phylogenetic tree construction, groups MYB TFs together based on similarity of the DNA binding domain

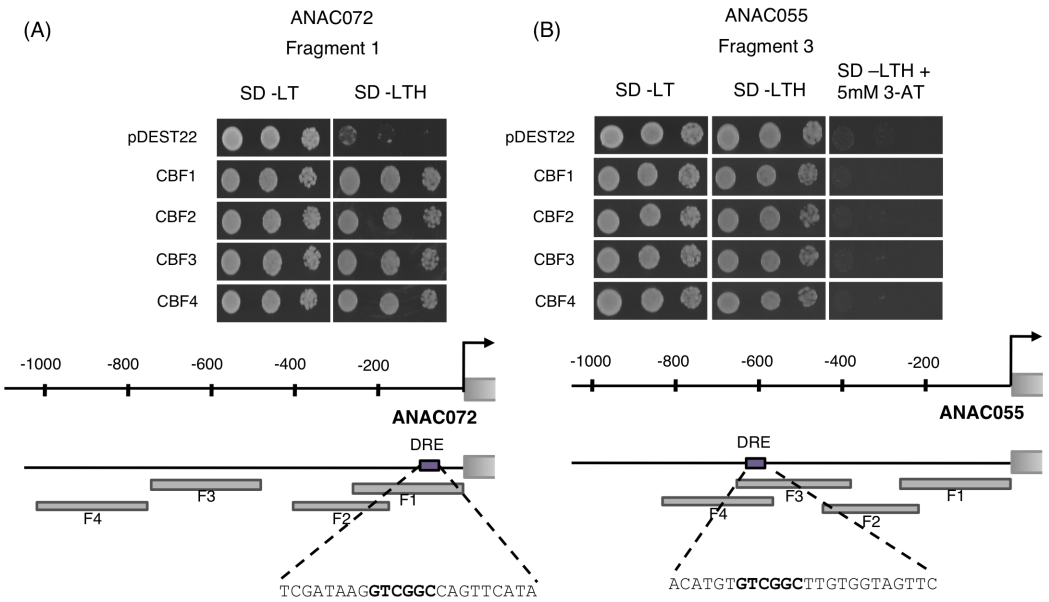


Figure 6.11: **CBF TFs that interact with *ANAC072* promoter.** (A) The C-repeat/dehydration-responsive element (DRE) binding factors (CBF1-4) bind to the promoter of *ANAC072*. Interactions between CBF1, CBF2, CBF3 and CBF4 and fragment 1 of the *ANAC072* promoter were confirmed by transforming bait strains with prey plasmids expressing the AD-CBF fusion proteins and assessing growth on media lacking histidine compared to bait strains transformed expressing AD alone (pDEST22). The only occurrence of the DRE (consensus GTCGGC, bold text) in the whole promoter of *ANAC072* is present 130 bp upstream of the TSS. (B) Testing the ability of CBF1-4 to bind to the promoter of *ANAC055*. Sequence analysis identified an occurrence of the DRE core sequence within *ANAC055* promoter fragment 3 (bold text). Diploid yeast cells containing promoter-bait and AD-CBF prey vectors did not display growth on selective media compared to control.

(Figure 6.13A). MYB2, MYB108, MYB21, MYB112 and MYB116 form part of a phylogenetic clade within the MYB family indicating that they have highly similar DNA binding domains. Protein sequence alignment of the interacting MYB TFs illustrates the conservation of the R2R3 DNA binding domain (Figure 6.13B). The lack of additional members of the MYB family to demonstrate positive interactions in the Y1H screen suggests that these MYB TFs bind to the promoters of the NACs in a selective sequence specific manner. Colonies indicating positive interactions were observed with *ANAC055* fragments 3 and 4 as bait sequences, indicating that the MYB binding site exists in the overlapping region. Sequence analysis identified a strong hit for a well-characterised MYB binding site (Transfac record, M00218) in this overlap. Even though this binding site is not associated with a MYB TF isolated in the Y1H screen, it still strongly suggests that the MYB TFs are interacting with this subsequence. Similar instances of this motif are also present in *ANAC019* promoter fragment 5 and *ANAC072* promoter fragment 4, which also display a positive interaction with the same MYBs (Figure 6.12A). An

alignment of the three predicted binding sites (Figure 6.12B) shows that they all share the core MYB recognition site, AAC, and match the core of a site that is known to interact with MYB2 (Abe *et al.*, 1997).

Studying expression patterns of MYB TFs in the AtGenExpress dataset reveals that some are coexpressed with NAC TFs under multiple conditions (Figure 6.1). Expression of *MYB2*, *MYB21*, *MYB108* and *MYB112* is all induced by virulent *P. syringae* (DC3000). Expression of *MYB2* and *MYB112* are both induced by osmotic stress and ABA. Only *MYB108* expression is induced by cold stress and only *MYB112* is induced by salt stress. *MYB116* does not appear to be activated by any stress covered by the AtGenExpress project. The binding of *MYB2* and *MYB108* to all three NAC promoters is particularly attractive because these TFs are believed to function in the same regulatory pathways as NAC TFs. MYB2 is heavily implicated in the ABA mediated regulation of salt- and drought- responsive genes (Abe *et al.*, 2003). MYB108 has been implicated in hormone signalling pathways and in the regulation of the response to infection by Botrytis and other pathogens (Mengiste *et al.*, 2003; Mandaokar and Browse, 2009). Furthermore, both *MYB2* and *MYB108* have expression patterns that overlap with those of *ANAC019*, *ANAC055* and *ANAC072* during infection by Botrytis and during senescence, based on high-resolution time-course data (Figure 6.14). Genes are more likely to have regulatory relationships if their expression has a similar temporal pattern and so the observation that *MYB2* and *MYB108* exhibit overlapping expression with the three NAC TFs in both experiments made them logical candidates for further study.

To investigate the predicted MYB binding site, mutated forms of *ANAC055* fragment 4 were generated, in which part of the putative MYB binding site was replaced with alternative sequence. Three mutant promoter constructs were created (m55F4a, m55F4b and m55F4c). The mutated bait vector was transformed into yeast carrying the individual AD-MYB TF prey vectors and plated out on selective media (Figure 6.15). The subsequent recombinant yeast strains displayed variable growth on selective media in the presence of 3-AT. Mutant m55F4a appeared to have the largest effect on growth compared to wild-type, and suggests that MYB2 and MYB108 do interact with the MYB binding site present in the promoters of the NAC TFs. Growth of co-transformants containing mutated promoter constructs was not completely abolished relative to the control, suggesting that the MYB TFs can still bind to the promoter fragment. This may reflect the presence of other motifs within the promoter that have affinity for MYB TFs. Studying the sequence surrounding the predicted MYB binding site, multiple occurrences of the MYB binding core sequence, AAC, are visible. It may be that the lack

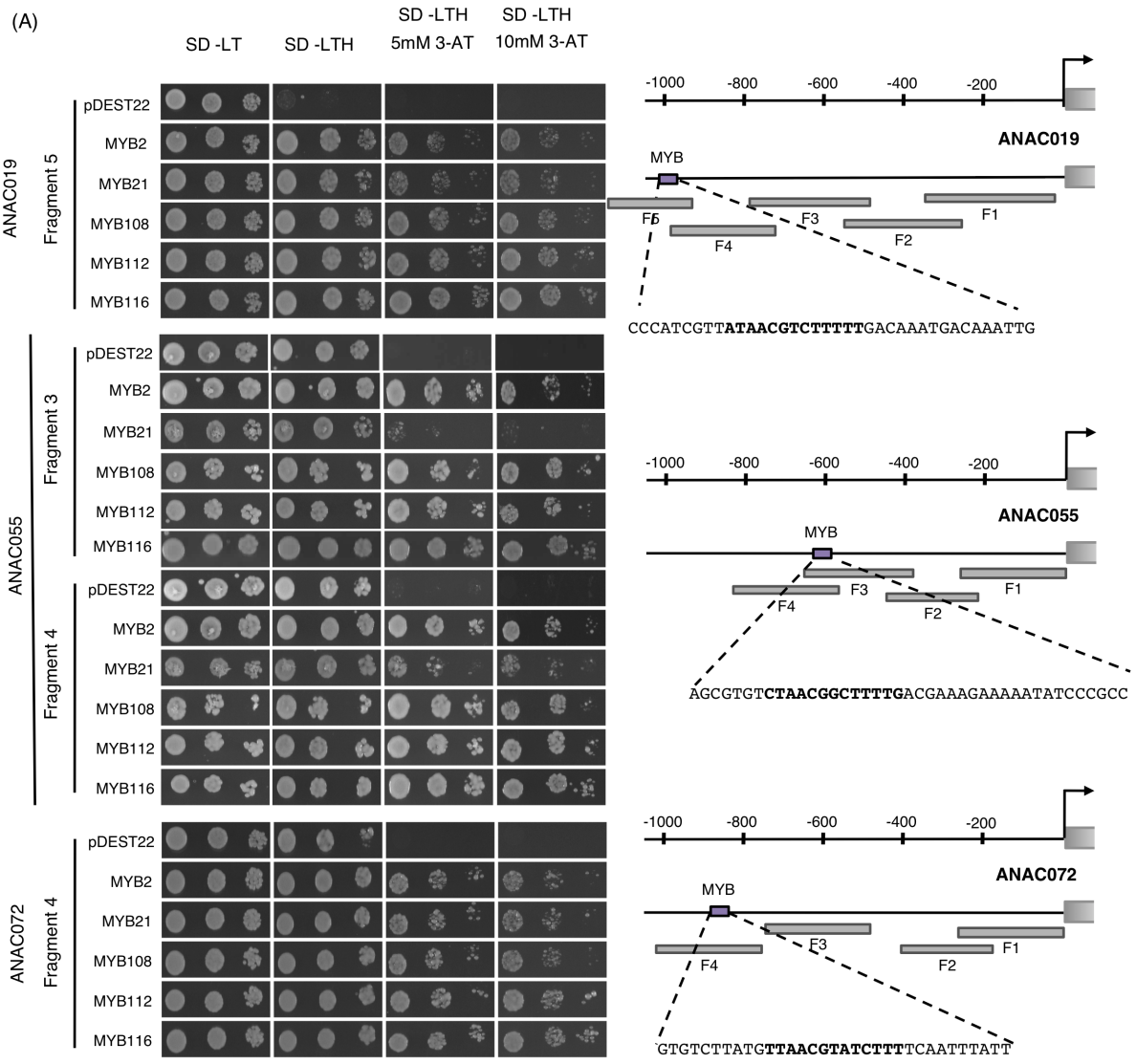


Figure 6.12: **MYB TFs interact with all promoters of *ANAC019*, *ANAC055* and *ANAC072*.** A collection of MYB TFs bind to the promoter of all three NAC genes tested in this study. (A) Left: Confirmation of interactions between a collection of MYB TFs and *ANAC019* promoter fragment 5, *ANAC055* promoter fragments 3 and 4, and *ANAC072* promoter fragment 4. Plasmids carrying each AD-MYB TF translational fusion and the control plasmid expressing AD alone (pDEST22) were transformed into yeast carrying the *HIS3* reporter gene under the control of each single promoter fragment. The transformants were examined for growth in the presence of increasing concentrations of 3-AT. Right: Diagram of each NAC promoter indicating the position of the putative MYB binding sites. The occurrence of the MYB motif is highlighted in bold. (B) Alignment of the sequences that match the motif describing the MYB binding site (Transfac record, M00218). Below the alignment is a binding site for MYB2 elucidated in the study by Abe *et al.* (2003).

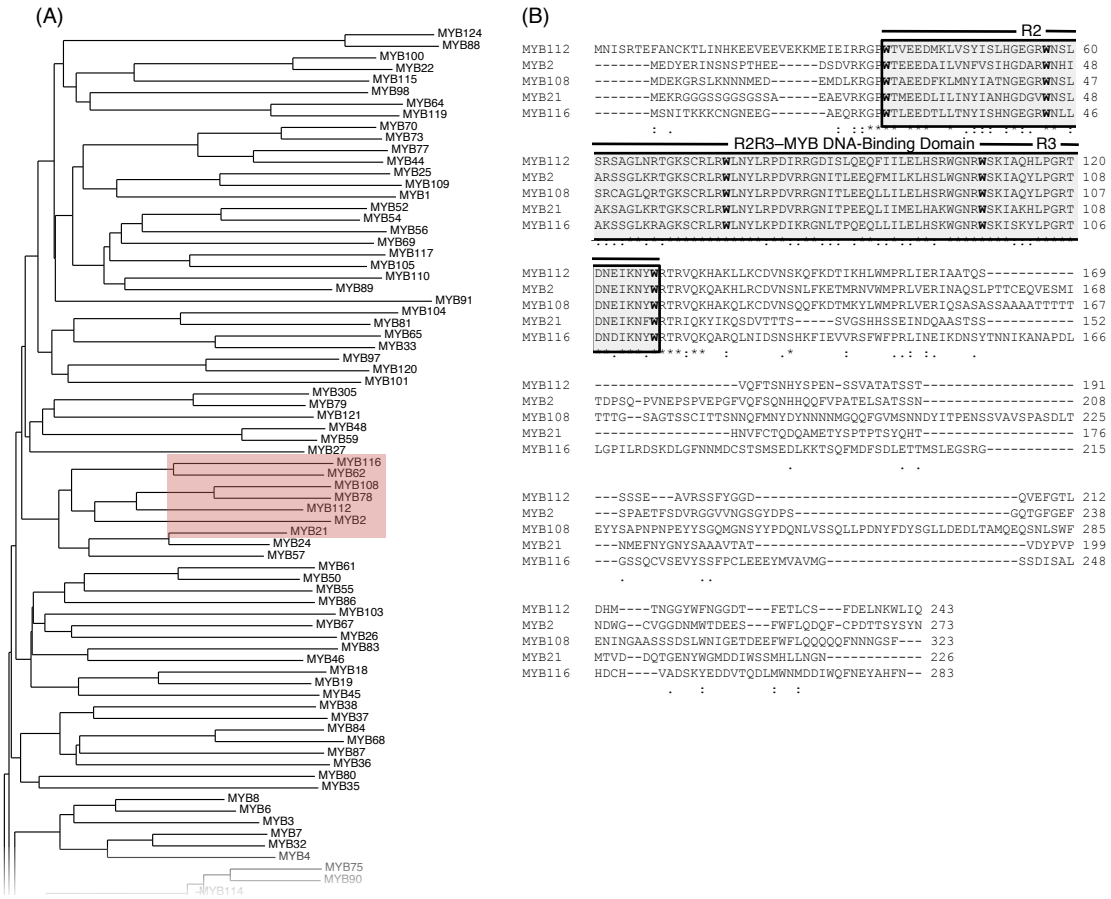


Figure 6.13: **Interacting MYB TFs belong to a phylogenetic clade based on similarity in DNA binding domain.** (A) Phylogenetic tree indicating the relationships between Arabidopsis MYB TFs. Shown is part of the tree generated from an alignment of all Arabidopsis MYB TF protein sequences performed using ClustalW. The MYB TFs that bind to the promoters of *ANAC019*, *ANAC055* and *ANAC072* form part of a phylogenetic clade (red box). (B) Comparison of amino acid sequences of MYB2, MYB21, MYB108, MYB112 and MYB116 using ClustalW. All MYB TFs that can interact with NAC promoters share a highly conserved R2R3 MYB DNA binding domain (boxed).

of a complete characterisation of any of the interacting MYB TFs binding specificities means that it is not possible to identify all of the MYB binding sites present in this promoter fragment or that interactions are dictated by binding mechanisms that are not yet fully understood. More mutant analysis is needed in order to completely abolish growth and as a result gain a more detailed understanding of the mechanisms of interaction taking place.

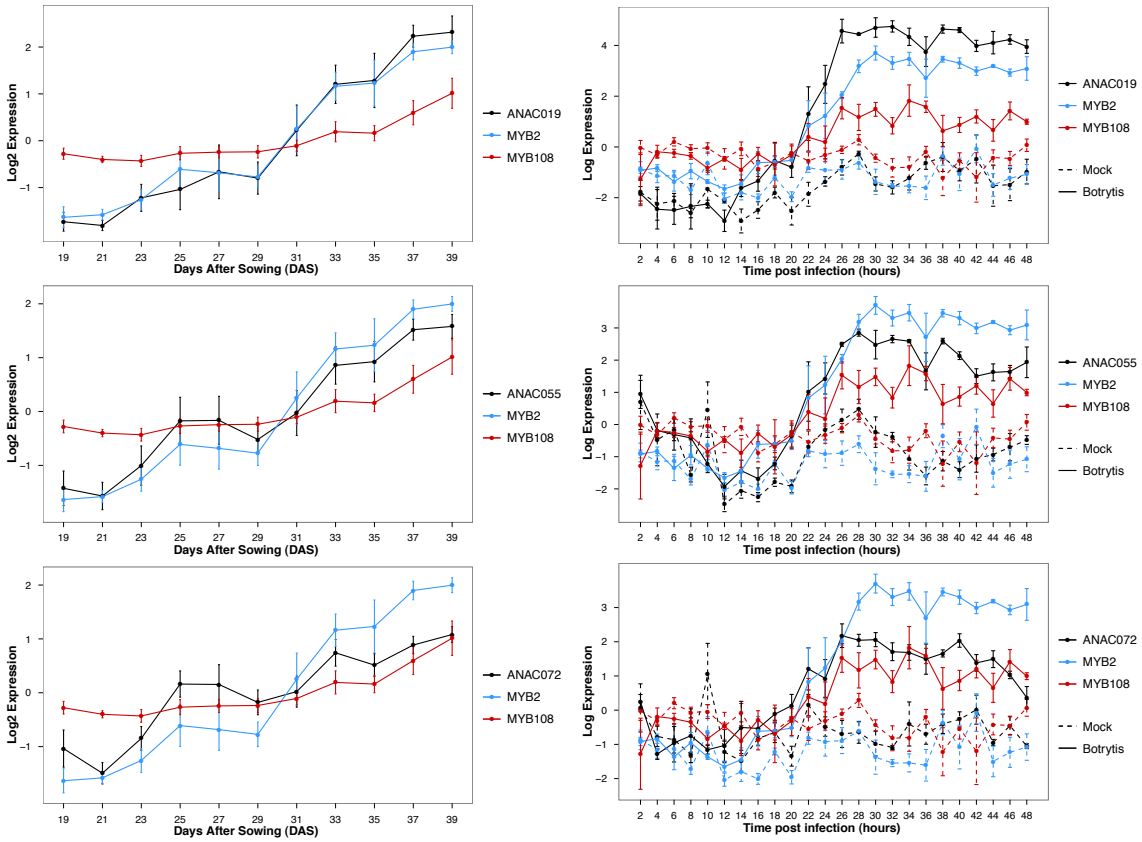


Figure 6.14: **Expression of *MYB2* and *MYB108* is positively correlated with that of *ANAC019*, *ANAC055* and *ANAC072* during time-course experiments.** Expression patterns of *MYB2* and *MYB108* during senescence (left) and response to Botrytis infection (right). Both MYB TFs and *ANAC019*, *ANAC055* and *ANAC072* have similar temporal expression profiles during both experiments.

6.2.5 Expression of *ANAC019*, *ANAC055* and *ANAC072* is perturbed in *MYB2* and *MYB108* Arabidopsis mutants

If MYB2 and MYB108 can bind to the promoters of *ANAC019*, *ANAC055* and *ANAC072*, and have overlapping expression time-courses in different contexts, then a natural question is, do these TFs regulate these genes *in vivo*? If this is the case then, if *MYB2* or *MYB108* are upregulated or downregulated, the expression of the NAC TFs may themselves be perturbed. To test this hypothesis, expression of the three NAC TFs was measured in *MYB2* and *MYB108* knockout and overexpressor lines following both infection with Botrytis and during dark-induced senescence. As the expression of *MYB2*, *MYB108*, and all three NAC genes is greatly enhanced in response to Botrytis infection and during senescence, these are logical conditions in which to probe potential regulatory interactions. It is more likely that the effect of perturbing the abundance of a regulator

```
ANAC055 F4 WT       TTTGTTGTTTGTCCCTCTCTCTGATCTTTAACGGCTTTTGACGAAAGAAAAATATCCCGCCTAAC
Mutant A (m55F4a)   TTTGTTGTTTGTCCCTCTCTCTGATCTTATTAAACTTTTGACGAAAGAAAAATATCCCGCCTAAC
Mutant B (m55F4b)   TTTGTTGTTTGTCCCTCTCTCTGATCTTAAAAAAAATTTGACGAAAGAAAAATATCCCGCCTAAC
Mutant C (m55F4c)   TTTGTTGTTTGTCCCTCTCTCTGATCTTTAAAAAAAAAAGACGAAAGAAAAATATCCCGCCTAAC
```

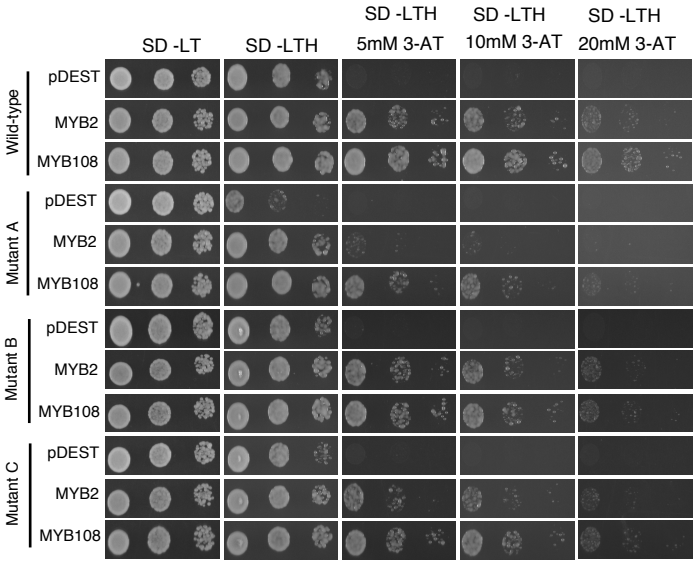


Figure 6.15: **Binding activity of MYB2 and MYB108.** Plasmids expressing the AD-MYB TF fusions were transformed into yeast carrying the reporter genes under control of fragment 4 of the *ANAC055* promoter containing either wild-type MYB binding site or one of three mutated motifs (m55F4a, m55F4b and m55F4c). Transformant growth assessed in the presence of dropout media lacking LTH + increasing levels of 3-AT. Strains containing the m55F4a promoter construct show reduced growth compared to wt, indicating MYB2 and MYB108 interact with that motif in wild-type bait strains.

will be observed when enhanced expression of the predicted targets is important as opposed to conditions where enhanced expression may not be observed.

As these MYB and NAC TFs are coexpressed during the Botrytis time-course, the same experimental procedure was used to assess the transcriptional differences between infected leaves from mutant and wild-type Arabidopsis plants. In order to capture the downstream effects of perturbing expression of each MYB TF, leaves from both mutant and wild-type plants were sampled shortly after the time of first differential expression and at the time where expression appears to peak in the time-course. Studying the Botrytis expression time-course, *MYB2* and *MYB108* begin to be upregulated at 24 hpi and 26 hpi respectively, while expression of both TFs appears to peak at 30 hpi. Sampling at the two respective timepoints should identify genes that are more likely to be direct targets of each MYB TF.

Senescence enhanced expression of the MYB and NAC genes was observed during the developmental senescence time-course. Developmental leaf senescence is a complex process, where the cells within a leaf are often at different stages in senescence (Buchanan-Wollaston *et al.*, 2005). This variability could make the identification of subtle changes in gene expression that are attributed to perturbation of a specific TF difficult. While senescence is a natural developmental process, it can also be induced by a range of different stresses, resulting in synchronised changes in the expression of senescence associated genes (SAGs). Dark-induced senescence is a common method for artificially inducing senescence and results in a more synchronised response across the whole leaf when compared to developmental senescence (Buchanan-Wollaston *et al.*, 2005). While there are differences in the sets of genes that are induced by developmental and dark-induced senescence, large overlap exists indicating that similar signalling pathways are required for both processes. A previous study (Buchanan-Wollaston *et al.*, 2005) has shown that *ANAC019*, *ANAC055* and *ANAC072* are upregulated during dark-induced senescence, indicating that as in developmental senescence, these TFs play important roles regulating the senescence process. Due to the increased uniformity of the transcriptional response across the whole leaf, differences in expression between mutant and wild-type leaves were assessed in response to dark-induced senescence. Samples were taken on two consecutive days following the onset of dark-induced senescence in the wild-type leaves (see Section 2.2.4).

Gene expression in wild-type, overexpression and knockout plants was analysed using the LimmaGUI Bioconductor package in order to identify differentially expressed genes. The significance of expression ratios of *ANAC019*, *ANAC055* and *ANAC072* in different MYB mutant backgrounds compared to wild-type are displayed in Figure 6.16, where the heatmap represents the signed corrected p-value. NAC genes with a ratio of wild-type to mutant significantly different to 1 with a corrected p-value <0.05 were considered differentially expressed. The analysis revealed that all three NAC TFs in our study were significantly upregulated in the *MYB108* overexpressor following infection by Botrytis. Expression levels of all three NACs were not significantly different in the *myb108* mutant, compared to wild-type at 26 hpi, but *ANAC019*, *ANAC055* and *ANAC072* displayed significantly reduced expression compared to wild-type at 30 hpi. No significant changes in expression of the three NACs were observed in the *myb2* mutant at 24 hpi, while *ANAC019* was the only NAC TF with a significant change in expression compared to wild-type at 30 hpi. These results suggest that MYB108 has the capacity to regulate all three NAC TFs during the response to Botrytis infection, while MYB2 only lies upstream of ANAC019 under the same conditions. The expression levels of the NACs following the onset of dark-induced senescence behaved similarly to

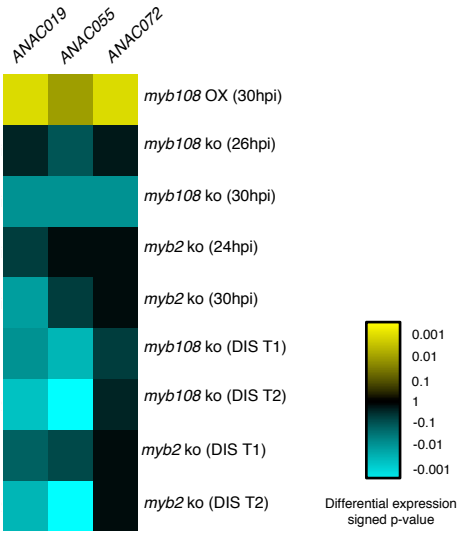


Figure 6.16: **Expression profiling of *ANAC019*, *ANAC055* and *ANAC072* in *MYB2* and *MYB108* mutant backgrounds under different stress conditions.** Shown are the significance of expression changes of *ANAC019*, *ANAC055* and *ANAC072* in *MYB108* overexpression line (*MYB108OX*), *myb108* knockout (ko) and *myb2* knockout. Heatmap representation of the significance of expression changes shows up-regulation (yellow) and down-regulation (cyan) in response to infection by Botrytis at two timepoints (hours post infection; hpi) and at two timepoints (T1, T2) that correspond to consecutive days following the induction of dark-induced senescence (DIS).

those in response to Botrytis infection; *ANAC019* and *ANAC055* were downregulated in the *myb108* mutant at both timepoints. Only *ANAC019* is significantly downregulated in the *myb2* mutant at the first timepoint, but both *ANAC019* and *ANAC055* are significantly downregulated at the second timepoint. Expression of *ANAC072* is unaffected in either mutant background either timepoint. As such, it is possible that MYB108 is positioned upstream of both ANAC019 and ANAC055 during the response to dark-induced senescence, while MYB2 appears to have more regulatory influence over *ANAC019* than *ANAC055*. Neither MYB2 or MYB108 is essential for the expression of *ANAC072* during dark-induced senescence, and it is possible that a double mutant may be needed to see an effect.

6.2.6 Global assessment of gene expression changes in *myb2* and *myb108* mutants during response to Botrytis infection and dark-induced senescence

To further understand the roles of MYB2 and MYB108 during the response to Botrytis and dark-induced senescence, the global differences in gene expression between wild-type and mutant plants were analysed. Genes with p-value <0.05 after adjustment for multiple testing and fold change of log2 >0.7 or <-0.7 were considered differentially expressed and subject to further analysis. Full lists of genes that are either significantly upregulated or downregulated in the *myb108* or *myb2* mutants following infection by Botrytis or during dark-induced senescence are shown in the electronic supplementary material.

6.2.6.1 Response to Botrytis infection

Microarray data analysis identified 77 genes differentially expressed in the *myb108* knockout 26 hpi with 51 upregulated and 26 downregulated compared to wild-type. Microarray data analysis identified 397 genes differentially expressed in the *myb108* knockout 30 hpi with 134 upregulated and 263 downregulated compared to wild-type. In order to examine which classes of genes are enriched in the sets of differentially expressed genes gene function enrichment analysis of Gene Ontology (GO) classifications was performed using BINGO (Maere *et al.*, 2005). As expected genes downregulated in infected *myb108* leaves were associated with biological processes linked with the response to pathogens, such as Response to Biotic Stimulus and Response to Stress. The GO term Response to Abiotic Stimulus is also overrepresented within the set of downregulated genes indicating that MYB108 regulates a broad range of stress responsive genes during the response to Botrytis infection.

In addition to the downregulation of *ANAC019*, *ANAC055* and *ANAC072*, two additional NAC TFs, *ANAC002/ATAF1* and *ANAC052* are also significantly downregulated. Although ANAC052 has not been explicitly linked to regulating stress responses, ANAC002/ATAF1 has been shown to regulate the responses to Botrytis and bacterial pathogens (Wang *et al.*, 2009). This suggests that MYB108 is a key regulatory hub, that regulates many NAC TFs that themselves play important roles in the response to pathogen infection.

Analysis of the microarray data comparing *myb2* plants with wild-type plants 24 hpi identified 213 genes differentially expressed with 103 upregulated and 110 downregulated in the *myb2* mutant. At the second timepoint, 30 hpi, there were 468 genes differentially

expressed with 210 upregulated and 258 downregulated in the *myb2* mutant compared to wild-type.

Until now, MYB2 has been only been implicated in abiotic stress signalling pathways, were it has been shown to control ABA mediated regulation of drought and salt responsive gene expression (Abe *et al.*, 2003). Functional analysis of the downregulated genes in the *myb2* mutants following Botrytis infection identified a number of enriched GO terms associated with stress responses such as Response to Stress and Response to Abiotic Stimulus. More specifically, the downregulated genes were overrepresented for GO terms describing responses to ABA, water deprivation and salt stress, further implicating MYB2 in the regulation of these processes (Figure 6.17). Table 6.2 lists the 21 genes that are listed in the GO category- Response to Abscisic Acid Stimulus. Within this list of ABA responsive genes, 13 are linked with the response to water deprivation,12 with the response to osmotic stress and 11 with the response to salt stress. The large degree of overlap suggests that MYB2 regulates these genes during the response to multiple abiotic stresses plus the response to Botrytis. Previous studies have demonstrated that MYB2 binds to the promoters and induces the expression of salt- and drought-responsive genes, *ADH1*, *RD22*, and *P5CS1* (Hoeren *et al.*, 1998; Abe *et al.*, 1997; Yoshiba *et al.*, 1999). *RD22* and *P5CS1* are present in the differentially expressed gene list for this experiment, while *ADH1* is >1.5 fold downregulated (log2 -0.66) with $p = 0.01$, even though it does fall just outside the imposed cutoff. This observation provides confidence that the microarray experiment is identifying relevant regulatory interactions.

While ANAC019 is not listed within the GO category, 'Response to Abscisic Acid Stimulus', multiple studies have identified that this TF is induced by ABA and functions as a positive regulator of ABA signalling. Given that MYB2 can bind to the promoter of *ANAC019*, which is downregulated in the *MYB2* knockout, it is natural to speculate that ABA-mediated signalling via ANAC019 is at least partially dependent on regulation by MYB2. The homeodomain TF, *HB12* is another ABA-related TF that is misregulated in the *MYB2* mutant. Y1H experiments demonstrated that HB12 can bind to the promoter of *ANAC019*, suggesting another route for the ABA induced expression of *ANAC019*.

The role of ABA in biotic stress responses is far from fully understood, and as been shown to have both positive and negative regulatory consequences (Ton *et al.*, 2009). The function of ABA in response to abiotic stress is more well defined, with elevated levels of ABA leading to positive regulation of genes that contribute towards stress tolerance. ABA mediated signalling during the response to fungal pathogens may reflect

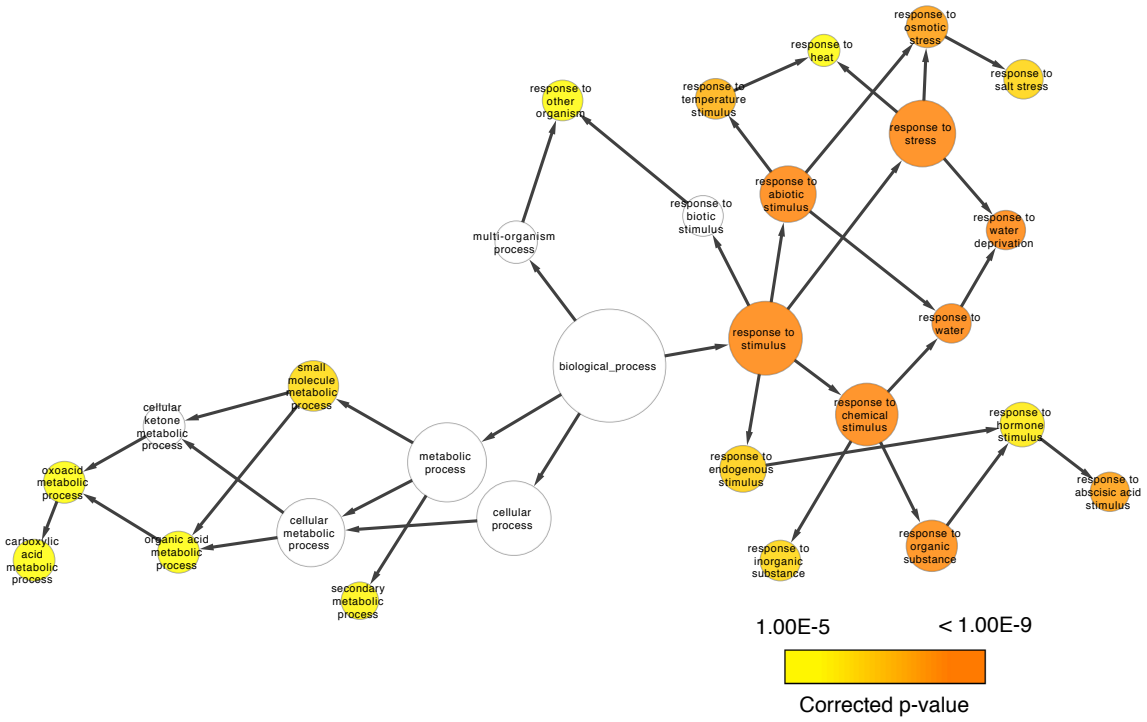


Figure 6.17: **GO categories overrepresented in genes that are downregulated in the *myb2* knockout during Botrytis infection.** The set of genes that were downregulated at either 24 hpi or 30 hpi were analysed for GO terms enrichment using BINGO. Shown are the most significantly overrepresented GO categories. The significance of GO enrichment is indicated by colour. P-values are subject to Benjamini and Hochberg correction. Edges indicate parent-child relationships between GO terms. The downregulated gene set is enriched for biological processes that MYB2 has previously been implicated in regulating, including salt stress, dehydration and ABA signalling.

induction of abiotic stress response pathways. It is likely that Botrytis infection imposes disruption and damage to cells similar to that experienced during salt, osmotic and drought stress. Given the number of genes that are important for stress responses that are misregulated in the *MYB2* knockout compared to wild-type during the response to Botrytis it suggests it is a key regulator of stress responses and more specifically ABA mediated stress responses.

