Tools for Exploring Non-Synonymous Variation in
Arabidopsis thaliana

by

Matthew Cumming

A thesis submitted in conformity with the requirements
for the degree of Master of Science
Department of Cell & Systems Biology
University of Toronto

© Copyright by Matthew Cumming 2018
Tools for Exploring Non-Synonymous Variation in

*Arabidopsis thaliana*

Matthew Cumming

Master of Science

Department of Cell & Systems Biology
University of Toronto

2018

Abstract

High throughput sequencing has opened the doors for investigators to probe genetic variation present in large populations of organisms. In plants, the 1001 Genomes Project (1001genomes.org) is one such effort that sought to characterize the extant worldwide variation in *Arabidopsis thaliana* for future analyses to compare and draw upon. In my thesis I began the construction of a synthetic yeast pathway to test the effects of natural non-synonymous variants on the activation of the *A. thaliana* abscisic acid (ABA) signaling pathway in the hopes of characterizing variants that modulate pathway output. I constructed a series of expression cassettes to confirm the correct expression and localization of our ABA members and laid the groundwork for future members to continue constructing the pathway. Additionally, I created a web application called The Variant Viewer, for investigators to query the 1001 Genomes database and view variants in any *A. thaliana* gene.
I couldn’t be happier with how my project has progressed over the last two years; that is definitely due to the support and guidance of a number of people.

Firstly, I would like to thank my supervisor Nicholas Provart, my graduate committee, and the Goring, Guttman, Desveaux, Chang, and Yoshioka research groups. Nick has always been supportive and encouraging of my interests and has always made many useful suggestions that helped me develop my investigative skills. I can easily say that without Nick as my supervisor I would not have enjoyed this experience nearly as much and would have learned far less. I would like to thank my committee, Shelley Lumba and Alan Moses, for helpful suggestions and guidance and for telling me what I needed to hear when I needed to hear it. I would like to thank our neighbour research groups for sharing equipment and lab space, I couldn’t have done anything without them! I would also like to thank Thanh Nguyen, Henry Hong, and Benjamin Scott, for keeping me from breaking wildly expensive equipment too badly.

Second, I would like to thank my lab mates Anna van Weringh, Michael Dong, Vincent Lau and Asher Pasha. As I joined the lab with little wet lab experience, Anna and Mike always took the time to share protocols and make suggestions or correct me where I was going wrong. On the bioinformatics side, Asher and Vince have pointed me in the right direction when I wasn’t sure how to approach a problem and have often provided me with easier solutions to problems when I was trying to re-invent the wheel.

Thirdly, I would like to thank all my friends in the RW and ESC buildings for constant support and a great atmosphere to learn in. I count myself lucky to have met such an incredibly diverse, intelligent and funny group to spend two years with. Specifically, I would like to thank Van Phan for teaching me great dance moves, Purva Karia for inspiring me to add skydiving to my bucket list, Anna van Weringh for pushing me down the sourdough rabbit hole, Armand Mirmiran for always having a great video or song handy, Bradley LaFlamme for being a constant source of dry sarcasm, and Derek Seto for being the loudest person I have ever met.

Finally, I would like to thank my family, and my partner Charlotte Dawson for supporting me during failures and long periods of difficulty and for all the things that loved ones do.
# Table of Contents

Acknowledgments ........................................................................................................ iii

Table of Contents ........................................................................................................ iv

List of Tables and Figures ............................................................................................. vi

Abbreviations ................................................................................................................ vii

## 1 Constructing a Synthetic Stress Pathway in Yeast ...................................................... 1

1.1 Introduction .............................................................................................................. 1

1.1.1 1001 *Arabidopsis thaliana* Genomes ................................................................. 1

1.1.2 The Difficulty of Screening Large Numbers of Variants ...................................... 1

1.1.3 ABA Stress Signaling .......................................................................................... 5

1.1.4 Project Goals ....................................................................................................... 8

1.2 Materials and Methods ........................................................................................... 8

1.2.1 Overall Pathway Design, and Components ........................................................ 8

1.2.2 Pathway Construction ......................................................................................... 11

1.2.3 Pathway Validation ............................................................................................ 13

1.3 Results and Discussion ............................................................................................ 15

1.3.1 Cloning and Transformation .............................................................................. 15

1.3.2 Western Blots ..................................................................................................... 16

1.3.3 Nuclear Localization .......................................................................................... 17

1.3.4 Testing for Pathway Activation .......................................................................... 18

1.4 Conclusions and Future Directions ......................................................................... 19

1.5 Figures ..................................................................................................................... 21

## 2 The Variant Viewer .................................................................................................... 26

2.1 Introduction .............................................................................................................. 26

2.1.1 Inspecting Variation .......................................................................................... 26

2.1.2 Project Goals ..................................................................................................... 28

iv
2.2 Application Design .................................................................................................................. 29
  2.2.1 Data Sources and Javascript Libraries ........................................................................ 29
  2.2.2 Application Logic ............................................................................................................. 30
  2.2.3 Natural Variation in the ABA signaling members ......................................................... 32
  2.2.4 Future Directions ............................................................................................................. 34

2.3 Figures .................................................................................................................................. 35

References ..................................................................................................................................... 40

Appendices .................................................................................................................................... 49

Permissions .................................................................................................................................... 53
List of Tables and Figures

Tables

Table 1 - *S. cerevisiae* strains constructed for this project........................................9

Table 2 - Yeast strains expressing combinations of SnRK2.6 and ABF2/ABI5 yield no fluorescent output.................................................................18

Figures

1.1 - The Core abscisic acid signaling pathway and our synthetic basic reconstruction........21

1.2 - Western blots of FLAG3X tagged ABA signaling proteins heterologously expressed in *S. cerevisiae*.................................................................22

1.3 - Nuclear localization of GFP tagged ABF2, ABI1, PYR1, and SnRK2.6 in *S. cerevisiae*...23

1.4 - Glucose and galactose induction of *S. cerevisiae* strains........................................24

1.5 - GAL4 trans activation domain fusions show no GFP expression following induction with Galactose.................................................................25

2.1 - Examples of existing tools for viewing *A. thaliana* variation.....................................35

2.2 - The Variant Viewer: an application for viewing variants from the 1001 Genomes Project.........................................................................................36

2.3 - Plotly.js generated scatter plots of protein frequencies against MSA of genes of interest.........................................................................................37

2.4 - SNPs in the coding regions of *A. thaliana* genes separated by PANTHER GO annotation.........................................................................................38

2.5 - An alignment of the reference isoforms of the ABFs (ABF1/ABF2/ABF3/ABF4) with their associated variants................................................................39
Abbreviations

ABA - abscisic acid

ABF - ABA Response Element Binding Factors

ABRE - ABA Response Element

AGI - Arabidopsis Genome Initiative

API - Application Programming Interface

ARF - Auxin Response Factor

BAR - Bio-Analytic Resource

bZIP - Basic Leucine Zipper

CDD - Conserved Domains Database

CE - Coupling Element

DRE - Drought Responsive Element

GFP - Green Fluorescent Protein

GUS - β-glucuronidase

IAA - Indole Acetic Acid

MAPK - Mitogen-Activated Protein Kinase

MEP - Methyerythritol 4-phosphate

NL - Nuclear Localizing

NLS - Nuclear Localization Sequence

PFAM - The Protein Families Database
PP2C - Protein Phosphatase 2C

PSI - Position Specific Iterated

PYR/PYL/RCAR - PYRABACTIN RESISTANCE1/PYR1-LIKE/REGULATORY COMPONENTS OF ABA RECEPTORS

SNP - Single Nucleotide Polymorphism

SnRK - Sucrose Non-Fermenting 1 Related Protein Kinase

SVG - Scaleable Vector Graphic

TAD - Transactivation Domain

TF - Transcription Factor

USER - Uracil-Specific Excision Reagent

VCF - Variant Call Format

YFP - Yellow Fluorescent Protein
1 Constructing a Synthetic Stress Pathway in Yeast

1.1 Introduction

1.1.1 1001 Arabidopsis thaliana Genomes

Whole genome sequencing of large populations is revolutionizing the way that we study natural variation within populations. For example, current efforts in human exome sequencing (where the entire complement of protein coding sequences in a patient is determined) hope for a world of ‘personalized medicine’ where links between genotype and disease phenotype can be reliably and inexpensively made allowing for targeted treatments to be based on a patient’s genetic background. Often investigators seek to create the same type of links for their own gene of interest, where the genetic basis for an observed phenotype is desired. In plants, the 1001 Genomes Project sought to characterize the extant natural variation in 1001 ecotypes of Arabidopsis thaliana distributed across the world through the high throughput sequencing of a subset of individuals (Alonso-Blanco et al., 2016). The project sampled 1135 populations of Arabidopsis thaliana from across Asia, Europe, North America and Africa, sequenced their genomes, and provided several tools for sequence analyses and investigation. These A. thaliana ecotypes represent divergent evolutionary histories and unique adaptations of these populations to their environments. If this unique coding sequence variation is present in molecular pathways that are known to control a response to a given stimulus, there exists the possibility that the variation is due to local adaptation, and is likely to confer a useful phenotypic output for the plant. Therefore, the suite of variation present in the 1135 accessions could be mined for insertion, deletion, truncation, and non-synonymous(missense) variants that might cause a desired phenotype given their presence in a gene of interest.

1.1.2 The Difficulty of Screening Large Numbers of Variants

Ideally, one could individually test the whole set of natural coding variation (including those that appear to have little likely effect) on the off chance that they show a favorable phenotype. If a single protein is the target, individual variants can be selected specifically based on how they might affect a protein’s function via their proximity to predicted catalytic domains, and known conserved residues, and then tested in planta in a common genetic background. However, the return on investment for this approach may not be realistic in that some of these mutants may not
appear to be high value targets, and the time to produce transformed *A. thaliana* can be prohibitive for large numbers of variants. Additionally, if a protein is not well studied and has few domains of obvious significance it might be hard to curate variants manually. Additionally, in the case of complex pathways with multiple redundant signaling nodes, or several tissue specific members, testing every variant in a network would be daunting. Finally, where the 1001 Genomes data set sometimes includes upwards of one hundred unique coding variants in a single protein, characterizing the full natural ‘mutome’ for a given pathway or protein is a large task. The need to prioritize experimental resources against the proportion of variant sequences that may give useful phenotypes is obvious.

An overabundance of variants for further study is a common problem in clinical research. Indeed, whole exome sequencing in humans has generated an abundance of variant data (~20,000 per individual), far above the number of SNPs identified in the 1001 Genomes Project (Raimondi *et al.*, 2016). Sifting through these data with a heuristic *in silico* approach is the obvious first step. When variants are sequenced specifically from populations of individuals with a certain disease, prioritizing those with a high observed frequency and using a variety of gene annotation data can be informative (Tranchevent, Leon-Charles; Moreau, 2012). Some complex methods draw from microarray, RNA-seq, protein-protein interaction, phylogenetic, and conserved domain data to attempt to classify SNPs and variants into categories based on their likelihood of causing functional change in disease pathways (Renkonen *et al.*, 2010; Tranchevent, Leon-Charles; Moreau, 2012; Raimondi *et al.*, 2016). However predictions made by these algorithms have high error rates and are limited by the annotation data, and learning data sets they draw from (Ipe *et al.*, 2017). Indeed, recent reviews of high throughput variant annotation methods highlight the need for high throughput functional assays to verify candidates (Ipe *et al.*, 2017). Additionally, changes in output of a specific disease pathway are likely the sum of a number of different variants each with a relatively minor effect (Renkonen *et al.*, 2010). In these cases, variants need to be tested for their effect in the context of other complementary variants to consider their combined chance of causing a disease phenotype. For example, Matthews *et al.* (2014) found that a yeast checkpoint effector kinase Rad-53s interaction with the regulatory subunit of a Dbf4 kinase is dependent on several conserved residues in their FHA and BRCT domains respectively (Matthews *et al.*, 2014). Using PSI Blast the authors found stretches of conserved residues contained in Rad-53 FHA domains able to bind Dbf4 BRCT domains but
found that because the conserved region was rather large, single mutants did not affect protein-protein interactions in a yeast two-hybrid assay. The authors found that one pair of mutants in Rad53, N112A/E129A was able to decrease binding affinity, but only through the introduction of a third would the interaction between Rad-53 and Dbf4 be totally prevented. This complex interplay would be completely missed in a system that singly tested variants; but this type of testing can be difficult to do on a large scale. Testing many combinations of variants needs to be done in a high throughput manner. For example, tests of protein activity could follow a forward genetics approach where screens of randomly mutated libraries of proteins are tested for differences in functional output, or reverse genetics where specific residues that are hypothesized to play a role in a protein's function are mutated in the background of variants they are correlated to occur with (as in the case of Rad-53 and Dbf4).