6.2.6.2 Response to dark-induced senescence

Analysis of the microarray data comparing *myb108* to wild-type plants following the initiation of dark-induced senescence revealed 1620 genes differentially expressed. Out of all differentially expressed genes 576 were upregulated and 1044 downregulated in *myb108* mutants. Microarray analysis of the comparison between knockout and wild-type sam-

AGI	Gene name	Short name
AT1G44170	Aldehyde dehydrogenase	ALDH3H1
AT1G54100	Aldehyde dehydrogenase	ALDH7B4
AT4G25000	Alpha-amylase-like	AMY1
AT1G20440	Cold-regulated 47	COR47/RD17
AT3G55610	Delta1-pyrroline-5-carboxylate synthase 2	P5CS2
AT2G39800	Delta1-pyrroline-5-carboxylate synthase 1	P5CS1
AT1G20450	Early responsive to dehydration 10	ERD10
AT1G31930	Extra-large GTP-binding protein 3	XLG3
AT3G61890	Homeobox 12	HB-12
AT2G46680	Homeobox 7	HB-7
AT1G01060	Late elongated hypocotyl	LHY
AT5G01540	Lectin receptor kinase A4.1	LECRKA4.1
AT5G59320	Lipid transfer protein 3	LTP3
AT5G59310	Lipid transfer protein 4	LTP4
AT1G55020	Lipoxygenase 1	LOX1
AT3G11410	Protein phosphatase 2CA	PP2CA
AT3G05880	Rare-cold-inducible 2A	RCI2A
AT5G66400	Responsive to ABA	RAB18
AT2G33380	Responsive to desiccation 20	RD20
AT5G25610	Responsive to desiccation 22	RD22
AT1G77180	SKIP	SKIP

Table 6.2: **ABA related genes downregulated in Botrytis infected *myb2* leaves.** Shown are all of the genes that are significantly downregulated at either 24 hpi or 30 hpi in the *myb2* knockout leaves compared to wild-type.

ples from the next day revealed a total of 4421 genes that a significantly misregulated, with 2147 upregulated and 2274 and downregulated in the knockout. The number of genes differentially expressed over this short time-course is considerably higher than during Botrytis infection and possibly reflects a lag in the response to dark-induced senescence between *myb108* knockout and wild-type leaves. The time since exposure to darkness at which to sample leaves was based on the initiation of senescence in the wild-type leaves only.

In order to aid analysis, new genes lists were generated by considering only genes that were significantly downregulated or upregulated at both timepoints. GO analysis of the 993 genes that are significantly downregulated at both timepoints revealed enrichment of GO terms similar to Botrytis infected *myb108* mutants including Response to Stress, Response to Biotic Stimulus and Response to Abiotic Stimulus. In addition, GO terms associated with JA are also enriched. MYB108 has been suggested to interact with JA signalling pathway both in stress conditions and during development (Mengiste *et al.*, 2003; Mandaokar and Browse, 2009). The JA signalling pathway is activated during senescence and so this finding suggests that MYB108 may function in JA signalling dur-

ing dark-induced senescence. GO categories related to senescence are also significantly enriched amongst the downregulated gene-set, indicating the reduction of MYB108 is delaying the onset of senescence and suggests that in wild-type leaves, MYB108 positively regulates SAGs. GO analysis of the 539 genes that are upregulated in *myb108* knockout leaves at both timepoints identifies biological processes, Response to Light Stimulus and Photosynthesis as enriched. Such processes are repressed during senescence and is therefore further evidence that senescence is delayed in *MYB108* knockout leaves.

Expression analysis of the *myb2* mutant during dark-induced senescence identified 138 genes differentially expressed when compared to wild-type plants the same time after being placed in darkness, with 27 genes upregulated and 111 genes downregulated. Analysis of the samples taken on the next day identified 1499 differentially regulated genes, with 601 genes upregulated and 898 genes downregulated in the knockout compared to wild-type. To simplify analysis genes that were significantly downregulated at both timepoints were combined to generate a more comprehensive set of downregulated genes. GO terms concerned with stress responses were enriched amongst the genes in the downregulated gene set from timepoint one, suggesting that MYB2 regulates stress related genes as the leaf begins to senesce. GO term Response to Abscisic Acid Stimulus was also enriched and further implicates MYB2 in ABA mediated signalling. As in any stress where *MYB2* expression is enhanced, it is likely that this TF is positively regulating ABA mediated responses. GO terms concerning senescence were also enriched in the downregulated gene set suggesting that MYB2 positively regulates senescence process.

6.2.6.3 Regulatory overlap

Given the large-scale transcriptional effects observed in *myb2* and *myb108* leaves when compared to wild-type behaviour suggests an important role for both MYB2 and MYB108 in both biotic and abiotic stress responses. If a gene is significantly downregulated in a given mutant during Botrytis infection and during dark-induced senescence it is natural to assume that the MYB TF is required to positively regulate its expression under wild-type conditions.

Consideration of all genes that were downregulated in *myb2* leaves in both treatments, the overlap between the two consists of 102 genes. GO analysis revealed that these genes were overrepresented for stress related processes. Interestingly, the overlapping gene set was not enriched for processes describing ABA signalling, even in light of the observation this term is overrepresented when considering in each stress in isolation. This suggests that MYB2 regulates different sets of genes in an ABA dependent manner conditional on the stress. As mentioned previously, the role of ABA during biotic stress is complicated

with negative and positive regulatory affects. The lack of a significant set of ABA related genes may reflect differences in ABA pathways that function in biotic and abiotic stresses.

Within the set of genes that are misregulated during both stresses are some well characterised stress genes. All genes that were covered by the Response to Stress GO category are listed in Table 6.3. Several of these genes are known to function in oxidative stress. The gene *AT2G21640* is a protein of unknown function, but is a known key oxidative stress marker gene, *DOX1* is involved in cellular protection against oxidative stress (Hamberg *et al.*, 1999), and *ATCAD8* encodes an oxidoreductase that is heavily induced in response to treatment with a variety of pathogens (Williamson *et al.*, 1995). *LOX1*, a well characterised lipoxygenase that plays important roles in plant defence and susceptibility to bacterial pathogens (Vellosillo *et al.*, 2007). *PEPR1* is a key receptor kinase that is essential for defence signalling (Yamaguchi *et al.*, 2010). *RAB18* is a well known marker gene for dehydration and cold acclimation (Lång and Palva, 1992). *RD20* is a heavily studied gene that is strongly induced by ABA, salt and drought and functions in the control of stomatal opening (Aubert *et al.*, 2010).

AGI	Gene name	Short name
AT2G15220	Plant basic secretory protein family protein	BSP
AT3G62550	Adenine nucleotide alpha hydrolases-like superfamily protein	N/A
AT2G15170	Plant basic secretory protein (bsp) family protein	BSP
AT2G21640	Encodes a protein of unknown function	N/A
AT3G01420	Plant alpha dioxygenase 1 (dox1)	DOX1
AT1G80920	Translocon at the outer envelope membrane of chloroplasts 12	TOC12
AT3G28740	Cytochrome p450 (cyp81d11)	CYP81D11
AT4G37990	Arabidopsis thaliana cinnamyl-alcohol dehydrogenase 8	ATCAD8
AT5G01600	Arabidopsis thaliana ferretin 1	FER1
AT4G14630	Germin-like protein 9	GLP9
AT1G55020	Arabidopsis lipoxygenase 1	LOX1
AT1G52890	Nac domain containing protein 19	ANAC019
AT2G40000	Ortholog of sugar beet hs1 pro-1 2	HSPRO2
AT1G73080	Pep1 receptor 1	PEPR1
AT5G66400	Responsive to aba 18	RAB18
AT2G33380	Responsive to desiccation 20	RD20
AT5G45890	Senescence-associated gene 12	SAG12
AT5G20150	SPX domain gene 1	SPX1
AT5G43580	Unusual serine protease inhibitor	UPI
AT2G38470	WRKY dna-binding protein 33	WRKY33

Table 6.3: **Stress related genes downregulated in *myb2* following infection by Botrytis and during dark-induced senescence.** Shown are all of the genes that are covered by the GO category, Response to Stress, that are significantly reduced expression in *myb2* leaves during the two stress treatments.

Consideration of all genes that were downregulated in *myb108* leaves in both treatments revealed 77 genes that were significantly downregulated in the knockout in both experiments. GO analysis revealed that these genes were overrepresented for stress related processes including Response to Biotic Stimulus, suggesting that MYB108 positively regulates a set of genes that function in response to biotic stress and abiotic stress, in keeping with the finding that MYB108 regulates the responses to both biotic and abiotic stress (Mengiste *et al.*, 2003). All genes that were covered by the Response to Stress GO category are listed in table 6.4. Among the downregulated stress-associated gene set are several well characterised defence related genes. *ACTIVATED DISEASE RESISTANCE 1* (*ADR1*) is has been demonstrated to be important regulator of defence responses, with *adr1* mutant plants displaying enhanced resistance against virulent pathogens. WRKY33 is a well characterised TF that is known to regulate responses to biotic and abiotic stress, including Botrytis (Zheng *et al.*, 2006) and abiotic stresses (Jiang and Deyholos, 2009).

AGI	Gene name	Short name
AT5G19890	Peroxidase superfamily protein	N/A
AT4G16260	Glycosyl hydrolase superfamily protein	DL4170C
AT1G01720	Nac domain containing protein 2	ANAC002/ATAF1
AT1G09940	Encodes glutamyl-trna reductase	HEMA2
AT1G33560	Activated disease resistance 1	ADR1
AT2G16500	Arginine decarboxylase 1	ADC1
AT5G01600	Ferretin 1	ATFER1
AT1G55020	Arabidopsis lipoxygenase 1	LOX1
AT1G52890	Nac domain containing protein 19	ANAC019
AT3G15500	Nac domain containing protein 55	ANAC055
AT2G42010	Phospholipase d beta 1	PLDBETA1
AT1G14870	Plant cadmium resistance 2	PCR2
AT1G74020	Strictosidine synthase 2	SS2
AT5G07350	Tudor-sn protein 1	TUDOR1
AT2G38470	WRKY dna-binding protein 33	WRKY33
AT5G49520	WRKY dna-binding protein 48	WRKY48

Table 6.4: **Stress related genes downregulated in *myb108* following infection by Botrytis and during dark-induced senescence.** Shown are all of the genes that are covered by the GO category, Response to Stress, that are significantly reduced expression in *myb108* leaves during the two stress treatments.

6.3 Discussion

Exposure to biotic and abiotic stress induces massive change in gene expression. Complex networks of TFs are major determinants of this genetic reprogramming, orchestrating the complex processes that lead to stress tolerance. Protein-DNA interactions underpin transcriptional regulatory networks by allowing TFs to regulate each other through TF-promoter interactions. Even though many TFs are highly similar at the protein level, they are often regulated in different ways, which in turn defines their functionality (Haake *et al.*, 2002; Kim *et al.*, 2005; Ha *et al.*, 2009).

The NAC TFs are a plant specific protein family that has been implicated in the regulation of a diverse array of stresses. Consisting of over 100 members, this large TF family possibly reflects the need to increase regulatory capacity and complexity of stress-associated GRNs, in order to fine tune the response to environmental stress and deal with others. *ANAC019*, *ANAC055* and *ANAC072* encode three highly related NAC TFs and have attracted considerable interest due to their link with the regulation of both biotic and abiotic stress.

In this chapter, using a AD-TF library consisting of approximately 1400 Arabidopsis TFs, a high-throughput Y1H assay was used to generate a detailed map of TF-DNA interactions across the core promoter regions of *ANAC019*, *ANAC055* and *ANAC072*. This analysis identified over 100 regulatory interactions between proteins and regulatory DNA, providing some of the most extensive characterisations of promoter sequences for any gene in Arabidopsis to date, and the first study of binding events that can occur along the promoters of these three important NAC TFs. The finding that many stress-associated TFs are represented amongst the set of positive interactions is intuitive and is likely to be relevant *in vivo*. The interacting TFs are associated with multiple pathways and suggests that *ANAC019*, *ANAC055* and *ANAC072* may act as convergence points for different signalling pathways. The broad range of binding and links to TFs associated with a diverse range of processes may not be surprising, as high-connectivity between networks is a characteristic of GRNs in other organisms (Deplancke *et al.*, 2006).

The Y1H assay revealed that members of a specific clade within the MYB TF family interacts with the promoters of *ANAC019*, *ANAC055* and *ANAC072*. Two members of this group, MYB2 and MYB108 have previously been implicated in the regulation of stress responses (Mengiste *et al.*, 2003; Abe *et al.*, 2003), while the study of expression behaviour showed that they exhibit overlapping expression patterns with all three NAC TFs. The regulatory importance of MYB2 and MYB108 with respect to the NAC

TFs was assessed *in vivo* using knockout lines in conjunction with microarrays with leaf samples that were infected with Botrytis or initiating dark-induced senescence. The result of this analysis suggested that MYB2 and MYB108 do influence the expression of *ANAC019*, *ANAC055* and *ANAC072*, along with many other stress related genes. Based on these observations it can be proposed that MYB2 and MYB108 are key nodes in the transcriptional networks that control senescence and the response to Botrytis in Arabidopsis. MYB2 and MYB108 regulate *ANAC019* and *ANAC055*, which then go on to regulatory roles downstream of the MYB TFs, expanding the networks regulatory capacity. Due to the widespread induction of NAC TFs in many different contexts, it is possible that regulation by MYB2 and MYB108 can be expanded to biotic and abiotic stress in general. These results confirm the potential of the Y1H system to identify biologically meaningful protein-DNA interactions.

It was noticeable that *ANAC072* was unaffected in the knockout lines during dark-induced senescence even when considering the large-scale effect on transcriptome brought about by absence of either MYB TF. Significant downregulation was observed at 30 hpi, but this may be an indirect affect. It is possible that MYB2 and MYB108 regulate *ANAC072* expression during other stress responses, such as osmotic stress or response to bacterial infection, where the expression of either MYB2 and/or MYB108 are also clearly enhanced. It is also possible that either of the MYB TFs are equally effective in regulating *ANAC072* expression as they may be functional homologs.

Many other potentially relevant TF-DNA interactions were identified across the NAC promoter regions tested, even though the regulatory scope and effect on target gene expression of these interacting TFs were not assessed in this study. One particularly promising set of interactions that may be relevant *in vivo* are those observed between TF that are key mediators of hormone signalling pathways. The phytohormone ABA is a central mediator of stress responses in plants whose signal is transduced through multiple different TFs, which interact with *cis*-regulatory elements such as the ABRE (Yoshida *et al.*, 2010). ABF TFs such as ABF3 and ABF4, which were shown to interact with the NAC promoters, are two such TFs that regulate transcription in an ABA-dependent manner. The ABRE is found within the vicinity of all ABF binding events. Another TF associated with ABA signalling is ABI4, an AP2 TF that interacts with GC-rich sequence. This GC-rich sequence, also known as the coupling element, has been shown to act in synergy with ABRE to promote transcription (Narusaka *et al.*, 2003). Interestingly, the relative position and order of these two elements within the promoters of *ANAC019* and *ANAC055* appear to be constrained, suggesting the location of these elements is of functional importance. Furthermore, the position of this pair

of motifs is within approximately 200 bp of the TSS. Previous studies in eukaryotes have demonstrated that this region is more likely to be functional than non-coding DNA positioned further upstream of the TSS (Tabach *et al.*, 2007). Given that both *ANAC019* and *ANAC055* are highly induced by elevated levels of ABA, this arrangement of motifs and the relevant TF interactions pose a plausible regulatory mechanism that should be tested in the future.

ORA59 was another key hormone signalling TF that bound to the promoters of all three NAC TFs. The location of the binding site for ORA59, which is a key integrator of JA and ET signalling pathways (Pré *et al.*, 2008), could be predicted within each promoter fragment with which it displayed an interaction. Again, the likely binding location was proximal to the TSS indicating that this proximity is of functional importance. Interestingly, unlike ABF binding events, which were distributed across the entire promoter regions tested, interactions with ORA59 were exclusively located within 200 bp of the TSS. Putative binding locations for both ABA and JA-ET associated TFs are located in close proximity to each other, and may represent a way to either facilitate protein-protein interactions or cause steric hindrance, which could have either positive or negative effects on transcription. Indeed, the ABA and JA-ET signalling pathways have been shown to interact in an antagonistic fashion (Anderson *et al.*, 2004).

Many other potential regulators of NAC genes can be postulated based on Y1H and expression data. GBF1, which bound to the promoters of both *ANAC019* and *ANAC055*, displays a similar pattern during senescence. Intriguingly, GBF1 has been shown to regulate the onset of senescence in Arabidopsis (Smykowski *et al.*, 2010), making this interaction a good candidate for further study. The interaction between PIF7 and all three NAC promoters is another regulatory event that should be probed further. PIF7 expression is anti-correlated with that of the NAC TFs during both Botrytis and senescence, tying in with the idea that it may be a repressor of stress-induced genes as demonstrated by previous studies with cold stress (Kidokoro *et al.*, 2009).

The CBF TF family was observed to bind the promoter of *ANAC072*. This was a unique binding event compared to the other NAC promoters, and there are additional links to suggest that this is not a coincidence. The CBF TF family are heavily characterised in terms of their key roles in responding to drought and low temperature stress (Sakuma *et al.*, 2002). From the AtGenExpress data, *CBF1-4* and *ANAC072* expression are clearly induced by cold stress. The CBFs appear to be induced rapidly by cold, implying they are early response regulators of the cold response; since *ANAC072* expression is induced at a later timepoint, it could be suggested that the CBFs may

regulate the expression of *ANAC072* during low temperature conditions.

Positive binding events are logical candidates for direct regulation, but the lack of TF binding can also provide potential insight into regulatory pathways. Whole key stress-associated TF families, such as the WRKYs, were not represented among the positive interactions that occurred along NAC promoters. It is possible that a particular constraint imposed by the yeast system has a negative impact on the binding ability of these whole families and highlights a potential problem with the Y1H system. On the other hand, it could reflect a true regulatory mechanism where members of specific families are excluded from interacting with NAC promoters.

Protein-DNA interaction data from Y1H experiments and expression perturbation experiments can be incorporated to generate a NAC centred transcriptional network (Figure 6.18). This network also integrates previously identified regulatory relationships between TFs and stress signalling pathways. ANAC019, ANAC055 and ANAC072 may act as nodes that integrate multiple hormone and stress signalling pathways. TFs that can bind to the promoters of the three NAC genes are associated with a diverse set of biotic and abiotic stress related pathways, while they also appear to act to integrate signals from multiple different hormone pathways. This high degree of connectivity suggests that ANAC019, ANAC055 and ANAC072 are general stress-related transcriptional regulators and indicates that they have a central role in regulating many stress responses. It is possible that these NAC TFs regulate overlapping sets of genes in different contexts, but little is known regarding the genes that are regulated by these TFs. Future efforts utilising techniques such as ChIP-SEQ and inducible overexpressor lines will reveal the network architecture downstream of these NAC TFs.

As discussed in section 5.3, Y1H has some caveats in that not all possible protein-DNA interactions can be identified. This is in part because the current AD-TF library does not have all correct Arabidopsis TF clones present in prey vector pDEST22. There are also technical limitations, such as the inability of the system to identify heterodimeric binding events. This combinatorial binding is a major source of regulatory complexity within eukaryotes and identification of such interactions will be vital for a complete understanding of transcriptional regulation.

Assessing the regulatory effect of TF-DNA binding *in vivo* is a difficult challenge with many associated factors. Indeed, the question as to what constitutes a biologically meaningful regulatory interaction is a source of debate within the field of gene regulation (Walhout, 2011). In this study, microarrays were used to assess the effects of

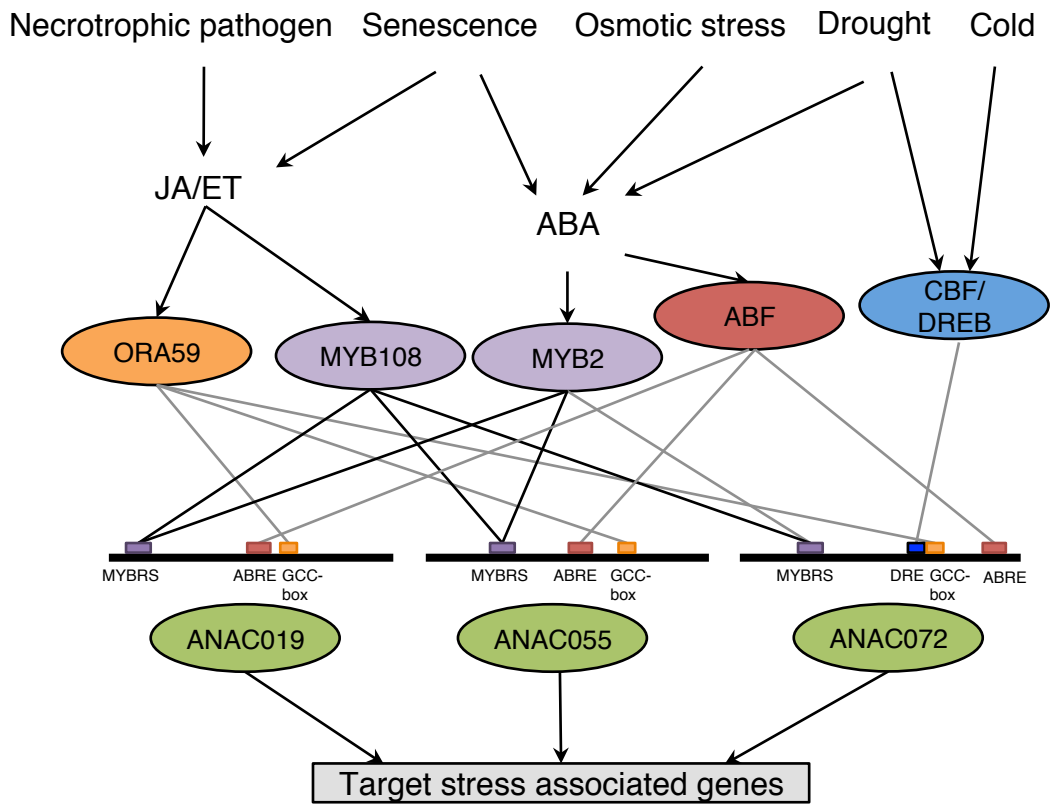


Figure 6.18: **Integrated NAC centred transcriptional network.** By integrating regulatory data from Y1H and expression perturbation experiments together with that curated from the literature it is possible to generate a preliminary NAC centred transcriptional regulatory network that operates under multiple stress conditions. TFs are shown as oval nodes, where colour indicate family membership. Edges with arrows indicate general regulatory association. Grey edges without arrows indicate DNA-binding. Black edges without arrows indicate DNA-binding and a regulatory association. *cis*-regulatory motifs that interact with specific TFs are shown within the promoters of the three NACs (not to scale): MYBRS, MYB recognition site; ABRE, ABA-responsive element; GCC-box; DRE, dehydration response element.

mutations in interacting TFs, MYB2 and MYB108, on the expression of putative target NAC TFs during the response to Botrytis infection and dark-induced senescence. The strength of this approach is that the effect of genetic perturbation is made under relevant environmental conditions, where the NAC TFs are known to be differentially expressed. Knockout lines have been used in similar studies to analyse the regulatory relationship between TFs and their targets based on protein-DNA interactions identified with Y1H (Brady *et al.*, 2011).

In this study, the regulatory effect of only two TFs was analysed; as such, future work should focus on probing the significance of other protein-DNA interactions identified

through the Y1H screen. Judging the significance of these interactions will not be straight-forward however as TFs can interact with genomic DNA and exhibit little regulatory effect, as demonstrated by studies with eukaryotic genomes (Hu *et al.*, 2007; yong Li *et al.*, 2008).

Expression perturbation experiments were able to identify the regulatory effect of MYB2 and MYB108 on expression of *ANAC019*, *ANAC055* and *ANAC072* under multiple different contexts; however it is still desirable to pin down the regulatory effects of MYB2 and MYB108 more cleanly. Inducible overexpression lines have been shown to be extremely useful in the the identification of genes that lie immediately downstream of a TF (Balazadeh *et al.*, 2010a). A similar approach using MYB2 or MYB108 under the control of an inducible promoter could confirm *ANAC019* and *ANAC055* as direct regulatory targets and identify new ones. The use of the chemical translation inhibitor cycloheximide, is another method for elucidating local network structure and could be used to analyse the significance of TF binding *in vivo*.

While the work presented in this chapter has contributed to our understanding of the regulatory code that controls gene expression in Arabidopsis it is important to remember that transcriptional regulation is only part of the regulatory mechanisms that controls stress responses, and it will be important to consider the additional levels of control that contribute towards regulation of gene expression. miRNAs have an important role in regulating gene expression, where they act to tightly control gene usage and finely tune regulatory programs (Flynt and Lai, 2008). A miRNA was shown to be critical for the regulation of *ANAC092* (Kim *et al.*, 2009) and it is unlikely that this is an isolated case in terms of stress signalling. Protein-protein interactions play a key role in regulating the functionality of TFs and therefore gene expression. Recent studies have identified high-connectivity between proteins involved in the regulation of transcription (Dreze *et al.*, 2011). Integrating multiple sources of regulation will expand the systems-level understanding of gene expression operating during response to stress in Arabidopsis.

Chapter 7

General Discussion

Plants respond to stress at both the cellular and molecular level by altering the expression of many genes via complex molecular signalling networks. A subset of these differentially expressed genes perform functional tasks that counteract the effects of stress by manipulating plant physiology or biochemistry. Due to this need for tight control, many of the stress induced genes encode TFs, which function collectively to regulate stress responses. These TFs act in complex networks in which TFs regulate other TFs by binding to *cis*-regulatory DNA present in the promoters of target genes. These sequence motifs form a regulatory code that is hard-wired into the genome and is ultimately responsible for regulating transcription. This thesis presented an investigation into the regulatory codes that are responsible for regulating stress responses in the model plant Arabidopsis.

Using high-quality microarray datasets generated within the PRESTA project it was possible to identify sets of genes that are differentially expressed over time during different stress treatments. These datasets reveal groups of genes that display similar expression patterns and therefore may be coregulated. In order to investigate these mechanisms the APPLES software suite was developed to analyse non-coding DNA that may contain elements of the regulatory code responsible for generating the observed expression patterns. A feature of the approach was the design and implementation of robust objects for the storage and manipulation of biological and theoretical entities that are relevant to sequence analysis. The APPLES design principles mean that the software is highly re-usable, which will facilitate investigation of biological sequences in the future.

The APPLES tools were used to identify regulatory motifs that may function to regulate stress responses in Arabidopsis. A novel motif that may be responsible for interacting with the drought related TF HSF3 was discovered among the promoters of HFS3 regulated genes. The importance of this motif was validated experimentally demonstrating

the potential power of sequence analysis. Using known motifs that describe previously characterised TF binding sites, it was possible to identify motifs that are associated with clusters of coregulated genes identified from time-course experiments that track transcriptional changes over time during senescence and in response to infection by Botrytis. Experimental approaches will be required to confirm the contributions of motifs towards regulating gene expression.

As an alternative approach to computational sequence analysis which highlights regulatory elements, the Y1H assay was used to experimentally identify interactions between TFs and non-coding DNA. The use of a TF library allowed for the testing of the interaction ability of over 1400 Arabidopsis TFs with a given DNA sequence in a single assay. This Y1H system was then used to perform a detailed mapping of protein-DNA interactions that can occur across the core promoters of genes that encode three highly-related stress inducible TFs, ANAC019, ANAC055 and ANAC072. Microarrays were used to assess the regulatory consequence of a subset of these interactions by perturbing the expression of interacting TFs and observing the effect on target gene expression during multiple stresses. This approach confirmed predicted regulatory relationships and therefore enhanced the current understanding of the transcriptional regulatory networks that operate during stress responses in Arabidopsis.

This study has revealed the potential for Y1H in combination with expression and motif analysis to provide multiple insights into our understanding of transcriptional regulation. As has been demonstrated, it is possible to generate binary protein-DNA interaction data, providing candidates that may directly regulate a target gene. These potential regulatory relationships can be analysed in follow up experiments where the context in which any regulatory interaction occurs can also be assessed. The usefulness of the interaction data depends on the sequence-specific binding nature of TFs. The experiments performed in this study demonstrate that for a particular binding event, only a subset of the members of that family demonstrate positive interaction, even though there are many more members of the family present in the AD-TF library. This observation suggests that the Y1H system is able to discriminate between the binding preferences that exist within TF families (e.g., only one clade of MYB TFs could interact with the promoters of *ANAC019*, *ANAC055* and *ANAC072*). Indeed, the power of the Y1H system to discriminate between subtle interaction preferences has been used to identify DNA-binding specificities for the *C. elegans* bHLH TF family (Grove *et al.*, 2009). The Y1H can therefore be used to elucidate the mechanisms of TF function divergence within families and identify targets specific to different family members. Due to this selective binding behaviour it would be natural to use the Y1H system to explore which TFs bind

to the motifs identified as candidates for regulating senescence and response to Botrytis infection.

While progress on the stress GRNs that operate in Arabidopsis has been made in this study, there is still a long way to go until we have a complete understanding of regulatory code of stress-related transcription. However, that is not to say that future studies will not greatly expand our understanding of these regulatory mechanisms, especially in the light of establishment of high-throughput technologies. Next-generation sequencing is one particular technology that is currently revolutionising molecular genetics; the ability to sequence DNA at an unprecedented rate compared to previous sequencing technology lends itself to multiple types of molecular biology techniques that all seek to answer subtly different questions.

Chromatin immunoprecipitation (ChIP) is a method that can identify interactions between a protein and specific regions of DNA *in vivo* (Kuo and Allis, 1999). ChIP can therefore be used to investigate TF functionality and identify regions of DNA that are likely to have a role in regulating gene expression. ChIP of TFs followed by deep-sequencing (ChIP-SEQ) is a method for analysing the genome-wide binding profile of a specific TF (Kaufmann *et al.*, 2010). Coupling ChIP to high-throughput sequencing provides an immensely powerful technique for analysing TF behaviour; not only does it identify direct targets of a particular TF, the isolated DNA can also be used to elucidate the sequence motif recognised by that TF. Combining these data, it is possible to map the GRN operating downstream of a particular TF.

An emerging technology that exploits high-throughput sequencing is DNase-SEQ (Boyle *et al.*, 2008). Treating chromatin with DNase I is a common and long-used technique for the identification of regulatory elements. DNase I preferentially digests DNA that is not bound by proteins, revealing footprints that distinguish between bound and unbound regions. High-throughput sequencing of DNase I hypersensitive sites suggest regions of DNA that are depleted of nucleosomes and are therefore likely to be active regulatory regions. This analysis can be expanded further to identify smaller sites within these hypersensitive regions that are protected from degradation as a result of TF binding (Hesselberth *et al.*, 2009; Boyle *et al.*, 2011). As such, DNase-SEQ can therefore be used used to probe higher-level regulation mediated by nucleosomes as well as lower-level transcriptional regulation mediated by TF and *cis*-regulatory elements.

Another consequence of the increased use of next generation sequencing will be the proliferation of fully sequenced plant genomes. The current number of sequenced plant

genomes is limited; in the future this will provide an excellent opportunity to use the power of evolutionary conservation to perform phlyogenetic footprinting analysis. Once the correct orthologs for stress-associated genes have been identified in multiple different species, the corresponding non-coding sequences can be analysed to identify motifs that are conserved across species. Motifs that are conserved are more likely to be functional as elements that are important for regulating gene expression will have been under selective pressure to remain at those loci within the genome.

Coupling the gene-centred Y1H system with protein-centred approaches that use next generation sequencing should go some way towards unleashing the regulatory code that controls stress responses, with the ultimate goal of exploiting this understanding of gene regulation to generate transgenic crops with improved yields and stress tolerance. As this regulatory code controls complex multigenic traits, it is natural to target these mechanisms. Indeed, there have already been numerous occasions where the manipulation of functionality in a single TF has been observed to have a major affect on plant physiology and behaviour (reviewed in Century *et al.* 2008). Perhaps the most prominent example of manipulating the regulatory code to increase crop productivity is in the domestication of teosinte to maize. A study of these two species identified the *TEOSINTE BRANCHED 1* (*TB1*) gene as the major source of evolutionary divergence (Doebley *et al.*, 1995). *TB1* encodes a TCP TF which represses the growth of lateral branches and is expressed at elevated levels in maize compared to teosinte (Doebley *et al.*, 1997). Further studies pinpointed the mutation to the upstream non-coding region this gene and not the coding sequence, indicating that modification of regulatory elements is responsible for the altered activity of TB1 (Wang *et al.*, 1999). This case study provides a fine example of the potential for manipulation of the regulatory code to result in major changes to aspects of plant biology, and provides assurance that similar small scale genome modifications can drive an increase in plant stress tolerance.

Appendix A

Computing similarity between two PSSMs

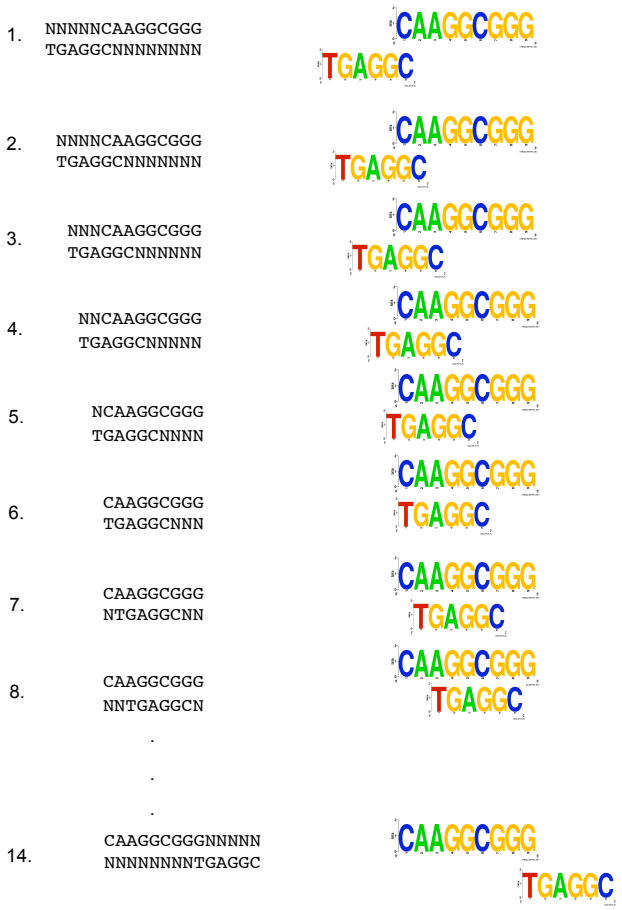


Figure A.1: **Assessing similarity between two PSSMs.** The KL distance can only be computed between an alignment of two PSSMs if they both have the same length. As in this example, where the two PSSMs do not meet this requirement, the ends of each PSSM are expanded with background nucleotide probabilities (N) until they are of the same length (left). The KL distance is then computed at each position between the two PSSMs.

Appendix B

Predicted direct targets of HSF3

Table B.1: Predicted direct targets of HSF3 based on pHSF3E enrichment. Also shown are instances of the motif in each promoter.

AGI	Chromosome	5' position	3' position	pHSF3E instance
AT4G10250	4	6370253	6370242	TCCAGAGACTTC
AT4G27670	4	13820075	13820064	TCGAGAAACTTC
AT1G52560	1	19575962	19575951	ACAAGAAGCATC
AT3G46230	3	16984955	16984966	TCAAGAAGCTTC
AT2G29500	2	12633927	12633938	TCCAGAAAATTC
AT2G29500	2	12633860	12633849	TCGAGAAATTTC
AT5G12020	5	3883047	3883058	TCGAGAAACGTC
AT5G12020	5	3882995	3882984	TCTAGAAATATC
AT3G07150	3	2264681	2264692	ACTAGAACCTTC
AT3G07150	3	2264623	2264612	ACTAGAAAACTC
AT1G53540	1	19980242	19980253	ACCAGAAACTTC
AT5G51440	5	20890701	20890690	TCGAGAAACATC
AT5G51440	5	20891070	20891081	TCTAGAACCATC
AT5G52640	5	21352273	21352262	TCCAGAAAGTTC
AT5G12030	5	3884854	3884865	TCGAGAAACGTC
AT5G12030	5	3884901	3884890	TCTAGAAATATC
AT4G25200	4	12916897	12916908	TCTAGAAACATC
AT4G25200	4	12916935	12916946	TCCAGAACATTC
AT1G03070	1	729795	729806	TCTAGAGACTTC
AT1G03070	1	729868	729879	TCAAGAACGTTC
AT3G12580	3	3993855	3993866	TCGAGAACGTTC
AT3G12580	3	3993956	3993967	TCGAGATGGGTC
AT1G07400	1	2274931	2274920	TCCAGAAAACTC
AT4G23680	4	12337573	12337584	GCTAGAATCTTC
AT5G59720	5	24062412	24062423	CCAAGAAAAATC
AT5G59720	5	24062486	24062475	TCTAAAACCTTC
AT3G09640	3	2956091	2956102	ACTAGAAGCTTC
AT1G54050	1	20180330	20180341	GCTAGAAAGTTC
AT1G54050	1	20180567	20180556	ACGAGATAAATC
AT2G47180	2	19370577	19370566	TCCAGAACTTTC
AT5G12110	5	3914387	3914398	ACTAGAAGCATC

Continued on next page

Table B.1 – *Continued from previous page*

AGI	Chromosome	5' position	3' position	pHSF3E instance
AT5G12110	5	3914362	3914351	TCTAGAGTGTTC
AT1G74310	1	27940254	27940265	ACCAGAGAATTC
AT1G74310	1	27940209	27940198	TCCAGATGAATC
AT1G59860	1	22031339	22031328	TCCAGAAAAGTC
AT1G21550	1	7554002	7553991	TCTAGAAACATC
AT3G10020	3	3091842	3091831	GCTAGAAACTTC
AT1G64200	1	23830150	23830161	ACTAGAACCATC
AT3G08970	3	2737415	2737426	TCGAGAAATTTC
AT3G08970	3	2737485	2737474	TCGAGAAAAATC
AT5G62730	5	25197416	25197427	TCTAGAGTTTTC
AT5G62730	5	25197181	25197192	ACTAGATTCATC
AT5G64510	5	25784380	25784391	TCCAGAACCATC
AT1G75750	1	28442432	28442443	TCTAGACTATTC
AT1G07500	1	2305145	2305156	TCTAGAAACTTA
AT1G07500	1	2305310	2305299	TCCAGTAAATTC
AT1G67360	1	25238243	25238232	TCCAGAAACATC
AT1G56600	1	21207271	21207260	TCTAGACTGTTC
AT1G56600	1	21207049	21207060	ACAAGACTTTTC
AT3G19270	3	6676976	6676987	TACAGAAAGGTC
AT3G19270	3	6676572	6676561	TCCACAATTTTC
AT5G58770	5	23734282	23734293	ACCAGAACCATC
AT2G19310	2	8370467	8370456	TCAAGAAAACTC
AT2G19310	2	8370442	8370453	TCAAGAGAATTC
AT4G17250	4	9670662	9670651	TCCAGAACTCTC
AT5G47830	5	19374389	19374378	GCTAGAATCCTC
AT5G47830	5	19374602	19374613	TCAAAAAAATTC
AT3G28270	3	10537699	10537710	ACAAGACATTTC
AT3G28270	3	10537846	10537857	TCTAGAAAGTGC
AT5G02430	5	526569	526580	TCTGGAAGTTTC
AT5G02430	5	526332	526321	ACAAAAAAAATC
AT3G26450	3	9683441	9683452	GCTAGAAACTTC
AT1G07350	1	2260583	2260594	TGTAGAAACATC
AT1G07350	1	2260150	2260139	TCCCGAAGCTTC
AT1G30220	1	10635601	10635612	ACTAGAAAAATC
AT3G24500	3	8918308	8918297	GCTAGAAAGATC
AT5G36690	5	14417314	14417325	ACAAGACATGTC
AT5G36690	5	14417478	14417467	TCTACAAAAATC
AT1G68620	1	25765693	25765682	TATAGAAAATTC
AT1G68620	1	25765478	25765467	TCTTGAACTTTC
AT5G62020	5	24916077	24916066	TCAAGAAACTTC
AT5G63500	5	25423356	25423367	ACGAGAAGCTTC
AT5G04340	5	1217437	1217426	TCTAGAATATTC
AT2G41090	2	17135434	17135445	ACGAGAAAAATC
AT3G51910	3	19264950	19264961	TCCAGAAGTTTC
AT3G51910	3	19265000	19265011	TCCAGAAGTTTC
AT3G10820	3	3387166	3387155	ACAAGAACTTTC
AT1G62660	1	23199345	23199334	TCTAGAATTTTC
AT4G11660	4	7042469	7042458	TCCAGAAGTTTC
AT1G49560	1	18342471	18342460	GCTAGAAGATTC

Appendix C

Known motif enrichment in clusters of genes derived from senescence and Botrytis time-courses

Table C.1: Known motifs enriched in senescence clusters.

Cluster number	Motif ID	Annotation	Enrichment p-value
1	M00435	PIF3	2.15E-07
1	S000482	SORLIP1AT	5.85E-05
1	S000145	ABREBNNAPA	0.000171818
1	S000272	GBOX10NT	1.26E-05
1	S000291	ABREMOTIFIIIOSRAB16B	0.000150157
1	S000345	HY5AT	6.07E-07
2	M00435	PIF3	0.000114776
2	S000448	WBOXGACAD1A	0.000940171
3	S000414	ABRELATERD1	0.000340425
3	M00435	PIF3	0.00016633
3	S000149	CCA1ATLHCB1	0.000944823
3	S000424	IBOXCORENT	0.00076894
3	S000163	O2F2BE2S1	0.000611398
3	S000291	ABREMOTIFIIIOSRAB16B	0.00010385
3	S000345	HY5AT	6.05E-05
6	S000163	O2F2BE2S1	2.38E-05
6	S000376	AMMORESVDCRNIA1	0.00015593
7	S000054	HSE	0.000916703
9	M00435	PIF3	0.000824493
9	S000089	PR2GCNT	7.52E-05
9	S000381	PIATGAPB	0.000357416
11	S000051	GT1MOTIFPSRBCS	0.000548481
13	S000460	CTRMCAMV35S	0.000364035
15	S000474	SITEIIATCYTC	3.98E-06

Continued on next page

Table C.1 – *Continued from previous page*

Cluster number	Motif ID	Annotation	Enrichment p-value
16	M00435	PIF3	0.000592945
16	S000145	ABREBNNAPA	0.000295532
16	S000408	MYB1AT	0.000123814
16	S000195	RE1ASPHYA3	8.42E-05
16	S000291	ABREMOTIFIIIOSRAB16B	0.000232001
16	S000345	HY5AT	0.000265948
17	M00361	CDC5	0.000989414
17	S000474	SITEIIATCYTC	8.50E-07
17	S000147	NONAMERATH4	8.84E-05
19	S000305	OBP1ATGST6	9.92E-05
21	S000372	GMHDLGMVSPB	7.05E-05
22	S000366	E2FANTRNR	1.44E-06
22	S000031	CELLCYCLESC	0.000272284
23	S000474	SITEIIATCYTC	7.21E-05
23	S000472	UP2ATMSD	6.88E-05
25	S000392	-10PEHVPSBD	0.000190689
26	S000192	ACIPVPAL2	0.000962159
29	M00435	PIF3	0.000277797
29	S000385	EVENINGAT	1.30E-07
29	S000424	IBOXCORENT	0.000148998
29	S000208	BOX1PVCHS15	0.000373798
29	S000345	HY5AT	9.71E-05
33	S000414	ABRELATERD1	0.000108706
33	M00435	PIF3	1.33E-05
33	S000385	EVENINGAT	0.000335454
33	S000081	POLASIG2	0.000326302
33	S000185	SURE2STPAT21	0.000269709
33	S000272	GBOX10NT	0.000627231
33	S000345	HY5AT	0.000822033
33	S000496	WRECSAA01	0.000548173
34	M00435	PIF3	0.000849033
34	M01021	ID1	3.40E-05
34	S000501	CGCGBOXAT	7.60E-05
34	S000282	CE3OSOSEM	0.000910379
35	S000261	CYTOSITECSHPRA	0.000592657
36	S000145	ABREBNNAPA	0.000306688
36	S000380	GBOXSORBCS1	0.00094476
36	S000345	HY5AT	0.000609855
37	M00435	PIF3	0.000533741
37	S000501	CGCGBOXAT	0.000384297
37	S000145	ABREBNNAPA	0.000684069
37	S000063	MARABOX1	0.000784702
37	S000034	CEREGLUBOX3PSLEGA	0.000316404
37	S000071	NONAMERMOTIFTAH3H4	0.000402139
37	S000487	SORLREP2AT	0.000922622
38	S000501	CGCGBOXAT	6.65E-10
38	S000018	ACGTSEED2	0.000877155
38	S000282	CE3OSOSEM	2.34E-05
39	M00359	bZIP911	7.24E-05
39	M00357	bZIP910	0.000843906
39	S000282	CE3OSOSEM	1.28E-05
39	S000301	CONSERVED11NTZMATP1	2.40E-06

Continued on next page

Table C.1 – *Continued from previous page*

Cluster number	Motif ID	Annotation	Enrichment p-value
40	M00949	AGL15	8.65E-06
41	S000414	ABRELATERD1	8.92E-10
41	ANAC019	ANAC019	2.42E-08
41	ANAC092	ANAC092	3.28E-10
42	S000414	ABRELATERD1	6.55E-05
42	M00359	bZIP911	0.000438643
42	S000501	CGCGBOXAT	2.99E-05
42	S000355	ACEATCHS	0.000967674
42	ANAC019	ANAC019	2.13E-10
42	ANAC092	ANAC092	2.48E-09
43	S000501	CGCGBOXAT	0.000329572
43	S000085	INTRONUPPER	2.04E-05
43	S000282	CE3OSOSEM	0.000411697
43	M00146	HSF1	0.000650684
44	S000385	EVENINGAT	0.000217662
44	S000063	MARABOX1	1.25E-05
44	S000064	MARARS	5.10E-05
44	S000067	MARTBOX	7.00E-05
44	ANAC019	ANAC019	7.08E-05
45	M00359	bZIP911	7.61E-05
45	M00435	PIF3	0.000223776
45	S000310	WBBOXPCWRKY1	0.000115943
45	S000081	POLASIG2	0.000126073
45	S000067	MARTBOX	0.000597865
45	ANAC092	ANAC092	2.80E-05
46	S000414	ABRELATERD1	2.53E-05
46	S000292	DPBFCOREDCDC3	7.35E-05
46	M00435	PIF3	0.000143764
46	S000145	ABREBNNAPA	1.84E-05
46	S000117	ABRE2HVA22	1.55E-05
46	S000206	GBOXRELOSAMY3	0.000183317
46	S000135	ABRE3HVA1	0.000796565
46	S000355	ACEATCHS	0.000286894
46	S000291	ABREMOTIFIIIOSRAB16B	1.22E-05
46	S000305	OBP1ATGST6	0.000658306
46	S000345	HY5AT	5.35E-06
47	S000063	MARABOX1	1.56E-07
47	S000135	ABRE3HVA1	0.000418225
47	S000208	BOX1PVCHS15	0.000661018
47	ANAC092	ANAC092	0.000447937
48	S000474	SITEIIATCYTC	0.000314304

Table C.2: Known motifs enriched in Botrytis clusters.

Cluster number	Motif ID	Annotation	Enrichment p-value
1	M00435	PIF3	0.000193079
2	M00441	GBF	3.20E-07
2	S-000345	HY5AT	3.12E-05
2	M00366	EMBP1	3.53E-05

Continued on next page

Table C.2 – *Continued from previous page*

Cluster number	Motif ID	Annotation	Enrichment p-value
2	M00654	OSBZ8	3.90E-05
2	M00367	HBP1A	9.67E-05
2	M00946	TGA1B	0.000102669
2	M00788	EMBP1	0.000303542
2	M00943	TAF1	0.000387611
2	M00442	ABF	0.000717907
4	S-000345	HY5AT	1.60E-08
4	M00442	ABF	2.42E-08
4	S-000291	ABREMOTIFIIIOSRAB16B	4.32E-08
4	M01585	PIF1	3.36E-07
4	M00441	GBF	4.78E-07
4	S-000287	SGBFGMGMAUX28	6.19E-07
4	M01584	HY5	9.19E-07
4	M00367	HBP1A	1.15E-06
4	M00366	EMBP1	2.31E-06
4	M00943	TAF1	2.79E-06
4	M00369	TAF1	7.81E-06
4	M00435	PIF3	1.15E-05
4	M00372	O2	2.00E-05
4	M00434	PIF3	2.86E-05
4	M00946	TGA1B	3.22E-05
4	S-000272	GBOX10NT	3.55E-05
4	M00400	ABF1	6.28E-05
4	S-000425	UPRMOTIFIAT	0.000107384
4	M00356	BZIP910	0.000270909
4	M00358	BZIP911	0.000377296
4	M00944	CPRF3	0.00038741
4	M00443	O2	0.00043399
4	S-000145	ABREBNNAPA	0.000484439
4	S-000163	O2F2BE2S1	0.000583535
4	M01065	ABZ1	0.000723705
4	S-000424	IBOXCORENT	0.000765352
4	M00371	CPRF2	0.00083261
5	S-000385	EVENINGAT	1.74E-10
5	M00788	EMBP1	8.71E-07
5	M00366	EMBP1	3.60E-06
5	M00369	TAF1	3.76E-06
5	M00943	TAF1	6.47E-06
5	M00367	HBP1A	4.48E-05
5	S-000117	ABRE2HVA22	7.64E-05
5	M01585	PIF1	8.04E-05
5	M00434	PIF3	0.000106388
5	S-000287	SGBFGMGMAUX28	0.000407308
5	M01065	ABZ1	0.000438198
5	M01584	HY5	0.000761692
5	M00442	ABF	0.000889975
5	M00441	GBF	0.000970203
6	M01585	PIF1	0.000646982
8	S-000206	GBOXRELOSAMY3	0.000148958
8	M00435	PIF3	0.000189853
8	M00943	TAF1	0.000765907
8	M00367	HBP1A	0.000798903

Continued on next page

Table C.2 – *Continued from previous page*

Cluster number	Motif ID	Annotation	Enrichment p-value
8	M01585	PIF1	0.00087982
8	M00369	TAF1	0.000962738
8	M00400	ABF1	0.001091499
9	S-000291	ABREMOTIFIIIOSRAB16B	1.47E-05
9	S-000424	IBOXCORENT	0.000212489
9	S-000474	SITEIIATCYTC	0.000232704
9	M00441	GBF	0.000367885
11	S-000474	SITEIIATCYTC	3.04E-06
11	S-000440	CDA1ATCAB2	0.000852323
12	M00435	PIF3	1.79E-05
12	M00089	ATHB1	3.60E-05
12	S-000500	SBOXATRBCS	0.00010177
13	S-000474	SITEIIATCYTC	4.15E-09
13	M00952	PCF5	0.000360743
13	S-000408	MYB1AT	0.001049202
14	S-000474	SITEIIATCYTC	3.50E-17
14	M00361	CDC5	0.000586374
15	S-000474	SITEIIATCYTC	1.22E-10
15	S-000225	BOXBPSAS1	1.04E-06
15	S-000472	UP2ATMSD	1.93E-06
15	M00948	PCF2	0.000502181
15	S-000351	GCBP2ZMGAPC4	0.000978286
16	S-000476	E2FCONSENSUS	3.20E-05
16	S-000310	WBBOXPCWRKY1	0.000178731
16	M00442	ABF	0.000384144
16	S-000287	SGBFGMGMAUX28	0.000626894
16	M00372	O2	0.000698445
16	M01187	STF1	0.000925406
16	M00366	EMBP1	0.00103951
16	M00435	PIF3	0.001115116
17	M00948	PCF2	0.000189945
17	S-000474	SITEIIATCYTC	0.000285557
17	M00944	CPRF3	0.000449639
17	S-000425	UPRMOTIFIAT	0.000493149
17	M00367	HBP1A	0.00083075
17	M00946	TGA1B	0.000945142
17	M00356	BZIP910	0.001075156
18	S-000474	SITEIIATCYTC	0.000107488
18	S-000208	BOX1PVCHS15	0.000750329
19	S-000474	SITEIIATCYTC	1.49E-14
19	M01582	AGL9	0.000571924
20	S-000474	SITEIIATCYTC	5.78E-26
20	S-000472	UP2ATMSD	2.47E-06
20	S-000225	BOXBPSAS1	2.72E-05
20	S-000404	CARGATCONSENSUS	0.000174691
20	M00968	ATMYB77	0.000285268
21	S-000474	SITEIIATCYTC	1.22E-12
21	S-000472	UP2ATMSD	2.93E-05
22	S-000474	SITEIIATCYTC	7.79E-08
22	M01057	ERF2	0.000100475
23	M00434	PIF3	0.000920818
23	M00010	O2	0.001009428

Continued on next page

Table C.2 – *Continued from previous page*

Cluster number	Motif ID	Annotation	Enrichment p-value
24	S-000310	WBBOXPCWRKY1	4.44E-07
24	M00735	ZAP1	2.93E-05
24	M00653	OCSBF1	0.000142998
24	M00372	O2	0.000958359
25	S-000063		0.000381571
26	M00735	ZAP1	2.72E-07
26	S-000310	WBBOXPCWRKY1	0.000174976
26	M01052	MYB80	0.000557116
26	S-000385	EVENINGAT	0.000773398
27	M00943	TAF1	6.72E-05
27	M00441	GBF	8.10E-05
27	M00369	TAF1	0.00012426
27	M00434	PIF3	0.000261077
27	S-000063		0.000419578
27	M00660	RITA1	0.000471019
27	M01136	DOF	0.00051886
27	M01584	HY5	0.000563722
27	ANAC092	ANAC092	0.000641
27	M00635	GT1	0.000608362
27	M00367	HBP1A	0.000849279
28	S-000067	MARTBOX	3.63E-06
28	S-000208	BOX1PVCHS15	0.000458367
28	M01136	DOF	0.000599631
28	S-000310	WBBOXPCWRKY1	0.000615639
28	S-000110	TATABOX3	0.00083907
28	M00735	ZAP1	0.001029423
29	M00366	EMBP1	5.17E-06
29	M00788	EMBP1	2.61E-05
29	M00441	GBF	2.71E-05
29	M01585	PIF1	7.26E-05
29	S-000310	WBBOXPCWRKY1	9.24E-05
29	S-000292	DPBFCOREDCDC3	0.000110634
29	M00372	O2	0.000145506
29	M00653	OCSBF1	0.000152635
29	M00442	ABF	0.000167053
29	M00376	TGA1A	0.000177909
29	S-000287	SGBFGMGMAUX28	0.000185831
29	M00660	RITA1	0.000195662
29	M00369	TAF1	0.000225695
29	M00367	HBP1A	0.000251018
29	M01065	ABZ1	0.000277507
29	M00944	CPRF3	0.000469522
29	M00735	ZAP1	0.000605628
29	M00435	PIF3	0.00071353
29	M00371	CPRF2	0.000904083
29	M00943	TAF1	0.000981836
29	M00359	BZIP911	0.000991395
30	M00946	TGA1B	0.000187697
30	M00697	HBP1B	0.000501302
30	M00441	GBF	0.00051849
30	M00367	HBP1A	0.000572465
30	M00788	EMBP1	0.000676315

Continued on next page

Table C.2 – *Continued from previous page*

Cluster number	Motif ID	Annotation	Enrichment p-value
33	M00442	ABF	1.16E-05
33	M00788	EMBP1	2.34E-05
33	M00367	HBP1A	3.64E-05
33	M00366	EMBP1	5.96E-05
33	M01584	HY5	8.92E-05
33	S-000282	CE3OSOSEM	0.000134652
33	M00441	GBF	0.000178049
33	M00653	OCSBF1	0.000367952
33	M00943	TAF1	0.000534585
33	M00946	TGA1B	0.000700757
33	M00369	TAF1	0.000817179
33	M00944	CPRF3	0.000906468
33	M01065	ABZ1	0.000952105
33	M00660	RITA1	0.001084277
34	M00366	EMBP1	7.61E-08
34	M01065	ABZ1	1.92E-07
34	M00946	TGA1B	2.29E-07
34	M00788	EMBP1	2.35E-07
34	M00434	PIF3	5.91E-07
34	M00372	O2	9.24E-07
34	M01584	HY5	7.56E-06
34	M00367	HBP1A	1.11E-05
34	M00943	TAF1	1.15E-05
34	M01156	BZR1	1.37E-05
34	S-000414	ABRELATERD1	2.18E-05
34	M00435	PIF3	2.64E-05
34	S-000117	ABRE2HVA22	2.84E-05
34	M01585	PIF1	4.04E-05
34	M00442	ABF	9.94E-05
34	S-000287	SGBFGMGMAUX28	0.000120555
34	M00660	RITA1	0.000325388
34	M00369	TAF1	0.000391049
34	S-000272	GBOX10NT	0.000470123
34	M00944	CPRF3	0.000897153
34	M00371	CPRF2	0.001108934
36	S-000067	MARTBOX	0.000448014
37	S-000025	AT1BOX	1.52E-05
37	S-000474	SITEIIATCYTC	4.94E-05
38	M00660	RITA1	5.22E-06
38	S-000433	WUSATAg	7.22E-05
38	M00653	OCSBF1	0.000162091
38	M01065	ABZ1	0.00021131
38	S-000063		0.000357574
38	S-000474	SITEIIATCYTC	0.000397552
38	M00943	TAF1	0.000496228
38	M00944	CPRF3	0.000515516
38	M00442	ABF	0.000590536
38	M00372	O2	0.000670745
38	M00371	CPRF2	0.000762263
38	S-000414	ABRELATERD1	0.000775493
39	S-000081		4.64E-05
39	S-000064	MARARS	8.36E-05

Continued on next page

Table C.2 – *Continued from previous page*

Cluster number	Motif ID	Annotation	Enrichment p-value
39	M01194	PDF2	0.000179418
39	M01136	DOF	0.000303054
39	S-000063		0.000428727
39	S-000034	CEREGLUBOX3PSLEGA	0.000463648
40	S-000108	TATABOX1	4.89E-05
40	S-000279	5256BOXLELAT5256	0.000553542
42	M00635	GT1	0.000686705
43	M01136	DOF	0.000171799
43	M01133	AG	0.001005556

Appendix D

List of NAC TFs displaying enhanced expression during senescence or Botrytis infection

Table D.1: NAC TFs that are significantly differently expressed during senescence and Botrytis time-courses.

AGI	Name	Time-course(s)
AT1G01720	ANAC002	Senescence/Botrytis
AT1G02220	ANAC003	Senescence/Botrytis
AT1G34190	ANAC017	Senescence
AT1G52880	ANAC018	Senescence
AT1G52890	ANAC019	Senescence/Botrytis
AT1G56010	ANAC021	Senescence
AT1G61110	ANAC025	Botrytis
AT1G69490	ANAC029	Senescence/Botrytis
AT1G77450	ANAC032	Senescence/Botrytis
AT2G02450	ANAC035	Botrytis
AT2G17040	ANAC036	Botrytis
AT2G24430	ANAC039	Botrytis
AT2G33480	ANAC041	Senescence/Botrytis
AT2G43000	ANAC042	Senescence/Botrytis
AT3G04060	ANAC046	Senescence/Botrytis
AT3G04070	ANAC047	Senescence/Botrytis
AT3G04420	ANAC048	Senescence
AT3G10490	ANAC052	Senescence
AT3G10500	ANAC053	Senescence/Botrytis
AT3G12910	NAC-homolog	Senescence/Botrytis
AT3G15170	ANAC054	Botrytis
AT3G15500	ANAC055	Senescence/Botrytis
AT3G15510	ANAC056	Senescence
AT3G29035	ANAC059	Senescence
AT3G49530	ANAC062	Senescence/Botrytis
AT3G56530	ANAC064	Botrytis

Continued on next page

Table D.1 – *Continued from previous page*

AGI	Name	Time-course(s)
AT4G10350	ANAC070	Senescence/Botrytis
AT4G27410	ANAC072	Senescence/Botrytis
AT4G28530	ANAC074	Senescence
AT4G35580	NTL9	Botrytis
AT5G04410	ANAC078	Senescence
AT5G07680	ANAC080	Senescence
AT5G08790	ANAC081	Botrytis
AT5G13180	ANAC083	Senescence/Botrytis
AT5G14000	ANAC084	Senescence
AT5G18270	ANAC087	Senescence/Botrytis
AT5G22290	ANAC089	Senescence
AT5G22290	ANAC089	Senescence
AT5G39610	ANAC092	Senescence/Botrytis
AT5G61430	ANAC100	Senescence
AT5G63790	ANAC102	Senescence/Botrytis
AT5G64060	ANAC103	Senescence
AT5G66300	ANAC105	Botrytis

Appendix E

Y1H colony growth indicating interactions between AD-TF and NAC gene promoters

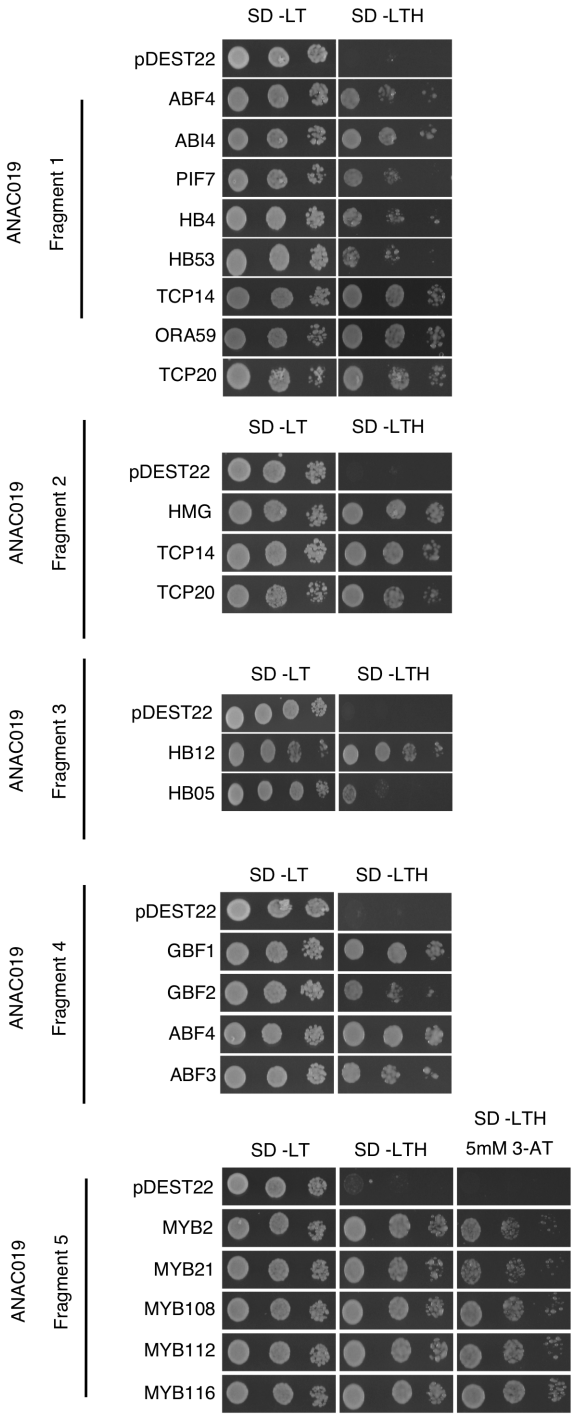


Figure E.1: **Identification of TFs that interact with the promoter fragments of *ANAC019*.** Shown are protein-DNA interactions revealed by Y1H assays. Interactions between promoter fragments and TFs originally identified in a mating based Y1H screen were re-tested by co-transforming yeast strain AH109 with appropriate pairs of bait and prey vectors, followed by growth on media lacking histidine and increasing concentrations of 3-AT. Autoactivation is assessed with control co-transformants containing promoter fragments plus empty pDEST22 vector alone (pDEST22). Protein-DNA interactions are confirmed by growth of promoter-TF co-transformants on more stringent selective media compared to controls.

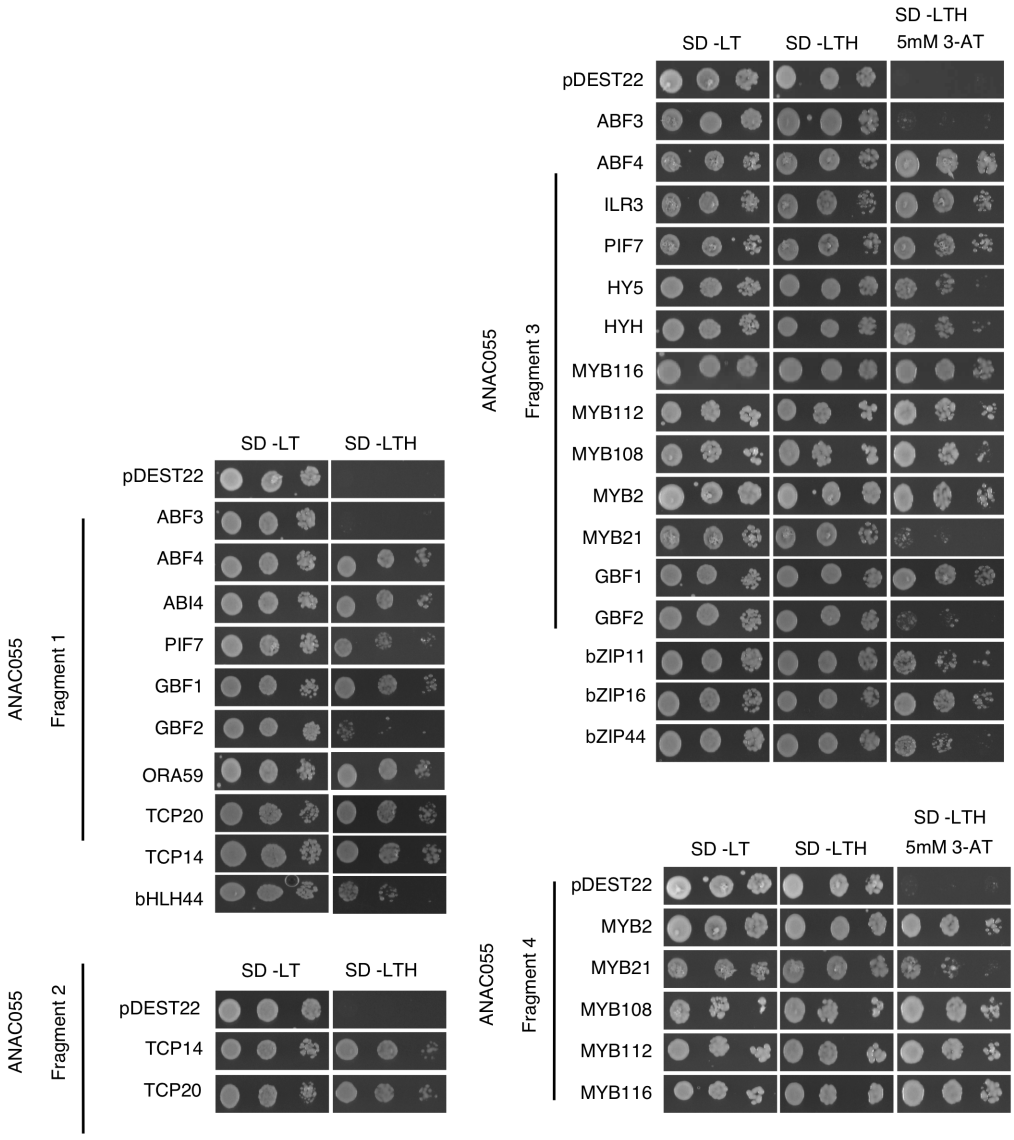


Figure E.2: **Identification of TFs that interact with the promoter fragments of *ANAC055*.** Shown are protein-DNA interactions revealed by Y1H assays. Interactions between promoter fragments and TFs originally identified in a mating based Y1H screen were re-tested by co-transforming yeast strain AH109 with appropriate pairs of bait and prey vectors, followed by growth on media lacking histidine and increasing concentrations of 3-AT. Autoactivation is assessed with control co-transformants containing promoter fragments plus empty pDEST22 vector alone (pDEST22). Protein-DNA interactions are confirmed by growth of promoter-TF co-transformants on more stringent selective media compared to controls.

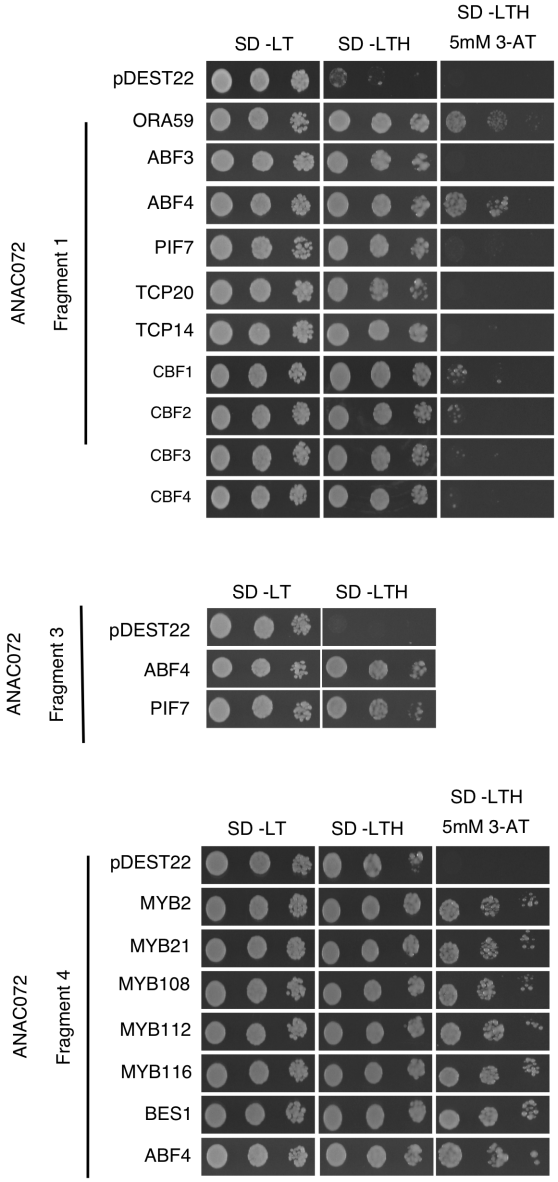


Figure E.3: **Identification of TFs that interact with the promoter fragments of *ANAC072*.** Shown are protein-DNA interactions revealed by Y1H assays. Interactions between promoter fragments and TFs originally identified in a mating based Y1H screen were re-tested by co-transforming yeast strain AH109 with appropriate pairs of bait and prey vectors, followed by growth on media lacking histidine and increasing concentrations of 3-AT. Autoactivation is assessed with control co-transformants containing promoter fragments plus empty pDEST22 vector alone (pDEST22). Protein-DNA interactions are confirmed by growth of promoter-TF co-transformants on more stringent selective media compared to controls.