Within the 1001 Genomes data the same problem exists: variants in pathways might be masked except in their specific context, making it difficult to ascribe any one variant's given impact. To screen these variants in planta an investigator would need to transform them all into a common genetic background along with a reporter and then grow up large populations of plants for screening. For example a recent study sought to characterize the spatiotemporal response of A. thaliana to varying ABA concentrations using a 6X ABRE synthetic GUS/GFP reporter (Wu et al., 2018). The 6X ABRE element was placed upstream of a minimal Cauliflower Mosaic Virus promoter (35S MP) and transformed into A. thaliana where whole plants could then be screened for their responses to differing ABA concentrations. To adapt a system such as this one for testing multiple variants each variant would need to be transformed into the correct mutant background in A. thaliana for each specific pathway member in conjunction with a synthetic reporter; and then screened using multiple techniques for differences in pathway activation. The amount of labour required effectively eliminates any possibility of doing large combinatorial screens of variants.

Indeed, in a situation where a gene contains 100+ missense variants, this is no small task. However, synthetic high throughput systems to detect changes in pathway activity given a natural variant have been created previously. Recently, the Nemhauser lab at the University of Washington created a synthetic A. thaliana degradation-based fluorescence assay in S. cerevisiae that allowed them to test natural variants (from the ~200 available accessions of the 1001 Genomes data at the time) of four proteins in the auxin signaling pathway (Wright et al., 2017).
The authors constitutively expressed two members of an auxin co-receptor complex, the Auxin Signaling F-Boxes TIR1 and AFB2 and the INDOLE-3-ACETIC ACID PROTEINS IAA1 and IAA17 in S. cerevisiae. In the absence of auxin, IAAs repress Auxin Response Factors (ARFs) that bind promoters of downstream genes, which in turn drive expression of plant development and environmental genes. In the presence of auxin, TIR1/AFB2 interacts with IAA1/17 and targets them for ubiquitination and subsequent degradation. Using IAA1/17-eYFP fusions the group measured the decrease in eYFP fluorescence using flow cytometry as a measure of AFB-mediated IAA degradation, and as a proxy for auxin pathway activation. By assessing variants using flow cytometry Nemhauser et al. identified both hyper and hypomorphic alleles in TIR1/AFB2, one of which was found to be hypermorphic for auxin sensitivity in four of five of the accessions of A. thaliana that contained it. Though a powerful technique for measuring variation in the auxin pathway, the assay relies on the degradation of IAA1/17 proteins for its fluorescent output, which might not be extensible to other pathways in plants where protein turnover doesn’t necessarily indicate pathway output.

Other assays use the direct output of a fluorescent reporter construct as their measure of pathway output. In this way, pathway output can be coupled to variable inputs and investigators can test a variety of pathway modifications using high throughput imaging. In S. cerevisiae the mating MAPK pathway is a well described pathway that has been tested in this way. Several approaches include: the combinatorial swapping of regulatory and catalytic domains (Peisajovich et al., 2010), the screen of promoters that tune pathway output or link existing networks (Bashor et al., 2010), and the screening for mutants that rescue pathway output given a stoichiometric imbalance (Kompella, Moses and Peisajovich, 2016). These synthetic modifications of yeast largely test endogenous yeast pathways, taking advantage of the high throughput assays to which yeast is amenable. Testing variants in plant pathways using 1001 Genomes variants in a high throughput context, could be accomplished in the same way though it would require heterologous expression of the relevant pathway in yeast similar to the approach of Nemhauser et al. above.

Here we intend to test the effects of natural non-synonymous variation in plant signaling proteins. Our focus will be the core abscisic acid hormone signaling pathway due to its control over a variety of essential plant growth, stress, and morphological phenotypes. Our ultimate goal is to create an ‘Arabidopsized’ yeast strain that expresses four core proteins from the ABA
pathway driving a green fluorescent protein output. To measure the effects of non-synonymous variation from natural accessions we will mutagenize pathway members with selected variants drawn from the 1001 Genomes Project and will measure the change in pathway output using flow cytometry. The results of this project will lead to the identification of candidate residues for further characterization in planta and will provide more information in terms of each of their specific output compared to previously used in silico methods at a relatively high throughput. Additionally, once the core ABA pathway is constructed other known interactors can be tested in this synthetic context to validate or explore their method of action. If successful, this method could be applied to other branches of the ABA signaling network, or other plant signaling networks ad nauseam.

1.1.3 ABA Stress Signaling

Abscisic acid (ABA) is a terpenoid that is synthesized through the cleavage of C_{40} carotenoids produced in the MEP pathway (Nambara and Marion-Poll, 2005). During stress, ABA rapidly accumulates through activation of 9-cis-epoxycarotenoid dioxygenase (NCED), the rate limiting enzyme in its biosynthesis. Rapid fluctuations in cellular ABA drive the transcriptional activation of many stress genes through its activation of a network of signaling molecules (Lumba et al., 2014).

ABA is bound by the PYR/PYL/RCAR (Pyrabactin Resistance 1 / Pyrabactin Resistance 1 Like / Regulatory Component of ABA receptor) START family of soluble ABA receptors (Park et al., 2009). PYR/PYLs bind ABA, stabilizing their interaction with group A type 2C protein phosphatases (PP2Cs) (Figure 1A). Under non stressed conditions PP2Cs dephosphorylate Subclass 2 Sucrose nonfermenting-1 related kinases (SnRK2s), repressing their ability to phosphorylate downstream targets (Fujii, Verslues and Zhu, 2011). ABA inducible genes usually contain the canonical ABA response element (ABRE) either in pairs or near coupling elements (CE) and drought response elements (DRE) (Guiltnan, Marcotte Jr. and Quatrano, 1990; Skriver et al., 1991; Hobo et al., 1999). ABREs are bound by basic leucine zipper ABA response element binding factors (ABFs) that require phosphorylation by the upstream SnRK2s for full activity (Takashi Furihata et al., 2006; Yoshida et al., 2010). Once ABFs are phosphorylated, they bind ABRE elements and recruit transcriptional machinery leading to a stress response.
In *A. thaliana* each protein in the core ABA signaling pathway is a member of a multi protein family, with robust redundancy in function. Soluble ABA receptors were simultaneously identified by 4 groups using ABA agonists, yeast two-hybrid screens of their binding partners, and their unintentional co-purification with PP2Cs (Ma *et al.*, 2009; S.-Y. Park *et al.*, 2009; Santiago *et al.*, 2009; Nishimura *et al.*, 2010). The 14 member PYR/PYL family (PYR1, and PYL1-13) exist in both monomeric and dimeric states where monomers have a high affinity for ABA, and dimers a ~10 fold lower affinity (Ma *et al.*, 2009; Szostkiewicz *et al.*, 2010; Gonzalez-Guzman *et al.*, 2012). PYR/PYLs bind ABA via a START domain composed of three alpha helices and a curved beta sheet (Miyazono *et al.*, 2009). Phe135, Leu144, and Val193 form the most critical contact points with ABA, though the interaction is additionally mediated by other hydrophobic residues in the binding pocket. PYR1 was identified through its mutant’s ability to germinate in the presence of the selective ABA agonist pyrabactin (Sang-Youl Park *et al.*, 2009) though even double mutants of PYR1 and PYL4 are sensitive to ABA through the redundancy of the remaining PYL receptors. Some PYR/PYLs such as PYR1 localize primarily to the nucleus while others are additionally found in the cytoplasm (Danquah *et al.*, 2014). The receptors have strong expression in leaves, stems, vascular tissue, guard cells, and roots, though this varies depending on the family member. Sextuple PYR/PYL *A. thaliana* mutants with inactive PYR1, PYL1,2,4,5, and 8 (a so-called “112458” mutant) are extremely susceptible to drought conditions, due to their constitutively open stomata, losing 60% of their fresh weight within an hour of movement to an environment with low humidity.

In *A. thaliana* the PP2Cs are members of an 80 protein family of which six (Group A) are known to repress stress gene expression in the absence of ABA (Schweighofer, Hirt and Meskiene, 2004). They are monomeric proteins composed of a C terminal catalytic domain with 11 sub domains. Group A PP2Cs are constitutively active in the absence of ABA and dephosphorylate SnRK2s at Ser/Thr residues in the kinase activation loop, unless inhibited by an interacting protein such as the PYR/PYLs (Umezawa *et al.*, 2009). Binding of PYR/PYLs to PP2Cs upon ABA sensing is mediated by two lid loops that close over the ABA binding pocket of PYR/PYLs (Miyazono *et al.*, 2012) creating an interaction surface where PYR/PYLs directly inhibit the catalytic domain of PP2Cs by inserting a Ser residue directly into the active site blocking a critical Glu residue. ABI1 was the first Group A PP2C shown to play a role in ABA responses, through the insensitivity of its dominant negative mutant alleles *abi1-1*, and *abi1-2* to
exogenously applied ABA in a germination screen (Koornneef, Reuling and Karssen, 1984). ABI1 localization to the nucleus is required for its inhibition of the SnRK2 family kinases via dephosphorylation (Moes et al., 2008). Overexpression of Group A PP2Cs, such as ABI1, in A. thaliana leads to impaired stomatal closure, and reduced ABA-dependent gene expression through their constitutive deactivation of SnRK2s (Santiago et al., 2009; Umezawa et al., 2009).

The SnRK2 family is made up of ten members (SnRK2.1 – 10) that have redundant functions (Fujii, Verslues and Zhu, 2011). They are composed of an N-terminal Ser/Thr kinase domain and a C terminal regulatory domain (Kulik et al., 2011). The regulatory domain has two subdomains; one of which is common to all SnRK2s and is required for osmotic responses; and an ABA domain required for ABA specific responses (Belin et al., 2006). Site directed mutagenesis of two serine residues in the activation loop of the kinase domain has shown that their phosphorylation is required for kinase activity (Boudsocq et al., 2007; Kulik et al., 2011). Upstream phosphorylation of these serine residues is hypothesized to occur via other SnRK2s or other kinases, though autophosphorylation has been proposed as a mechanism for their activation (Fujii et al., 2009). Single mutants of the aptly named Open Stomata 1 (SnRK2.6) result in partial resistance to ABA resulting in constitutively open stomata and a drought sensitive phenotype. Triple mutants of SnRK2.2/2.3/2.6 show further ABA insensitive phenotypes in germination and stomatal closure, and wilt with small decreases in humidity similar to sextuple PYR/PYL mutants (Fujita et al., 2009). Decuple mutants (snrk2.1/2/3/4/5/6/7/8/9/10) are hypersensitive to osmotic stress, and show more extreme phenotypes compared to SnRK2.2/2.3/2.6 mutants (Fujii, Verslues and Zhu, 2011). SnRK2s phosphorylate ABFs at conserved LxRxS/T motifs when PP2C is bound by PYR/PYL (Sirichandra et al., 2010). SnRK2s also phosphorylate other proteins critical for drought and osmotic responses such as potassium channels, and other ion channels that function in guard cells highlighting their diverse and critical roles in ABA dependent responses.

The first ABF, Abscisic acid Insensitive 5 (ABI5), was identified by the ability of its mutants to germinate on ABA containing media (Finkelstein, 1994). Additional ABFs (ABF 1-4) were identified in yeast one-hybrid assays where the previously identified abscisic acid response element (ABRE) was used as a bait sequence (Choi et al., 2000; Uno et al., 2000). ABFs bind to conserved G/ABRE (GGACACGTGCG) motifs in cis elements of ABA dependent genes and activate downstream stress responses. They are composed of a C terminal bZIP domain, and N
terminal regulatory domain containing Ser/Thr residues that are phosphorylated by SnRK2s upon ABA accumulation. ABFs form hetero- and homodimers and have large overlap in their gene activation profiles (Yoshida et al., 2010). Upon their phosphorylation, ABFs bind ABREs and activate transcriptional cascades driving the expression of hundreds of genes that affect all aspects of plant growth (Yoshida et al., 2015). A. thaliana quadruple mutants (abf1/abf2/abf3/abf4) display large decreases in ABA dependent gene expression in aerial and root regions, and aberrant responses to stress. ABF2, ABF4 and ABF3, have overlapping tissue specific expression, and functionality and can form homo- and hetero dimers with one another in plant nuclei (Yoshida et al., 2010). ABF2 is uniquely well characterized with a number of deletion mutants, and phosphomimic constructs having been tested in planta and in protoplasts (Fujii et al., 2009).

The relatively simple hierarchy of the initial ABA response is an attractive option for a first pass at a synthetic reconstruction. With four critical components, each with a number of redundant members, we have a wide set of possible interactions and variations to draw from. Also, because of the large amount of functional analysis performed on the pathway we might have an easier means of interpreting the biological significance of our results in the context of previous work.

1.1.4 Project Goals

To achieve the larger goal of an “Arabidopsized” yeast this section of my project specifically tackled the initial steps of defining a system whereby one could express different A. thaliana proteins and protein fusions in yeast under a variety of contexts and to establish the methods that would be required to test their ability to drive the output of a fluorescent reporter. More specifically I sought to create strains of yeast expressing a promRD29B-GFP reporter construct, ABF2, SnRK2.6, ABI1, and PYR1, and validate their expression and potential activation of our fluorescent GFP reporter using a yeast strain containing promRD29B-GFP, ABF2 and SnRK2.6 cassette.

1.2 Materials and Methods

1.2.1 Overall Pathway Design, and Components

To address the long-term goal of assembling a full ABA pathway we would require the expression of up to four nuclear localized proteins, in addition to a reporter construct. To start,
we decided to recapitulate the minimal core pathway *in vitro* reconstruction performed by Fujii et al. 2009, using ABF2, SnRK2.6 and RD29B-reporter constructs as a proof of concept for pathway activation, with the eventual goal of co-expressing ABI1 and PYR1 in yeast as well. Since ABF2 requires phosphorylation for activity *in planta*; I performed site directed mutagenesis to create aspartic acid and alanine substitutions at four N-terminal conserved Ser/Thr phosphorylation sites in the hopes that this mutant would be constitutively active in yeast (T. Furihata et al., 2006). Additionally, ABF2/AREB1’s N terminal region (residues 1-61) has previously been shown to have transactivation activity in *A. thaliana* (Fujita et al., 2005).

Internal deletions of residues 62 – 317 yield AREB1ΔQT, a mutant that constitutively activates RD29B-LUC reporters when transfected into *A. thaliana* T87 protoplasts. We decided to also test the activity of the ABF2 internal deletion mutant in yeast to see if it displays the same constitutive activity in our synthetic

**Table 1. *S. cerevisiae* strains constructed for this project**

<table>
<thead>
<tr>
<th>Strain ID</th>
<th>Background</th>
<th>Cassette</th>
<th>EasyClone Vector</th>
<th>Intended Use / Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>YPL01</td>
<td>CB008 (W303A MATa, far1a, his3, trp1, leu2, ura3)</td>
<td>None</td>
<td>pCIB2226</td>
<td>Strain used for all construction</td>
</tr>
<tr>
<td>YPL05</td>
<td>X-4:Empty Vector</td>
<td>X-2: Empty Vector</td>
<td>pCIB2226</td>
<td>Empty vector control</td>
</tr>
<tr>
<td>YPL12</td>
<td>X-2: Empty Vector</td>
<td>X-2: Empty Vector</td>
<td>pCIB2180</td>
<td>Empty vector control</td>
</tr>
<tr>
<td>YPL24</td>
<td>XI-2: Empty Vector</td>
<td>XI-2: Empty Vector</td>
<td>pCIB2224</td>
<td>Empty vector control</td>
</tr>
<tr>
<td>YPL25</td>
<td>XI-3: Empty Vector</td>
<td>XI-3: Empty Vector</td>
<td>pCIB2195</td>
<td>Empty vector control</td>
</tr>
<tr>
<td>YPL03</td>
<td>X-2:pRD29B-GFP</td>
<td>X-2:pRD29B</td>
<td>pCIB2188</td>
<td>Reporter Strain used for all pathway construction</td>
</tr>
<tr>
<td>YPL07</td>
<td>X-4:pADH1-ABF2-SV40NLS</td>
<td>X-4:pADH1-ABF2-SV40NLS</td>
<td>pCIB2226</td>
<td>Negative for pathway construction</td>
</tr>
<tr>
<td>YPL09</td>
<td>X-2:pRD29B-GFP</td>
<td>X-4:pADH1-ABF2-SV40NLS</td>
<td>pCIB2226</td>
<td>Strong expression</td>
</tr>
<tr>
<td>YPL14</td>
<td>X-2:pRD29B-GFP</td>
<td>X-4:pcYC1-ABF2-SV40NLS</td>
<td>pCIB2226</td>
<td>Weak / medium expression</td>
</tr>
<tr>
<td>YPL22</td>
<td>X-2:pRD29B-GFP</td>
<td>X-4:pGPD-ABF2-SV40NLS</td>
<td>pCIB2226</td>
<td>Very strong expression</td>
</tr>
<tr>
<td>YPL23</td>
<td>X-2:pRD29B-GFP</td>
<td>X-4:pADH1-ABF2-SV40NLS-FLAG3X</td>
<td>pCIB2226</td>
<td>Western blots</td>
</tr>
<tr>
<td>YPL15</td>
<td>X-2:pRD29B-GFP</td>
<td>X-4:pGPD-ABF2-SV40NLS-FLAG3X</td>
<td>pCIB2226</td>
<td>Western blots</td>
</tr>
<tr>
<td>YPL33</td>
<td>X-2:pRD29B-GFP</td>
<td>X-4:pGA1-ABF2-FLAG3X</td>
<td>pCIB2226</td>
<td>Western blots</td>
</tr>
<tr>
<td>YPL52</td>
<td>X-2:pRD29B-GFP</td>
<td>X-4:pGA1-ABF2delellaQT-SV40NLS</td>
<td>pCIB2226</td>
<td>Test of constitutive activation in yeast</td>
</tr>
<tr>
<td>YPL13</td>
<td>X-4:pADH-GFP-SV40NLS</td>
<td>X-4:pADH-GFP-SV40NLS</td>
<td>pCIB2226</td>
<td>Positive GFP control</td>
</tr>
<tr>
<td>YPL33</td>
<td>X-2:pRD29B-GFP</td>
<td>X-4:pGA1-ABF2delellaQT-FLAG3X</td>
<td>pCIB2226</td>
<td>Western blots</td>
</tr>
<tr>
<td>YPL30</td>
<td>X-2:pRD29B-GFP</td>
<td>X-4:pGA1-ABF2delellaQT-SV40NLS-FLAG3X</td>
<td>pCIB2226</td>
<td>Western blots</td>
</tr>
<tr>
<td>YPL34</td>
<td>X-2:pRD29B-GFP</td>
<td>X-4:pGA1-ABF2mutant-SV40NLS-FLAG3X-tCYC</td>
<td>pCIB2226</td>
<td>Negative control for pathway activation in yeast</td>
</tr>
<tr>
<td>YPL35</td>
<td>X-2:pRD29B-GFP</td>
<td>X-4:pGA1-ABF2mutant-FLAG3X-tCYC</td>
<td>pCIB2226</td>
<td>Western blots</td>
</tr>
<tr>
<td>YPL36</td>
<td>X-2:pRD29B-GFP</td>
<td>X-4:pGA1-ABF2mimic-SV40NLS-tCYC</td>
<td>pCIB2226</td>
<td>Western blots</td>
</tr>
<tr>
<td>YPL37</td>
<td>X-2:pRD29B-GFP</td>
<td>X-4:pGA1-ABF2mimic-FLAG3X-tCYC</td>
<td>pCIB2226</td>
<td>Western blots</td>
</tr>
<tr>
<td>YPL17</td>
<td>XI-2:pADH1-ABI1-SV40NLS-FLAG3X</td>
<td>XI-2:pADH1-ABI1-SV40NLS-FLAG3X</td>
<td>pCIB2226</td>
<td>Very strong expression</td>
</tr>
<tr>
<td>YPL18</td>
<td>XI-2:pCYC1-ABI1-FLAG3X</td>
<td>XI-2:pCYC1-ABI1-FLAG3X</td>
<td>pCIB2226</td>
<td>Weak / medium expression</td>
</tr>
<tr>
<td>YPL19</td>
<td>XI-3:pADH1-PRY1-SV40NLS-FLAG3X</td>
<td>XI-3:pADH1-PRY1-SV40NLS-FLAG3X</td>
<td>pCIB2195</td>
<td>Positive GFP control</td>
</tr>
<tr>
<td>YPL20</td>
<td>XI-3:pCYC1-PRY1-FLAG3X</td>
<td>XI-3:pCYC1-PRY1-FLAG3X</td>
<td>pCIB2195</td>
<td>Western blots</td>
</tr>
<tr>
<td>YPL28</td>
<td>X-2:pRD29B-GFP</td>
<td>X-4:pADH1-ABF2-FLAG3X-tCYC, pGAL1-SnRk2.6-FLAG3X-tADH</td>
<td>pCIB2226</td>
<td>No pathway activation, Western blots</td>
</tr>
<tr>
<td>YPL29</td>
<td>X-2:pRD29B-GFP</td>
<td>X-4:pADH1-ABF2-FLAG3X-tCYC, pGAL1-SnRk2.6-SV40NLS-FLAG3X-tADH</td>
<td>pCIB2226</td>
<td>No pathway activation, Western blots</td>
</tr>
<tr>
<td>YPL38</td>
<td>X-2:pRD29B-GFP</td>
<td>X-4:pADH1-ABF2-SV40NLS-FLAG3X-tCYC, pGAL1-SnRk2.6-SV40NLS-FLAG3X-tADH</td>
<td>pCIB2226</td>
<td>No pathway activation, Western blots</td>
</tr>
<tr>
<td>YPL39</td>
<td>X-2:pRD29B-GFP</td>
<td>X-4:pADH1-ABF2-SV40NLS-FLAG3X-tCYC, pGAL1-SnRk2.6-SV40NLS-FLAG3X-tADH</td>
<td>pCIB2226</td>
<td>Active pathway under GAL induction, Western blots</td>
</tr>
<tr>
<td>YPL40</td>
<td>X-4:pADH1-ABF2-SV40NLS-GFP</td>
<td>X-4:pADH1-ABF2-SV40NLS-GFP</td>
<td>pCIB2226</td>
<td>Nuclear Localization</td>
</tr>
<tr>
<td>YPL41</td>
<td>X-4:pADH1-PRY1-SV40NLS-GFP</td>
<td>X-4:pADH1-PRY1-SV40NLS-GFP</td>
<td>pCIB2226</td>
<td>Nuclear Localization</td>
</tr>
<tr>
<td>YPL42</td>
<td>X-4:pADH1-ABI1-SV40NLS-GFP</td>
<td>X-4:pADH1-ABI1-SV40NLS-GFP</td>
<td>pCIB2226</td>
<td>Nuclear Localization</td>
</tr>
<tr>
<td>YPL43</td>
<td>X-4::pGAL1-SnRK2.6-SV40NLS-GFP</td>
<td>pCIB2226</td>
<td>Nuclear Localization</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------------</td>
<td>---------</td>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td>YPL44</td>
<td>X-4::pADH1-ABF2-GFP</td>
<td>pCIB2226</td>
<td>Nuclear Localization Negative</td>
<td></td>
</tr>
<tr>
<td>YPL45</td>
<td>X-4::pADH1-PYR1-GFP</td>
<td>pCIB2226</td>
<td>Nuclear Localization Negative</td>
<td></td>
</tr>
<tr>
<td>YPL46</td>
<td>X-4::pADH1-ABI1-GFP</td>
<td>pCIB2226</td>
<td>Nuclear Localization Negative</td>
<td></td>
</tr>
<tr>
<td>YPL47</td>
<td>X-4::pGAL1-SnRK2.6-GFP</td>
<td>pCIB2226</td>
<td>Nuclear Localization Negative</td>
<td></td>
</tr>
<tr>
<td>YPL66</td>
<td>X-2::pRD29B-GFP</td>
<td>XI-2::pGAL1-SnRK2.6</td>
<td>pCIB2226</td>
<td>Galactose inducible SnRK2.6</td>
</tr>
<tr>
<td>YPL67</td>
<td>X-2::pRD29B-GFP</td>
<td>XI-2::pGAL1-SnRK2.6</td>
<td>pCIB2224</td>
<td>Galactose inducible SnRK2.6</td>
</tr>
<tr>
<td>YPL71</td>
<td>YPL66</td>
<td>XI-2::pADH1-ABF2</td>
<td>pCIB2224</td>
<td>Constitutive non-fusion ABF2</td>
</tr>
<tr>
<td>YPL72</td>
<td>YPL66</td>
<td>XI-2::pADH1-ABI5</td>
<td>pCIB2224</td>
<td>Constitutive non-fusion ABI5</td>
</tr>
<tr>
<td>YPL73</td>
<td>YPL69</td>
<td>pDEST-AD-ABI5</td>
<td>pDEST-AD</td>
<td>Y1H Constitutive non-fusion ABF2</td>
</tr>
<tr>
<td>YPL74</td>
<td>YPL69</td>
<td>pDEST-AD-ABF2</td>
<td>pDEST-AD</td>
<td>Y1H Constitutive non-fusion ABI5</td>
</tr>
<tr>
<td>YPL75</td>
<td>YPL69</td>
<td>X-4::pADH1-ABF2</td>
<td>pCIB2226</td>
<td>Constitutive non-fusion ABF2</td>
</tr>
<tr>
<td>YPL76</td>
<td>YPL69</td>
<td>X-4::pADH1-GALTAD-ABF2</td>
<td>pCIB2226</td>
<td>Constitutive GAL4-TAD fusion ABF2</td>
</tr>
</tbody>
</table>