Appendix F

Publications using content from this thesis.

The Plant Cell, Vol. 23: 873–894, March 2011, www.plantcell.org © 2011 American Society of Plant Biologists

LARGE-SCALE BIOLOGY ARTICLE

High-Resolution Temporal Profiling of Transcripts during *Arabidopsis* Leaf Senescence Reveals a Distinct Chronology of Processes and Regulation [C][W][OA]

Emily Breeze,[a,1] Elizabeth Harrison,[a,1] Stuart McHattie,[a,b,1] Linda Hughes,[a,2] Richard Hickman,[a,b] Claire Hill,[a] Steven Kiddle,[a,b] Youn-sung Kim,[a,3] Christopher A. Penfold,[b] Dafyd Jenkins,[b] Cunjin Zhang,[a] Karl Morris,[a] Carol Jenner,[a] Stephen Jackson,[a] Brian Thomas,[a] Alexandra Tabrett,[a] Roxane Legaie,[b] Jonathan D. Moore,[b] David L. Wild,[b] Sascha Ott,[b] David Rand,[b] Jim Beynon,[a,b] Katherine Denby,[a,b] Andrew Mead,[a] and Vicky Buchanan-Wollaston,[a,b,4]

[a] School of Life Sciences, University of Warwick, Wellesbourne, Warwick CV35 9EF, United Kingdom
[b] Warwick Systems Biology, University of Warwick, Coventry CV4 7AL, United Kingdom

Leaf senescence is an essential developmental process that impacts dramatically on crop yields and involves altered regulation of thousands of genes and many metabolic and signaling pathways, resulting in major changes in the leaf. The regulation of senescence is complex, and although senescence regulatory genes have been characterized, there is little information on how these function in the global control of the process. We used microarray analysis to obtain a high-resolution time-course profile of gene expression during development of a single leaf over a 3-week period to senescence. A complex experimental design approach and a combination of methods were used to extract high-quality replicated data and to identify differentially expressed genes. The multiple time points enable the use of highly informative clustering to reveal distinct time points at which signaling and metabolic pathways change. Analysis of motif enrichment, as well as comparison of transcription factor (TF) families showing altered expression over the time course, identify clear groups of TFs active at different stages of leaf development and senescence. These data enable connection of metabolic processes, signaling pathways, and specific TF activity, which will underpin the development of network models to elucidate the process of senescence.

INTRODUCTION

During leaf senescence, the plant recovers and recycles valuable nutrient components that have been incorporated during growth that would otherwise be lost when the leaf dies or is shed. Efficient senescence is essential to maximize viability in the next season or generation, but premature senescence, a protective mechanism employed when plants are stressed, results in reduced yield and quality of crop plants. During the senescence process, viability of cells within the leaf is actively maintained until maximum remobilization has occurred (Hörtensteiner and Feller, 2002). This requires meticulous control of each step of the process, regulated by internal and external signals via a series of interlinking signaling pathways involving gene expression changes and influenced by the balance of hormones and metabolites. Thus, senescence is a very complex process involving the expression of thousands of genes and many signaling pathways (Buchanan-Wollaston et al., 2005; van der Graaff et al., 2006). Elucidation of the relative influences of each pathway and the crosstalk between them is crucial in identifying the key regulatory genes that control senescence.

To date, genes with a role in leaf senescence have been identified either by forward genetic screening to find mutants with altered senescence rates followed by cloning of the genes involved or by using reverse genetics for functional analysis of genes that show differential expression during senescence (reviewed by Lim et al., 2007). Many of these altered senescence phenotypes occur as a result of altered hormone signaling, such as reduced ethylene signaling (Grbic and Bleecker, 1995) or increased cytokinin signaling (Kim et al., 2006), both of which result in delayed senescence. However, traditional molecular biology approaches in which one gene or mutant at a time is

[1] These authors contributed equally to this work.
[2] Current address: Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, United Kingdom.
[3] Current address: Research and Development Center of GenDocs, 544-1 B-Station Bongmyung-Dong, Yusong-Gu, Daijon, Republic of Korea, 305-301.
[4] Address correspondence to vicky.b-wollaston@warwick.ac.uk.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Vicky Buchanan-Wollaston (vicky.b-wollaston@warwick.ac.uk).
[C]Some figures in this article are displayed in color online but in black and white in the print edition.
[W]Online version contains Web-only data.
[OA]Open Access articles can be viewed online without a subscription.
www.plantcell.org/cgi/doi/10.1105/tpc.111.083345

identified and analyzed have resulted in interesting information but have generally failed to reveal a global picture of senescence regulatory networks, including likely feed-forward, feedback, and crosstalk mechanisms. To understand a system as complex as senescence, where the influence of many external and internal signals is balanced to allow controlled disassociation and dispersal of cellular components, it is essential to study the system in its entirety rather than focus on small parts. The first step in this global analysis is to identify the dynamic changes that are occurring in transcript levels as senescence progresses. Obviously, transcripts are only one part of the regulatory process; factors such as RNA stability, translation rates, protein processing and stability, metabolite concentrations, and many others will have essential roles in the fine-scale moderation of cellular activity. However, transcription plays a key role in regulating both senescence and hormone signaling; therefore, identification of regulatory networks based on transcript levels is an ideal starting point in identifying key switch points in senescence.

Here, we use high-resolution time series microarray data, collected over many time points during the development of the leaf, to identify and characterize the gene expression changes during the different steps that make up the senescence process. The resulting detailed measurement of transcript levels for 22 time points during the developmental process is highly valuable for the investigation of numerous complex processes, such as the discovery of metabolic pathway switches, the identification of key regulatory genes that are active at different time points, and the inference of gene regulatory networks. Analysis of the expression patterns has enabled us to propose a detailed chronology of transcriptional and functional changes during leaf senescence. Promoter motif and transcription factor (TF) analysis has revealed a progression of regulatory genes that influence gene expression at different times during development. Finally, a preliminary model, generated with selected genes from the array data, is presented to illustrate the value of this data set for future network inference.

RESULTS

Growth and Biochemical Changes during Senescence

All measurements in this study were made on samples collected from leaf 7, chosen because senescence and mobilization of constituents from this leaf occur concurrently with flower development and silique filling in our growth conditions and are thus likely to be controlled by developmental signals. Each sample was harvested from an independent *Arabidopsis thaliana* plant at each time point (Figure 1A), and samples were not pooled for any of the analyses. *Arabidopsis* Col-0 plants were grown in controlled conditions until leaf 7 was ~50% of its final size (19 d after sowing [DAS]). This leaf was harvested at defined time points until 39 DAS when it was visibly senescent (~50% of leaf area being yellow, Figure 1B). Samples were taken in the morning (7 h into the light period) and afternoon (14 h into the light period) every other day, resulting in 22 time points in total. Sampling was carried out at these two time points each day to allow us to distinguish genes that are altered in a diurnal rhythm, as well as being differentially expressed over time; the times were selected based on likely maximum changes in expression. Plants started flowering from around 21 DAS. Leaf 7 started to show yellowing at the tip at around 31 DAS and was 25 to 50% yellow by 37 DAS. By the final sample time (39 DAS), the plants were fully flowering, and siliques were filling. Physiological parameters were measured in the morning samples only (i.e., 11 time points). Sampled leaves reached full expansion by 23 DAS (Figure 1C). However, leaf weight increased significantly between time points up to 25 DAS ($P < 0.01$) and continued to increase, reaching a maximum at 31 DAS when the first signs of yellowing were visible and then declined rapidly after 37 DAS ($P < 0.05$). Once the leaf is fully expanded, weight may continue to increase due to synthesis of macromolecules, expansion of organelles, and water uptake. Similarly, loss of fresh weight is primarily due to the decline in macromolecules and the loss of water as the leaf begins to dry.

Protein and chlorophyll levels are often used as markers for the progression of senescence since both are degraded during the senescence process. Levels of total chlorophyll and protein were measured (Figure 1D). Chlorophyll levels did not change significantly until after 31 DAS, when levels started to fall ($P < 0.001$ from maximum). However, relative protein levels started to drop considerably earlier at 23 DAS ($P < 0.05$ from maximal), which is before the time at which maximum leaf weight is reached. This implies that the leaf weight increase seen up to 31 DAS is not due to new protein synthesis but is probably due to increased water content and possibly continuing increases in cell wall density and membrane and other structural developments. Levels of the small and large subunit (SSU and LSU, respectively) of the photosynthetic carbon-fixation enzyme ribulose-1,5-bis-phosphate carboxylase/oxygenase (Rubisco) increased to maximum at 23 DAS (LSU) and 25 DAS (SSU) and then fell steadily during senescence ($P < 0.001$ from maximum; Figure 1E). Rubisco is abundant in a mature green leaf and has been suggested to have some role as a storage protein (Staswick, 1994). Early degradation of this protein may provide building blocks for synthesis of additional proteins required for senescence without affecting the rate of photosynthesis.

Senescence results in activation of signaling pathways involving the stress-related plant hormones salicylic acid (SA), jasmonic acid (JA), and ABA (Weaver et al., 1998; Morris et al., 2000; He et al., 2002). Levels of these three hormones were measured in the leaf 7 developmental time series and showed phased increases during senescence (Figure 2). SA levels were high in immature leaves, gradually decreased to minimal levels at 31 DAS ($P < 0.001$ from initial maximum) and then rose significantly ($P < 0.05$) from a relatively late stage (35 DAS). ABA levels significantly increased earlier at around 31 DAS ($P < 0.05$), with a subsequent increase to maximum at 39 DAS; JA levels showed a complex pattern with peaks at 25, 33, and 39 DAS.

Microarray Analysis over Multiple Time Points Identified Thousands of Differentially Expressed Genes

Four biological replicates for each time point were used for RNA isolation (88 samples in total), each hybridized as four technical replicates to the two channel microarrays. The resulting gene

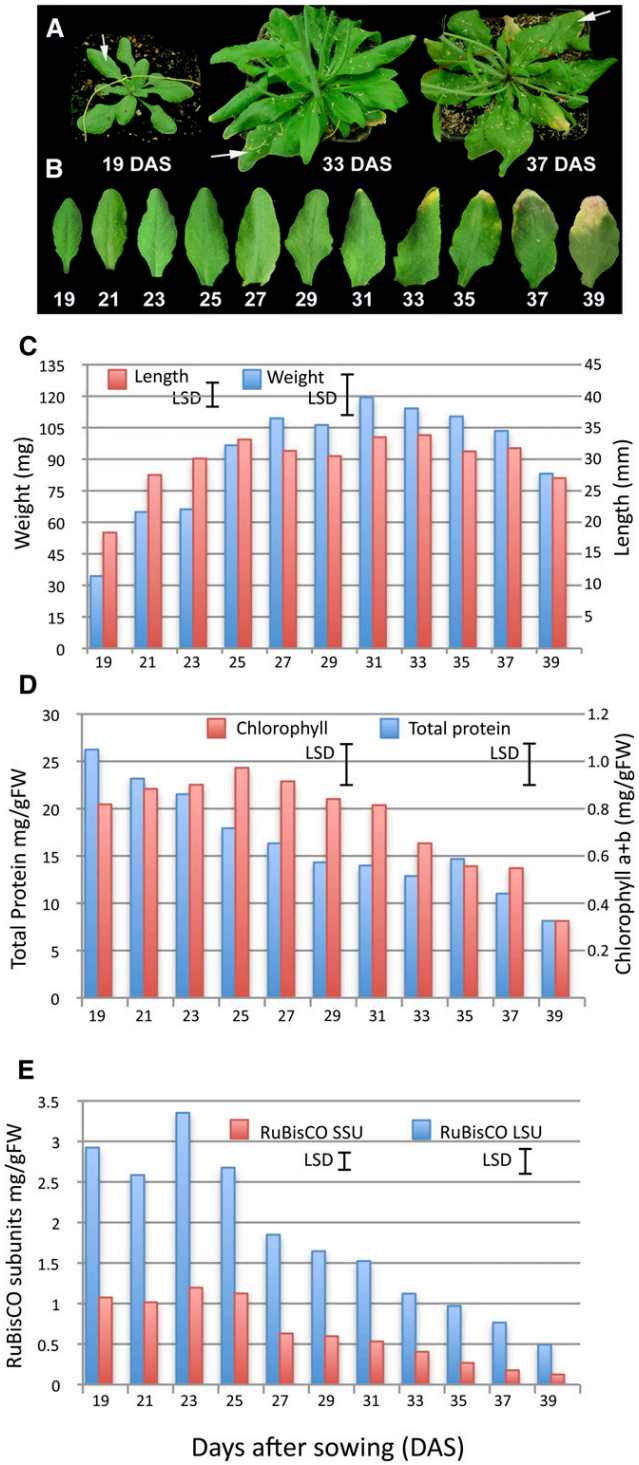


Figure 1. Plant Growth Parameters and Protein and Chlorophyll Measurements.

(A) Appearance of the *Arabidopsis* plants at three different stages of development, 19, 33, and 37 DAS. White arrows indicate leaf 7, the sampled leaf from each plant.

(B) An example of leaf 7 harvested from plants at 19 to 39 DAS (picture shows the morning sample only).

expression profiles were analyzed, and time point means were extracted using a local adaptation of the MAANOVA (MicroArray ANalysis Of VAriance) package, which quality checks and normalizes the data and produces data files containing predicted means for each gene; in essence, a single normalized value for each gene for each biological replicate measured at each time point (Wu et al., 2003; Churchill, 2004). Two different data sets were obtained in this way following the completed analysis: the first contained predicted mean values for each of the four biological replicates at each of the 22 distinct time points, therefore including time-of-day variation, whereas the second contained predicted mean values for eight biological replicates for each of 11 d, the values calculated by omitting the time-of-day and day/time-of-day interaction effects from the fitted model. These data sets were both used in the further analysis described below.

F tests, constructed from the variance estimates obtained from the MAANOVA model-fitting process, were used to assess each gene for significant changes in gene expression between time points. The model fitting allowed separate assessments of the variation due to differences between days (averaged across time-of-day samples), differences between time of day (averaged across days), and the interaction between these terms. Significance levels for all tests were adjusted across genes for multiple testing using a step-down false discovery rate (FDR) controlling procedure (Westfall et al., 1998; Benjamini and Liu, 1999), resulting in 8878 genes showing significant ($P < 0.05$) variation due to day of sampling (19–39 DAS). Additional genes were identified as showing significant ($P < 0.05$) variation due to the time of day or the interaction between day of sampling and the time of day, and the numbers of genes having significant test results for combinations of these terms are summarized in a Venn diagram, together with sample expression profiles for each combination (see Supplemental Figure 1 online). The selection process that was used to identify the list of differentially expressed genes used in all further analyses is described in the Methods section and combined information about the adjusted significance levels of the statistical tests with visual examination of expression patterns. The final list of genes used for the analysis described below contains probes for 6323 genes (see Supplemental Data Set 1 online).

We have generated a web tool that illustrates the expression levels of each individual probe on the Complete *Arabidopsis thaliana* MicroArray (CATMA; Allemeersch et al., 2005) array

(C) Length (mm, red bars) and weight (mg, blue bars) of the sampled leaves over the time course. Least significant differences (LSD; 5%, 71 [length] and 99 [weight] *df*) calculated based on the minimum sample size of 6 (length) and consistent sample size of 10 (weight) for comparing pairs of means is shown for each variable, calculated from the ANOVA.

(D) Total protein (blue bars) and chlorophyll *a+b* (red bars) levels were measured in leaf samples at each stage of development. LSD (5%, 42 [protein] and 43 [chlorophyll] *df*) is shown for both variables. FW, fresh weight.

(E) Levels of the large (LSU) and small (SSU) subunits of Rubisco were estimated from stained polyacrylamide gels. LSD (5%, 42 *df*) is shown for both variables. Values shown in **(D)** and **(E)** represent the means of five independent biological replicates per time point.

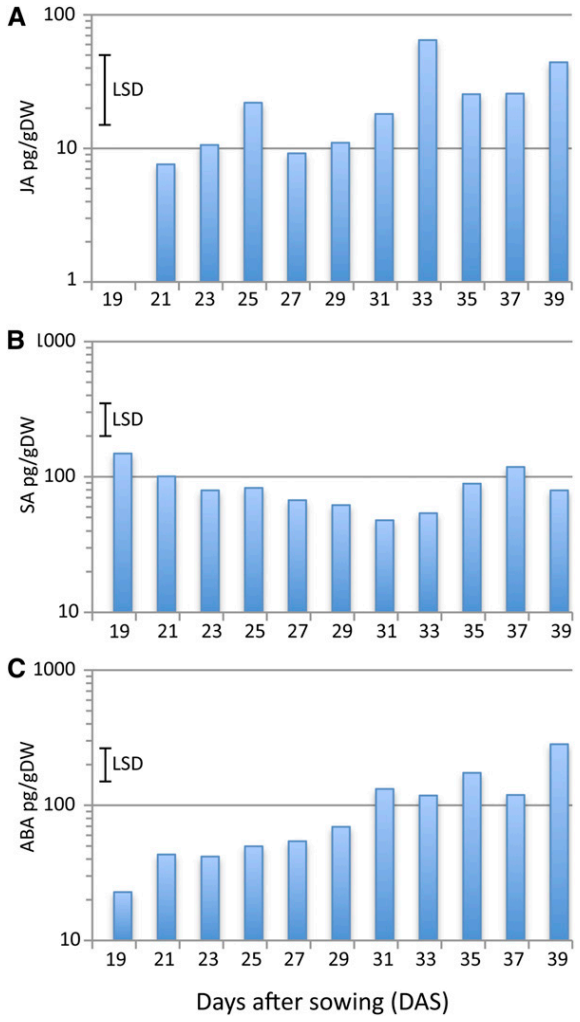


Figure 2. Hormone Levels during Leaf Development.

Levels of JA **(A)**, SA **(B)**, and ABA **(C)** were measured in leaf 7 harvested at different times during plant development. LSD (5%, 41 *df*) for comparisons between pairs of means are shown for each hormone, calculated from the ANOVA of log10-transformed data. Values represent the means of five independent biological replicates per time point. DW, dry weight.

[See online article for color version of this figure.]

using the two alternative summaries of the senescence data. Expression patterns for each gene in the 22– and 11–time point data can be viewed in the Data section at http://go.warwick.ac.uk/presta.

Clustering Genes by Expression Pattern Illustrates the Extensive Metabolic Changes Occurring during Leaf Senescence

The 6323 differentially expressed genes were clustered using the time series-clustering software SplineCluster (Heard et al., 2006). Clustering analysis of both the 22– and 11–time point data was performed, and 48 and 74 clusters were obtained respectively. Supplemental Data Set 1 online shows the cluster number for each differentially expressed gene in both the 11– and 22–time point clusters. For reasons of space, only the 22–time point clusters are analyzed in this article (Figure 3A). The heat map (Figure 3B) indicates that there are changes in gene expression at each time point but that there are several time points at which an obvious step change in the transcriptome occurs. Overall, the major switch in gene expression in leaf 7, both in genes upregulated and downregulated, occurs between 29 and 33 DAS, and the genes identified as differentially expressed can be divided into two major groups, genes in clusters 1 through 24, which are downregulated during this period, and genes in clusters 27 through 48, which are upregulated. Some of the clusters in the center of the heat map show a more complex pattern; for example, cluster 26 genes are downregulated initially and then increase in expression, and genes in clusters 27, 28, and 29 show an initial increase followed by a decrease in expression (Figures 3A and 3B).

A clear diurnal variation in expression is seen with many of the differentially expressed genes, which show higher expression in either the morning or the afternoon samples from the same sample day. Other genes show a distinct morning to night rhythm that did not alter significantly over the 22 d of the experiment. These genes were not selected as being differentially expressed over time but were identified in the MAANOVA analysis as being significantly affected by the time of day term only (1086 genes; see Supplemental Figure 1 online). These genes show clearly that the clock does not decline as late senescence is reached. Morning genes include well-characterized clock genes such as *LATE ELONGATED HYPOCOTYL* (*LHY*), *CIRCADIAN CLOCK ASSOCIATED1* (*CCA1*), and *PSEUDO-RESPONSE REGULATOR 7* (*PRR7;* Harmer, 2009) as well as genes involved in light signaling such as *PHYTOCHROME A* (*PHYA*), *CRYPTOCHROME 1* (*CRY1*), and *PIF4*, a phytochrome-interacting factor (see Supplemental Figure 2A online). Afternoon genes include *EARLY FLOWERING 4* (*ELF4*) and *PHYTOCLOCK1* (*PCL1*), both of which are negatively regulated by *CCA1* and *LHY* (Hazen et al., 2005; Kikis et al., 2005; see Supplemental Figure 2B online).

The 48 clusters of genes identified from SplineCluster analysis of the 22–time point data were analyzed using the gene ontology (GO) enrichment tool BiNGO (Maere et al., 2005). Initially, the two groups of genes showing either decreasing (clusters 1–24) or increasing (clusters 27–48) expression during leaf development were analyzed for overrepresented functions using the GoSlim Plants annotation. The BiNGO-derived graph (Figure 4) illustrates the most highly significant enrichment of specific functions. Downregulated genes are significantly enriched for genes linked to the plastid and thylakoid, and with functions in metabolic processes, particularly photosynthesis and carbohydrate and amino acid metabolism. A more detailed investigation using all GO terms (see Supplemental Table 1 online) shows overrepresentation for genes involved in chloroplast activity such as photosystem (PS) I and II, carbon fixation, chlorophyll (tetrapyrrole) biosynthesis, and amino acid metabolism. All these functions are essential for a growing and active leaf but are downregulated during senescence, when cellular structures are dismantled.

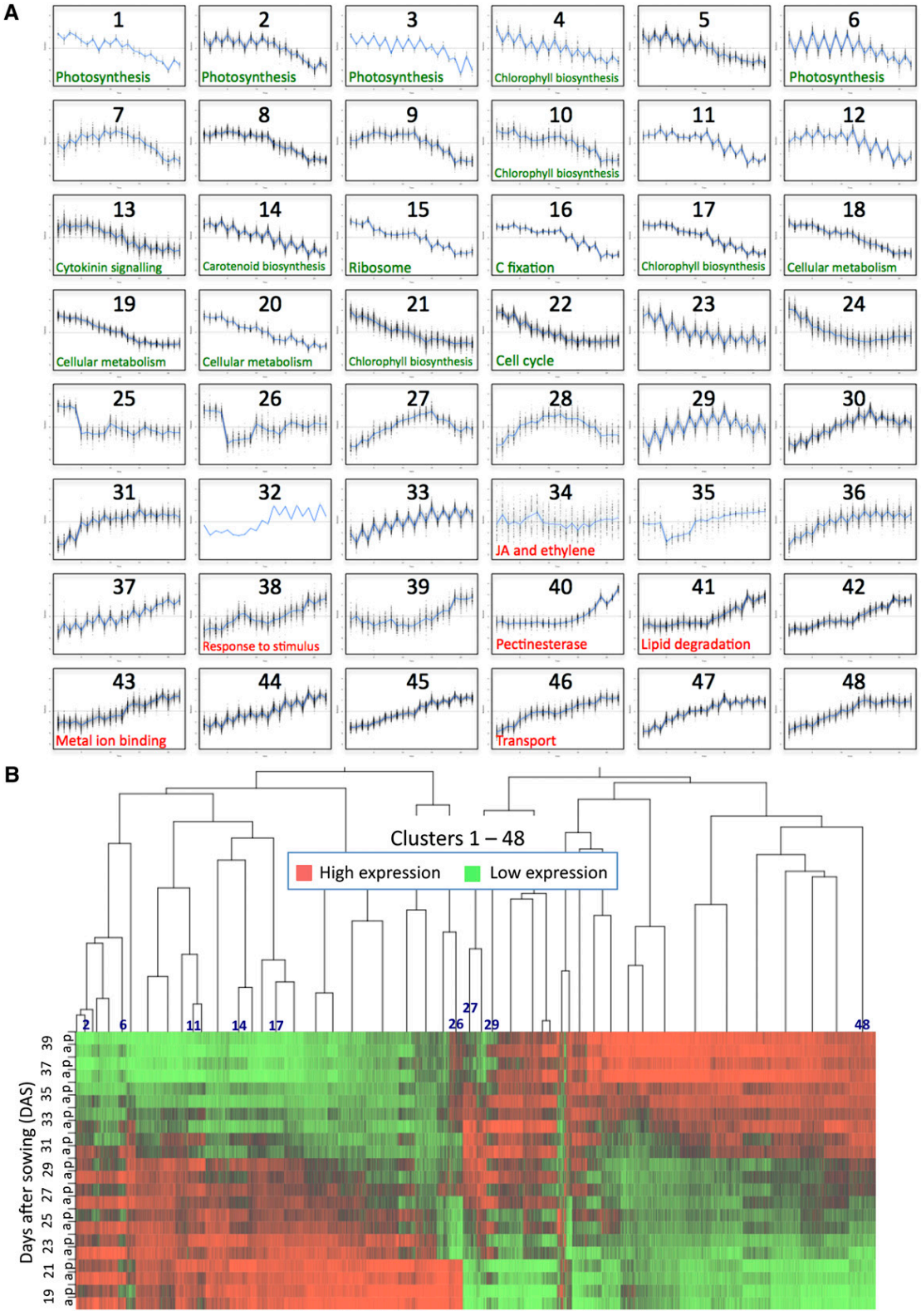


Figure 3. SplineCluster Analysis of Differentially Expressed Genes.

Upregulated genes show a very different picture, illustrating clearly the protective steps the plant takes to respond to the stress generated by the degradative and mobilization functions that occur during senescence (Figure 4). Only two cellular component terms are overrepresented, peroxisome and vacuole. Within the molecular function annotations, only transporter, protein binding, and transcription are overrepresented, whereas there is significant enrichment for stress response and catabolic processes in the biological process terms. In more detail (see Supplemental Table 1 online), enrichment is seen for genes involved in response to stimulus, particularly ABA and ethylene, and many stress responses such as osmotic, salt, and water stress. Enrichment of genes involved in metal ion binding is interesting. Many of these genes (64/222) encode zinc finger (C3H4-type RING) proteins, which may be involved in targeting specific proteins for ubiquitination and degradation. Other zinc binding proteins present have DNA binding activity and may act as TFs. Also, copper chaperones, metallothioneins, calcium binding proteins, and metal ion transporters are represented, which may illustrate the importance of the remobilization of valuable metal ions. Autophagy genes are a significant group; 15 *Arabidopsis* genes involved in autophagy are upregulated during senescence, showing the key role of autophagy in the controlled degradation of cellular components.

This global analysis is highly informative, as it shows broad classes of genes altered in expression during senescence and indicates the processes that are changing. Analysis of enriched GO terms in individual clusters should help to elucidate the chronology of gene expression and associated metabolic activities (Figure 3, see Supplemental Data Set 2 online). Not surprisingly, there was a very strong representation of photosynthesis-related genes in many of the clusters of downregulated genes. Clusters 2, 3, and 6, all of which show a strong diurnal variation with higher morning expression (Figure 3A), are highly enriched with photosynthesis genes, particularly those for the light reaction. Cluster 16, which shows less diurnal change, contains genes encoding enzymes such as Rubisco that are involved in carbon fixation. Clusters 4, 10, 17, and 21 are enriched for chlorophyll biosynthesis genes, and clusters 18 and 19 contain genes involved in cellular biosynthesis such as those for amino acid, polysaccharide, and lipid metabolism. Cluster 15 contains many genes encoding ribosomal proteins. Downregulation of these groups of genes reflects the shutdown of cellular biosynthetic activity as senescence occurs, and the coregulation is an indication of the organized control of this process. Cluster 13 is enriched for cytokinin signaling genes; a reduced level of cytokinin is a key signal that initiates the senescence process (Noodén et al., 1990).

GO terms enriched in the clusters of genes showing increased expression during senescence are less informative than those for the downregulated gene clusters. Certain clusters are enriched for stress-related genes, e.g., genes involved in JA and ethylene signaling are overrepresented in cluster 34. Other clusters are enriched for genes involved with macromolecule degradation, such as clusters 40 and 41 containing genes involved in carbohydrate and lipid degradation, respectively. Metal ion binding, particularly calcium binding, is overrepresented in cluster 43 and transporter genes in cluster 46.

Distinct Pathways Become Active at Different Times during Senescence

Although SplineCluster is useful in identifying groups of genes that are coexpressed and hence may be coregulated across the entire time series, it is not easy to divide these clusters according to their time of differential expression because the overall pattern of expression is the driving factor for cluster membership. To identify an ordering of events, the rate of change of gene expression (gradient) was inferred using Gaussian process (GP) regression applied to the 11–time point data set (described in detail in Supplemental Methods 1 online). Where data are sufficiently time resolved, this method can be used to identify the time points at which the gradient of a gene's expression profile is significantly positive (increased), significantly negative (decreased), or not statistically different from zero (steady), whereas for less resolved data, it will identify times of significant change to the derivative of the global trend. The results are illustrated using the well-characterized, senescence-enhanced gene *SAG12* (Figure 5). Expression profiles (Figure 5A) are used to train a GP model of gene expression (Figure 5B), after which a GP model of the gradient is obtained (Figure 5C) and used to identify whether the gradient at any time is sufficiently far from zero at three different significance thresholds (Figure 5D). A numeric representation of Figure 5C is shown in Figure 5E and suggests that *SAG12* expression first becomes significantly enhanced around 31 DAS. This method can also be used to show when the gene expression gradient is maximal, i.e., the time of most rapid change. For example, the maximum change of the expression for *SAG12* occurs between 33 and 35 DAS (Figure 5E).

After examining the results for a number of genes, a significance stringency of two standard deviations was taken to

Figure 3. (continued).

SplineCluster analysis was performed on the 22–time point data using normalized data for the 6323 differentially expressed gene probes (average of the four biological replicates).

(A) Forty-eight clusters were obtained. The blue line on each plot represents the mean expression profile for the cluster. The genes present in each cluster may be viewed in Supplemental Data Set 1 online. Selected enriched GO terms (data shown in Supplemental Table 1 online) are indicated, in green for downregulated and red for upregulated genes.

(B) The heat map illustrates the expression profiles for genes in each cluster, with red representing high expression and green representing low expression. Morning (a) and afternoon (p) data are shown for 19 to 39 DAS. Darker shades show intermediate levels of expression. A few cluster positions are identified to compare with the cluster profiles shown in **(A)**.

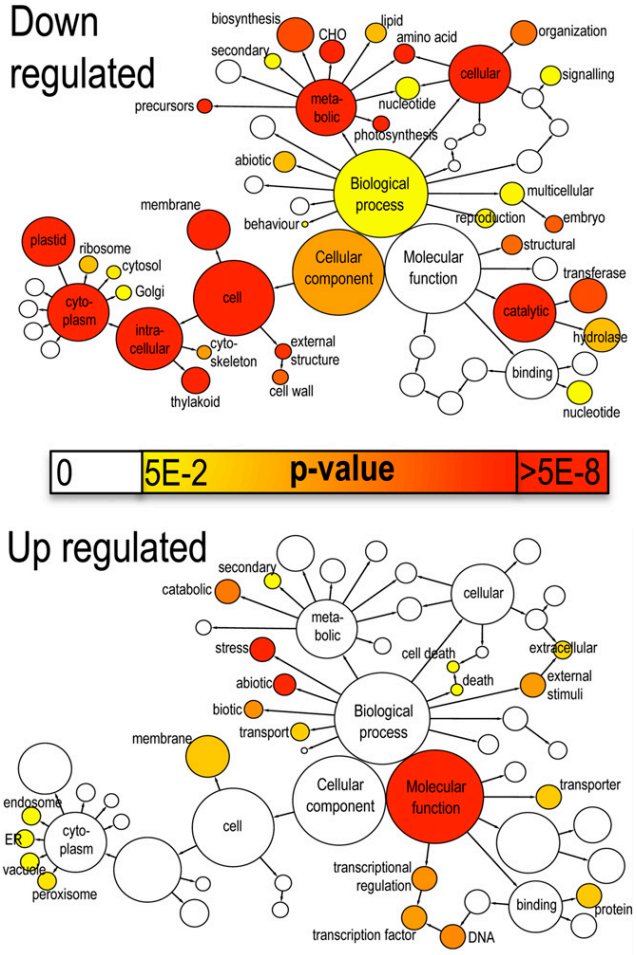


Figure 4. Enriched GO Terms in Genes Upregulated and Downregulated during Senescence.

The network graphs show BiNGO visualization of the overrepresented GO terms for the combined clusters of genes either downregulated (clusters 1–24, 2849 genes) or upregulated (clusters 27–48, 3292 genes) during senescence. Categories in GoSlimPlants (Maere et al., 2005) were used to simplify this analysis and the same nodes are shown on both graphs. Uncolored nodes are not overrepresented, but they may be the parents of overrepresented terms. Colored nodes represent GO terms that are significantly overrepresented (Benjamini and Hochberg corrected P value < 0.05), with the shade indicating significance as shown in the color bar. A more detailed analysis of the GO categories is shown in Supplemental Table 1 online. ER, endoplasmic reticulum.

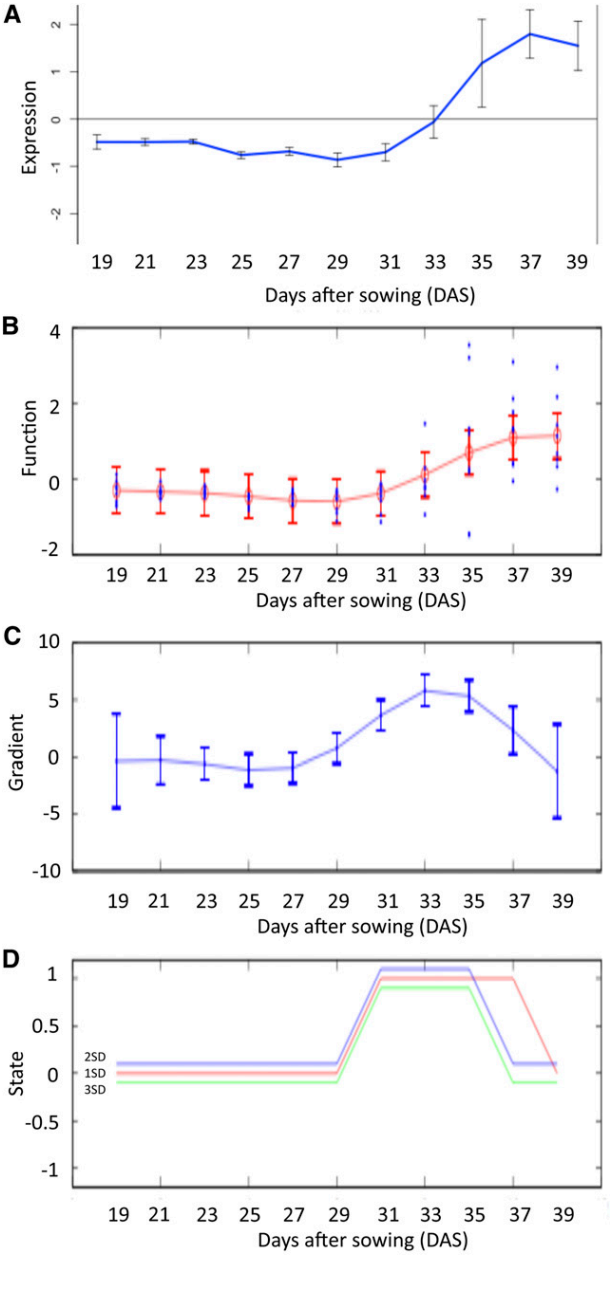
represent a sufficient distance from zero and was used to generate discrete representations of the state of a gene (Figure 5D) for the 6323 differentially expressed genes. The resulting data were then sorted according to the time of first differential expression to identify 19 clusters (see Supplemental Data Set 3 online). GO term enrichment within the 19 clusters or subsets provided more clarity on the cellular and metabolic activities, showing step changes at each time point during the experiment than was gained from the cluster analysis described above (Figure 6, see Supplemental Data Set 4 online).