We also decided to place constitutively active mutants under inducible promoters so that we could modulate them on or off to act as a positive regulatory control if desired.

We decided to integrate genes into the yeast genome to avoid the burden of multiple high copy plasmids inside a single yeast cell, and having to maintain yeast in multiply selective media; both of which can reduce *S. cerevisiae* growth rates (Karim, Curran and Alper, 2013). Integrated genes often have more stable (though albeit overall lower) expression cell to cell in yeast which allows us to vary expression primarily using different promoter strengths (Lee *et al.*, 2015). However, integration into the yeast genome is more difficult than plasmid transformation and occurs at a lower efficiency, while also sometimes leading to defects in host growth. I investigated/tested several cloning methods (Gateway™, Golden Gate, and other Type IIS RE), and yeast expression plasmids (pRS series). Each had its drawback and advantages: Gateway™ has expensive enzymes and multi part assembly can be difficult, restriction enzymes for Type IIS (BioBrick) and Golden Gate assemblies tended to have low fidelity, so repeated cloning efforts were frustrating.

Ultimately, to construct our synthetic ABA signaling pathway we took advantage of the existing EasyClone 2.0 yeast expression toolkit (Stovicek *et al.*, 2015). The toolkit has a set of vectors that integrate into the yeast genome at locations specifically selected to have stable expression levels demonstrated to not affect yeast growth. The vectors use a variety of selective markers so that strains can be constructed using whichever dominant (ble, KanMX, hphMX, natMX, amdSYM, dsdAMX) or auxotrophic (URA,HIS,LEU,LYS) selectable marker is convenient. The kit also provides primer sequences for checking correct integration into any of the sites, removing the need for primer optimization. Finally, the vectors have identical USER cassette
integration sequences so that gene expression cassettes can be swapped between vectors with ease allowing for the rapid production of different strains and protein fusions.

All yeast expression cassettes were constructed from interchangeable “part” sequences using seamless USER cloning (Geu-Flores et al., 2007). Each cassette is assembled from a nicked and linearized EasyClone 2.0 vector which generates 7–12 bp overhangs, a compatible overhang flanked promoter sequence, a coding sequence, and a C-terminal tag (FLAG3X, SV40NLS, GFP). Flexible Gly-Gly-Ser-Gly-Gly-Ser polylinker sequences are inserted between protein coding regions and C-terminal tags to limit the chance of misfolding due to C-terminal epitope additions. C terminal epitopes were designed to be interchangeable as we intend to perform western blots to validate protein expression, and GFP protein fusions to validate the localization of our proteins in yeast. Since the pCfB series vectors allow for the expression of two proteins on sense and antisense strands all cassettes were inserted facing the tCYC terminator except for bidirectional constructs where SnRK2.6 faces the tADH terminator. C-terminal FLAG3X, SV40NLS and GFP tags are interchangeable, but stacking tags requires sub-cloning; for example, a SV40NLS tagged coding sequence is required to produce protein-SV40NLS-FLAG3X, and protein-SV40NLS-GFP cassettes. Finally, to attempt to validate the ability of ABF2 to bind the pRD29B element in our yeast strains we created ABF2 N terminal GAL4 transactivation domain fusions; this was done to verify that ABF2 can bind our promoter in yeast and drive transcription by recruiting CB008’s endogenous transcriptional machinery.

1.2.2 Pathway Construction

Amplifications destined for USER cassettes are performed using Phusion U Hot Start Polymerase (Thermo Fisher Scientific, Catalog #: F555S) according to the manufacturers recommended conditions in a Biometra T300 Thermocycler. E. coli and yeast colony PCR are performed using NEB Taq DNA Polymerase in ThermoPol® buffer using the manufacturers recommended protocol. ABF2 (AT1G45249), and SnRK2.6 (AT4G33950), ABI1 (AT4G26080) and PYR1 (AT4G17870) coding sequences were amplified from TAIR ORF cDNA clones (www.arabidopsis.org, U16135, SnRK2.6; PDEST-AD074C02, ABF2; GC104649, ABI1; PYR1, U15941) using directional uracil containing primers (https://www.idtdna.com/) required for USER cloning. Yeast ADH1, GAL1, CYC1 and GPD (TDH3) and the GFP ORF sequences were amplified from pRS series shuttle plasmids that were a generous gift from the Chang Lab at
the University of Toronto. The yeast GAL4 TADThe RD29B promoter was amplified from pDEST-AD-2u (Gaudinier et al., 2011) which was a generous gift of the Brady lab. Columbia-0 DNA using primers flanking the 1 kb region directly upstream of the RD29B coding region (Fujii et al., 2009; Ma et al., 2009; Umezawa et al., 2013). C-terminal SV40NLS (Kalderon et al., 1984) and FLAG3X epitope sequences were created by annealing equal molar ratios of ssDNA Ulramer™ oligos (https://www.idtdna.com/) such that overhangs were compatible with the pCfB series vectors. Site directed mutagenesis reactions (point mutants and deletions) are performed using inverse PCR with USER primers designed using AMUSER 1.0 (http://www.cbs.dtu.dk/services/AMUSER/) with an extension time of 2 min + 1 min / kb.

pCfB series EasyClone 2.0 destination plasmids are sequentially linearized with AsiSI (NEB, R0630S) nicked with Nb.BsmI (NEB, R0706S) in 1X NEB 3.1 buffer. PCR products and linearized-nicked vectors are cleaned with Presto™ PCR Cleanup columns (FroggaBio, DF300), eluted in elution buffer and diluted to ~50ng / uL. USER reactions are performed in 0.5X Thermofisher HF Phusion buffer, with a 3:1 vector to insert(s) ratio for 700 bp+ insert fragments and a ~12:1 ratio for FLAG3X and SV40NLS C-terminal tags, using 0.5 uL USER enzyme mix per 10uL reaction (NEB, M5505S). USER reactions are performed according to the EasyClone 2.0 kits protocol and are transformed directly into sub-cloning efficiency DH5α (Thermo Fisher Scientific, Product #: 18265017) and plated on 150 ug/mL LB-Ampicillin plates. Colonies are screened with universal colony PCR primers flanking the USER cassette site, or internal sequence specific primers where required. Plasmids are purified from overnight cultures using PrestoTM Mini Plasmid Kits (FroggaBio), and all overlaps are checked by Sanger sequencing (www.tcag.ca/) and aligned to template using MAFFT (www.ebi.ac.uk/Tools/msa/mafft) in Benchling (www.benchling.com) for correct orientation and identity of inserts. Plasmid harbouring E. coli colonies are stored in 25% glycerol at -80°C.

Haploid CB008 (W303 MATa, far1Δ, his3, trp1, leu2, ura3) were a generous gift from the Chang / Peisajovich Laboratory at the University of Toronto and were used for all yeast manipulations. Yeast are transformed using the Frozen EZ II Yeast Transformation Kit (www.zymoresearch.com) and are grown at 30°C for 3-6 days on selective plates (400 mg/ mL Geneticin, 200 mg/ mL Hygromycin B, 300mg/ mL bleomycin as needed) until colonies are ~2mm in diameter. Yeast DNA is prepared from successful transformants by boiling a small amount of a yeast colony in 20uL 0.02mM NaOH at 96°C for 15 minutes, and pelleting cell
debris by centrifuging at 2700g for 5 minutes at room temperature. 2 uL of the solution is used as template in a 20 uL yeast colony PCR using primers that cross the integration border and are specific for each integration location. Positive transformants are inoculated into 2X YPD, grown overnight, and stored at -80°C in 15% glycerol.

1.2.3 Pathway Validation
1.2.3.1 Western Blots

To attempt to validate the expression of our A. thaliana proteins in yeast we performed western blots using FLAG3X epitope fusions of nuclear localized and non-nuclear localized proteins under the control of constitutive ADH1, CYC1, GPD (ABF2, ABI1, PYR1) and inducible (GAL1) promoters (SnRK2.6, ABF2 mutants). To prepare yeast protein lysates for blotting we use an adapted protocol provided by the Moses Lab at the University of Toronto (Zalatan et al., 2012). A single colony is grown overnight in YEPD media (1% yeast extract, 2% peptone, 2% glucose/dextrose) diluted to an OD$_{600}$ of 0.2 in YEPD (YEP + 2% Galactose for GAL1 inducible cassettes) and grown for approximately 4-6 hours to an OD$_{600}$ of between 0.6 and 0.8. Cultures are pelleted at 1600g for 2 minutes at 4°C to harvest yeast. The supernatant is discarded, and the pellet is frozen in liquid nitrogen, resuspended in 100uL Lysis Buffer (0.1M NaOH, 0.05 EDTA, 2% SDS, 2% β-mercaptoethanol), and incubated for 20 minutes at 90°C. The lysate is neutralized with 2.5 uL of 4M acetic acid, vortexed for 30 seconds and incubated for 10 minutes at 90°C. The supernatant (approximately 100 uL) is diluted with 25uL of 5X Loading Buffer (0.25M Tric-HCL pH6.8, 50% Glycerol, 0.05% Bromophenol Blue), centrifuged at 1600g for 2 minutes at room temperature and the supernatant is either stored at -20°C or 15uL loaded directly onto an SDS PAGE gel. Samples are run at 150V for 1 hour on a 12% SDS resolving gel, transferred to nitrocellulose membranes using the OWL™ Semidy electroblotting system and are blocked in PBS-T + 5% milk (2% BSA in previous blots) overnight. Membranes are incubated with a 1:30,000 dilution primary rabbit monoclonal Anti-FLAG3X (CST DYDDDDK Tag, #2368) antibody in PBS-T + 1% milk for 1 hr, washed 4X with PBS-T for 1 hour and incubated and washed again with Anti-Rabbit HRP- linked Antibody (CST, #7074). Blots are incubated in Amersham ECL Select Western Blotting Detection Reagent and are auto optimal exposed on a Bio Rad ChemiDoc™ MP Imaging System.
1.2.3.2 Nuclear Localization

To validate correct localization of our heterologous *A. thaliana* proteins to the yeast nucleus we created C-terminal tagged GFP fusions of our four pathway components under the control of constitutive ADH1 (ABF2, ABI1, PYR1) and inducible GAL1 (SnRK2.6) promoters. Cells are cultured in the same manner as is used for western blotting except that SD media is used to limit background fluorescence of cultures. Cultures are pelleted to concentrate cells and are resuspended in ~100 uL of media. A single drop of cell suspension is placed on a concanavalin A coated coverslip, the coverslip is gently tapped down to immobilize yeast, and GFP fluorescence, and DIC images are recorded using a Leica TCS SP5 Confocal Microscope with a 488nm Argon excitation laser at 20% using the GFP bandpass filter, and 2 frame averages at 200 Hz.