Downregulated Gene Clusters Show Step Changes in Cellular Dismantling

It is clear that there are progressive changes in genes being downregulated as senescence progresses, and these are highly informative in indicating changes in metabolic pathways. Genes downregulated from the first time point (19 DAS, cluster D1; see Supplemental Data Sets 3 and 4 online), before the leaf is fully expanded, are enriched for genes involved in amino acid metabolism, including those for biosynthesis of Arg, Trp, Lys, and Gln. Genes involved in tRNA aminoacylation and over 30 ribosomal protein genes are downregulated at 21 DAS (cluster D2), indicating that expansion of the ribosomal content of the cells has slowed down. This suggests that large-scale de novo protein synthesis has ceased and that leaf cells are fully developed and equipped for activity.

Many tetrapyrrole or chlorophyll biosynthesis genes are first downregulated at 23 DAS (cluster D3), including the two genes encoding HEMA (glutamyl- tRNA reductase), which catalyzes the rate-limiting and first committed step in tetrapyrrole biosynthesis, and two genes encoding the D subunit of Mg-chelatase, part of the enzyme that diverts the tetrapyrrole pathway toward chlorophyll biosynthesis (Tanaka and Tanaka, 2007). Thus, the requirement for de novo chlorophyll biosynthesis appears to cease at 23 DAS, indicating that all chloroplasts are fully developed. Three genes involved in a branch of the carotenoid biosynthesis pathway (*LUT1*, *2*, and *5*) show a correlated drop in expression at this stage. These genes encode enzymes in the pathway leading from trans-lycopene via α-carotene to lutein, the major carotenoid component in the leaf with an important role in light-harvesting complex-II structure and function and in photoprotection (Kim and DellaPenna, 2006). In addition, expression of three cytokinin-inducible transcription repressors (response regulators *ARR4*, *6*, and *7)* that mediate a negative feedback loop in cytokinin signaling (Hwang and Sheen, 2001) also drops at this time point.

At the next stage (25 DAS, cluster D4), there is significant overrepresentation of genes involved in fixation of carbon dioxide or the Calvin cycle, including two Rubisco small subunit genes and sedoheptulose bisphosphatase, a key enzyme involved in the regeneration of the CO_2 acceptor molecule, ribulose-1,5-bisphosphate. The reduction in expression of the two Rubisco small subunit genes correlates with the reduction in protein levels shown in Figure 1E and indicates that photosynthetic activity probably starts to drop at this stage. At 27 DAS (cluster D5), expression of genes involved in Gly metabolism declines, including Gly decarboxylase and Ser trans hydroxymethyl transferase 1, both involved in photorespiration, which presumably is also less important as photosynthesis becomes less active. Interestingly, five genes designated as *HIGH CHLOROPHYLL FLUORESCENCE PHENOTYPE* (*HCF101*, *109*, *152*, *173*, and *208*) are downregulated together at 27 DAS. Several such genes have been shown to have a role in maintaining the stability of chloroplast-encoded transcripts (Meurer et al., 1996; Meierhoff et al., 2003), and it may be that reduced expression of these genes enables enhanced degradation of photosynthetically related transcripts in the chloroplast. Finally, gene clusters that show expression that declines at 29, 31, and 33 DAS



E

DAS	19	21	23	25	27	29	31	33	35	37	39
1 SD	0	0	0	0	0	0	1	1	1	1	0
2 SD	0	0	0	0	0	0	1	1	1	0	0
3 SD	0	0	0	0	0	0	1	1	1	0	0
Gradient	-0.35	-0.26	-0.60	-1.12	-0.94	0.77	3.65	5.81	5.34	2.32	-1.23

Figure 5. Gradient Analysis of *SAG12* to Identify the Time of First Significant Change in Expression.

(A) Expression levels from microarray data (output from Gene Viewer). The blue line shows the mean of the eight replicates (n = 8; error bars = standard deviation, SD).

(B) Expression levels used by the GP regression, showing the eight biological replicates and the 95% confidence interval.

(C) GP model of the gradient showing 95% confidence interval.

(D) Change in gradient measured at each time point, shown at three different significance values: 3 SD, 2 SD, and 1 SD. Positive value of 1 shows an increased expression, 0 shows no significant change in expression, and −1 shows a significant decrease in expression.

(E) Data output for each significance value, i.e., whether gradient is significantly positive or steady at each time point, with the actual gradient value shown below.

[See online article for color version of this figure.]

(clusters D6, D7, and D8) are significantly overrepresented for photosynthesis-related genes. Cluster D7 includes the gene-encoding TF GOLDEN2-LIKE (GLK2) that, together with its functional homolog GLK1, has been shown to coordinate expression of the photosynthesis apparatus genes in *Arabidopsis* (Waters et al., 2009). Inducible expression of *GLK2* resulted in significantly increased expression of many photosynthesis-related genes (Waters et al., 2009), including those for the PSII chlorophyll binding proteins LHCB 2.2, 4.2, and 6 that are found in the same cluster of downregulated genes as *GLK2*, together with many others encoding subunits of the PSI and PSII complexes. The observation that expression of many photosynthesis-related genes is maintained until this late stage of development implies that there must be a continued requirement of these transcripts to retain sufficient energy production for the senescence process to occur.

Upregulated Gene Clusters Illustrate the Complexity of the Senescence Process and Reveal Novel Groups of Coregulated Genes

Genes that show increased expression at different time points during senescence were divided into clusters based on the time of first significant increase, but these clusters were also subdivided further depending on the subsequent expression patterns (see Supplemental Data Sets 3 and 4 online). This separation revealed additional enriched GO terms, as shown in Supplemental Data Set 4 online and Figure 6.

Many autophagy-related (*ATG*) genes are enhanced in expression from the start of the experiment (cluster U1), indicating that there may be a role for these proteins even before the leaf is fully expanded. Autophagy has a key role in the senescence process, and accelerated senescence has been observed in a number of autophagy-defective mutants (Doelling et al., 2002; Hanaoka et al., 2002; Yoshimoto et al., 2004). Nine of the 15 upregulated autophagy genes show increased expression from the first time point, with five others upregulated at 21 or 23 DAS and one, *ATG7,* being upregulated at 29 DAS. Investigation of the overall expression patterns of the autophagy genes shows four genes, *ATG7*, *ATG8H*, *ATG8A*, and *ATG8B* that show correlated and rapidly increased expression between 29 and 31 DAS (see Supplemental Figure 3A online). In yeast, ATG7p has been shown to be required for activation of ATG8p to allow conjugation with phosphatidylethanolamine (Ichimura et al., 2000), and the resulting ATG8p-phosphatidylethanolamine conjugates

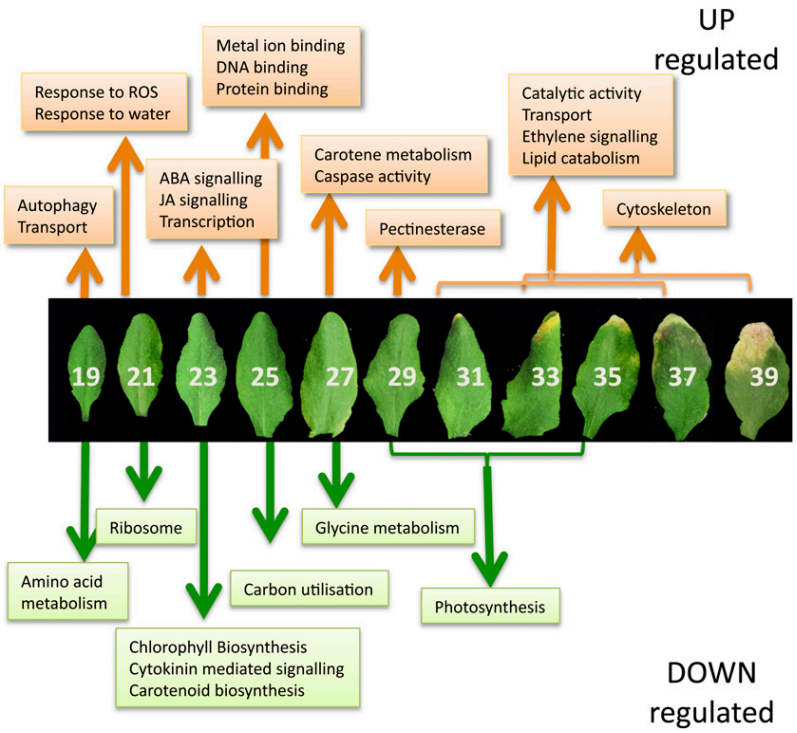


Figure 6. Metabolic Processes Initiated or Repressed at Different Time Points during Development.

Enriched GO terms were identified using BiNGO (Maere et al., 2005) in groups of genes that show first significant upregulation or downregulation at each time point during leaf development and senescence. ROS, reactive oxygen species; white numbers indicate DAS.

assist in the formation of the autophagosome. Thus, the timing of expression of *ATG7* (around 29 DAS) may be the key control point for autophagy activation in senescing leaf cells.

Genes induced at 21 DAS (cluster U2) are enriched for response to oxidative stress. These include TFs such as *DREB2A*, a key regulator in drought and heat stress responses (Sakuma et al., 2006), and *LSD1*, a zinc finger that monitors superoxide levels and regulates cell death (Epple et al., 2003). Increased expression of the mitogen-activated protein kinase MPK7, which is also induced by hydrogen peroxide treatment and enhances plant defense responses (Dóczi et al., 2007), and heat shock proteins such as HSP70, a stress-enhanced heat shock chaperone with a protective role, also indicates that the plant is protecting itself from the detrimental effects of oxidative stress caused by the early stages of senescence. Genes involved in response to water deprivation are also overrepresented; *DREB2A* described above, *Arabidopsis HISTIDINE KINASE3* (*AHK3*), a stress-responsive gene that has been shown to influence cytokinin control of leaf longevity (Kim et al., 2006), dehydration-responsive genes such as *ERD1* and *ERD14*, and *RAB18*, *ABF2*, and other ABA-responsive genes.

Genes involved in both JA and ABA responses are clearly overrepresented in those whose expression increases at 23 DAS (cluster U3). Many of these show a peak of activity at this time point followed by a drop, and this subset is highly enriched for JA biosynthetic genes (cluster U3_1). JA-related genes upregulated at this time point include genes required for JA biosynthesis such as two lipoxygenases, two allene oxide cyclase genes, *AOC1* and *AOC4*, and 12-oxophytodienoate reductase. This increase correlates with a peak in levels of JA detected at 25 DAS (Figure 2). Also, genes implicated in controlling JA responses are upregulated, including the TF *MYC2*, and jasmonate ZIM-Domain genes, *JAZ1*, *JAZ6*,and *JAZ8*. JAZ proteins are repressors of JA signaling, binding to MYC2 and preventing its action (reviewed in Staswick, 2008). MYC2 is also involved in expression of ABA response genes, and this may be the cause of the increased expression of ABA-related genes at this time point. ABA levels only show a large increase later in senescence (Figure 2), but several ABA-signaling genes (e.g., *ABI1* and *AFP1*) and dehydration response genes whose expression is induced by ABA (e.g., *RD20* and *RD26*; Fujita et al., 2004; Choudhury and Lahiri, 2011) are upregulated at 25 DAS. This suggests a potential coordination of JA and ABA responses at this early stage of senescence.

Many genes encoding TFs are first upregulated at 23 DAS, including four WRKY factors, eight NAC domain (for *Petunia hybrida* NAM and for *Arabidopsis* ATAF1, ATAF2, and CUC2) proteins, 10 zinc finger proteins, and the Nuclear Factor Y subunit NF-YA4, which has been implicated in regulating endoplasmic reticulum stress (Liu and Howell, 2010). Many of these genes show an increased expression followed by a fall in expression later in senescence (cluster U3_3), whereas others

show a continued increase from this time until later in senescence (see Supplemental Data Set 3 online). These are likely candidates for the control of later senescence-related processes.

At 25 DAS, the cluster of genes upregulated (cluster U4) is enriched for metal ion binding proteins, including many genes encoding DNA binding proteins, TFs, calcium-signaling genes, etc. The subgroup of this cluster that shows a pattern of increased followed by decreased expression (cluster U4_1) is enriched for genes with a protein binding function. There are five C3HC4-type RING finger protein binding genes in this group, which presumably have a role in regulating specific protein levels via the ubiquitination pathway.

At 27 DAS, there is an interesting overrepresentation of genes involved in carotene metabolism (cluster U5). The three genes involved are a β-carotene hydroxylase and two carotenoid cleavage dioxygenase genes, *CCD7* and *CCD8.* Carotenoids are precursors of signaling molecules that regulate shoot branching in *Arabidopsis*, and *CCD7* and *CCD8* mutants, *max3* and *max4*, respectively, show increased lateral branching (Ongaro and Leyser, 2008). These genes are involved in the production of a strigolactone-related signaling molecule (Gomez-Roldan et al., 2008). Interestingly, another gene that has a shoot-branching role, *MAX2*, was originally identified as *ORE9*, encoding an F box leucine-rich repeat protein required for normal leaf senescence (Woo et al., 2001). Mutants in *MAX2* show increased branching, indicating that this protein is a regulator of the strigolactone signal. The *ORE9*/*MAX2* gene also shows senescence-enhanced expression and is upregulated at 21 DAS. Enhanced expression of all three genes that regulate shoot branching in a senescing leaf and the fact that a mutant in *ORE9*/*MAX2* shows delayed senescence indicate that there may be a role for the novel strigolactone-like hormone in regulating an aspect of leaf senescence.

Caspase activity is also an enriched GO term in genes upregulated at 27 DAS due to increased expression of two of the nine *Arabidopsis* genes encoding potential caspase counterparts, metacaspases *MC6* and *MC9*. The other seven *Arabidopsis* metacaspase genes do not show differential expression during senescence. In other organisms, caspases play an essential role in controlling and executing programmed cell death (PCD), and two metacaspase genes, *MC1* and *MC2*, have been shown recently to control pathogen-induced PCD in *Arabidopsis* (Coll et al., 2010). Links between senescence and plant PCD are tenuous, but the coregulated expression of these two caspase-like genes at this time point may indicate that they have a role in the degradative processes and/or cell death that occur in leaf senescence. The autophagy gene *ATG7* also shows initial enhanced expression at this time point as described above, and this may be a significant link showing that, under our growth conditions, the first degradative processes of senescence are initiated at 27 DAS.

At the next time point (29 DAS, cluster U6), senescence-related degradation processes are shown by enrichment in genes for cell wall degradation. Six upregulated genes encode pectinesterase, involved in the degradation of plant cell wall pectin components. These enzymes and others such as xylosidase, glucosyl hydrolase, β-glucosidase, pectate lyase, and pectin methylesterase inhibitor, may have a role in controlling the degradation of cell wall components and releasing sugars for respiration (Lee et al., 2007). Similarly, genes upregulated at 31 DAS (cluster U7) are highly enriched for catalytic activity, reflecting the considerable degradation that is underway. These include additional carbohydrate-degrading enzymes such as pectinesterases, glycosyl and glucosyl transferases, and polygalacturonase, and several proteases including the well-characterized senescence-enhanced Cys protease *SAG12*, which may have a role in chloroplast degradation (Martínez et al., 2008). Two upregulated genes, *LACS6*, encoding a long-chain acyl-CoA synthetase (Shockey et al., 2002), and *ACX1* (acyl-CoA oxidase), encoding the enzyme that catalyzes the first step in fatty acid β-oxidation in the peroxisome (Fulda et al., 2002), are involved in mobilizing membrane lipids via β-oxidation, likely to provide an energy source to fuel the senescence process. This could act as an initiator of precursors for jasmonate biosynthesis because levels of this hormone increase after this time point (Figure 2).

By the later time points in this experiment, after 31 DAS, the senescing leaf becomes more and more heterogeneous, with some cells within the leaf being at a more advanced stage of senescence than others and more variability between biological replicates. This means that there is less clarity in the functions of different groupings of genes that are differentially expressed at each time point after 31 DAS. GO term enrichment analysis of the 31, 33, and 35 DAS groups combined (clusters U7, U8, and U9) illustrates the degradation and mobilization of nutrients, with 44% of these genes involved in catalytic activity, with lipid catabolism highly represented, and 10% involved in transport. Response to chemical stimulus is also high, with two of the three *Arabidopsis* genes annotated detection of ethylene stimulus, i.e., ethylene *ETR1* and *ACC OXIDASE2* (*ACO2*) being upregulated late in senescence, indicating that the ethylene regulation of senescence may have a significant role at this time.

A surprising group of genes identified by this analysis is downregulated for most of the time course followed by a significant increase in expression at 35 or 37 DAS (clusters U8_1 and U9). This group is highly enriched for genes involved in the cytoskeleton (see Supplemental Figure 3B online), with members of the α-tubulin family (*TUA2*, *4*, and *5)*, actin genes, *ACT3* and *ACT11*, and two aurora genes (*AURORA1* and *AURORA2)* encoding kinase proteins that have a role in histone phosphorylation and have been reported to be associated with microtubule spindles and abundantly transcribed only in dividing cells (Demidov et al., 2005, 2009). Downregulation of this group of genes after completion of the cell division and expansion stages of leaf development is to be expected, but the increase in expression at the end of senescence is unanticipated. AURORA1 has been shown to phosphorylate histone H3 at Ser10 (Demidov et al., 2009), and, in mammalian cells, this modification has been suggested to have a crucial role in transcription and apoptosis as well as in cell division (Prigent and Dimitrov, 2003). These proteins may alter chromatin structure in late senescence to allow DNA fragmentation and eventual degradation. Histone modification and chromatin restructuring is a key regulator in *Arabidopsis* stress responses (Kim et al., 2010), and H3 phosphorylation increased in response to salinity, osmotic stress, and ABA treatment of cultured cells (Sokol et al., 2007).

The increased expression of actin and tubulin genes late in senescence could reflect an autophagy role (Monastyrska et al., 2009). Evidence from yeast and mammalian systems indicates that efficient autophagy requires microtubule action to facilitate autophagosome movement, and actin microfilaments have a role in selective types of autophagy in yeast. The *Arabidopsis* autophagy *ATG8* gene family shows significant homology to mammalian microtubule binding proteins and bind to microtubules in vitro (Ketelaar et al., 2004). Therefore, plant autophagy may involve the action of microtubules and microfilaments, explaining the increased expression of these genes late in senescence when the autophagic degradation of cellular compounds is active.

Chlorophyll degradation is a key step in the senescence process, and several of the genes involved are under transcriptional control (Hörtensteiner, 2009). The *STAYGREEN* gene, *SGR1* regulates the first step in the dismantling of chlorophyll from the chlorophyll binding proteins. Key genes involved in chlorophyll degradation, *SGR1, SGR2*, *NYC1* (chlorophyll *b* reductase), and *PaO* (pheophorbide a oxygenase), all show enhanced expression during senescence. All of these genes increase in expression during early time points, leveling out between 25 and 29 DAS, followed by a sudden increase in expression after 29 DAS (see Supplemental Figure 3C online). It is likely that it is the expression of *SGR* after 29 DAS that initiates the dismantling of the protein chlorophyll complexes, releasing chlorophyll for detoxification.

Additional information can be gained from the gradient analysis described above if the time of maximum gradient is also considered. For example, statistical analysis of gene clusters based on time of first differential expression indicated that photosynthetic genes were overrepresented in clusters showing downregulation at 29 to 33 DAS. However, if all the downregulated genes annotated as photosynthesis are examined, many of these show initial significant downregulation earlier in the time series (see Supplemental Figure 4A online), but the size of the clusters at these time points means this annotation does not show up as being significantly enriched. When the maximum absolute gradient for each of the photosynthesis genes was calculated (see Supplemental Figure 4B online), the vast majority of photosynthesis genes showed the most rapid drop in expression between 31 and 35 DAS, confirming the observation that photosynthesis-related gene activity is maintained until late in the leaf's development (Figure 6).

The maximum absolute gradient analysis was also applied to investigate genes responding to JA and ABA stimulus. In both cases, the majority of genes were first significantly upregulated at 21 and 23 DAS, early in senescence (see Supplemental Figures 4C and 4E online). However, although several of the genes had a maximum gradient early, at 23 DAS, there were also many showing maximum gradient much later in the time series, up to 35 and 37 DAS for the JA response genes and 33 and 35 DAS for the ABA response genes (see Supplemental Figures 4D and 4F online). This correlates with the data on the levels of JA and ABA shown in Figure 2 where a maximum level of both hormones is measured late in the process (increasing at 33 and 31 DAS, respectively). The timing of expression of specific hormone biosynthesis genes (see Supplemental Figure 5 online) clearly illustrates the rapid increase in JA biosynthesis genes between 23 and 25 DAS, whereas ABA and SA biosynthesis genes show a later increase in expression with a maximum at the final stage of senescence. Although ethylene levels were not measured during the time course, the ethylene biosynthesis genes *ACS2* and *ACS7* also show increased expression from around 29 to 31 DAS, with a steady increase as senescence progresses. Thus, ABA, ethylene, and probably SA synthesis appear to be coordinately regulated in senescence, whereas JA synthesis shows a different pattern. Interestingly, some JA biosynthesis and signaling genes are only expressed at the early time point (e.g., *OPR3*), whereas others (e.g., *LOX3*) are also upregulated late, presumably enabling the accumulation of JA later during senescence.

Our detailed expression profiles and novel tools have enabled us to distinguish biological processes initiated at different stages of senescence and hence tease apart some of the components of this complex phenomenon. We now have a timeline that can be built upon to link these different processes and to identify the overarching regulatory mechanisms as well as candidate genes for specific senescence processes.

TF Binding Motifs Show Specific Enrichment in Differentially Expressed Gene Clusters

The SplineCluster analysis of differentially expressed genes (Figure 3) identified groups of genes that exhibit similar expression profiles and thus may be coregulated. Analysis of such coregulated gene sets should help pinpoint potential TF binding motifs important for gene expression during leaf development. To gain an initial understanding of the regulatory mechanisms of genes differentially expressed during senescence, promoters corresponding to 500 bp upstream of the predicted transcription start site of genes in each cluster were screened for overrepresentation of known TF binding motifs.

This analysis shows clearly that certain sequence motifs are selectively enriched in clusters that exhibit similar expression patterns (Figure 7, data shown in Supplemental Data Set 5 online), and there is an obvious difference in the range of motifs distributed over the different clusters. Consistent with the GO term analysis results, several of the downregulated clusters (clusters 1–24) are significantly enriched for sequence motifs associated with the regulation of photosynthesis and cell growth. For example, the G box variant motif is linked with the regulation of photosynthetic genes and responses to light (Martínez-García et al., 2000) and the TCP motif binds members of the TCP family of TFs, which have been implicated in the regulation of growth and cell division (Li et al., 2005). Binding sites for E2F TFs, key regulators of cell proliferation (Ramirez-Parra et al., 2003) are enriched in cluster 22, which is consistent with this cluster being enriched with genes annotated with the GO term cell cycle. Genes in this cluster are downregulated from the start of the measured time course, and this would be expected since cell division has ceased before the leaf is fully expanded.

Sequence regions upstream of genes in upregulated clusters (clusters 27–48) contain a number of sequence motifs that can bind TF families that are themselves upregulated during senescence. For example, NAC domain and WRKY TFs constitute a

large proportion of the senescence-regulated TFs and are known to play significant roles in regulating leaf senescence in *Arabidopsis* (Miao et al., 2004; Guo and Gan, 2006; Kim et al., 2009). Binding sites for NAC and WRKY TF families are overrepresented in several upregulated clusters sharing similar expression profiles. Sequence motifs associated with stress responses are also enriched. The heat shock element is overrepresented in a single cluster, and several heat shock factors are upregulated during senescence. The CGCG motif, which has been implicated as a calcium-signaling element in a range of stresses, is enriched in several upregulated clusters. This motif has been shown to bind CAMTA TFs (Yang and Poovaiah, 2002) involved in signaling responses to wounding, cold, and other stresses (Walley et al., 2007; Doherty et al., 2009). The ABA-responsive element (ABRE) is overrepresented in multiple upregulated clusters, and enrichment correlates with the observed increase in levels of ABA during senescence. ABRE-binding factors are known to activate target genes in an ABA-dependent manner (Nakashima et al., 2006). The ABRE contains an ACGT-core and, therefore, is a subset of the G box sequence (CACGTG). However, the pattern of overrepresentation of these two similar motifs across the senescence clusters is different, suggesting that divergent functional roles can be identified. G box–like motifs can bind many members of the bZIP and bHLH TF superfamilies (Toledo-Ortiz et al., 2003; Jakoby et al., 2002), and TFs from both these families are upregulated during senescence.

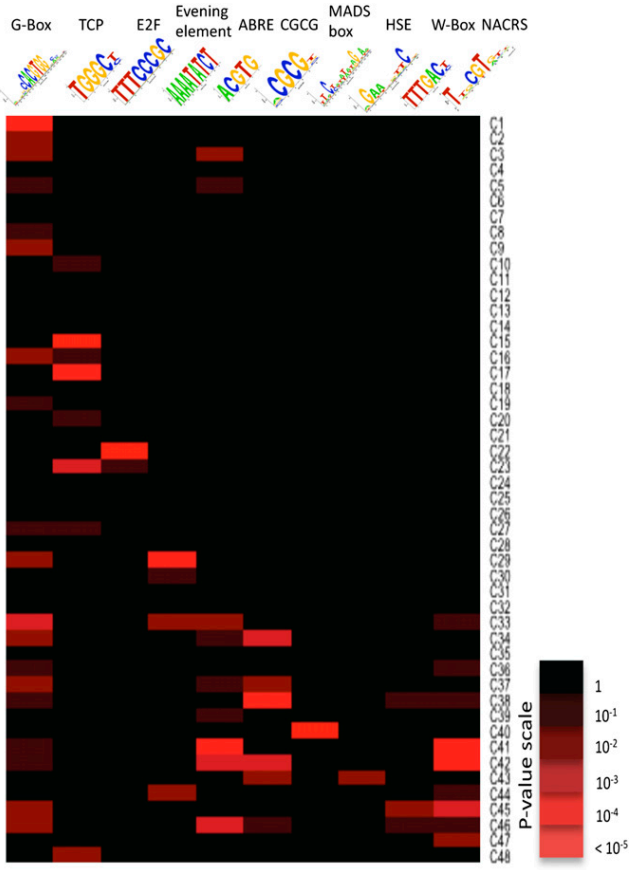


Figure 7. Over-Representation of Known TF Binding Motifs in Promoters of Coexpressed Genes.

Logo representations of known TF binding motifs are on the horizontal axis, and expression profile for each cluster (see Figure 3) is on the vertical axis. Colored boxes represent pairs of motif and expression cluster with a significant statistical link. Shown are a limited number of representative motifs and clusters (see Supplemental Data Set 5 online for full results).

TF Families Are Active at Different Times during Senescence

To complement the analysis of TF binding motifs above, we investigated whether specific families of TFs were differentially expressed at particular times during senescence. The groups of genes identified as having the same initial timing of differential expression by the GP gradient tool analysis were further analyzed to identify time periods when families were overrepresented for genes with a positive or negative gradient (i.e., expression significantly increasing or decreasing). A heatmap, mapped to the significance of each family's activity, is shown in Figure 8, with the numerical data shown in Supplemental Data Set 6 online.

A number of TF families were significantly overrepresented for upregulated genes (adjusted $P < 0.01$), indicating a large amount of similar transcriptional activity and potential coregulation within these families. Specific members of the bZIP family have been shown to participate in defense against pathogens, development, stress treatments such as cold and drought, ABA signaling, and phenylpropanoid biosynthesis (Weisshaar and Jenkins, 1998; Jakoby et al., 2002). Another significantly overrepresented upregulated family is the large C3H superfamily, of which little is known of the function of many of its members. Several subfamilies related to the CCAAT box binding factor family were also significantly upregulated; factors in this family form the heterotrimeric NF-Y binding complex (consisting of NF-YA, NF-YB, and NF-YC subunits) that has been shown to influence flowering time and stress responses in plants (Wenkel et al., 2006; Liu and Howell, 2010). Interestingly, it is the NF-YA subunits specifically that are enriched in the senescence-enhanced gene lists, with nine of the 10 genes in the genome showing increased expression during senescence. In comparison, few of the NF-YB and NF-YC genes are altered in expression, with only three NF-YB genes and one NF-YC gene being upregulated. This implies that it may be the regulated expression of the NF-YA subunit that controls the activity of the NF-Y complex during senescence.

The large NAC family also had a significant early overrepresentation, with over 30 of the members of this family being altered in expression at various times during senescence. Members of this family are known to have a large number of regulatory interactions with a diverse range of biological processes, including senescence, defense, and abiotic stress (Olsen et al., 2005, and references quoted above). Several other families show upregulated transcription later in the time course. The WRKY family shows an overrepresentation, with many members of this family being upregulated from around 25 DAS. WRKY TFs have been shown to be important for senescence (Robatzek and Somssich, 2001; Miao et al., 2004); others are induced by

infection by viruses or bacteria (Eulgem et al., 2000) and are downstream of defense-signaling mitogen-activated protein kinase pathways and involved in the regulation of SA- and JA-dependent defense signaling pathways (Ülker and Somssich, 2004; Eulgem and Somssich, 2007). The large AP2-EREBP family becomes significantly overrepresented around 27 DAS; members of this family are induced in several cases by hormones such as JA, SA, and ethylene, along with other signals related to pathogens, wounding, and abiotic stresses, and have influence on other stress and disease resistance pathways (Kizis et al., 2001; Gutterson and Reuber, 2004). Therefore, cascades of cellular information flow during the progress of leaf senescence can be predicted by this analysis, such as upregulation of NAC or WRKY genes influencing various hormone responses, followed by upregulation of AP2-EREBP TFs. This knowledge is key for future modeling of senescence transcriptional networks.

ANAC092 Target Genes Are Highly Enriched in Clusters Overrepresented for NAC Binding Motifs

The motif and TF analyses described above pinpoint NAC domain genes as being of key importance in regulation of senescence and we follow this observation up in more detail as an example of the increased understanding that this data set provides. A recent publication (Balazadeh et al., 2010) describes an elegant experiment using inducible expression of *ANAC092* to identify likely target genes. Of the 170 genes identified in that study as being upregulated after induced expression of ANAC092, 102 of these are senescence enhanced in our time course experiment; of these, 75%, including *ANAC092* itself, are to be found in the clusters enriched for NAC domain motifs (clusters 41, 42, 44, and 45; see Supplemental Data Set 7 online). This provides clear evidence that the detection of enriched motifs within clusters is providing biologically relevant information and also indicates that ANAC092, probably together with other NAC domain proteins, has an influential role in regulating the expression of many genes at this stage of senescence. This information might be used in preliminary modeling experiments to predict interactions that regulate gene expression in these clusters.

DISCUSSION

In this article, we describe a high-resolution, highly replicated time-course analysis of gene expression during *Arabidopsis* leaf development from before complete expansion to full senescence. Over this time, the leaf develops from a sink that is importing nutrients for growth into an active source organ, performing maximum photosynthesis and exporting fixed carbon for further growth of the plant. This is followed, relatively rapidly in this short-lived plant, by the initiation of senescence whereby the leaf is converted from a source of photosynthetic carbon to a source of valuable macromolecules such as nitrogen, phosphorus. and minerals, as cellular components become degraded and mobilized from the leaf. Thus, in this short time period, the leaf undergoes enormous changes in metabolism and transport of metabolites.

To obtain insight into the timing and potential coregulation of the changes in genes and pathways in the complex process of senescence, it is essential to sample highly controlled replicate leaves and to measure at many time points. This is also essential if these data are to be used for network inference analysis. In the experiment reported here, we harvested the same leaf (leaf 7) from individual plants at different times over 3 weeks. Previous

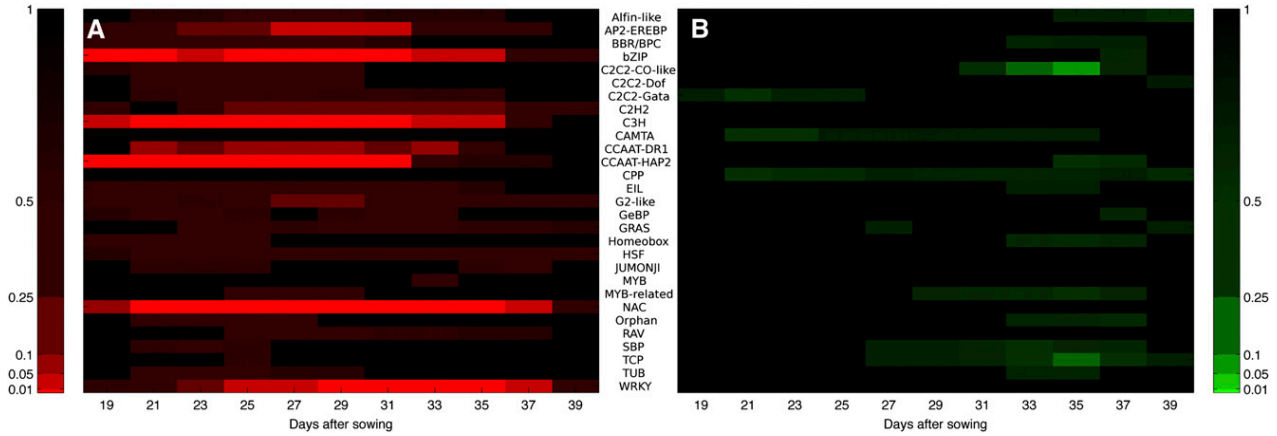


Figure 8. TF Families Significantly Over-Represented with Positive or Negative Gene Expression Gradients Highlighting Distinct Periods of Activity.

(A) A number of TF families significantly upregulated early, including the NAC and bZip families (19–21 DAS), remaining strongly upregulated throughout the experiment, and a small number of families only upregulated toward the middle and end of the experiment (27+ DAS), such as WRKY, AP2, and G2-like.

(B) A small number of weakly significant TF families downregulated early, and several families (C2C2-CO-like and TCP) significantly downregulated toward the end of the experiment (33+ DAS).

Color bars indicate P value (after FDR correction), with a range of significance thresholds (0.01, 0.05, 0.1, 0.25, and 0.5). Numerical data used to derive this figure are shown in Supplemental Data Set 6 online.

transcript profiling studies on developmental leaf senescence in *Arabidopsis* have analyzed a more limited number of time points (e.g., Buchanan-Wollaston et al., 2005; van der Graaff et al., 2006). Analysis of gene expression changes in field-grown *Populus* leaves has been performed over several time points during late summer and autumn (Andersson et al., 2004). In these examples, pooled leaves were used as biological samples; and replicates were limited; although these experiments gave a picture of the overall changes in gene expression that occur during senescence, they provide little information on the timing of the changes during the process. Senescing leaves are, by their nature, quite variable, particularly in mid or late senescence; therefore, to accurately determine coregulation of genes through senescence, it is essential to control and allow estimation of other sources of variability, such as may be caused by plant-to-plant differences and environmental factors by using a carefully designed sampling strategy.

Methods Developed to Enable Large-Scale, Two-Color Microarray Analysis and Identification of Differentially Expressed Genes

This large-scale microarray experiment with both biological and technical replication across multiple time points required the development of a novel and complex design approach to take full advantage of the two-color microarray system, providing efficient estimation of the differences in response between adjacent time points while still allowing effective comparison of all samples. The highly replicated experiment allowed the application of stringent statistical analysis to identify and characterize genes differentially expressed at different time points during senescence, which has not been possible with data from previous studies. The quantity of data generated necessitated the adaptation of existing analysis methods/algorithms, as well as the development of some new analysis tools.