1.2.3.3 Flow Cytometry

To measure pathway activation a single colony is inoculated in YEP-Raf (2% Raffinose) media and grown overnight. Cultures are pelleted and resuspended in PBS and are grown at 30°C with shaking at 225RPM for 2.5 hours in a 96 well deep block. Cell growth is arrested with cycloheximide after 2.5 hours of growth and 10,000 events are measured using a MACSQuant® VYB flow cytometer (Miltenyi Biotec Inc.) using the 488nm excitation laser and the 525/550nm bandpass filter. Data is analyzed using the flowCore, and flowStats (Meur, Hahne and Ellis, 2007; Hahne and Gopalakrishnan, 2014) packages in R. Yeast events are gated using Forward Scatter and Side Scatter channels automatically using the lymphGate function in the flowStats package capturing between 70 and 85% of events as yeast.

1.2.3.4 Testing ABF2’s activation of pRD29B-GFP

To test if non-fusion ABF2 or GAL4-TAD-ABF2 fusions were able to activate our reporter construct we grew yeast strains harbouring untagged GAL1-SnRK2.6 cassettes and constitutively expressed non-fusion ABF2 or GAL4-TAD-ABF2 fusions under the control of the ADH1 promoter. We also tested the ability of ABI5 (an ABF related to ABF2) to activate our reporter, in order to determine if a different transcription factor might be able to activate our system. Yeast cells were cultured in the same manner as for western blotting in Raffinose(+) (non-inductive) and Galactose(+) and were imaged using a Leica MZ16F Fluorescence Stereomicroscope, with a 488nm Argon excitation laser, on the ET-GFP excitation/emission or
transmission filters, under 230x magnification (10x eyepiece, 11.5x zoom, 2x planapochromatic lens) with a Leica DFC7000T camera.

1.3 Results and Discussion

1.3.1 Cloning and Transformation

Once optimization of our USER cloning system was complete I successfully constructed a series of cassettes for expressing ABF2, ABI1, and PYR1 proteins, and integrated them into CB008 (Table 1). My initial efforts to transform constitutively (ADH1) expressed SnRK2.6 into CB008 were not successful for several reasons; bleomycin selective plates were repeatedly contaminated (despite filter sterilization of bleomycin and a lack of contamination on all other plates), or had few transformants that failed to be confirmed by yeast colony PCR. I attempted to transform CB008 with constitutive SnRK2.6 cassettes under the control of ADH1, CYC1, and GDP promoters in several different integrating constructs (different integration locations, different selective markers) with no success and so decided to use inducible GAL1 promoters to drive SnRK2.6 expression in the system. I successfully transformed SnRK2.6 at two different genomic locations (with two different selectable markers) to allow future students to use a more flexible set of selection markers (including non EasyClone 2.0 plasmids) as desired. Additionally, since a single integration event is easier to perform than two sequential ones I created a bidirectional cassette for both ABF2 and SnRK2.6 and successfully transformed it into CB008. This way, to compare subsets of specific missense mutants a future student would need only modify one plasmid template using a single inverse PCR reaction and could use a single set of standardized sequencing primers making downstream analyses easier. However, since none of my downstream pathway activation experiments yielded any GFP signal, these plasmids would need to be modified to be used in a final functional system.

All pathway construction intended to test for activation of ABA is done in the background of YPL3 (except negative controls) as indicated in Table 1. However, I have additional strains containing singly integrated ABF2 / SnRK2.6 so that future students can test different reporter constructs as desired.
1.3.2 Western Blots

I was able to confirm the expression of pADH1-ABF2 in both nuclear localized (NL) and non-NL FLAG3X tagged constructs (NL, 50.2 kDa; non-NL 48.9, kDa), though not under the control of the GPD promoter (Figure 1.2). I was additionally able to confirm the expression of pADH1 driven ABI1 (52.2 kDa expected) and PYR1 (25 kDa expected) NL constructs though not under the control of the CYC1 promoter. In Figure 1.2A a very strong band is evident in the pADH1-ABI1-SV40NLS-FLAG3X lane at ~63 kDa (not present in other lanes) but a 52.2 kDa band is not present. However, in 1.3B a 52.2 kDa band is present, in addition to the 63 kDa band suggesting that perhaps ABI1 is being post-translationally modified; the PP2C’s are known to be regulated by 26S proteasome degradation in planta, with ABI1’s ubiquitination being mediated by the U-box E3 ligases PUB12/13 (Kong et al., 2015). A single ubiquitin molecule would add ~8.5 kDa to the expected size of ABI1 for a band of approximately 60.7 kDa, similar to what is depicted in Figure 1.2A and B. Future analyses could test for ABI1 ubiquitination in our yeast system using a 26S proteasome inhibitor such as MG132 (Moes et al., 2008), to see if it causes an accumulation of polyubiquitinated ABI1 not being degraded in yeast. Ubiquitinated ABI1 might not be active and could impede SnRK2.6 dephosphorylation causing a lack of correct pathway response in future analyses.

CYC1 and GPD are weak and high strength constitutive promoters respectively (Partow et al., 2010), so it’s conceivable that CYC1 is not very active under my culture conditions, but it’s unclear as to why I don’t observe a very strong band for pGPD-ABF2. None of my CYC1 cassettes appear to be expressing protein (ABF2, ABI1, PYR1) so perhaps there is an issue with that promoter fragment that is common to all three (though sequencing results confirm it is in the correct orientation). Our extraction protocol doesn’t include any protease inhibitors in the lysis stage, so perhaps a small amount of ABF2 is being degraded in the case of pCYC1, but there is an absence of degradation banding in the pCYC1-PYR1/ABI1 and pGPD-ABF2 lanes. I think that a likely cause is unequal sample loading due to uneven OD_{600} measurements prior to loading and differences in original culture OD_{600} prior to lysis. For example, in Figure 1.3A and B (my first two attempts) band intensities and background vary greatly from lane to lane whereas in Fig 1.3C the background is more even between the two samples shown (full gel in Appendix). Additionally, I used the Amersham ECL Select Western Blotting Detection Reagent which is used to detect very weak to medium amounts of protein, Figure 1.2A, and B use 1:10,000 and
1:30,000 dilutions of primary and secondary antibody respectively, whereas Figure 1.2C is
1:100,000 and 1:300,000 (the maximum dilution suggested by the kit), so some background
might be just due to an overabundance of protein for the sensitivity of our kit.

1.3.3 Nuclear Localization

GFP fusions of NL and non-NL ABF2, NL PYR1 display fluorescence in the nucleus, whereas
non-NL is observed for SnRK2.6 non-NL fusions (Figure 1.3). Indeed, ABF2 possesses a
bipartite NL signal in its DNA binding domain so perhaps it does not require an SV40NLS to
localize in yeast. PYR1 nuclear localization appears far tighter in the presence of an SV40NLS
sequence (Figure 1.3D, E, H, I compared to A). SnRK2.6 non-NL constructs appear to show
either very strong cytoplasmic and nuclear expression or just nuclear expression. My positive
pADH1-GFP-SV40NLS shows an intense signal in the nucleus but also shows GFP signal in the
cytoplasm; cytoplasmic – nuclear shuttling is a known method of transcriptional regulation in
yeast that may account for this effect, though alternatively the GFP moiety itself may account for
the mis localization by partially disrupting SnRK2.6 structure (Huh et al., 2003). A time series
following SnRK2.6-GFP expression could determine if SnRK2.6 localization shifts over time,
which could contribute to variability in the output of the system at different timepoints, requiring
strong standardization of experimental methods in a final system. We observed no fluorescence
for both ABI1 fusions and for SnRK2.6 NL fusions (Figure 1.3). Perhaps in ABI1 C terminal
fusions of GFP or SV40NLS destabilize ABI1 or GFP folding is unable to occur because ABI1 is
badly folded itself. Given our Western blot (Figure 1.2) results for ABI1 it is conceivable that
ABI1 is being ubiquitinated and degraded and therefore does not exhibit any GFP signal.
SnRK2.6 may have the same issue except it appears to be dependent on the presence of the
SV40NLS. N terminal SnRK2.6 and ABI1 fusions (MBP and 10X His) have been used
previously for kinase / phosphatase assays (Belin et al., 2006; Fujii et al., 2009) suggesting that
N terminal tagged versions of these proteins display proper localization and activity. To verify
that these proteins are truly nuclear localizing a follow up analyses using a cytoplasmic
expressed fluorescent marker (RFP, mCherry), or fixation followed by DAPI / Hoechst staining
of nuclei would ensure that the restricted GFP signal is indeed nuclear localized.
1.3.4 Testing for Pathway Activation

My initial tests did not show GFP signal from any of my tested constructs compared to positive (ADH1-GFP) and negative (CB008, pRD29B-GFP) controls following glucose or galactose induction for 2.5 hours (Figure 4). To test if the lack of pathway activation was due to growing conditions; I modified culture time (1hr to overnight), media (Synthetic vs YPD), induction period (1hr to overnight), media volume (200µL to 5mL) but never observed any fluorescence. Previous analyses of ABF2’s ability to bind the ABRE and RD29B have used synthetic trimers of the wheat Em1A element in a yeast one-hybrid assay, where ABF2 is fused to a yeast GAL4 TAD, that recruits the yeast transcriptional machinery, upstream of a reporter element (Choi et al., 2000).

Table 2. Yeast strains expressing combinations of SnRK2.6 and ABF2/ABI5 yield no fluorescent output.

<table>
<thead>
<tr>
<th>Pathway Activation</th>
<th>SnRK2.6</th>
<th>ABF2 / ABI5</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>SnRK2.6</td>
<td>ABF2</td>
</tr>
<tr>
<td>No</td>
<td>SnRK2.6</td>
<td>GALTAD-ABF2</td>
</tr>
<tr>
<td>No</td>
<td>SnRK2.6</td>
<td>ABI5</td>
</tr>
<tr>
<td>No</td>
<td>SnRK2.6-FLAG3X</td>
<td>GALTAD-ABF2-SV40NLS-FLAG3X</td>
</tr>
<tr>
<td>No</td>
<td>SnRK2.6-SV40NLS-FLAG3X</td>
<td>GALTAD-ABF2-FLAG3X</td>
</tr>
<tr>
<td>No</td>
<td>SnRK2.6</td>
<td>pDEST-AD-ABI5</td>
</tr>
<tr>
<td>No</td>
<td>SnRK2.6</td>
<td>pDEST-AD-ABF2</td>
</tr>
</tbody>
</table>

Since none of our previous tests showed any activation we were concerned that perhaps our C-Terminal tags were modifying ABF2’s ability to bind RD29B or that ABF2 did not show the same ability to recruit the transcriptional machinery. To test this, I created a strain expressing untagged ABF2 in the background of GAL1-SnRK2.6 and promRD29B-GFP and additionally constructed an N-Terminal fusion of ABF2 with GAL4’s TAD. I also constructed matching ABI5, ABI5-TAD constructs to test if ABI5 was able to activate our reporter but none of the constructs yielded any GFP signal (Figure 5). ABI5 and ABF2 have been shown to bind ABREs in planta and through yeast one-hybrid assays previously (Uno et al., 2000; Nakamura, Lynch and Finkelstein, 2001; Koornneef et al., 2002) though only ABF2 has been shown to directly bind RD29B whereas ABI5 transactivates through its interaction with other regulators. The inability of our GAL4-TAD-ABF2 fusion to drive our reporter suggests that full length RD29B may not be an optimal promoter for ABF2 to bind in yeast. Indeed, previous yeast one-hybrid assays have used synthetic minimal promoters or sections of A. thaliana promoters containing
ABREs and CE repeats as baits, and have found that 3’ flanking residues matter greatly and cause a significant difference in the ability of TF to bind ABI1 promoters relative to RD29B (Wu et al., 2018).