The Bioconductor package MAANOVA was adapted to meet the specifications of the CATMA arrays, providing monitoring of slide and data quality and using information from the four technical replicates of each sample to remove the influence of the occasional outliers. The mixed model-fitting algorithm then enabled the estimation and testing of the differences caused by the treatment factors (day, time of day, the interaction [combined effect] of these factors, and the biological replicates), allowing for the complex design structure and the sources of variability (between slides and between dyes) imposed by using the two-color microarray system. There are considerable advantages to the experimental and analysis approaches used. The use of two-color arrays allows direct comparisons of samples between key time points and, through careful design of the pairs of samples compared on each array, across all time points by indirect association. By contrast, many applications of two-color arrays compare experimental samples by calculating the ratio of expression responses of each with a control sample hybridized on every slide, thus halving the amount of useful data obtained per slide (or doubling the cost of obtaining the same data).

Identification of genes showing interesting differential expression patterns was achieved by first assessing the significance of the variation due to different model terms (day, time of day, and the interaction between them) relative to the between-biological replicate variation using the MAANOVA analysis, and then conducting a visual inspection of gene expression responses over time for genes giving a less significant test result ($0.0001 < P < 0.05$) for the effect of day. Of course, high levels of biological (between-plant) variability can lead to large changes in gene expression not being identified as statistically significant and, hence, genes not identified as being differentially expressed. Further exploration of approaches to control this biological variation, through both the statistical design of future experiments and the development of novel analysis methods, is important for future successful research.

The mixed model-fitting algorithm implemented in the MAANOVA package allows separation of the variation due to different sources within the treatment combinations (i.e., day, time of day, the interaction between these factors, and biological replicates), and hence the identification of genes showing different generic patterns of differential expression (see Supplemental Figure 1 online). An advantage of this approach, over a simple comparison of the responses across all 22 time points, is that genes showing only diurnal (time of day) variation can be easily identified and ignored in subsequent modeling of potential senescence-related gene networks, with those showing combined effects of day and time of day also easily identified and included. A disadvantage is that the analysis does not formally include any allowance for the ordering of the samples through time (e.g., the effect of day is essentially just an assessment of the average variability between the 11 mean values, and reordering the days would not change the test statistic and hence level of significance). A further development of this approach, allowing both separation of effects within a factorial treatment structure and estimation of the underlying shape of response over time, possibly following the approach proposed by Eastwood et al. (2008), should lead to a more reliable identification of genes showing important patterns of differential expression, although issues with high levels of biological variability would still result in some false-negative test results. Better estimation of the shape of expression profiles could also contribute to improved clustering of genes with similar shapes of expression profiles. In the absence of such a modeling approach, the approach used here, combining the highly significant results of the formal analysis with a visual inspection, is likely to result in the identification of most of the important genes showing differential expression related to senescence, while minimizing the number of false positives.

Analysis of Differentially Expressed Genes Revealed a Chronology of Processes and Signals

Analysis of individual clusters identified in the SplineCluster analysis, particularly those for downregulated genes, identified groups of genes involved in a common process such as photosynthesis, chlorophyll metabolism, etc. It seems likely that genes involved in the same process, with similar expression profiles, are coregulated rather than simply coexpressed during senescence, and this prediction is strengthened by the promoter motif analysis.

The GP gradient analysis, developed to enable more effective dissection of gene expression changes over time, identified

groups of genes that showed their first significant change in expression between the same pair of adjacent time points. The resulting clusters present a highly informative picture of the timeline of senescence, showing when individual pathways are upregulated or downregulated (Figure 6). Knowledge of such timing will prove a powerful tool for separation of pathways into groups to allow identification of upstream genes that control them.

Comparing the two approaches used to cluster the differentially expressed genes, it is clear that they will generate different sets of clusters. SplineCluster groups genes with overall profile shapes that are similar based on the fitted regression coefficients, which should therefore mean that genes in the same cluster will have similar changes in expression between every pair of adjacent time points. However, the approach does not take any account of the biological (between-plant) variability, so that only the initial filter will determine the significance associated with the overall differential expression, and so a gene with highly significant variation in expression could be clustered with one just breaking the significance threshold. By contrast, the GP gradient analysis groups genes that have the first significant changes in expression in the same direction at the same time. However, unless the gradient information is also taken into account, these may not always be showing the most dramatic change in expression at the same time point. Therefore, both approaches have value in identifying coregulated genes, but both have the potential to inappropriately group genes.

The clustering results have been discussed in detail above and have identified groupings of genes that had not been observed previously with more limited time series data. For example, it is clear that the extensive overall reduction in expression of genes involved in chloroplast activities occurs via a timed process. Chlorophyll biosynthesis genes are downregulated before carbon fixation genes, and these are downregulated well before the majority of key genes encoding proteins involved in photosynthesis, including chlorophyll binding proteins and components of PSI and II. Autophagy genes are enhanced from the start, but the level of the key gene *ATG7* starts to rise at 29 DAS. This is also the time at which chlorophyll degradation genes show a rapid induction of expression. Other metabolic pathways such as strigolactone synthesis, hormone biosynthesis, cell wall degradation, cytoskeleton, and microtubule activity, to name just a few, are implicated at different times during the senescence process.

Microarray Data Analysis Tools Used to Develop and Test Hypotheses for Transcriptional Control during Senescence

Analysis of the core promoters of coexpressed genes revealed potential regulatory sequence motifs that are likely to contribute to the coregulation of genes involved in the senescence process. Known sequence motifs are enriched in the promoters of genes that share similar expression profiles and correlate with the biological processes associated with such genes. The importance of the NAC and, to a lesser extent, the WRKY TF families in promoting senescence in *Arabidopsis* is illustrated through the specific and highly significant enrichment for potential binding sites for these regulators in the promoters of certain clusters of upregulated genes. In addition, most of the genes implicated as under the control of the senescence-enhanced ANAC092 TF (Balazadeh et al., 2010) occur in these same clusters, showing the importance of this TF and other NAC family members in regulating gene expression during senescence. The identity of specific TFs that target these known motifs is unknown, and further bioinformatic analysis and modeling as well as laboratory experiments is required to characterize them fully. Regulation via known *cis*-regulatory elements is not sufficient to explain the expression patterns of all genes, and unknown sequence motifs likely contribute toward regulating specific groups of genes within the senescence process.

In comparison with previous senescence gene expression studies, the study reported here collected highly replicated gene expression responses at a high temporal resolution across the period during which the senescence response develops. Thus, these data are more suitable than previously collected data sets

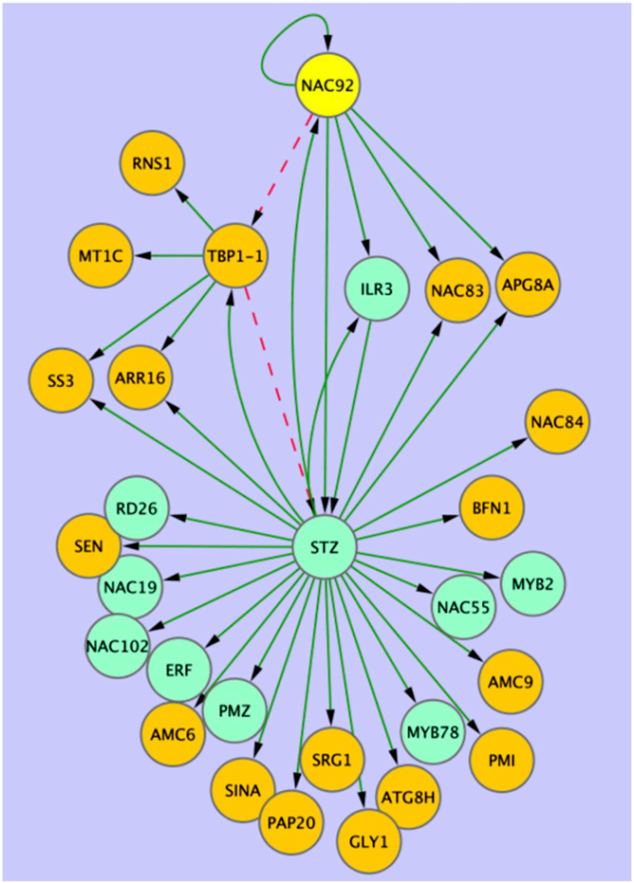


Figure 9. Network Model Inferred from Microarray Data.

Variational Bayesian state space modeling was used to generate a network model using senescence-enhanced genes selected from SplineClusters 41, 42, 44, or 45 in the microarray data. Genes showing induced expression in an *ANAC092* (yellow node) inducible overexpression experiment (see Supplemental Data Set 6 online; Balazadeh et al., 2010; orange nodes) were combined with selected TFs from the same clusters (green nodes). Green edges represent positive interactions, while red dashed edges predict negative effects. Genes are identified in Supplemental Data Set 7 online.

for applying statistical analyses aimed at predicting the gene regulatory networks operating during senescence. This is of immense value for the next step in the development of a model for the regulation of leaf senescence. However, even this substantial microarray experiment still imposes data limitations that make the application of network inference nontrivial; gene expression measurements of just 88 biological samples are insufficient to accurately model the regulation of thousands of genes. As the number of possible networks grows superexponentially with the number of genes involved, the correlations in the patterns of response across these 88 samples are certainly insufficient to accurately identify a unique network model describing the regulation of the 6326 genes identified as differentially expressed. This is known as the "curse of dimensionality" (Bellman, 1961). Therefore, the design of future studies needs to consider the balance between experimental cost and informative data for network inference, especially as a greater number of samples brings diminishing returns with respect to the number of additional genes it allows a researcher to model. It is also important to balance the need for good information about the biological (between-plant) variability, technical replicate variability, and temporal changes in gene expression during the senescence process.

The data presented here can be used to produce constrained network models for many small sets of genes. An example of the type of network model that can be inferred from these data is shown (Figure 9). A variational Bayesian state space modeling method (Beal et al., 2005) was applied to a selection of genes using the 11–time point data series (see Supplemental Table 2 online). Genes selected were present in clusters enriched for the NAC motif (i.e., clusters 41–45) and also present in the *ANAC092* overexpression data described above (Balazadeh et al., 2010; see Supplemental Data Set 7 online) or were annotated as TFs. The resulting model correctly predicts a positive influence of ANAC092, either direct or indirect, on multiple known downstream target genes (Figure 9, orange nodes) The model also makes several new hypotheses for experimental testing. For example, it predicts the influence of a zinc finger protein (STZ) in the expression of *ANAC092* and its downstream genes. Although no experimental evidence exists for this regulatory link, both *STZ* and *ANAC092* are induced in expression during salt stress, and knock out mutants in both genes have enhanced tolerance to stress in *Arabidopsis* (Mittler, et al., 2006; Balazadeh et al., 2010). Regulation by ANAC092 of several other TFs known to be stress related is also predicted in this network model. ANAC019, ANAC055, and RD26 (ANAC072) have all been shown to have a role in drought stress (Tran et al., 2004); MYB2 has a role in ABA signaling and salt stress (Abe et al., 2003; Yoo et al., 2005), and PMZ, a zinc finger protein, has a role in stress-induced senescence (Breeze et al., 2008). The gene regulatory network model also predicts feedback and feed-forward connections between *ANAC092*, *STZ*, and *TBP1-1*, which encodes a telomere binding protein. These types of interactions are crucial for the robustness of gene regulatory networks and would be almost impossible to predict from biological data alone. Thus, this model, generated with a small subset of the array data, correctly predicts known gene-gene interactions and generates complex novel predictions for experimental testing.

Many different models can be obtained with different collections of genes, and these transcriptional network models can be expanded using information on coregulated pathways and promoter motif analysis to identify likely downstream targets of key TFs. This model development and experimental testing is underway to generate validated gene regulatory network models underlying senescence.

METHODS

Plant Growth

Arabidopsis thaliana plants were grown as described in Breeze et al. (2008). Leaf 7 was tagged with thread 18 DAS. Sampling of leaf 7 started at 19 DAS and continued every other day until full senescence was reached (39 DAS). Leaves were harvested twice on each sampling day, 7 and 14 h into the light period. This resulted in samples being obtained at 22 distinct time points. At each time point, leaf 7 was sampled from 20 plants among the 720 being grown in the controlled-environment growth room, the plants being randomly selected to avoid any potential effects of position within the growth room. Leaves were rapidly weighed and photographed with a size scale before being frozen in individual tubes in liquid nitrogen. Leaf length was estimated against this scale from the photographs.

Protein and Chlorophyll Measurements

Total protein was extracted from five individual leaf samples by grinding the sample in liquid nitrogen before the addition of 500 μL of extraction buffer (50 mM lithium phosphate [pH 7.2], 1 mM monoiodoacetic acid, 120 mM mercaptoethanol, 5% [v/v] glycerol, 1 mM PMSF, and 0.2% lithium dodecyl sulfate). At this stage, a 100-μL aliquot of the extract was taken for chlorophyll analysis. The protein extract was boiled for 45 s and centrifuged for 20 min at 12,800*g*. Total protein was measured using the RC DC protein assay (Bio-Rad) according to the manufacturer's instructions. In addition, levels of the small and large subunits of Rubisco were assessed by diluting the protein extracts to normalize for leaf weight and then running an equal volume of each extract (equivalent to 0.5 mg of fresh tissue) on polyacrylamide gels (Invitrogen Novex 4-12% Bis-Tris gel), staining with Coomassie blue, and scanning the relevant protein band. Protein levels were assessed densitometrically using image analysis software (GeneTools; Syngene) against a calibration curve of bovine serum albumin (LSU) and lysozyme (SSU).

Chlorophyll was measured from five individual leaves using the total protein extracts. The chlorophyll was extracted using 80% acetone, vortexed, and then stored at −20°C for 1 h in the dark. The samples were then centrifuged for 3 min at 12,800*g*, and the absorbance of 1 mL was measured at 663 and 646 nm. Chlorophyll concentrations were calculated using the equations: total chlorophyll (mg/L) = $20.2A_{646} + 8.02A_{663}$, chlorophyll *a* (mg/L) = $13.19A_{663} - 2.57A_{646}$, and chlorophyll *b* (mg/L) = $22.1A_{646} - 5.26A_{663}$.

Hormone Measurements

The hormones ABA, SA, and JA were measured in five individual leaves. Each leaf was freeze dried and 10 mg of freeze dried tissue was used for hormone extraction as described in Forcat et al. (2008). Analysis was performed using an HPLC-ESI/MS-MS.

Statistical Treatment of Leaf Morphological and Biochemical Measurements

Leaf morphology (weight and length) and biochemical assay (total protein, chlorophyll *a+b*, Rubisco LSU and SSU, and hormone) data were

subjected to ANOVA to assess for differences in response over the time course using GenStat (VSN International). Data from hormone assays were subjected to a $\log_{10}$ transformation (including the addition of a small constant to cope with zero observations) prior to analysis to satisfy the assumption of homogeneity of variance. LSDs were calculated at a 5% significance level to allow easy comparison of differences between adjacent time points. Significant effects noted in the results relate to either F tests for the overall variability over time or *t* tests for comparisons between adjacent time points.

Microarray Analysis

RNA Preparation and Labeling

Total RNA was isolated from four individual leaves from each sampled time point (arbitrarily labeled as biological replicates A, B, C, and D) using TRIzol reagent (Invitrogen), purified with RNeasy columns (Qiagen), and amplified using the MessageAmp II aRNA Amplification kit (Ambion) in accordance with the kit protocol with a single round of amplification. Cy3- and Cy5-labeled cDNA probes were prepared by reverse transcribing 5 μg of aRNA with Cy3- or Cy5-dCTP (GE Healthcare) and a modified dNTP mix (10 mM each dATP, dGTP, and dTTP; 2 mM dCTP) using random primers (Invitrogen) and SuperScript II reverse transcriptase (Invitrogen), with the inclusion of RNase inhibitor (RNaseOUT; Invitrogen) and DTT. Labeled probes were purified using QiaQuick PCR Purification columns (Qiagen), freeze-dried, and resuspended in 50 μL of hybridization buffer (25% formamide, 5× SSC, 0.1% SDS, and 0.5μg/μL yeast tRNA; Invitrogen).

Microarray Experiments

The microarray experiments were performed using the CATMA (version 3) microarray (Allemeersch et al., 2005; http://www.catma.org). CATMA probe annotations were updated using the TAIR9 release: oligo sequences of CATMA array probes were mapped to individual mRNA sequences of transcripts from the TAIR9 genome assembly using BLASTn (Altschul et al., 1997), with e-value cutoff of 0.01. Additionally, results were filtered to exclude alignments shorter than 30 bp or with less than 80% sequence identity. The best matching gene model (by e-value of hit to transcript) was identified for each probe. In addition, probe sequences were mapped to TAIR9 genomic DNA to clarify cases where a probe had been designed to a region of an earlier genome assembly now unannotated in TAIR9.

A novel experimental design strategy (A. Mead, unpublished data), based on the principle of the "loop design" (Kerr and Churchill, 2001), was developed to enable efficient extraction of information about key sample comparisons using a two-color hybridization experimental system. With 88 distinct samples (four biological replicates at each of 22 time points) to be compared, the experimental design included 176 two-color microarray slides, allowing four technical replicates of each sample to be observed. The detailed structure of the design, indicating how pairs of treatments were allocated to arrays, is described in Supplemental Methods 1 online, with an illustrative diagram shown in Supplemental Figure 6 online. According to a randomization of this experimental design, pairs of labeled samples were hybridized to slides overnight at 42°C. Following hybridization, slides were washed and scanned using an Affymetrix 428 array scanner at 532 nm (Cy3) and 635 nm (Cy5). Scanned data were quantified using Imagene 7.5.0 software (BioDiscovery, Inc.).

MAANOVA Analysis

A local adaptation of the MAANOVA package (Wu et al., 2003) was used to analyze the quantified microarray data, providing data quality assurance, slide normalization through LOWESS data transformation, mixed model fitting, and identification of genes showing significant differential expression via F tests of fixed (treatment) terms included within the model. MAANOVA was selected to analyze the data because it is able to provide an accurate analysis of the effects on gene expression of multiple sources of variation (both fixed, treatment, terms, and random sources of background variation) in the experimental design, harnessing the power of direct comparisons between pairs of samples obtained using two-channel microarrays (Churchill, 2004). Full details of the data quality checking procedures, of the mixed model fitting approach to describe the observed gene expression data, and of the construction of F tests for fixed treatment terms are given in Supplemental Methods 1 online. Having fitted the mixed model to each gene, predicted means were calculated for each of the 88 samples, assuming the full treatment model (effects of day, time of day, the interaction between them, and the nested biological replicates) to produce a four-replicate 22–time point data set for each gene, or assuming a reduced treatment model (effects of day and the nested biological replicates) to produce an eight-replicate 11–time point data set for each gene. These data sets were then used in subsequent analyses.

Selection of Differentially Expressed Genes

The most significant differentially expressed genes were identified initially and this was followed by visual analysis of genes close to the borderline of significance. First, CATMA probes with no corresponding gene model in the TAIR9 annotation were ignored; also, replicate CATMA probes were removed (the most gene-specific probe being identified in each case). In total, 4989 genes had an adjusted day main effect F test P-value < 0.0001 (after multiple testing correction via a step-down FDR-controlling procedure (Westfall et al., 1998; Benjamini and Liu, 1999), equivalent to responses showing a significant test result at an FDR of $P < 0.0001$), and these were included in the initial list of differentially expressed genes. The patterns of expression of all genes with an adjusted day main effect F test P-value between 0.0001 and 0.05 (further responses showing a significant test result at an FDR of $P < 0.05$) were then screened visually to remove any showing either a small or a very variable change in expression over time. The final list of 6323 differentially expressed genes, with adjusted F test statistics, is shown in Supplemental Table 1 online.

Gene Expression Profile Clustering

Clustering of coregulated genes was performed by the application of SplineCluster (Heard et al., 2006), a Bayesian model-based hierarchical clustering algorithm for time series data, using the mean of the biological replicates for each gene. Recent functionality added to SplineCluster, (Heard, 2011) improves the gene allocation to clusters. Where a gene has become an outlier for its allocated cluster, it is reallocated to alternative clusters to maximize the log marginal likelihood once more. This option was used on all SplineCluster analyses presented in this article. The 22–time point data (averaged across the four biological replicates) was clustered using a prior precision of 5×10^{-4}, while the other data set composed of 11 time points was averaged across all eight morning and afternoon biological replicates before being clustered using a prior precision of 1×10^{-4}. These prior precisions were selected as they produce ~50 clusters for each of the two data sets.

GP Gradient Analysis

To identify an ordering of events, the rate of change of gene expression (gradient) was inferred using a GP regression approach (see Supplemental Methods 1 online), which has the notable advantage of incorporating all biological replicates. Furthermore, since the marginal distribution of a

GP is itself a Gaussian distribution, the probability that the gradient (at any particular time) lies sufficiently far from zero may be calculated analytically. When data are sufficiently time resolved, the GP model may therefore be used to identify times when the gradient of a gene expression profile is significantly positive (increased), negative (decreased), or not statistically different from zero (steady), whereas for less time-resolved data, it may identify times of significant change to the global trend.

GO Analysis

GO annotation analysis on gene clusters was performed using the BiNGO 2.3 plugin tool in Cytoscape version 2.6 with GO_full and GO_slim categories, as described by Maere et al. (2005). Over-represented GO_Full categories were identified using a hypergeometric test with a significance threshold of 0.05 after a Benjamini and Hochberg FDR correction (Benjamini and Hochberg, 1995).

Promoter Analysis

Plant position-specific scoring matrices (PSSMs) were collected from the TRANSFAC database, version 2010.3, (Matys et al., 2006) and the PLACE database (Higo et al., 1999). This set was supplemented with PSSMs for a heat shock element (TRANSFAC matrix record M00146) and two NAC TF binding sites (Olsen et al., 2005) since these important motifs were absent from the databases. PSSMs were clustered, and a representative of each cluster was chosen for screening. Promoter regions corresponding to 500 bp upstream of the transcription start site were retrieved from the Ensembl Plants sequence database (release 50).

For any given PSSM and promoter, we scanned the sequence and computed a matrix similarity score (Kel et al., 2003) at each position on both strands. P values for each score were computed from a score distribution obtained by applying the PSSM to a random sequence of 100 million bases in length generated by a 3rd order Markov model learned from the whole Arabidopsis genome. We took the top k nonoverlapping hits and performed the binomial test for the occurrence of k sites with observed n values within a sequence of length 500 bp. The parameter k is optimized within the range 1 to 5 for minimum binomial P-value. This allows detection of binding sites without a fixed threshold per binding site. Using a threshold ($P < 0.05$), the presence or absence of a PSSM was scored for each promoter based on the binomial probability.

For each PSSM, its frequency in promoters of each cluster was compared with its occurrence in all promoters in the entire genome. Motif enrichment was calculated using the hypergeometric distribution (phyper function in the R stats package). Hypergeometric P-values were corrected for the number of clusters tested using Bonferroni correction. Corrected P-values ≤ 0.05 were considered significant. Sequence logos were generated using code modified from Lenhard and Wasserman (2002). Sequence analysis was performed within the APPLES software framework (S. Ott, unpublished data).

TF Family Analysis

Gene expression activity was analyzed for 1733 TFs, grouped into 50 families defined in the *Arabidopsis thaliana* Transcription Factor Database, AtTFDB (Palaniswamy et al., 2006; 1843 TF, 50 families as of June 2010). Of these, 1733 were probes on the CATMA array using the GP gradient model. Families overrepresented for genes with significantly positive or negative gradients at each time point, using all genes within the experiment as a reference, were identified using the hypergeometric distribution (computed using the *hypgeomdist* function in MS Excel 12.2.3) with Benjamini and Hochberg FDR correction. A heatmap of adjusted P-values, using five levels of significance (0.01, 0.05, 0.1, 0.25, and 0.5) was then generated, using only those values that correspond to overrepresented counts (e.g., the proportion of positive/negative gradient TFs for a given family is larger than the proportion of positive/negative gradient genes in the entire data set for each time point).

Variational Bayesian State Space Modeling

Data for eight biological replicates from the 11–time point data series were used to generate a network model using the method published in Beal et al., 2005. ANAC092 was used as the gene on which to base the model and two groups of genes were selected to accompany it. First, several genes were selected that showed rapidly increased expression following induced expression of *ANAC092* in green leaves (from Balazadeh et al., 2010), many of which were also in clusters enriched for NAC domains. Therefore, these are likely to be direct or indirect targets of *ANAC092* activation. Second, a group of TFs that show coexpression with *ANAC092* selected from the clusters 41 through 45 was included. Ten models were run from different random seeds and connections occurring in more than 50% of models at a confidence level of >95% were included in the network shown in Figure 9.

Data Repository

The microarray data used in this article have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and have been given a GEO Series accession number, GSE22982.

Accession Numbers

Arabidopsis gene names and identifiers referred to in this article are: ANAC092 (At5g39610), *WRKY53* (At4g23810), *ANAC029* (At1g69490), *SAG12* (At5g45890), *LHY* (At1g01060), *CCA1* (At2g46830), *PRR7* (At5g02810), *PHYA* (At1g09570), *CRY1* (At4g08920), *PIF4* (At2g43010), *ELF4* (At2g40080), *PCL1* (At3g46640), *LUT1* (At3g53130), *LUT2* (At5g57030), LUT5 *(*At1g31800), *ARR4* (At1g10470), ARR6 (At5g62920), ARR7 (At1g19050), sedoheptulose bisphosphatase (At3g55800), *HCF101* (At3g24430), *HCF109* (At5g36170), *HCF152* (At3g09650), *HCF173* (At1g16720), *HCF208* (At5g52110), GLK2 (At5g44190), GLK1 (At2g20570), LHCB2.2 (At2g05070), LHCB4.2 (At3g08940), LHCB6 (At1g15820), *ATG7* (At5g45900), *ATG8H* (At3g06420), *ATG8A* (At4g21980), *ATG8B* (At4g04620), DREB2A (At5g05410), LSD1 (At4g20380), AtMPK7 (At2g18170), HSP70 (At3g12580), (*AHK3* (At1g27320), *ERD1* (At5g51070), *ERD14* (At1g76180), *RAB18* (At5g66400), *ABF2* (At1g45249), *AOC1* (At3g25760), *AOC4* (At1g13280), 12-oxophytodienoate reductase (At2g06050), *MYC2* (At1g32640), *JAZ1* (At1g19180), *JAZ6* (At1g72450), *JAZ8* (At1g30135), COI1 (At2g39940), *LOX3* (At1g17420), *ABI1* (At4g26080), *AFP1* (At1g69260), *RD20* (At2g33380), *RD26* (At4g27410), NF-YA4 (At2g34720), *CCD7* (At2g44990), *CCD8* (At4g32810), *ORE9* (At2g42620), *AtMC6* (At1g79320), *AtMC9* (At5g04200), *LACS6* (At3g05970), *ACX1* (At4g16760), *ETR1* (At1g66340), *ACO2* (At1g62380), *TUA2* (At1g50010), *TUA4* (At1g04820), *TUA5 (*At5g19780), *ACT3* (At3g53750), *ACT11* (At3g12110), *AtAURORA1 (*At4g32830), *AtAURORA2* (At2g25880), *SGR1/NYE1* (At4g22920), *SGR2* (At4g11910), *NYC1* (At4g13250), and *PaO* (At3g44880).

Author Contributions

All authors had a role in discussion of results. V.B.-W., J.B., K.D., B.T., S.J., D.R., S.O., and D.L.W. designed the research; E.B., E.H., C.Z., K.M., A.T., and C.J. did the experimental work; A.M. designed the array experiment; S.M., L.H., and A.M. developed microarray data extraction and primary analysis methods; S.M. designed Gene Viewer; S.M., D.L.W., and S.K. developed and applied clustering methods; Y.-s.K., C.H., V.B.-W., E.B., R.H., C.A.P., and D.J. performed data analysis and

biological interpretation; C.A.P. and D.L.W. developed the gradient analysis tool; R.H. and S.O. did the promoter motif analysis; D.J. developed the TF family analysis method; J.D.M. and R.L. organized data handling and web page; and V.B.-W. wrote the majority of the paper with input from E.B., E.H., S.M., A.M., C.H., Y.-s.K., R.H., S.O., D.J., C.A.P., S.K., K.D., J.D.M., and D.L.W.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Venn Diagram Summarizing Numbers of Genes Showing Significantly Different Expression for Different Combinations of Treatment Terms in the MAANOVA Fixed Model.

Supplemental Figure 2. Expression Patterns of Selected Genes Showing Time-of-Day Changes Only.

Supplemental Figure 3. Expression Patterns of Selected Genes during Leaf Senescence.

Supplemental Figure 4. Gradient Analysis on Selected Groups of Genes.

Supplemental Figure 5. Expression Patterns of Selected Hormone Biosynthesis Genes.

Supplemental Figure 6. Representation of the Experimental Design for the Microarray Experiment.

Supplemental Table 1. Enriched GO Terms in Genes Downregulated or Upregulated during Senescence.

Supplemental Table 2. Genes Used in the Variational Bayesian State Space Modeling Method Model Shown in Figure 9.

Supplemental Data Set 1. Genes Differentially Expressed during Senescence.

Supplemental Data Set 2. Enriched GO Terms in Each Cluster Shown in Figure 3.

Supplemental Data Set 3. Gradient Data for Differentially Expressed Genes.

Supplemental Data Set 4. Cluster Patterns Identified from the Gradient Analysis Showing Enriched GO terms.

Supplemental Data Set 5. Known DNA Sequence Motif Enrichment.

Supplemental Data Set 6. TF Family Analysis Data.

Supplemental Data Set 7. Senescence-Enhanced Genes Differentially Expressed in Inducible *ANAC092* Line.

Supplemental Methods 1. Microarray Experiments and GP Regression and Gradient.

ACKNOWLEDGMENTS

We thank Mary Coates (University of Warwick) for helping with manuscript preparation and Miriam Gifford (University of Warwick) for critical reading. We thank Mark Bennett (Imperial College) and Prof. Murray Grant (Exeter University) for performing the hormone measurements. E.B., E.H., C.Z., C.J., and K.M. were funded for this work by a Biotechnology and Biological Sciences Research Council (BBSRC) core strategic grant to Warwick HRI; S.M., S.K., and R.H. are funded by the Engineering and Physical Sciences Research Council/BBSRC-funded Warwick Systems Biology Doctoral Training Centre; L.H. was funded by a BBSRC studentship; J.B., K.D., V.B.-W., D.R., D.L.W., S.O., C.H., Y.-s.K., C.A.P., D.J., J.D.M., R.L., and A.T. are part of the BBRSC-funded grant Plant Response to Environmental Stress Arabidopsis (BB/F005806/1).

Received January 21, 2011; revised January 21, 2011; accepted February 28, 2011; published March 29, 2011.

REFERENCES

Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003). *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell **15:** 63–78.

Allemeersch, J., et al. (2005). Benchmarking the CATMA microarray. A novel tool for Arabidopsis transcriptome analysis. Plant Physiol. **137:** 588–601.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Andersson, A., et al. (2004). A transcriptional timetable of autumn senescence. Genome Biol. **5:** R24.

Balazadeh, S., Siddiqui, H., Allu, A.D., Matallana-Ramirez, L.P., Caldana, C., Mehrnia, M., Zanor, M.I., Köhler, B., and Mueller-Roeber, B. (2010). A gene regulatory network controlled by the NAC transcription factor ANAC092/AtNAC2/ORE1 during salt-promoted senescence. Plant J. **62:** 250–264.

Beal, M.J., Falciani, F., Ghahramani, Z., Rangel, C., and Wild, D.L. (2005). A Bayesian approach to reconstructing genetic regulatory networks with hidden factors. Bioinformatics **21:** 349–356.

Bellman, R. (1961). Adaptive Control Processes: A Guided Tour. (Princeton, NJ: Princeton University Press).

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B Stat. Methodol. **57:** 289–300.

Benjamini, Y., and Liu, W. (1999). A step-down multiple testing procedure that controls the false discovery rate under independence. J. Stat. Planning Inference **82:** 163–170.

Breeze, E., Harrison, E., Page, T., Warner, N., Shen, C., Zhang, C., and Buchanan-Wollaston, V. (2008). Transcriptional regulation of plant senescence: From functional genomics to systems biology. Plant Biol. (Stuttg.) **10** (Suppl. 1), 99–109.

Buchanan-Wollaston, V., Page, T., Harrison, E., Breeze, E., Lim, P.O., Nam, H.G., Lin, J.F., Wu, S.H., Swidzinski, J., Ishizaki, K., and Leaver, C.J. (2005). Comparative transcriptome analysis reveals significant differences in gene expression and signaling pathways between developmental and dark/starvation-induced senescence in Arabidopsis. Plant J. **42:** 567–585.

Choudhury, A., and Lahiri, A. (2011). Comparative analysis of abscisic acid-regulated transcriptomes in Arabidopsis. Plant Biol. (Stuttg.) **13:** 28–35.

Churchill, G.A. (2004). Using ANOVA to analyze microarray data. Biotechniques **37:** 173–175, 177.

Coll, N.S., Vercammen, D., Smidler, A., Clover, C., Van Breusegem, F., Dangl, J.L., and Epple, P. (2010). Arabidopsis type I metacaspases control cell death. Science **330:** 1393–1397.

Demidov, D., Hesse, S., Tewes, A., Rutten, T., Fuchs, J., Ashtiyani, R.K., Lein, S., Fischer, A., Reuter, G., and Houben, A. (2009). Aurora1 phosphorylation activity on histone H3 and its cross-talk with other post-translational histone modifications in Arabidopsis. Plant J. **59:** 221–230.

Demidov, D., Van Damme, D., Geelen, D., Blattner, F.R., and Houben, A. (2005). Identification and dynamics of two classes of aurora-like kinases in Arabidopsis and other plants. Plant Cell **17:** 836–848.

Dóczi, R., Brader, G., Pettkó-Szandtner, A., Rajh, I., Djamei, A.,

Pitzschke, A., Teige, M., and Hirt, H. (2007). The Arabidopsis mitogen-activated protein kinase kinase *MKK3* is upstream of group C mitogen-activated protein kinases and participates in pathogen signaling. Plant Cell **19:** 3266–3279.

Doelling, J.H., Walker, J.M., Friedman, E.M., Thompson, A.R., and Vierstra, R.D. (2002). The APG8/12-activating enzyme APG7 is required for proper nutrient recycling and senescence in Arabidopsis thaliana. J. Biol. Chem. **277:** 33105–33114.

Doherty, C.J., Van Buskirk, H.A., Myers, S.J., and Thomashow, M.F. (2009). Roles for Arabidopsis CAMTA transcription factors in cold-regulated gene expression and freezing tolerance. Plant Cell **21:** 972–984.

Eastwood, D.C., Mead, A., Sergeant, M.J., and Burton, K.S. (2008). Statistical modelling of transcript profiles of differentially regulated genes. BMC Mol. Biol. **9:** 66.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. **30:** 207–210.

Epple, P., Mack, A.A., Morris, V.R., and Dangl, J.L. (2003). Antagonistic control of oxidative stress-induced cell death in Arabidopsis by two related, plant-specific zinc finger proteins. Proc. Natl. Acad. Sci. USA **100:** 6831–6836.

Eulgem, T., Rushton, P.J., Robatzek, S., and Somssich, I.E. (2000). The WRKY superfamily of plant transcription factors. Trends Plant Sci. **5:** 199–206.

Eulgem, T., and Somssich, I.E. (2007). Networks of WRKY transcription factors in defense signaling. Curr. Opin. Plant Biol. **10:** 366–371.