Perhaps future tests should use synthetic promoters or sections of RD29B as their promoter element. A future student could swap out the RD29B promoter for a short synthetic promoter composed of ABRE repeats similar to that used in Choi et al. (2000), or Wu et al. (2018). Another potential positive control fusion could be ABF2 fused to LexA’s DBD to test if ABF2 can recruit transcriptional machinery in yeast when bound to a LexA promoter-GFP or a bacterial B42-TAD-ABF2 (where B42 is the TAD used in the DupLEXA kit) fusion to see if ABF2 can bind an ABRE and drive expression with the aid of a strong TAD (Origene-Technologies, no date). Hopefully, one of these additional tests would elucidate if ABF2 is a suitable candidate for pursuing a synthetic pathway reconstruction. The addition of a TAD to ABF2, however, may partially eliminate our ability to measure the effect of variants on ABF2’s interactions with other proteins.

1.4 Conclusions and Future Directions

In this portion of my project I have shown that we are able to quickly produce sets of cassettes for expressing members of the ABA pathway in yeast. I have additionally shown that PYR1, ABF2, and SnRK2.6 are translated and localize to the nucleus via confocal microscopy as desired and that ABI1, PYR1, and ABF2 are being translated via western blotting. Additionally, I have created a set of strains for future students to continue construction our ‘Arabidopsized’ yeast, and have made strides into determining what components might be required for a future system. The protocols I have assembled for flow cytometer work, my scripts for analyzing our flow data and my cloning techniques can be used as well in the future to continue this work in the Provart Lab.

Unfortunately, I was unsuccessful in assembling an ‘Arabidopsized’ yeast. As I suggested previously, efforts to use different ABF2 TAD fusions could be used but these essentially recapitulate other kits such as the Grow N Glow yeast one-hybrid assays (https://www.mobitec.com/cms/products/bio/01_2H_sys/gfp_ohs.html). Perhaps the best solution would be to simply use a commercial kit for the reporter output and use my constructed cassettes to express upstream pathway members (SnRK2.6, ABI1, PYR/PYLs). This would
partially eliminate the ability to test the downstream ABF transcription factors as the effects of non-synonymous changes to ABF2 might not be relevant in an ABF fusion protein due to changes in its binding affinity for partners conferred by fusion proteins. Luckily, the ABA regulatory network is large and has many members that can be tested.
Figure 1.1. The Core abscisic acid signaling pathway and our synthetic basic reconstruction. A. Modified from Furihata et al. (2006) (used with permission) this diagram shows pathway states in the presence and absence of ABA and with a well characterized abi1-1 mutant that impedes the ability of RCAR/PYR/PYL receptors to bind the PP2C ABI1 and inhibit its phosphatase activity. B. The base synthetic reconstruction, in the absence of SnRK2.6 constitutively expressed (via ADH1) ABF2 is unable to bind ABRE elements contained in the A. thaliana drought responsive promoter of RD29B similar to the interaction between ABFs and SnRK2 kinases in A. thaliana nuclei (Sirichandra et al., 2010). C. Intended activity of the synthetic pathway, in the presence of galactose GAL4 binds the GAL1 promoter and drives strong expression of SnRK2.6 which phosphorylates ABF2 causing it to bind RD29B and drive transcription of the synthetic pathway.
Figure 1.2. Western blots of FLAG3X tagged ABA signaling proteins heterologously expressed in S. cerevisiae. Proteins were extracted from 5ml of 0.6 – 0.8 OD$_{600}$ S. cerevisiae using an NaOH protein lysis protocol and were solubilized in SDS sample buffer. Lysates were resolved by SDS-PAGE, transferred to a nitrocellulose membrane blocked in 2% BSA overnight, incubated with 1:10,000 primary antibody and 1:30,000 secondary antibody in panels A,B and 1:30,000 and 1:300,000 in panel C. Membranes were then incubated in Amersham ECL Select Western Blotting Detection Reagent and were “Optimal Auto Exposed” on a BioRad ChemiDoc™ imaging system. Red arrows indicate expected band sizes in lanes where a protein was expected but a band was not visible.
Figure 1.3. Nuclear localization of GFP tagged ABF2, ABI1, PYR1, and SnRK2.6 in *S. cerevisiae*. Strains were grown overnight in YNB+Gal media, 15 μL of yeast media suspension was placed on a #1.5 concanavalin A coated coverslip, a microscope slide was pressed down gently but firmly to immobilize yeast, and images were acquired using a Leica TCS SP5 confocal microscope at a 63X magnification with an Argon 488nm excitation laser (20%) with a GFP bandpass filter. DIC and GFP images were sequentially acquired and overlaid using the LAS X software suite. GFP images are the result of 2 frame averages. Panels A,B are negative strain controls, panel C is constitutively expressed GFP. Panels D,E,F,G are nuclear localized constructs, whereas panels H,I,J,K are non-nuclear localized constructs. Inset photo on the bottom left is Figure 3A from Fehrenbacher *et al.* (2005) (used with permission) showing the nuclei and mitochondria in DAPI stained yeast (Fehrenbacher, 2005).
Figure 1.4. Glucose and galactose induction of *S. cerevisiae* strains. Yeast strains were grown overnight in YEP + Raf media, washed with PBS, diluted in SD+Glu, or SD+Gal to an initial OD$_{600}$ of 0.05 in a 1mL deep block media and grown @ 30°C for 2.5 hours at 225 RPM. All strains are in the pRD29B-GFP background except CB008 and GFP-SV40NLS. All ABF2 constructs are under the control of ADH1 unless indicated, all SnRK2.6 are under the control of pGAL1. Yeast growth was arrested with cycloheximide, and GFP fluorescence of 10,000 events was recorded on a MACSQuant® VYB (Miltenyi Biotec) using the blue 488nm excitation laser and the 525/550 nm filter (B1). Events were gated using the lymphGate function from the flowStats package in R (using a norm2Filter). Relative fluorescence is the mean fluorescence of a population divided by the mean fluorescence of CB008. Each bar is the average of 8 samples taken on two successive sampling dates.
Figure 1.5. GAL4 trans activation domain fusions show no GFP expression following induction with Galactose. Panels A,B show constitutively active GFP expression under galactose (inductive) and raffinose (non-inductive) as a positive control. Panels C,D,E show CB008 with integrated SnRK2.6-GFP constructs which show strong expression and nuclear localization after 4 hours of growth in galactose media. F,G are negative control strains containing only our reporter element. Panels H,I show CB008 with integrated GAL1-SnRK2.6 and ADH1-GALTAD-ABF2 cassettes and is representative of all the images for the strains in Table 2.
2 The Variant Viewer

2.1 Introduction

2.1.1 Inspecting Variation

The 1001 Genomes Project (1001genomes.org) has made available a wealth of information and tools for investigators to obtain population data for their gene or region of interest in *A. thaliana*. As *A. thaliana* is the system where most inferences regarding plant physiology have been made, it is often used as the first step in an analysis of any gene to be eventually modified in crop species. Due to its ease of genetic manipulation, fast growth time, and relatively small genome it is amenable to all kinds of assays not available in other plants (Provart et al., 2016).

Additionally, *A. thaliana* has a wealth of germplasm accumulated over the years characterizing ecotypes grown in different regions (Europe, Asia, North America, Africa) under different climatic regimes, and selection pressures. For these reasons the 1001 Genomes Consortium selected a set of 1135 ecotypes of *A. thaliana* that were more representative or diverse than previous samples and compiled their genomes into a single publicly available resource for plant biologists (Alonso-Blanco et al., 2016). The data set contains ~10.7 million biallelic single nucleotide polymorphisms and ~1.5 million indels for researchers to query. Of the 14 million variants ~1.4 million are missense variants representing one missense variant per 88 bases or 52 missense variants per protein coding gene on average. Interrogating these variants bioinformatically can be an informative method where other methods have been previously exhausted.

Indeed, large collections of ecotypes have been used previously to identify previously unknown regulators of ABA signaling. For example, Kalladan et al. (2017) sought to determine how ABA accumulation specifically during stress is determined in *A. thaliana* given that similar water stress conditions can yield different levels of endogenous ABA depending on the ecotype queried, suggesting an additional layer of regulation (Kalladan et al., 2017). Their GWAS analysis uncovered 44 candidate genes that were tested for differences in ABA accumulation under drought stress, six of which had never been previously linked to ABA responses. Interestingly, the analyses found that ABA accumulation was not higher in accessions that were sampled from regions with frequent droughts. The authors suggested that this was because variation in genes like the ones they identified were acting as an additional regulatory layer.
making further ABA accumulation unnecessary. This kind of result lends credence for the need of tools to analyze variation in these systems at a broader level than previously done before.

Included in the 1001 Genomes online suite are useful tools to subset FASTA pseudogenomes and variant call format (http://tools.1001genomes.org/pseudogenomes, http://tools.1001genomes.org/vcfsubset) data sets by region or gene, to view polymorphic sites with annotations in a tabular format (http://tools.1001genomes.org/polymorph), and to plot variants onto a single gene coding sequence (http://1001proteomes.masc-proteomics.org/). These tools help an individual filter their gene of interest for specific classes of variants (missense SNPs, silent SNPs, indels, truncations) and then download the raw data for further analyses (Figure 2.1). More recently the 1001 Genomes Project exposed their application programming interface (API) (http://tools.1001genomes.org/api/index.html) for 3rd party apps to request data on the fly as requested by a user. For example, the ePlant’s Molecule Viewer (Waese et al., 2017) web application in the Bio- Analytic Resource for Plant Biology (http://bar.utoronto.ca/) allows a user to view the missense variants (retrieved from an older 1001 Proteomes (Joshi et al., 2012) data feed, encompassing 452 ecotypes) plotted against the reference gene model of a gene of interest while also showing a JSmol 3D model with PFAM (Finn et al., 2014) or CDD (Marchler-bauer et al., 2018) domains from online data sources highlighted. Individuals can also view a genomic locus with different splice isoforms and translated protein sequences with overlaid missense and silent variants in the ePlant Sequence Viewer (based on Araport.org’s(Krishnakumar et al., 2015) JBrowse instance), and can convert from DNA modification to protein modification across a locus. These tools allow users to view different levels of variation across the 1001 Genomes data while overlaying genome annotation to give a holistic view of a variant’s possible impact.

However, researchers are sometimes interested in comparing families of genes in *A. thaliana* to attempt to elucidate the functions of unknown members by looking for differences in active sites or known functional domains. Often, researchers will perform a multiple sequence alignment using their algorithm of choice and compare conserved regions to see if the presence or absence of a residue at a given site known to confer activity in a protein correlates with a specific level of catalytic activity in an unknown protein. For example, the Christendat Lab previously characterized the shikimate kinase(SK) family (proteins involved in the production of aromatic amino acids) and compared the conservation of SK shikimate binding sites within the family
Fucile, Falconer and Christendat, 2008). They found that there were differences in shikimate binding sites between SK and SK Like proteins in the family and that this was reflected in their binding affinity for the shikimate molecule. Ultimately, they found that positive selection drove diversification and neofunctionalization of the SKL proteins. Since a multiple sequence alignment of these binding sites showed them to be positively selected they might be able to find new variants at these sites that confer differences in activity of those plants using missense variants from the 1001 Genomes Project, but this might only be apparent in an overview showing an alignment of the SK and SKLs in context with the variants. Another tool on the BAR called Gene Slider (Waese et al., 2016) (http://bar.utoronto.ca/geneslider/) allows a researcher to view conserved regions across species, but not across gene families within a species.

There is no simple online tool that shows variants across A. thaliana gene alignments and also plots population wide variation simultaneously, as it is largely a process that must be done using desktop applications or command line scripts unfamiliar to biologists. Another level of complexity often not captured is that of alternative splicing. Ding et al. (2014) found that under salt stress there is an increase in differential alternative splicing in genes that are related to the stress applied moving their transcripts from non-functional or less functional isoforms to active functional isoforms (Ding et al., 2014). If an investigator wished to explore variation in these alternatively spliced genes they would need to compare the correct gene isoforms to fully understand what a given variants impact might be in a specific ecotype. Given that these tools each have different strengths, a slightly different tool that allows a researcher to quickly overview A. thaliana genes with additional annotation such as the PFAM and CDD data contained in the BAR Molecule Viewer might assist researchers in obtaining a general overview of the variation in their genes of interest.

2.1.2 Project Goals

In this project I sought to create a simple web app, referred to hereafter as the Variant Viewer (http://bar.utoronto.ca/~mcumming/Variant_viewer/) for investigating the 1001 Genomes data set, that would allow researchers to view population level variation across splice isoforms, and across alignments of coding sequences, with a useful level of annotation for orienting oneself. The application is not intended to act as a complicated evolutionary analysis tool; those analyses are best done with a careful idea of the ultimate goals and curation of the sequences used. This
web application is intended to act as a simple means for investigators to get a broad idea of the
variation present in their gene of interest, without having to navigate to several different data
sources to collate the information on their own and then to move on with further analyses.

2.2 Application Design

2.2.1 Data Sources and Javascript Libraries

To construct gene models, and plot variants, I request data from several sources when a user
enters an Arabidopsis Genome Initiative identifier (AGI ID). First the Araport11 gene structure
for that AGI ID is requested from an internal BAR web service
which returns gene models corresponding to the genomic positions for each exon, intron, and
untranslated region. Next, a similar internal BAR web service
(http://bar.utoronto.ca/webservices/bar_araport/get_protein_sequence_by_identifier.php?locus=
AGI_ID) returns the protein sequence for the selected AGI ID isoform, followed by a request to
the 1001 Genomes API (http://tools.1001genomes.org/api/v1.1/effects.json?type=snps;accs=all;
gid=AGI_ID) requesting Variant Call Format data annotated by SNPeff (Cingolani et al., 2012)
for the selected AGI ID. All AGI ID isoforms that have valid gene models have corresponding
protein sequences (for example AT1G45249 returns 10 gene isoforms, so AT1G45249.12 will
not have a corresponding protein sequence), but not all gene isoforms have corresponding variant
entries in the 1001 genomes database. This is because the variants are assigned to the transcript
that they map most closely to instead of to the representative gene model (for example
AT1G45249 isoforms 1,2,3 all have mapped variants but only AT1G45249.3 is the
representative model). If the 1001 Genomes variant query is successful, a final set of requests are
made to two BAR webservices (http://bar.utoronto.ca/eplant/cgi-bin/CDDannot.cgi,
http://bar.utoronto.ca/eplant/cgi-bin/PfamAnnot.cgi) that return PFAM (Finn et al., 2014) and
CDD (Marchler-bauer et al., 2018) domains for the AGI ID’s protein sequence. CDD requests
return a domain annotation and individual residues for a location in a protein query (for example
“DNA binding domain: K235, L238, I241”), whereas PFAM requests return a domain,
annotation, Expectation Score and Start / end locations (for example “DNA Binding Domain: [Start: 235, End: 241, Expect 2E-45, Annot: Basic Leucine Zipper]”). It should be noted that
since different isoforms of genes have different coding sequences, and each are submitted
separately as queries, some isoforms may return domains while others may not. All returned data are in JSON format and are saved to objects that can be exported by the user after submitting their query if they wish to have the original information for their reference.

My initial efforts to plot variants were done using Python and R, where all the variant call format (Daneczek et al., 2011) (VCF) file parsing was done by Python and the plotting/data manipulation was performed via R. This method worked for a single user who had coding knowledge and was comfortable manipulating large data sets but was opaque to the non-Python/R fluent investigator. However, partway into my project the 1001 Genomes released an API that allowed for requests to be made to their data set. This allowed me to build a client-facing web application instead of asking a user to learn a scripting language. The application interface is built using “vanilla” Javascript, and a Bootstrap 4 (https://getbootstrap.com) template and style layout. Graphics and tables are added to the web app’s “document object model” using d3.js (https://d3js.org). The variant and coding sequence plot is rendered using the open source plotly.js (https://plot.ly/javascript) library which allows the user to interact with the convenient plotly.js interface (zooming, export, point selection etc.).

The web application is split into five main components, a controller script that determines what is being displayed to the user (main.js), a data script that asynchronously requests and formats data (varData.js), two plotting scripts that render images into graphics for the user (plots.js and overview.js), and a backend R script (R-Development-Core-Team, 2008) that aligns proteins and maps their locations (alignProteins.R). The scripts can be accessed by looking at my github account (https://github.com/daea/Variant_viewer).

2.2.2 Application Logic

Upon navigating to the Variant Viewer the user is greeted with a simple Bootstrap 4 responsive layout and a splash page that provides basic information about the app and a sample gene model (Figure 2.2). The title bar provides links to several databases and viewers for information on their gene of interest. The side input panel (Figure 2.2, Item 1) contains a dropdown suggestion search box that will autocomplete AGI ID and gene names and restrict the user to entering only valid gene IDs. When users have entered a valid AGI ID they click the “Add Gene” button to select a gene and request relevant data from the data sources mentioned in 2.2.1 above. If an improper AGI ID is entered the application will prompt users with an alert box and they will be
given a chance to re-enter an AGI ID. The “Selected genes” panel (Figure 2.2, Item 2, 4) is then populated with that gene and the gene models are drawn in the main app panel to the right. Underneath the Selected genes panel is an export panel that allows users to export their data if desired (Figure 2.2, Item 3). As users “Add genes” to their selected genes list, the main panel draws the respective gene isoforms and sends a request to the 1001 Genomes API for the variant data (Figure 2.2, Item 5). The gene models are scalable vector graphics (SVG) images that dynamically scale to fill the page width when they are rendered. At this time the width of the gene models is not responsive (they do not change size when the page is resized). Gene models are plotted in three colours to represent exons, untranslated regions and coding regions (Figure 2.2, Item 7). All isoforms of a gene model are scaled to the total length of a gene on the chromosome, so the different features line up and allow the user to easily see which features are retained in alternative splice variants. The user can hover over the gene models to view chromosomal position for the type of feature they are interested in (Figure 2.2, Item 6).

When the 1001 Genomes data has been retrieved and parsed, the application plots variants onto the gene models. The positions of variants on the horizontal are calculated relative to the width of the parent gene model to reflect the genomic position of a variant in the gene. Variants are currently plotted in two colours, red and black, for missense variants and all other variants respectively. The colour of the circle fill (red or black) currently reflects the number of ecotypes (out of 1135) that contain the respective variant. Similar to the gene models, when users hover over the variants they are provided with a tooltip that displays information about that variant such as the positions, severity, type, codon change (if applicable), the variant code as called by SNPeff and the number of ecotypes that contain the variant. If a user selects multiple genes, new requests are made, and gene models are rendered for each gene.

After users have selected several genes, if desired, they may submit the genes for alignment. At this point the AGI IDs and protein sequences are submitted to a server-side PHP script (authored by Vincent Lau and Asher Pasha) that passes the data to an R script which performs a multiple sequence alignment (MSA) using the DECIPHER (Wright, 2016) function AlignSeqs. The R script then returns a matrix that relates the new position in the MSA to the actual positions of the amino acids in their coding sequences. This matrix is then used to map the variant positions in the requested 1001 Genomes data to the multiple sequence alignment (for example in an alignment of two proteins position 3 in an original CDS may correspond to position 8 in the
MSA due to the insertion of gaps prior to position 3). The application then renders a plotly.js scatter plot above the table of gene models which shows the number of ecotypes containing a given variant relative to their position in the multiple sequence alignment (Figure 2.3). The bottom subplot displays PFAM and CDD domains corresponding to each of the plotted genes as coloured boxes overlaying the MSA. Additionally, the MSA can be displayed by clicking the “Show MSA” entry in the figure legend, which then displays the residues for each aligned coding sequence (Figure 2.3C). A user is able to zoom in on the plot and highlight variants by selecting them, export the graph as a portable network graphics (PNG) file or edit the graph using the online plotly.js interface (Figure 2.3B). Additionally, by clicking on legend entries a user can choose to remove missense, synonymous or premature stop codons to filter the displayed data as desired.

2.2.3 Natural Variation in the ABA signaling members

To compare the variation present in the core ABA signaling pathway members, the variants present in the members of the ABF, SnRK2, Group A PP2C and PYR/PYL/RCAR members were retrieved from the 1001 Genomes Project site using Polymorph 1001. The PANTHER Database (http://www.pantherdb.org/) was queried for A. thaliana genes with annotations as abiotic stress related, kinases, phosphatases, receptors, and transcription factors (Mi, Muruganujan and Thomas, 2013). The AGI IDs were extracted from the retrieved lists and the missense variants were retrieved for each AGI in the list. A boxplot of the variants separated by GO annotation reveals a lower number of SNPs in the core ABA signaling members than in the wider sets of proteins sharing their same annotations overall (Figure 4). However, a closer inspection of the individual proteins shows that PP2C phosphatases and the SnRK2 Ser/Thr (blue points within the genome sets phosphatases, kinases Figure 2.4), may have slightly higher , and lower numbers of variants than the genome wide sets respectively suggesting that each signaling hub in the ABA core pathway might be evolving antagonistically compared to proteins with similar functions.

To compare potential variants contained within the ABF family the four representative isoforms of ABF1, ABF2, ABF3, ABF4 (AT1G49720.2, AT1G45249.3, AT4G34000.1, AT3G19290.3 respectively) were aligned using the Variant Viewer (Figure 5A). Missense variants and non-synonymous variants largely occur across the entirety of the ABFs, but are totally absent within
the DNA binding domain and at key leucine zipper residues in their bZIP domains (Jakoby et al., 2002). This suggests that these key residues are totally conserved across ecotypes and ABF family members and are the bare minimum required residues for these bZIP transcription factors to function.

Two residues are contained within the bZIP domain but do not fall directly on the essential leucine dimerization residues or others noted to heavily affect dimerization (Vinson, Hai and Boyd, 1993). There are no missense variants that directly overlap the conserved RxxS/T phosphorylation sites in the N terminus of the protein, but there is a low frequency (5 ecotypes) early stop codon in ABF2 at site 62 (Q to *) which would result in a 61-residue truncated ABF2 protein. Coincidentally, truncations of ABF2 at site 62 have been previously tested in cultured T87 A. thaliana protoplasts and have been shown to be able to strongly activate reporters when fused to a non-native DNA binding domain such as GAL4, however the truncation should totally abolish the activity of ABF2 as it occurs early in the first exon of ABF2 and removes the bZIP domain in the C terminus of the protein. A second truncation in ABF2 is found at site 263 (Gly to *, 3 ecotypes), which still places it before the bZIP domain rendering ABF2 non-functional. Null mutants of ABF2 have no discernable growth or drought stress phenotype due to the overlapping redundancy of ABF2, ABF3 and ABF4 (Yoshida et al., 2010). Additionally, the accessions containing these early truncations are all Swedish in origin and are found at a higher latitude where drought may be less common. However, ABF2 has more specific localization to roots, has an essential role in glucose signaling and in regulating early seedling development by driving the expression of hexokinases which initiate glycolysis (Kim et al., 2004), and has a narrower role in osmotic stress responses alongside ABF3. It would be interesting to investigate variants in proteins that complement the truncated ABF2s in those ecotypes to see if those populations are modulating these responses in a novel way to account for the loss of ABF2. For example, ABF3 in terms of osmotic stress, or other transcription factors that drive the expression of glucose metabolism genes. In this way an investigator might uncover a locally adapted response to a novel variant that would otherwise go unnoticed if one only considered Col 0.
2.2.4 Future Directions

Although many of the basic features of the app have been implemented as of this moment, the app is a work in progress. There are several small tweaks, and larger features that I am planning to implement. In terms of small modifications, I plan to:

- modify the variants points plotted on the domain models to scale their size according to the log(number of ecotypes) that harbour them
- modify the variant scatter plot to show any variants that are contained in the coding region relevant to that isoform and not just the ones that map to that transcript, this is to show as complete a picture as possible to the user
- add common gene names to the all the panels so users can more easily identify genes when multiple are selected
- resize gene models on window resize for easier interaction with the application
- add a legend to the gene models and colour the variants according to severity of the change instead of red and black as displayed now
- add the ability to collapse gene isoforms that are not of interest for a more compact view
- add a flag that identifies the reference isoform for each gene model
- add hover information for the domains plotted in the MSA (in progress)

In terms of larger features, I hope to implement:

- a homolog suggestion pane which suggests A. thaliana gene family members for the user to add to their alignment
- the ability for a user to upload their own multiple sequence alignment; our MSA is performed using the default values for DECIPHER in R and is not a high quality hand curated alignment and should be treated as an in place convenience for a quick overview
- a better formatted data export package, currently the data is a poorly organized raw export and not user friendly