Forcat, S., Bennett, M.H., Mansfield, J.W., and Grant, M.R. (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:** 16.

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

Fulda, M., Shockey, J., Werber, M., Wolter, F.P., and Heinz, E. (2002). Two long-chain acyl-CoA synthetases from *Arabidopsis thaliana* involved in peroxisomal fatty acid β-oxidation. Plant J. **32:** 93–103.

Gomez-Roldan, V., et al. (2008). Strigolactone inhibition of shoot branching. Nature **455:** 189–194.

Grbic, V., and Bleecker, A.B. (1995). Ethylene regulates the timing of leaf senescence in Arabidopsis. Plant J. **8:** 595–602.

Guo, Y., and Gan, S. (2006). *AtNAP*, a NAC family transcription factor, has an important role in leaf senescence. Plant J. **46:** 601–612.

Gutterson, N., and Reuber, T.L. (2004). Regulation of disease resistance pathways by AP2/ERF transcription factors. Curr. Opin. Plant Biol. **7:** 465–471.

Hanaoka, H., Noda, T., Shirano, Y., Kato, T., Hayashi, H., Shibata, D., Tabata, S., and Ohsumi, Y. (2002). Leaf senescence and starvation-induced chlorosis are accelerated by the disruption of an Arabidopsis autophagy gene. Plant Physiol. **129:** 1181–1193.

Harmer, S.L. (2009). The circadian system in higher plants. Annu. Rev. Plant Biol. **60:** 357–377.

Hazen, S.P., Schultz, T.F., Pruneda-Paz, J.L., Borevitz, J.O., Ecker, J.R., and Kay, S.A. (2005). *LUX ARRHYTHMO* encodes a Myb domain protein essential for circadian rhythms. Proc. Natl. Acad. Sci. USA **102:** 10387–10392.

He, Y., Fukushige, H., Hildebrand, D.F., and Gan, S. (2002). Evidence supporting a role of jasmonic acid in Arabidopsis leaf senescence. Plant Physiol. **128:** 876–884.

Heard, N.A. (2011). Iterative reclassification in agglomerative clustering. J. Comput. Graph. Stat. http://dx.doi.org/10.1198/jcgs.2011.09111

Heard, N.A., Holmes, C.C., and Stephens, D.A. (2006). A quantitative study of gene regulation involved in the immune response of anopheline mosquitoes: An application of Bayesian hierarchical clustering of curves. JASA **101:** 18–29.

Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res. **27:** 297–300.

Hörtensteiner, S. (2009). Stay-green regulates chlorophyll and chlorophyll-binding protein degradation during senescence. Trends Plant Sci. **14:** 155–162.

Hörtensteiner, S., and Feller, U. (2002). Nitrogen metabolism and remobilization during senescence. J. Exp. Bot. **53:** 927–937.

Hwang, I., and Sheen, J. (2001). Two-component circuitry in Arabidopsis cytokinin signal transduction. Nature **413:** 383–389.

Ichimura, Y., Kirisako, T., Takao, T., Satomi, Y., Shimonishi, Y., Ishihara, N., Mizushima, N., Tanida, I., Kominami, E., Ohsumi, M., Noda, T., and Ohsumi, Y. (2000). A ubiquitin-like system mediates protein lipidation. Nature **408:** 488–492.

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., and Parcy, F.; bZIP Research Group. (2002). bZIP transcription factors in *Arabidopsis.* Trends Plant Sci. **7:** 106–111.

Kel, A.E., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V., and Wingender, E. (2003). MATCHTM: A tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res. **13:** 3576–3579.

Kerr, M.K., and Churchill, G.A. (2001). Experimental design for gene expression microarrays. Biostatistics **2:** 183–201.

Ketelaar, T., Voss, C., Dimmock, S.A., Thumm, M., and Hussey, P.J. (2004). Arabidopsis homologues of the autophagy protein Atg8 are a novel family of microtubule binding proteins. FEBS Lett. **567:** 302–306.

Kizis, D., Lumbreras, V., and Pagès, M. (2001). Role of AP2/EREBP transcription factors in gene regulation during abiotic stress. FEBS Lett. **498:** 187–189.

Kikis, E.A., Khanna, R., and Quail, P.H. (2005). *ELF4* is a phytochrome-regulated component of a negative-feedback loop involving the central oscillator components *CCA1* and *LHY.* Plant J. **44:** 300–313.

Kim, H.J., Ryu, H., Hong, S.H., Woo, H.R., Lim, P.O., Lee, I.C., Sheen, J., Nam, H.G., and Hwang, I. (2006). Cytokinin-mediated control of leaf longevity by *AHK3* through phosphorylation of ARR2 in Arabidopsis. Proc. Natl. Acad. Sci. USA **103:** 814–819.

Kim, J.H., Woo, H.R., Kim, J., Lim, P.O., Lee, I.C., Choi, S.H., Hwang, D., and Nam, H.G. (2009). Trifurcate feed-forward regulation of age-dependent cell death involving miR164 in Arabidopsis. Science **323:** 1053–1057.

Kim, J.M., and DellaPenna, D. (2006). Defining the primary route for lutein synthesis in plants: The role of Arabidopsis carotenoid β-ring hydroxylase CYP97A3. Proc. Natl. Acad. Sci. USA **103:** 3474–3479.

Kim, J.M., To, T.K., Nishioka, T., and Seki, M. (2010). Chromatin regulation functions in plant abiotic stress responses. Plant Cell Environ. **33:** 604–611.

Lee, E.J., Matsumura, Y., Soga, K., Hoson, T., and Koizumi, N. (2007). Glycosyl hydrolases of cell wall are induced by sugar starvation in Arabidopsis. Plant Cell Physiol. **48:** 405–413.

Lenhard, B., and Wasserman, W.W. (2002). TFBS: Computational framework for transcription factor binding site analysis. Bioinformatics **18:** 1135–1136.

Li, C., Potuschak, T., Colón-Carmona, A., Gutiérrez, R.A., and

Doerner, P. (2005). Arabidopsis TCP20 links regulation of growth and cell division control pathways. Proc. Natl. Acad. Sci. USA **102:** 12978–12983.

Lim, P.O., Kim, H.J., and Nam, H.G. (2007). Leaf senescence. Annu. Rev. Plant Biol. **58:** 115–136.

Liu, J.X., and Howell, S.H. (2010). *bZIP28* and NF-Y transcription factors are activated by ER stress and assemble into a transcriptional complex to regulate stress response genes in Arabidopsis. Plant Cell **22:** 782–796.

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics **21:** 3448–3449.

Martínez, D.E., Costa, M.L., and Guiamet, J.J. (2008). Senescence-associated degradation of chloroplast proteins inside and outside the organelle. Plant Biol. (Stuttg.) **10** (Suppl. 1), 15–22.

Martínez-García, J.F., Huq, E., and Quail, P.H. (2000). Direct targeting of light signals to a promoter element-bound transcription factor. Science **288:** 859–863.

Matys, V., et al. (2006). TRANSFAC and its module TRANSCompel: Transcriptional gene regulation in eukaryotes. Nucleic Acids Res. **34** (Database issue): D108–D110.

Meierhoff, K., Felder, S., Nakamura, T., Bechtold, N., and Schuster, G. (2003). HCF152, an *Arabidopsis* RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast *psbB-psbT-psbH-petB-petD* RNAs. Plant Cell **15:** 1480–1495.

Meurer, J., Meierhoff, K., and Westhoff, P. (1996). Isolation of high-chlorophyll-fluorescence mutants of *Arabidopsis thaliana* and their characterisation by spectroscopy, immunoblotting and northern hybridisation. Planta **198:** 385–396.

Miao, Y., Laun, T., Zimmermann, P., and Zentgraf, U. (2004). Targets of the *WRKY53* transcription factor and its role during leaf senescence in Arabidopsis. Plant Mol. Biol. **55:** 853–867.

Mittler, R., Kim, Y., Song, L., Coutu, J., Coutu, A., Ciftci-Yilmaz, S., Lee, H., Stevenson, B., and Zhu, J.K. (2006). Gain- and loss-of-function mutations in Zat10 enhance the tolerance of plants to abiotic stress. FEBS Lett. **580:** 6537–6542.

Monastyrska, I., Rieter, E., Klionsky, D.J., and Reggiori, F. (2009). Multiple roles of the cytoskeleton in autophagy. Biol. Rev. Camb. Philos. Soc. **84:** 431–448.

Morris, K.A.H., MacKerness, S.A., Page, T., John, C.F., Murphy, A.M., Carr, J.P., and Buchanan-Wollaston, V. (2000). Salicylic acid has a role in regulating gene expression during leaf senescence. Plant J. **23:** 677–685.

Nakashima, K., Fujita, Y., Katsura, K., Maruyama, K., Narusaka, Y., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2006). Transcriptional regulation of ABI3- and ABA-responsive genes including RD29B and RD29A in seeds, germinating embryos, and seedlings of Arabidopsis. Plant Mol. Biol. **60:** 51–68.

Noodén, L.D., Singh, S., and Letham, D.S. (1990). Correlation of xylem sap cytokinin levels with monocarpic senescence in soybean. Plant Physiol. **93:** 33–39.

Olsen, A.N., Ernst, H.A., Leggio, L.L., and Skriver, K. (2005). NAC transcription factors: Structurally distinct, functionally diverse. Trends Plant Sci. **10:** 79–87.

Ongaro, V., and Leyser, O. (2008). Hormonal control of shoot branching. J. Exp. Bot. **59:** 67–74.

Palaniswamy, S.K., James, S., Sun, H., Lamb, R.S., Davuluri, R.V., and Grotewold, E. (2006). AGRIS and AtRegNet. A platform to link cis-regulatory elements and transcription factors into regulatory networks. Plant Physiol. **140:** 818–829.

Prigent, C., and Dimitrov, S. (2003). Phosphorylation of serine 10 in histone H3, what for? J. Cell Sci. **116:** 3677–3685.

Ramirez-Parra, E., Fründt, C., and Gutierrez, C. (2003). A genome-wide identification of E2F-regulated genes in Arabidopsis. Plant J. **33:** 801–811.

Robatzek, S., and Somssich, I.E. (2001). A new member of the Arabidopsis WRKY transcription factor family, AtWRKY6, is associated with both senescence- and defence-related processes. Plant J. **28:** 123–133.

Sakuma, Y., Maruyama, K., Qin, F., Osakabe, Y., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2006). Dual function of an Arabidopsis transcription factor *DREB2A* in water-stress-responsive and heat-stress-responsive gene expression. Proc. Natl. Acad. Sci. USA **103:** 18822–18827.

Shockey, J.M., Fulda, M.S., and Browse, J.A. (2002). Arabidopsis contains nine long-chain acyl-coenzyme a synthetase genes that participate in fatty acid and glycerolipid metabolism. Plant Physiol. **129:** 1710–1722.

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

Staswick, P.E. (1994). Storage proteins of vegetative plant tissues. Annu. Rev. Plant Physiol. Plant Mol. Biol. **45:** 303–322.

Staswick, P.E. (2008). JAZing up jasmonate signaling. Trends Plant Sci. **13:** 66–71.

Tanaka, R., and Tanaka, A. (2007). Tetrapyrrole biosynthesis in higher plants. Annu. Rev. Plant Biol. **58:** 321–346.

Toledo-Ortiz, G., Huq, E., and Quail, P.H. (2003). The Arabidopsis basic/helix-loop-helix transcription factor family. Plant Cell **15:** 1749–1770.

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

Ülker, B., and Somssich, I.E. (2004). WRKY transcription factors: From DNA binding towards biological function. Curr. Opin. Plant Biol. **7:** 491–498.

van der Graaff, E., Schwacke, R., Schneider, A., Desimone, M., Flügge, U.I., and Kunze, R. (2006). Transcription analysis of Arabidopsis membrane transporters and hormone pathways during developmental and induced leaf senescence. Plant Physiol. **141:** 776–792.

Walley, J.W., Coughlan, S., Hudson, M.E., Covington, M.F., Kaspi, R., Banu, G., Harmer, S.L., and Dehesh, K. (2007). Mechanical stress induces biotic and abiotic stress responses via a novel cis-element. PLoS Genet. **3:** 1800–1812.

Waters, M.T., Wang, P., Korkaric, M., Capper, R.G., Saunders, N.J., and Langdale, J.A. (2009). GLK transcription factors coordinate expression of the photosynthetic apparatus in Arabidopsis. Plant Cell **21:** 1109–1128.

Weaver, L.M., Gan, S., Quirino, B., and Amasino, R.M. (1998). A comparison of the expression patterns of several senescence-associated genes in response to stress and hormone treatment. Plant Mol. Biol. **37:** 455–469.

Weisshaar, B., and Jenkins, G.I. (1998). Phenylpropanoid biosynthesis and its regulation. Curr. Opin. Plant Biol. **1:** 251–257.

Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierec, J., Samach, A., and Coupland, G. (2006). *CONSTANS* and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. Plant Cell **18:** 2971–2984.

Westfall, P.H., Krishen, A., and Young, S.S. (1998). Using prior

information to allocate significance levels for multiple endpoints. Stat. Med. **17:** 2107–2119.

Woo, H.R., Chung, K.M., Park, J.-H., Oh, S.A., Ahn, T., Hong, S.H., Jang, S.K., and Nam, H.G. (2001). ORE9, an F-box protein that regulates leaf senescence in Arabidopsis. Plant Cell **13:** 1779–1790.

Wu, H., Kerr, K., Cui, X., and Churchill, G. (2003). MAANOVA: A software package for the analysis of spotted cDNA microarray experiments. In The Analysis of Gene Expression Data: Methods and Software, G. Parmigiani, E. Garett, R. Irizarry, and S. Zeger, eds. (New York: Springer), pp. 313–341.

Yang, T., and Poovaiah, B.W. (2002). A calmodulin-binding/CGCG box DNA-binding protein family involved in multiple signaling pathways in plants. J. Biol. Chem. **277:** 45049–45058.

Yoo, J.H., et al. (2005). Direct interaction of a divergent CaM isoform and the transcription factor, MYB2, enhances salt tolerance in Arabidopsis. J. Biol. Chem. **280:** 3697–3706.

Yoshimoto, K., Hanaoka, H., Sato, S., Kato, T., Tabata, S., Noda, T., and Ohsumi, Y. (2004). Processing of ATG8s, ubiquitin-like proteins, and their deconjugation by ATG4s are essential for plant autophagy. Plant Cell **16:** 2967–2983.

Bibliography

Abdeen, A., Schnell, J., and Miki, B. (2010). Transcriptome analysis reveals absence of unintended effects in drought-tolerant transgenic plants overexpressing the transcription factor ABF3. *BMC Genomics*, **11**, 69.

Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D., and Shinozaki, K. (1997). Role of arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. *The Plant Cell*, **9**(10), 1859–68.

Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003). Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *The Plant Cell*, **15**(1), 63–78.

AbuQamar, S., Chen, X., Dhawan, R., Bluhm, B., Salmeron, J., Lam, S., Dietrich, R. A., and Mengiste, T. (2006). Expression profiling and mutant analysis reveals complex regulatory networks involved in Arabidopsis response to Botrytis infection. *Plant J*, **48**(1), 28–44.

Aggarwal, P., Padmanabhan, B., Bhat, A., Sarvepalli, K., Sadhale, P. P., and Nath, U. (2011). The TCP4 transcription factor of Arabidopsis blocks cell division in yeast at G1S transition. *Biochem Biophys Res Commun*, **410**(2), 276–81.

Aida, M., Ishida, T., Fukaki, H., Fujisawa, H., and Tasaka, M. (1997). Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant. *The Plant Cell*, **9**(6), 841–57.

Alonso, J. M., Hirayama, T., Roman, G., Nourizadeh, S., and Ecker, J. R. (1999). EIN2, a bifunctional transducer of ethylene and stress responses in Arabidopsis. *Science*, **284**(5423), 2148–52.

Anderson, J. P., Badruzsaufari, E., Schenk, P. M., Manners, J. M., Desmond, O. J., Ehlert, C., Maclean, D. J., Ebert, P. R., and Kazan, K. (2004). Antagonistic interaction between abscisic acid and jasmonate-ethylene signaling pathways modulates defense gene expression and disease resistance in Arabidopsis. *The Plant Cell*, **16**(12), 3460–79.

Arda, H. E., Taubert, S., MacNeil, L. T., Conine, C. C., Tsuda, B., Gilst, M. V., Sequerra, R., Doucette-Stamm, L., Yamamoto, K. R., and Walhout, A. J. M. (2010). Functional modularity of nuclear hormone receptors in a Caenorhabditis elegans metabolic gene regulatory network. *Mol Syst Biol*, **6**, 367.

Asselbergh, B., Vleesschauwer, D. D., and Höfte, M. (2008). Global switches and fine-tuning-ABA modulates plant pathogen defense. *Mol Plant Microbe Interact*, **21**(6), 709–19.

Aubert, Y., Vile, D., Pervent, M., Aldon, D., Ranty, B., Simonneau, T., Vavasseur, A., and Galaud, J.-P. (2010). RD20, a stress-inducible caleosin, participates in stomatal control, transpiration and drought tolerance in Arabidopsis thaliana. *Plant Cell Physiol*, **51**(12), 1975–87.

Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research*, **34**(Web Server issue), W369–73.

Balazadeh, S., Siddiqui, H., Allu, A. D., Matallana-Ramirez, L. P., Caldana, C., Mehrnia, M., Zanor, M.-I., Köhler, B., and Mueller-Roeber, B. (2010a). A gene regulatory network controlled by the NAC transcription factor ANAC092/AtNAC2/ORE1 during salt-promoted senescence. *Plant J*, **62**(2), 250–64.

Balazadeh, S., Wu, A., and Mueller-Roeber, B. (2010b). Salt-triggered expression of the ANAC092-dependent senescence regulon in Arabidopsis thaliana. *Plant Signal Behav*, **5**(6).

Balazadeh, S., Kwasniewski, M., Caldana, C., Mehrnia, M., Zanor, M. I., Xue, G.-P., and Mueller-Roeber, B. (2011). ORS1, an H2O2-Responsive NAC Transcription Factor, Controls Senescence in Arabidopsis thaliana. *Molecular Plant*, **4**(2), 346–60.

Baniwal, S. K., Bharti, K., Chan, K. Y., Fauth, M., Ganguli, A., Kotak, S., Mishra, S. K., Nover, L., Port, M., Scharf, K.-D., Tripp, J., Weber, C., Zielinski, D., and von Koskull-Döring, P. (2004). Heat stress response in plants: a complex game with chaperones and more than twenty heat stress transcription factors. *J Biosci*, **29**(4), 471–87.

Beal, M. J., Falciani, F., Ghahramani, Z., Rangel, C., and Wild, D. L. (2005). A Bayesian approach to reconstructing genetic regulatory networks with hidden factors. *Bioinformatics*, **21**(3), 349–56.

Bednarek, P., Pislewska-Bednarek, M., Svatos, A., Schneider, B., Doubsky, J., Mansurova, M., Humphry, M., Consonni, C., Panstruga, R., Sanchez-Vallet, A., Molina, A., and Schulze-Lefert, P. (2009). A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science*, **323**(5910), 101–6.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, **57**(1), 289–300.

Bilgin, D. D., Zavala, J. A., Zhu, J., Clough, S. J., Ort, D. R., and DeLucia, E. H. (2010). Biotic stress globally downregulates photosynthesis genes. *Plant Cell Environ*, **33**(10), 1597–613.

Böttcher, C., Westphal, L., Schmotz, C., Prade, E., Scheel, D., and Glawischnig, E. (2009). The multifunctional enzyme CYP71B15 (PHYTOALEXIN DEFICIENT3) converts cysteine-indole-3-acetonitrile to camalexin in the indole-3-acetonitrile metabolic network of Arabidopsis thaliana. *The Plant Cell*, **21**(6), 1830–45.

Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., Furey, T. S., and Crawford, G. E. (2008). High-resolution mapping and characterization of open chromatin across the genome. *Cell*, **132**(2), 311–22.

Boyle, A. P., Song, L., Lee, B.-K., London, D., Keefe, D., Birney, E., Iyer, V. R., Crawford, G. E., and Furey, T. S. (2011). High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. *Genome Res*, **21**(3), 456–64.

Brady, S. M., Zhang, L., Megraw, M., Martinez, N. J., Jiang, E., Yi, C. S., Liu, W., Zeng, A., Taylor-Teeples, M., Kim, D., Ahnert, S., Ohler, U., Ware, D., Walhout, A. J. M., and Benfey, P. N. (2011). A stele-enriched gene regulatory network in the Arabidopsis root. *Mol Syst Biol*, **7**(459).

Breeze, E., Harrison, E., McHattie, S., Hughes, L., Hickman, R., Hill, C., Kiddle, S., Kim, Y.-S., Penfold, C. A., Jenkins, D., Zhang, C., Morris, K., Jenner, C., Jackson, S., Thomas, B., Tabrett, A., Legaie, R., Moore, J. D., Wild, D. L., Ott, S., Rand, D., Beynon, J., Denby, K., Mead, A., and Buchanan-Wollaston, V. (2011). High-resolution temporal profiling of transcripts during Arabidopsis leaf senescence reveals a distinct chronology of processes and regulation. *The Plant Cell*, **23**(3), 873–94.

Bu, Q., Jiang, H., Li, C.-B., Zhai, Q., Zhang, J., Wu, X., Sun, J., Xie, Q., and Li, C. (2008). Role of the Arabidopsis thaliana NAC transcription factors ANAC019 and ANAC055 in regulating jasmonic acid-signaled defense responses. *Cell Res*, **18**(7), 756–67.

Buchanan-Wollaston, V. (1997). The molecular biology of leaf senescence. *Journal of Experimental Botany*, **42**(8), 181–199.

Buchanan-Wollaston, V., Page, T., Harrison, E., Breeze, E., Lim, P. O., Nam, H. G., Lin, J.-F., Wu, S.-H., Swidzinski, J., Ishizaki, K., and Leaver, C. J. (2005). Comparative transcriptome analysis reveals significant differences in gene expression and signalling pathways between developmental and dark/starvation-induced senescence in Arabidopsis. *Plant J*, **42**(4), 567–85.

Carviel, J. L., Al-Daoud, F., Neumann, M., Mohammad, A., Provart, N. J., Moeder, W., Yoshioka, K., and Cameron, R. K. (2009). Forward and reverse genetics to identify genes involved in the age-related resistance response in Arabidopsis thaliana. *Mol Plant Pathol*, **10**(5), 621–34.

Cashmore, A. (1990). Mutation of either G box or I box sequences profoundly affects expression from the Arabidopsis rbcS-1A promoter. *The EMBO Journal*, **9**(6), 1717–26.

Castrillo, G., Turck, F., Leveugle, M., Lecharny, A., Carbonero, P., Coupland, G., Paz-Ares, J., and Oñate-Sánchez, L. (2011). Speeding cis-trans regulation discovery by phylogenomic analyses coupled with screenings of an arrayed library of Arabidopsis transcription factors. *PLoS ONE*, **6**(6), e21524.

Century, K., Reuber, T. L., and Ratcliffe, O. J. (2008). Regulating the regulators: the future prospects for transcription-factor-based agricultural biotechnology products. *Plant Physiol*, **147**(1), 20–9.

Chen, C. Y. and Schwartz, R. J. (1995). Identification of novel DNA binding targets and regulatory domains of a murine tinman homeodomain factor, nkx-2.5. *J Biol Chem*, **270**(26), 15628–33.

Chen, H., Hwang, J. E., Lim, C. J., Kim, D. Y., Lee, S. Y., and Lim, C. O. (2010). Arabidopsis DREB2C functions as a transcriptional activator of HsfA3 during the heat stress response. *Biochem Biophys Res Commun*, **401**(2), 238–44.

Chen, M.-K., Hsu, W.-H., Lee, P.-F., Thiruvengadam, M., Chen, H.-I., and Yang, C.-H. (2011). The MADS box gene, FOREVER YOUNG FLOWER, acts as a repressor controlling floral organ senescence and abscission in Arabidopsis. *Plant J*, **68**(1), 168–85.

Chini, A., Fonseca, S., Fernández, G., Adie, B., Chico, J. M., Lorenzo, O., García-Casado, G., López-Vidriero, I., Lozano, F. M., Ponce, M. R., Micol, J. L., and Solano, R. (2007). The JAZ family of repressors is the missing link in jasmonate signalling. *Nature*, **448**(7154), 666–71.

Choi, H., Hong, J., Ha, J., Kang, J., and Kim, S. Y. (2000). ABFs, a family of ABA-responsive element binding factors. *J Biol Chem*, **275**(3), 1723–30.

Clay, N. K., Adio, A. M., Denoux, C., Jander, G., and Ausubel, F. M. (2009). Glucosinolate metabolites required for an Arabidopsis innate immune response. *Science*, **323**(5910), 95–101.

Cooper, S. J., Trinklein, N. D., Anton, E. D., Nguyen, L., and Myers, R. M. (2006). Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res*, **16**(1), 1–10.

Cubas, P., Lauter, N., Doebley, J., and Coen, E. (1999). The TCP domain: a motif found in proteins regulating plant growth and development. *Plant J*, **18**(2), 215–22.

Cutler, S. R., Rodriguez, P. L., Finkelstein, R. R., and Abrams, S. R. (2010). Abscisic acid: emergence of a core signaling network. *ANNUAL REVIEW OF PLANT BIOLOGY*, **61**, 651–79.

de Torres-Zabala, M., Truman, W., Bennett, M. H., Lafforgue, G., Mansfield, J. W., Egea, P. R., Bögre, L., and Grant, M. (2007). Pseudomonas syringae pv. tomato hijacks the Arabidopsis abscisic acid signalling pathway to cause disease. *The EMBO Journal*, **26**(5), 1434–43.

Denby, K. J., Kumar, P., and Kliebenstein, D. J. (2004). Identification of Botrytis cinerea susceptibility loci in Arabidopsis thaliana. *Plant J*, **38**(3), 473–86.

Deplancke, B., Dupuy, D., Vidal, M., and Walhout, A. J. M. (2004). A gateway-compatible yeast one-hybrid system. *Genome Res*, **14**(10B), 2093–101.

Deplancke, B., Mukhopadhyay, A., Ao, W., Elewa, A. M., Grove, C. A., Martinez, N. J., Sequerra, R., Doucette-Stamm, L., Reece-Hoyes, J. S., Hope, I. A., Tissenbaum, H. A., Mango, S. E., and Walhout, A. J. M. (2006). A gene-centered C. elegans protein-DNA interaction network. *Cell*, **125**(6), 1193–205.

Djamei, A., Pitzschke, A., Nakagami, H., Rajh, I., and Hirt, H. (2007). Trojan horse strategy in Agrobacterium transformation: abusing MAPK defense signaling. *Science*, **318**(5849), 453–6.

Dobi, K. C. and Winston, F. (2007). Analysis of transcriptional activation at a distance in Saccharomyces cerevisiae. *Mol Cell Biol*, **27**(15), 5575–86.

Doebley, J., Stec, A., and Gustus, C. (1995). teosinte branched1 and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics*, **141**(1), 333–46.

Doebley, J., Stec, A., and Hubbard, L. (1997). The evolution of apical dominance in maize. *Nature*, **386**(6624), 485–8.

Dombrecht, B., Xue, G. P., Sprague, S. J., Kirkegaard, J. A., Ross, J. J., Reid, J. B., Fitt, G. P., Sewelam, N., Schenk, P. M., Manners, J. M., and Kazan, K. (2007). MYC2 differentially modulates diverse jasmonate-dependent functions in Arabidopsis. *The Plant Cell*, **19**(7), 2225–45.

Dong, X. (2004). NPR1, all things considered. *Curr Opin Plant Biol*, **7**(5), 547–52.

Dreze, M., Arabidopsis, and Consortium, I. M. (2011). Evidence for network evolution in an Arabidopsis interactome map. *Science*, **333**(6042), 601–7.

Ernst, H., Olsen, A., Skriver, K., Larsen, S., and Leggio, L. (2004). Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors. *EMBO Reports*, **5**(3), 297–303.

Fang, Y., You, J., Xie, K., Xie, W., and Xiong, L. (2008). Systematic sequence analysis and identification of tissue-specific or stress-responsive genes of NAC transcription factor family in rice. *Mol Genet Genomics*, **280**(6), 547–63.

Feder, M. E. and Hofmann, G. E. (1999). Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annu Rev Physiol*, **61**, 243–82.

Ferrari, S., Plotnikova, J. M., Lorenzo, G. D., and Ausubel, F. M. (2003). Arabidopsis local resistance to Botrytis cinerea involves salicylic acid and camalexin and requires EDS4 and PAD2, but not SID2, EDS5 or PAD4. *Plant J*, **35**(2), 193–205.

Finkelstein, R. R., Wang, M. L., Lynch, T. J., Rao, S., and Goodman, H. M. (1998). The Arabidopsis abscisic acid response locus ABI4 encodes an APETALA 2 domain protein. *The Plant Cell*, **10**(6), 1043–54.

Flors, V., Ton, J., van Doorn, R., Jakab, G., García-Agustín, P., and Mauch-Mani, B. (2008). Interplay between JA, SA and ABA signalling during basal and induced resistance against Pseudomonas syringae and Alternaria brassicicola. *Plant J*, **54**(1), 81–92.

Flynt, A. S. and Lai, E. C. (2008). Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nature Reviews Genetics*, **9**(11), 831–42.

Fujii, H., Chiou, T.-J., Lin, S.-I., Aung, K., and Zhu, J.-K. (2005). A miRNA involved in phosphate-starvation response in Arabidopsis. *Curr Biol*, **15**(22), 2038–43.

Fujita, M., Fujita, Y., Maruyama, K., Seki, M., Hiratsu, K., Ohme-Takagi, M., Tran, L.-S. P., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2004). A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. *Plant J*, **39**(6), 863–76.

Gao, Q.-M., Venugopal, S., Navarre, D., and Kachroo, A. (2011). Low oleic acid-derived repression of jasmonic acid-inducible defense responses requires the WRKY50 and WRKY51 proteins. *Plant Physiol*, **155**(1), 464–76.

Gepstein, S. and Thimann, K. V. (1980). Changes in the abscisic acid content of oat leaves during senescence. *PNAS*, **77**(4), 2050–3.

Giraud, E., Ng, S., Carrie, C., Duncan, O., Low, J., Lee, C. P., Aken, O. V., Millar, A. H., Murcha, M., and Whelan, J. (2010). TCP transcription factors link the regulation of genes encoding mitochondrial proteins with the circadian clock in Arabidopsis thaliana. *The Plant Cell*, **22**(12), 3921–34.

Glazebrook, J. (2005). Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu Rev Phytopathol*, **43**, 205–27.

Goda, H., Sasaki, E., Akiyama, K., Maruyama-Nakashita, A., Nakabayashi, K., Li, W., Ogawa, M., Yamauchi, Y., Preston, J., Aoki, K., Kiba, T., Takatsuto, S., Fujioka, S., Asami, T., Nakano, T., Kato, H., Mizuno, T., Sakakibara, H., Yamaguchi, S., Nambara, E., Kamiya, Y., Takahashi, H., Hirai, M. Y., Sakurai, T., Shinozaki, K., Saito, K., Yoshida, S., and Shimada, Y. (2008). The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *Plant J*, **55**(3), 526–42.

Godoy, M., Franco-Zorrilla, J. M., Pérez-Pérez, J., Oliveros, J. C., Lorenzo, O., and Solano, R. (2011). Improved protein-binding microarrays for the identification of DNA-binding specificities of transcription factors. *Plant J*, **66**(4), 700–11.

Govrin, E. M. and Levine, A. (2000). The hypersensitive response facilitates plant infection by the necrotrophic pathogen Botrytis cinerea. *Curr Biol*, **10**(13), 751–7.

Grove, C. A., Masi, F. D., Barrasa, M. I., Newburger, D. E., Alkema, M. J., Bulyk, M. L., and Walhout, A. J. M. (2009). A multiparameter network reveals extensive divergence between C. elegans bHLH transcription factors. *Cell*, **138**(2), 314–27.

Guo, Y. and Gan, S. (2006). AtNAP, a NAC family transcription factor, has an important role in leaf senescence. *Plant J*, **46**(4), 601–12.

Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., and Noble, W. S. (2007). Quantifying similarity between motifs. *Genome Biol*, **8**(2), R24.

Ha, M., Kim, E.-D., and Chen, Z. J. (2009). Duplicate genes increase expression diversity in closely related species and allopolyploids. *PNAS*, **106**(7), 2295–300.

Haake, V., Cook, D., Riechmann, J. L., Pineda, O., Thomashow, M. F., and Zhang, J. Z. (2002). Transcription factor CBF4 is a regulator of drought adaptation in Arabidopsis. *Plant Physiol*, **130**(2), 639–48.

Hamberg, M., Sanz, A., and Castresana, C. (1999). alpha-oxidation of fatty acids in higher plants. Identification of a pathogen-inducible oxygenase (piox) as an alpha-dioxygenase and biosynthesis of 2-hydroperoxylinolenic acid. *J Biol Chem*, **274**(35), 24503–13.

Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., Hannett, N. M., Tagne, J.-B., Reynolds, D. B., Yoo, J., Jennings, E. G., Zeitlinger, J., Pokholok, D. K., Kellis, M., Rolfe, P. A., Takusagawa, K. T., Lander, E. S., Gifford, D. K., Fraenkel, E., and Young, R. A. (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature*, **431**(7004), 99–104.

Hardtke, C. S., Gohda, K., Osterlund, M. T., Oyama, T., Okada, K., and Deng, X. W. (2000). HY5 stability and activity in arabidopsis is regulated by phosphorylation in its COP1 binding domain. *The EMBO Journal*, **19**(18), 4997–5006.

He, X.-J., Mu, R.-L., Cao, W.-H., Zhang, Z.-G., Zhang, J.-S., and Chen, S.-Y. (2005). AtNAC2, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development. *Plant J*, **44**(6), 903–16.

Heard, N. and Holmes, C. (2006). A Quantitative Study of Gene Regulation Involved in the Immune Response of Anopheline Mosquitoes. *Journal of the American Statistical Association*, **101**(473), 18–29.

Hertz, G. Z. and Stormo, G. D. (1999). Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics*, **15**(7-8), 563–77.

Hervé, C., Dabos, P., Bardet, C., Jauneau, A., Auriac, M. C., Ramboer, A., Lacout, F., and Tremousaygue, D. (2009). In vivo interference with AtTCP20 function induces severe plant growth alterations and deregulates the expression of many genes important for development. *Plant Physiol*, **149**(3), 1462–77.

Hesselberth, J. R., Chen, X., Zhang, Z., Sabo, P. J., Sandstrom, R., Reynolds, A. P., Thurman, R. E., Neph, S., Kuehn, M. S., Noble, W. S., Fields, S., and Stamatoyannopoulos, J. A. (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat Methods*, **6**(4), 283–9.

Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Research*, **27**(1), 297–300.

Hoeren, F. U., Dolferus, R., Wu, Y., Peacock, W. J., and Dennis, E. S. (1998). Evidence for a role for AtMYB2 in the induction of the Arabidopsis alcohol dehydrogenase gene (ADH1) by low oxygen. *Genetics*, **149**(2), 479–90.

Hu, H., Dai, M., Yao, J., Xiao, B., Li, X., Zhang, Q., and Xiong, L. (2006). Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. *PNAS*, **103**(35), 12987–92.

Hu, Z., Killion, P. J., and Iyer, V. R. (2007). Genetic reconstruction of a functional transcriptional regulatory network. *Nature Genetics*, **39**(5), 683–7.

Hubbard, T. J. P., Aken, B. L., Ayling, S., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Clarke, L., Coates, G., Fairley, S., Fitzgerald, S., Fernandez-Banet, J., Gordon, L., Graf, S., Haider, S., Hammond, M., Holland, R., Howe, K., Jenkinson, A., Johnson, N., Kahari, A., Keefe, D., Keenan, S., Kinsella, R., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Megy, K., Meidl, P., Overduin, B., Parker, A., Pritchard, B., Rios, D., Schuster, M., Slater, G., Smedley, D., Spooner, W., Spudich, G., Trevanion, S., Vilella, A., Vogel, J., White, S., Wilder, S., Zadissa, A., Birney, E., Cunningham, F., Curwen, V., Durbin, R., Fernandez-Suarez, X. M., Herrero, J., Kasprzyk, A., Proctor, G., Smith, J., Searle, S., and Flicek, P. (2009). Ensembl 2009. *Nucleic Acids Res*, **37**(Database issue), D690–7.

Hughes, J. D., Estep, P. W., Tavazoie, S., and Church, G. M. (2000). Computational identification of cis-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae. *J Mol Biol*, **296**(5), 1205–14.

Jiang, Y. and Deyholos, M. K. (2009). Functional characterization of Arabidopsis NaCl-inducible WRKY25 and WRKY33 transcription factors in abiotic stresses. *Plant Mol Biol*, **69**(1-2), 91–105.

Jung, C., Lyou, S. H., Yeu, S., Kim, M. A., Rhee, S., Kim, M., Lee, J. S., Choi, Y. D., and Cheong, J.-J. (2007). Microarray-based screening of jasmonate-responsive genes in Arabidopsis thaliana. *Plant Cell Reports*, **26**(7), 1053–63.

Kato, H., Motomura, T., Komeda, Y., Saito, T., and Kato, A. (2010). Overexpression of the NAC transcription factor family gene ANAC036 results in a dwarf phenotype in Arabidopsis thaliana. *J Plant Physiol*, **167**(7), 571–7.

Katsir, L., Chung, H. S., Koo, A. J. K., and Howe, G. A. (2008). Jasmonate signaling: a conserved mechanism of hormone sensing. *Curr Opin Plant Biol*, **11**(4), 428–35.

Kaufmann, K., Muiño, J. M., Østerås, M., Farinelli, L., Krajewski, P., and Angenent, G. C. (2010). Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). *Nat Protoc*, **5**(3), 457–72.

Kel, A., Gossling, E., and Reuter, I. (2003). MATCHTM: a tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Research*, **31**(13), 3576–3579.

Kerchev, P. I., Pellny, T. K., Vivancos, P. D., Kiddle, G., Hedden, P., Driscoll, S., Vanacker, H., Verrier, P., Hancock, R. D., and Foyer, C. H. (2011). The Transcription Factor ABI4 Is Required for the Ascorbic Acid-Dependent Regulation of Growth and Regulation of Jasmonate-Dependent Defense Signaling Pathways in Arabidopsis. *The Plant Cell*, **23**(9), 3319–3334.

Kidokoro, S., Maruyama, K., Nakashima, K., Imura, Y., Narusaka, Y., Shinwari, Z. K., Osakabe, Y., Fujita, Y., Mizoi, J., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2009). The phytochrome-interacting factor PIF7 negatively regulates DREB1 expression under circadian control in Arabidopsis. *Plant Physiol*, **151**(4), 2046–57.

Kieffer, M., Master, V., Waites, R., and Davies, B. (2011). TCP14 and TCP15 affect internode length and leaf shape in Arabidopsis. *Plant J*, **68**(1), 147–58.

Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., D'Angelo, C., Bornberg-Bauer, E., Kudla, J., and Harter, K. (2007). The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J*, **50**(2), 347–63.

Kim, H., Park, B., Yoo, J., Jung, M., Lee, S., and Han, H. (2007). Identification of a Calmodulin-binding NAC Protein as a Transcriptional Repressor in Arabidopsis. *Journal of Biological Chemistry*, **282**(50), 36292–302.

Kim, H. S., Yu, Y., Snesrud, E. C., Moy, L. P., Linford, L. D., Haas, B. J., Nierman, W. C., and Quackenbush, J. (2005). Transcriptional divergence of the duplicated oxidative stress-responsive genes in the Arabidopsis genome. *Plant J*, **41**(2), 212–20.

Kim, J.-B., Kang, J.-Y., and Kim, S. Y. (2004). Over-expression of a transcription factor regulating ABA-responsive gene expression confers multiple stress tolerance. *Plant Biotechnology Journal*, **2**(5), 459–66.

Kim, J. H., Woo, H. R., Kim, J., Lim, P. O., Lee, I. C., Choi, S. H., Hwang, D., and Nam, H. G. (2009). Trifurcate feed-forward regulation of age-dependent cell death involving miR164 in Arabidopsis. *Science*, **323**(5917), 1053–7.

Kosugi, S. and Ohashi, Y. (2002). DNA binding and dimerization specificity and potential targets for the TCP protein family. *Plant J*, **30**(3), 337–48.

Kreps, J. A., Wu, Y., Chang, H.-S., Zhu, T., Wang, X., and Harper, J. F. (2002). Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress. *Plant Physiol*, **130**(4), 2129–41.

Kundaje, A., Middendorf, M., Gao, F., Wiggins, C., and Leslie, C. (2005). Combining sequence and time series expression data to learn transcriptional modules. *IEEE/ACM transactions on computational biology and bioinformatics*, **2**(3), 194–202.

Kuo, M. H. and Allis, C. D. (1999). In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods*, **19**(3), 425–33.

Lång, V. and Palva, E. T. (1992). The expression of a rab-related gene, rab18, is induced by abscisic acid during the cold acclimation process of Arabidopsis thaliana (L.) Heynh. *Plant Mol Biol*, **20**(5), 951–62.

Laurie-Berry, N., Joardar, V., Street, I. H., and Kunkel, B. N. (2006). The Arabidopsis thaliana JASMONATE INSENSITIVE 1 gene is required for suppression of salicylic acid-dependent defenses during infection by Pseudomonas syringae. *Mol Plant Microbe Interact*, **19**(7), 789–800.

Lehmann, J., Atzorn, R., Brückner, C., Reinbothe, S., Leopold, J., Wasternack, C., and Parthier, B. (1995). Accumulation of jasmonate, abscisic acid, specific transcripts and proteins in osmotically stressed barley leaf segments. *Planta*, **197**(1), 156–162.

Lenhard, B. and Wasserman, W. W. (2002). TFBS: Computational framework for transcription factor binding site analysis. *Bioinformatics*, **18**(8), 1135–6.

Leon, T. Y. Y., Ngan, E. S. W., Poon, H.-C., So, M.-T., Lui, V. C. H., Tam, P. K. H., and Garcia-Barcelo, M. M. (2009). Transcriptional regulation of RET by Nkx2-1, Phox2b, Sox10, and Pax3. *J Pediatr Surg*, **44**(10), 1904–12.

Li, C., Potuschak, T., Colón-Carmona, A., Gutiérrez, R. A., and Doerner, P. (2005). Arabidopsis TCP20 links regulation of growth and cell division control pathways. *PNAS*, **102**(36), 12978–83.

Li, J., Brader, G., and Palva, E. T. (2004). The WRKY70 transcription factor: a node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense. *The Plant Cell*, **16**(2), 319–31.

Li, J. J. and Herskowitz, I. (1993). Isolation of ORC6, a component of the yeast origin recognition complex by a one-hybrid system. *Science*, **262**(5141), 1870–4.

Lim, P. O., Lee, I. C., Kim, J., Kim, H. J., Ryu, J. S., Woo, H. R., and Nam, H. G. (2010). Auxin response factor 2 (ARF2) plays a major role in regulating auxin-mediated leaf longevity. *Journal of Experimental Botany*, **61**(5), 1419–30.

Lohmann, C., Eggers-Schumacher, G., Wunderlich, M., and Schöffl, F. (2004). Two different heat shock transcription factors regulate immediate early expression of stress genes in Arabidopsis. *Mol Genet Genomics*, **271**(1), 11–21.

Lopez-Molina, L., Mongrand, S., and Chua, N. H. (2001). A postgermination developmental arrest checkpoint is mediated by abscisic acid and requires the ABI5 transcription factor in Arabidopsis. *PNAS*, **98**(8), 4782–7.

Lorenzo, O., Piqueras, R., Sánchez-Serrano, J. J., and Solano, R. (2003). ETHYLENE RESPONSE FACTOR1 integrates signals from ethylene and jasmonate pathways in plant defense. *The Plant Cell*, **15**(1), 165–78.

Lorenzo, O., Chico, J. M., Sánchez-Serrano, J. J., and Solano, R. (2004). JASMONATE-INSENSITIVE1 encodes a MYC transcription factor essential to discriminate between different jasmonate-regulated defense responses in Arabidopsis. *The Plant Cell*, **16**(7), 1938–50.

Lu, P.-L., Chen, N.-Z., An, R., Su, Z., Qi, B.-S., Ren, F., Chen, J., and Wang, X.-C. (2007). A novel drought-inducible gene, ATAF1, encodes a NAC family protein that negatively regulates the expression of stress-responsive genes in Arabidopsis. *Plant Mol Biol*, **63**(2), 289–305.

Ma, S., Gong, Q., and Bohnert, H. J. (2006). Dissecting salt stress pathways. *Journal of Experimental Botany*, **57**(5), 1097–107.

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, **21**(16), 3448–9.

Mandaokar, A. and Browse, J. (2009). MYB108 acts together with MYB24 to regulate jasmonate-mediated stamen maturation in Arabidopsis. *Plant Physiol*, **149**(2), 851–62.

Mao, G., Meng, X., Liu, Y., Zheng, Z., Chen, Z., and Zhang, S. (2011). Phosphorylation of a WRKY transcription factor by two pathogen-responsive MAPKs drives phytoalexin biosynthesis in Arabidopsis. *The Plant Cell*, **23**(4), 1639–53.

Martinez, N. J., Ow, M. C., Barrasa, M. I., Hammell, M., Sequerra, R., Doucette-Stamm, L., Roth, F. P., Ambros, V. R., and Walhout, A. J. M. (2008). A C. elegans genome-scale microRNA network contains composite feedback motifs with high flux capacity. *Genes & Development*, **22**(18), 2535–49.

Martínez-García, J. F., Huq, E., and Quail, P. H. (2000). Direct targeting of light signals to a promoter element-bound transcription factor. *Science*, **288**(5467), 859–63.

Matys, V., Kel-Margoulis, O. V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A. E., and Wingender, E. (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Research*, **34**(Database issue), D108–10.

McHattie, S. D. (2011). *Modelling transcriptional networks in plant senescence*. Ph.D. thesis, The University of Warwick.

Melotto, M., Underwood, W., Koczan, J., Nomura, K., and He, S. Y. (2006). Plant stomata function in innate immunity against bacterial invasion. *Cell*, **126**(5), 969–80.

Mengiste, T., Chen, X., Salmeron, J., and Dietrich, R. (2003). 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), 2551–65.

Menkens, A. E., Schindler, U., and Cashmore, A. R. (1995). The G-box: a ubiquitous regulatory DNA element in plants bound by the GBF family of bZIP proteins. *Trends Biochem Sci*, **20**(12), 506–10.

Miao, Y., Laun, T., Zimmermann, P., and Zentgraf, U. (2004). Targets of the WRKY53 transcription factor and its role during leaf senescence in Arabidopsis. *Plant Mol Biol*, **55**(6), 853–67.

Mitsuda, N., Ikeda, M., Takada, S., Takiguchi, Y., Kondou, Y., Yoshizumi, T., Fujita, M., Shinozaki, K., Matsui, M., and Ohme-Takagi, M. (2010). Efficient yeast one-/two-hybrid screening using a library composed only of transcription factors in Arabidopsis thaliana. *Plant Cell Physiol*, **51**(12), 2145–51.

Mochida, K., Yoshida, T., Sakurai, T., Yamaguchi-Shinozaki, K., Shinozaki, K., and Tran, L.-S. P. (2009). In silico analysis of transcription factor repertoire and prediction of stress responsive transcription factors in soybean. *DNA Res*, **16**(6), 353–69.

Mohr, P. G. and Cahill, D. M. (2007). Suppression by ABA of salicylic acid and lignin accumulation and the expression of multiple genes, in Arabidopsis infected with Pseudomonas syringae pv. tomato. *Funct Integr Genomics*, **7**(3), 181–91.

Morishita, T., Kojima, Y., Maruta, T., Nishizawa-Yokoi, A., Yabuta, Y., and Shigeoka, S. (2009). Arabidopsis NAC transcription factor, ANAC078, regulates flavonoid biosynthesis under high-light. *Plant Cell Physiol*, **50**(12), 2210–22.

Morohashi, K., Xie, Z., and Grotewold, E. (2009). Gene-specific and genome-wide ChIP approaches to study plant transcriptional networks. *Methods Mol Biol*, **553**, 3–12.

Morris, K., MacKerness, S. A., Page, T., John, C. F., Murphy, A. M., Carr, J. P., and Buchanan-Wollaston, V. (2000). Salicylic acid has a role in regulating gene expression during leaf senescence. *Plant J*, **23**(5), 677–85.

Nakashima, K., Takasaki, H., Mizoi, J., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2011). NAC transcription factors in plant abiotic stress responses. *Biochimica et biophysica acta*.

Narusaka, Y., Nakashima, K., Shinwari, Z. K., Sakuma, Y., Furihata, T., Abe, H., Narusaka, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003). Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of Arabidopsis rd29A gene in response to dehydration and high-salinity stresses. *Plant J*, **34**(2), 137–48.

Ndamukong, I., Abdallat, A. A., Thurow, C., Fode, B., Zander, M., Weigel, R., and Gatz, C. (2007). SA-inducible Arabidopsis glutaredoxin interacts with TGA factors and suppresses JA-responsive PDF1.2 transcription. *Plant J*, **50**(1), 128–39.

Neill, S. J., Desikan, R., Clarke, A., Hurst, R. D., and Hancock, J. T. (2002). Hydrogen peroxide and nitric oxide as signalling molecules in plants. *Journal of Experimental Botany*, **53**(372), 1237–47.

Noyes, M. B., Christensen, R. G., Wakabayashi, A., Stormo, G. D., Brodsky, M. H., and Wolfe, S. A. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell*, **133**(7), 1277–89.

Oh, S. A., Park, J. H., Lee, G. I., Paek, K. H., Park, S. K., and Nam, H. G. (1997). Identification of three genetic loci controlling leaf senescence in Arabidopsis thaliana. *Plant J*, **12**(3), 527–35.

Okamuro, J. K., Caster, B., Villarroel, R., Montagu, M. V., and Jofuku, K. D. (1997). The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in Arabidopsis. *PNAS*, **94**(13), 7076–81.

Oliphant, A. R., Brandl, C. J., and Struhl, K. (1989). Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein. *Mol Cell Biol*, **9**(7), 2944–9.

Olsen, A., Ernst, H., Leggio, L., and Skriver, K. (2005). DNA-binding specificity and molecular functions of NAC transcription factors. *Plant Science*, **169**(4), 785–797.

Oñate-Sánchez, L. and Singh, K. B. (2002). Identification of Arabidopsis ethylene-responsive element binding factors with distinct induction kinetics after pathogen infection. *Plant Physiol*, **128**(4), 1313–22.

Ooka, H., Satoh, K., Doi, K., Nagata, T., Otomo, Y., Murakami, K., Matsubara, K., Osato, N., Kawai, J., Carninci, P., Hayashizaki, Y., Suzuki, K., Kojima, K., Takahara, Y., Yamamoto, K., and Kikuchi, S. (2003). Comprehensive analysis of NAC family genes in Oryza sativa and Arabidopsis thaliana. *DNA Res*, **10**(6), 239–47.

Ou, B., Yin, K.-Q., Liu, S.-N., Yang, Y., Gu, T., Hui, J. M. W., Zhang, L., Miao, J., Kondou, Y., Matsui, M., Gu, H.-Y., and Qu, L.-J. (2011). A high-throughput screening system for Arabidopsis transcription factors and its application to Med25-dependent transcriptional regulation. *Molecular plant*, **4**(3), 546–55.

Pauwels, L., Morreel, K., Witte, E. D., Lammertyn, F., Montagu, M. V., Boerjan, W., Inzé, D., and Goossens, A. (2008). Mapping methyl jasmonate-mediated transcriptional reprogramming of metabolism and cell cycle progression in cultured Arabidopsis cells. *PNAS*, **105**(4), 1380–5.

Pauwels, L., Inzé, D., and Goossens, A. (2009). Jasmonate-inducible gene: What does it mean? *Trends Plant Sci*, **14**(2), 87–91.

Pavesi, G., Mereghetti, P., Mauri, G., and Pesole, G. (2004). Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res*, **32**(Web Server issue), W199–203.

Pfluger, J. and Wagner, D. (2007). Histone modifications and dynamic regulation of genome accessibility in plants. *Current Opinion in Plant Biology*, **10**(6), 645–52.

Picot, E., Krusche, P., Tiskin, A., Carré, I., and Ott, S. (2010). Evolutionary analysis of regulatory sequences (EARS) in plants. *Plant J*, **64**(1), 165–76.

Pieterse, C. M. J., Leon-Reyes, A., der Ent, S. V., and Wees, S. C. M. V. (2009). Networking by small-molecule hormones in plant immunity. *Nat Chem Biol*, **5**(5), 308–16.

Pontier, D., Gan, S., Amasino, R., and Roby, D. (1999). Markers for hypersensitive response and senescence show distinct patterns of expression. *Plant Mol Biol*.

Pré, M., Atallah, M., Champion, A., Vos, M. D., Pieterse, C. M. J., and Memelink, J. (2008). The AP2/ERF domain transcription factor ORA59 integrates jasmonic acid and ethylene signals in plant defense. *Plant Physiol*, **147**(3), 1347–57.

Pruneda-Paz, J. L., Breton, G., Para, A., and Kay, S. A. (2009). A functional genomics approach reveals CHE as a component of the Arabidopsis circadian clock. *Science*, **323**(5920), 1481–5.

Quirino, B. F., Normanly, J., and Amasino, R. M. (1999). Diverse range of gene activity during Arabidopsis thaliana leaf senescence includes pathogen-independent induction of defense-related genes. *Plant Mol Biol*, **40**(2), 267–78.

Ramírez, V., Agorio, A., Coego, A., García-Andrade, J., Hernández, M. J., Balaguer, B., Ouwerkerk, P. B. F., Zarra, I., and Vera, P. (2011). MYB46 modulates disease susceptibility to Botrytis cinerea in Arabidopsis. *Plant Physiol*, **155**(4), 1920–35.

Ramirez-Parra, E., Fründt, C., and Gutierrez, C. (2003). A genome-wide identification of E2F-regulated genes in Arabidopsis. *Plant J*, **33**(4), 801–11.

Riechmann, J., Heard, J., Martin, G., Reuber, L., and Z, C. (2000). Arabidopsis Transcription Factors: Genome-Wide Comparative Analysis Among Eukaryotes. *Science*, **290**(5499), 2105–10.

Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., Davletova, S., and Mittler, R. (2004). When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress. *Plant Physiol*, **134**(4), 1683–96.

Robatzek, S. and Somssich, I. E. (2002). Targets of AtWRKY6 regulation during plant senescence and pathogen defense. *Genes Dev*, **16**(9), 1139–49.

Rushton, P. J., Bokowiec, M. T., Han, S., Zhang, H., Brannock, J. F., Chen, X., Laudeman, T. W., and Timko, M. P. (2008). Tobacco transcription factors: novel insights into transcriptional regulation in the Solanaceae. *Plant Physiol*, **147**(1), 280–95.

Rushton, P. J., Somssich, I. E., Ringler, P., and Shen, Q. J. (2010). WRKY transcription factors. *Trends in Plant Science*, **15**(5), 247–58.

Sakuma, Y., Liu, Q., Dubouzet, J. G., Abe, H., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2002). DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem Biophys Res Commun*, **290**(3), 998–1009.

Sasaki-Sekimoto, Y., Taki, N., Obayashi, T., Aono, M., Matsumoto, F., Sakurai, N., Suzuki, H., Hirai, M. Y., Noji, M., Saito, K., Masuda, T., ichiro Takamiya, K., Shibata, D., and Ohta, H. (2005). Coordinated activation of metabolic pathways for antioxidants and defence compounds by jasmonates and their roles in stress tolerance in Arabidopsis. *Plant J*, **44**(4), 653–68.

Schenk, P. M., Kazan, K., Wilson, I., Anderson, J. P., Richmond, T., Somerville, S. C., and Manners, J. M. (2000). Coordinated plant defense responses in Arabidopsis revealed by microarray analysis. *PNAS*, **97**(21), 11655–60.

Schneider, T. D. and Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res*, **18**(20), 6097–100.

Schöffl, F., Prändl, R., and Reindl, A. (1998). Regulation of the heat-shock response. *Plant Physiol*, **117**(4), 1135–41.

Schommer, C., Palatnik, J. F., Aggarwal, P., Chételat, A., Cubas, P., Farmer, E. E., Nath, U., and Weigel, D. (2008). Control of jasmonate biosynthesis and senescence by miR319 targets. *PLoS Biol*, **6**(9), e230.

Sclep, G., Allemeersch, J., Liechti, R., Meyer, B. D., Beynon, J., Bhalerao, R., Moreau, Y., Nietfeld, W., Renou, J.-P., Reymond, P., Kuiper, M. T., and Hilson, P. (2007). CATMA, a comprehensive genome-scale resource for silencing and transcript profiling of Arabidopsis genes. *BMC Bioinformatics*, **8**, 400.

Segal, E., Yelensky, R., and Koller, D. (2003). Genome-wide discovery of transcriptional modules from DNA sequence and gene expression. *Bioinformatics*, **19 Suppl 1**, i273–82.

Seki, M., Narusaka, M., Abe, H., Kasuga, M., Yamaguchi-Shinozaki, K., Carninci, P., Hayashizaki, Y., and Shinozaki, K. (2001). Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. *The Plant Cell*, **13**(1), 61–72.

Seki, M., Ishida, J., Narusaka, M., Fujita, M., Nanjo, T., Umezawa, T., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Yamaguchi-Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y., and Shinozaki, K. (2002). Monitoring the expression pattern of around 7,000 Arabidopsis genes under ABA treatments using a full-length cDNA microarray. *Funct Integr Genomics*, **2**(6), 282–91.

Skirycz, A., Reichelt, M., Burow, M., Birkemeyer, C., Rolcik, J., Kopka, J., Zanor, M. I., Gershenzon, J., Strnad, M., Szopa, J., Mueller-Roeber, B., and Witt, I. (2006). DOF transcription factor AtDof1.1 (OBP2) is part of a regulatory network controlling glucosinolate biosynthesis in Arabidopsis. *Plant J*, **47**(1), 10–24.

Smykowski, A., Zimmermann, P., and Zentgraf, U. (2010). G-Box binding Factor1 reduces CATALASE2 expression and regulates the onset of leaf senescence in Arabidopsis. *Plant Physiol*, **153**(3), 1321–31.

Solano, R., Stepanova, A., Chao, Q., and Ecker, J. R. (1998). Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. *Genes & Development*, **12**(23), 3703–14.

Souer, E., van Houwelingen, A., Kloos, D., Mol, J., and Koes, R. (1996). The no apical meristem gene of Petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. *Cell*, **85**(2), 159–70.

Spoel, S. H., Johnson, J. S., and Dong, X. (2007). Regulation of tradeoffs between plant defenses against pathogens with different lifestyles. *PNAS*, **104**(47), 18842–7.

Spoel, S. H., Mou, Z., Tada, Y., Spivey, N. W., Genschik, P., and Dong, X. (2009). Proteasome-mediated turnover of the transcription coactivator NPR1 plays dual roles in regulating plant immunity. *Cell*, **137**(5), 860–72.

Stajich, J. E., Block, D., Boulez, K., Brenner, S. E., Chervitz, S. A., Dagdigian, C., Fuellen, G., Gilbert, J. G. R., Korf, I., Lapp, H., Lehväslaiho, H., Matsalla, C., Mungall, C. J., Osborne, B. I., Pocock, M. R., Schattner, P., Senger, M., Stein, L. D., Stupka, E., Wilkinson, M. D., and Birney, E. (2002). The Bioperl toolkit: Perl modules for the life sciences. *Genome Res*, **12**(10), 1611–8.

Sunkar, R., Chinnusamy, V., Zhu, J., and Zhu, J.-K. (2007). Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. *Trends in Plant Science*, **12**(7), 301–9.

Suzuki, N., Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., and Mittler, R. (2005). Enhanced tolerance to environmental stress in transgenic plants expressing the transcriptional coactivator multiprotein bridging factor 1c. *Plant Physiol*, **139**(3), 1313–22.

Swindell, W. R. (2006). The association among gene expression responses to nine abiotic stress treatments in Arabidopsis thaliana. *Genetics*, **174**(4), 1811–24.

Swindell, W. R., Huebner, M., and Weber, A. P. (2007). Transcriptional profiling of Arabidopsis heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. *BMC Genomics*, **8**, 125.

Tabach, Y., Brosh, R., Buganim, Y., Reiner, A., Zuk, O., Yitzhaky, A., Koudritsky, M., Rotter, V., and Domany, E. (2007). Wide-scale analysis of human functional transcription factor binding reveals a strong bias towards the transcription start site. *PLoS ONE*, **2**(8), e807.

Takahashi, S., Seki, M., Ishida, J., Satou, M., Sakurai, T., Narusaka, M., Kamiya, A., Nakajima, M., Enju, A., Akiyama, K., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2004). Monitoring the expression profiles of genes induced by hyperosmotic, high salinity, and oxidative stress and abscisic acid treatment in Arabidopsis cell culture using a full-length cDNA microarray. *Plant Mol Biol*, **56**(1), 29–55.

Teichmann, S. A. and Babu, M. M. (2004). Gene regulatory network growth by duplication. *Nature Genetics*, **36**(5), 492–6.

Thomma, B. P., Eggermont, K., Penninckx, I. A., Mauch-Mani, B., Vogelsang, R., Cammue, B. P., and Broekaert, W. F. (1998). Separate jasmonate-dependent and salicylate-dependent defense-response pathways in Arabidopsis are essential for resistance to distinct microbial pathogens. *PNAS*, **95**(25), 15107–11.

Thomma, B. P., Eggermont, K., Tierens, K. F., and Broekaert, W. F. (1999). Requirement of functional ethylene-insensitive 2 gene for efficient resistance of Arabidopsis to infection by Botrytis cinerea. *Plant Physiol*, **121**(4), 1093–102.

Ton, J., Flors, V., and Mauch-Mani, B. (2009). The multifaceted role of ABA in disease resistance. *Trends Plant Sci*, **14**(6), 310–7.

Tournier, B., Sanchez-Ballesta, M. T., Jones, B., Pesquet, E., Regad, F., Latché, A., Pech, J.-C., and Bouzayen, M. (2003). New members of the tomato ERF family show specific expression pattern and diverse DNA-binding capacity to the GCC box element. *FEBS Lett*, **550**(1-3), 149–54.

Tran, L.-S. P., Nakashima, K., Sakuma, Y., Simpson, S. D., Fujita, Y., Maruyama, K., Fujita, M., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2004). Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. *The Plant Cell*, **16**(9), 2481–98.

Tran, L.-S. P., Nakashima, K., Sakuma, Y., Osakabe, Y., Qin, F., Simpson, S. D., Maruyama, K., Fujita, Y., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2007). Co-expression of the stress-inducible zinc finger homeodomain ZFHD1 and NAC transcription factors enhances expression of the ERD1 gene in Arabidopsis. *Plant J*, **49**(1), 46–63.

Trémousaygue, D., Garnier, L., Bardet, C., Dabos, P., Hervé, C., and Lescure, B. (2003). Internal telomeric repeats and 'TCP domain' protein-binding sites co-operate to regulate gene expression in Arabidopsis thaliana cycling cells. *Plant J*, **33**(6), 957–66.

Ulker, B., Mukhtar, M. S., and Somssich, I. E. (2007). The WRKY70 transcription factor of Arabidopsis influences both the plant senescence and defense signaling pathways. *Planta*, **226**(1), 125–37.

van der Graaff, E., Schwacke, R., Schneider, A., Desimone, M., Flügge, U.-I., and Kunze, R. (2006). Transcription analysis of arabidopsis membrane transporters and hormone pathways during developmental and induced leaf senescence. *Plant Physiol*, **141**(2), 776–92.

Vellosillo, T., Martínez, M., López, M. A., Vicente, J., Cascón, T., Dolan, L., Hamberg, M., and Castresana, C. (2007). Oxylipins produced by the 9-lipoxygenase pathway in Arabidopsis regulate lateral root development and defense responses through a specific signaling cascade. *The Plant Cell*, **19**(3), 831–46.

Vicente, M. R.-S. and Plasencia, J. (2011). Salicylic acid beyond defence: its role in plant growth and development. *Journal of Experimental Botany*, **62**(10), 3321–38.

Walhout, A. J. (2011). What does biologically meaningful mean? A perspective on gene regulatory network validation. *Genome Biol*, **12**(4), 109.

Walhout, A. J. M. (2006). Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. *Genome Res*, **16**(12), 1445–54.

Wang, R. L., Stec, A., Hey, J., Lukens, L., and Doebley, J. (1999). The limits of selection during maize domestication. *Nature*, **398**(6724), 236–9.

Wang, X., Basnayake, B. M. V. S., Zhang, H., Li, G., Li, W., Virk, N., Mengiste, T., and Song, F. (2009). The Arabidopsis ATAF1, a NAC transcription factor, is a negative regulator of defense responses against necrotrophic fungal and bacterial pathogens. *Mol Plant Microbe Interact*, **22**(10), 1227–38.

Waters, M. T. and Langdale, J. A. (2009). The making of a chloroplast. *The EMBO Journal*, **28**(19), 2861–73.

Weaver, L. M., Gan, S., Quirino, B., and Amasino, R. M. (1998). A comparison of the expression patterns of several senescence-associated genes in response to stress and hormone treatment. *Plant Mol Biol*, **37**(3), 455–69.

Weltmeier, F., Ehlert, A., Mayer, C. S., Dietrich, K., Wang, X., Schütze, K., Alonso, R., Harter, K., Vicente-Carbajosa, J., and Dröge-Laser, W. (2006). Combinatorial control of Arabidopsis proline dehydrogenase transcription by specific heterodimerisation of bZIP transcription factors. *The EMBO Journal*, **25**(13), 3133–43.

Wettenhall, J. M. and Smyth, G. K. (2004). limmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinformatics*, **20**(18), 3705–6.

Williamson, J. D., Stoop, J. M., Massel, M. O., Conkling, M. A., and Pharr, D. M. (1995). Sequence analysis of a mannitol dehydrogenase cDNA from plants reveals a function for the pathogenesis-related protein ELI3. *PNAS*, **92**(16), 7148–52.

Windram, O. P. (2010). *Using a Systems Biology Approach to Elucidate Transcriptional Networks Regulating Plant Defence*. Ph.D. thesis, The University of Warwick.

Woo, H. R., Kim, J. H., Nam, H. G., and Lim, P. O. (2004). The delayed leaf senescence mutants of Arabidopsis, ore1, ore3, and ore9 are tolerant to oxidative stress. *Plant Cell Physiol*, **45**(7), 923–32.

Woo, H. R., Kim, J. H., Kim, J., Kim, J., Lee, U., Song, I.-J., Kim, J.-H., Lee, H.-Y., Nam, H. G., and Lim, P. O. (2010). The RAV1 transcription factor positively regulates leaf senescence in Arabidopsis. *Journal of Experimental Botany*, **61**(14), 3947–57.

Wu, Y., Deng, Z., Lai, J., Zhang, Y., Yang, C., Yin, B., Zhao, Q., Zhang, L., Li, Y., Yang, C., and Xie, Q. (2009). Dual function of Arabidopsis ATAF1 in abiotic and biotic stress responses. *Cell Res*, **19**(11), 1279–90.

Xie, Q., Frugis, G., Colgan, D., and Chua, N. H. (2000). Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. *Genes Dev*, **14**(23), 3024–36.

Xu, Y., Chang, P., Liu, D., Narasimhan, M. L., Raghothama, K. G., Hasegawa, P. M., and Bressan, R. A. (1994). Plant Defense Genes Are Synergistically Induced by Ethylene and Methyl Jasmonate. *The Plant Cell*, **6**(8), 1077–1085.

Yamaguchi, Y., Huffaker, A., Bryan, A. C., Tax, F. E., and Ryan, C. A. (2010). PEPR2 is a second receptor for the Pep1 and Pep2 peptides and contributes to defense responses in Arabidopsis. *The Plant Cell*, **22**(2), 508–22.

yong Li, X., MacArthur, S., Bourgon, R., Nix, D., Pollard, D. A., Iyer, V. N., Hechmer, A., Simirenko, L., Stapleton, M., Hendriks, C. L. L., Chu, H. C., Ogawa, N., Inwood, W., Sementchenko, V., Beaton, A., Weiszmann, R., Celniker, S. E., Knowles, D. W., Gingeras, T., Speed, T. P., Eisen, M. B., and Biggin, M. D. (2008). Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. *PLoS Biol*, **6**(2), e27.

Yoshiba, Y., Nanjo, T., Miura, S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1999). Stress-responsive and developmental regulation of Delta(1)-pyrroline-5-carboxylate synthetase 1 (P5CS1) gene expression in Arabidopsis thaliana. *Biochem Biophys Res Commun*, **261**(3), 766–72.

Yoshida, T., Fujita, Y., Sayama, H., Kidokoro, S., Maruyama, K., Mizoi, J., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2010). AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. *Plant J*, **61**(4), 672–85.

youn Kang, J., in Choi, H., young Im, M., and Kim, S. Y. (2002). Arabidopsis basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling. *The Plant Cell*, **14**(2), 343–57.

Zarei, A., Körbes, A. P., Younessi, P., Montiel, G., Champion, A., and Memelink, J. (2011). Two GCC boxes and AP2/ERF-domain transcription factor ORA59 in jasmonate/ethylene-mediated activation of the PDF1.2 promoter in Arabidopsis. *Plant Mol Biol*, **75**(4-5), 321–31.

Zeevaart, J. and Creelman, R. (1988). Metabolism and physiology of abscisic acid. *Annu. Rev. Plant Physiol. Plant Mol. Biol*, **39**, 439–473.

Zeller, G., Henz, S. R., Widmer, C. K., Sachsenberg, T., Rätsch, G., Weigel, D., and Laubinger, S. (2009). Stress-induced changes in the Arabidopsis thaliana transcriptome analyzed using whole-genome tiling arrays. *Plant J*, **58**(6), 1068–82.

Zhang, L., Li, Z., Quan, R., Li, G., Wang, R., and Huang, R. (2011). An AP2 Domain-Containing Gene, ESE1, Targeted by the Ethylene Signaling Component EIN3 Is Important for the Salt Response in Arabidopsis. *Plant Physiol*, **157**(2), 854–65.

Zheng, Z., Qamar, S. A., Chen, Z., and Mengiste, T. (2006). Arabidopsis WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens. *Plant J*, **48**(4), 592–605.

Zhong, R., Lee, C., and Ye, Z.-H. (2010). Global Analysis of Direct Targets of Secondary Wall NAC Master Switches in Arabidopsis. *Mol Plant*, **3**(6), 1087–103.

Zhu, Q., Zhang, J., Gao, X., Tong, J., Xiao, L., Li, W., and Zhang, H. (2010). The Arabidopsis AP2/ERF transcription factor RAP2.6 participates in ABA, salt and osmotic stress responses. *Gene*, **457**(1-2), 1–12.

Zou, C., Sun, K., Mackaluso, J. D., Seddon, A. E., Jin, R., Thomashow, M. F., and Shiu, S.-H. (2011). Cis-regulatory code of stress-responsive transcription in Arabidopsis thaliana. *PNAS*, **108**(36), 14992–7.