Additionally, given that the app is a recent creation it has not undergone significant user testing, I expect that in the future as users begin to use the app there will be minor bug fixes, and edge cases that need to be addressed on an ad hoc basis. Hopefully the tool helps researchers to think big regarding their possible experiments, and to take advantage of the great 1001 Genomes resource for possible hypothesis generation.
2.3 Figures

Figure 2.1. Examples of existing tools for viewing *A. thaliana* variation. **A.** Polymorph1001 allows users to search for variants in their gene or region of interest and to filter results by different variant classifiers. **B.** 1001 Proteomes viewer displays a coding sequence and has variant pins (underneath the numbered positions) that correspond to missense variants. **C.** ePlant’s Molecule Viewer app provides an intuitive interface for viewing variants, including pins that scale to the number of ecotypes and highlighted domains in the context of the (predicted) 3D structure of the protein.
Figure 2.2. The Variant Viewer: an application for viewing variants from the 1001 Genomes Project. Graphics are rendered using d3.js (https://d3js.org/) and plotly.js (https://plot.ly/javascript). Variants are pulled on the fly from the 1001 Genomes Project API (http://tools.1001genomes.org/api) for each gene entered into the dropdown search box (1). Selected genes (2) are displayed as simple gene models (5,7) and variants are plotted onto models to give an overview of the variants in the gene region. Hover events are triggered allowing users to inspect individual exons, UTRs, coding sequences, and variants. Users are also able to export data for their records (4).
Figure 2.3. Plotly.js Generated scatter plots of protein frequencies against MSA of genes of interest. A. variants are plotted for ABF2 (AT1G45249.3) the lower panel shows the returned CDD domains for the ABF2 bZIP domain. Hover events are triggered for each point in the scatter plot allowing users to inspect individual variants. B. when users zoom in on the bZIP domain they can toggle to display the MSA and can see the known residues that govern dimerization between bZIP domains of ABF2 and the variants that are nearby. C, a MSA of SK1, SK2, SKL1, SKL2 shows highlighted residues corresponding to the shikimate binding, ADP binding sites, and magnesium binding domains as retrieved from the CDD. Points on the scatter plot change shape depending on the identity of the protein to which they belong.
Figure 2.4. SNPs in the coding regions of *A. thaliana* genes separated by PANTHER GO annotation. A general set of abiotic stress proteins and “Core ABA Members” (highlighted in blue) are shown in addition to genome wide sets of kinases, phosphatases, receptors etc. The abiotic stress and core ABA members have accumulated a median of approximately ~45 missense variants per protein, slightly lower than the ~52 missense variants per protein genome wide, but lower than the median for the more inclusive sets of each protein functional class.
Figure 2.5. An alignment of the reference isoforms of the ABFs (ABF1/ABF2/ABF3/ABF4) with their associated variants. **A.** An overview of an alignment of the aligned ABFs. **B.** A close-up of the DNA binding domain shows synonymous variants but lacks any missense variants compared to the rest of the protein sequence (red box). Missense variants are present in between critical dimerization residues of the bZIP domain but do not occur at any critical hydrophobic residues (black box).
References


Kalladan, R. et al. (2017) ‘Natural variation identifies genes affecting drought-induced abscisic acid accumulation in *Arabidopsis thaliana*, *Proceedings of the National Academy of Sciences*,


Umezawa, T. et al. (2013) ‘Genetics and phosphoproteomics reveal a protein phosphorylation


Wright, E. S. (2016) ‘Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R’, 8, pp. 352–359.


Appendices

Figure A1. EasyClone2.0 plasmid pCfB2226-ADH1-ABF2-SV40NLS-FLAG3X. Singly C terminal tagged multipart assemblies are performed at the ADH1 terminator-ADH1 Promoter, ADH1 Promoter-ABF2, and SV40NLS-CYC1 terminator junctions. USER overlaps of approximately 8 – 12 bp (depending on sequence) are generated at junctions and splice together in single pot reactions. Promoter, CDS, FLAG3X and NLS “parts” are re-usable between plasmids. For example, to construct pCfB2226-GAL1-SnRK2.6-SV40NLS-tCYC one only needs to replace the yeast ADH1 promoter, and ABF2 with GAL1 promoter and a SnRK2.6 PCR product made with Uracil tagged residues in an assembly reaction. Sequencing is performed across the ADH1 terminator and CYC1 terminator junctions into the promoter and coding sequences to confirm orientation as suggested by the EasyClone 2.0 kit. Image generated with SnapGene Viewer.
Figure A2. EasyClone2.0 plasmid pCfB2188-pRD29B-GFP-tCYC. The RD29B promoter is cloned upstream of a GFP coding sequence facing the tCYC terminator. This plasmid is integrated into YPL3 and is our reporter construct for all strains tested for pathway activation. Image generated with SnapGene Viewer.
### Table A1. Primers used in this Project

<table>
<thead>
<tr>
<th>Type of Primer</th>
<th>Annotation</th>
<th>Primer ID</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annealing Oligo</td>
<td>Gly-Gly-Ser/SV40NLS Polylinker Element</td>
<td>GV2R-Polylinker-ABF2-Top</td>
<td>ACGCGAAGACCGAGAAGAGCTGATGAAACGGAGCG</td>
</tr>
<tr>
<td>Annealing Oligo</td>
<td>Gly-Gly-Ser/SV40NLS Polylinker Element</td>
<td>GG2.1F-Poly-SV40NLS-Bot</td>
<td>TCGAACCGCTGCTTCTTGGTCTCCGGGCTCTGGCT</td>
</tr>
<tr>
<td>Annealing Ulramer Oligo</td>
<td>Gly-Gly-Ser/Flag3X Polylinker Element</td>
<td>GG2.1F-Polylinker-FLAG3X-Top</td>
<td>ACGCGAAGAGATTATATAAGGAGACAGACGACGACAATAGATGGAACGAGACGAG</td>
</tr>
<tr>
<td>Annealing Ulramer Oligo</td>
<td>Gly-Gly-Ser/Flag3X Polylinker Element</td>
<td>GV2R-Polylinker-FLAG3X Bottom</td>
<td>TCACTTGCTACGCTACCTTTAATACCTCTTGTACATCGTCATCTTTTATAATCTGC</td>
</tr>
<tr>
<td>Cloning</td>
<td>GFP Reporter Element</td>
<td>GFP-GV2R-v1</td>
<td>CGAGCAATTAGTTAATATGATGAGGAGATATTG</td>
</tr>
<tr>
<td>Cloning</td>
<td>GFP Reporter Element</td>
<td>GFP-GP2F-v1</td>
<td>ATCTGGTCAIAAAATATGTGTAAGAGGAGAGAGAAG</td>
</tr>
<tr>
<td>Cloning</td>
<td>Reporter Element</td>
<td>PG2R-Rd29b-v1</td>
<td>ATGACAGAATCAAAAGTCGTTG</td>
</tr>
<tr>
<td>Cloning</td>
<td>Reporter Element</td>
<td>P2VF-Rd29b-v1</td>
<td>CTCGGCAAGATTCTGCAGATATAGTCAACAGA</td>
</tr>
<tr>
<td>Cloning</td>
<td>ABF2 Coding Sequence</td>
<td>ABF2-GG2.1R</td>
<td>TCGAACCGCCCGAGCAAGTCGCTG</td>
</tr>
<tr>
<td>Cloning</td>
<td>ABF2 Coding Sequence</td>
<td>ABF2-GP2F</td>
<td>ATCTGGTCAIAAAATATGTGTAAGAGGAGAGAAG</td>
</tr>
<tr>
<td>Cloning</td>
<td>ADH1 Promoter</td>
<td>ADH1-PG2R</td>
<td>ATGACAGAATCTGATGAGGAGATATTG</td>
</tr>
<tr>
<td>Cloning</td>
<td>ADH1 Promoter</td>
<td>ADH1-PV2F</td>
<td>CTCGGCAAGATTCTGCAGATATAGTCAACAGA</td>
</tr>
<tr>
<td>Cloning</td>
<td>SnRK2.6 Coding Sequence</td>
<td>GG2.1R-SnRK2-6_v1</td>
<td>TCGAACCGCCCGAGCAAGTCGCTG</td>
</tr>
<tr>
<td>Cloning</td>
<td>SnRK2.6 Coding Sequence</td>
<td>GP2F-SnRK2-6_v1</td>
<td>ATCTGGTCAIAAAATATGTGTAAGAGGAGAGAAG</td>
</tr>
<tr>
<td>Cloning</td>
<td>CYC1 Promoter</td>
<td>PG2R-CYC-v1</td>
<td>ATGACAGAATCTGATGAGGAGATATTG</td>
</tr>
<tr>
<td>Cloning</td>
<td>CYC1 Promoter</td>
<td>PV2F-CYC-v1</td>
<td>CTCGGCAAGATTCTGCAGATATAGTCAACAGA</td>
</tr>
<tr>
<td>Cloning</td>
<td>GPD1 Promoter</td>
<td>PG2R-GPD-v1</td>
<td>ATGACAGAATCTGATGAGGAGATATTG</td>
</tr>
<tr>
<td>Cloning</td>
<td>GPD1 Promoter</td>
<td>PV2F-GP2F-v1</td>
<td>CTCGGCAAGATTCTGCAGATATAGTCAACAGA</td>
</tr>
<tr>
<td>Cloning</td>
<td>GFP Coding Sequence (Positive Control)</td>
<td>GP2F-GG2.1R</td>
<td>TCGAACCGCCCGAGCAAGTCGCTG</td>
</tr>
<tr>
<td>Cloning</td>
<td>GFP Coding Sequence (Positive Control)</td>
<td>GP2F-GP2F</td>
<td>ATCTGGTCAIAAAATATGTGTAAGAGGAGAGAAG</td>
</tr>
<tr>
<td>Cloning</td>
<td>FLAGX Fusion Overhang Primer</td>
<td>GG2.1R-ADH1-ABF2-SV40NLS-Polylinker</td>
<td>TCGAACCGCCCGAGCAAGTCGCTG</td>
</tr>
<tr>
<td>Cloning</td>
<td>FLAGX Fusion Overhang Primer</td>
<td>PV2F-ADH1-ABF2-SV40NLS-55</td>
<td>CTCGGCAAGATTCTGCAGATATAGTCAACAGA</td>
</tr>
<tr>
<td>Cloning</td>
<td>ABH1 Coding Sequence</td>
<td>ABH1-GG2.1R</td>
<td>TCGAACCGCCCGAGCAAGTCGCTG</td>
</tr>
<tr>
<td>Cloning</td>
<td>ABF2 Internal Deletion SDM (inverse PCR)</td>
<td>ABF2deltaQT_DEL1</td>
<td>ATCTGGTCAIAAAATATGTGTAAGAGGAGAGAAG</td>
</tr>
<tr>
<td>Cloning</td>
<td>ABF2 Internal Deletion SDM (inverse PCR)</td>
<td>ABF2deltaQT_DEL2</td>
<td>AAGAATGTTATGATGAGGAGAGGAGAG</td>
</tr>
<tr>
<td>Cloning</td>
<td>PYR1 Coding Sequence</td>
<td>PYR1-GG2.1R</td>
<td>TCGAACCGCCCGAGCAAGTCGCTG</td>
</tr>
<tr>
<td>Cloning</td>
<td>ABH1 Coding Sequence</td>
<td>ABH1-GP2F</td>
<td>ATCTGGTCAIAAAATATGTGTAAGAGGAGAGAAG</td>
</tr>
<tr>
<td>Cloning</td>
<td>PFR1 Coding Sequence</td>
<td>PFR1-GP2F</td>
<td>ATCTGGTCAIAAAATATGTGTAAGAGGAGAGAAG</td>
</tr>
<tr>
<td>Cloning</td>
<td>ABF2 Site Directed Mutagenesis Primers</td>
<td>S26A_SDM_2</td>
<td>AGGATGAAATATGATGAGGAGAGGAGAG</td>
</tr>
<tr>
<td>Cloning</td>
<td>ABF2 Site Directed Mutagenesis Primers</td>
<td>S26A_SDM_2</td>
<td>AGGATGAAATATGATGAGGAGAGGAGAG</td>
</tr>
<tr>
<td>Cloning</td>
<td>ABF2 Site Directed Mutagenesis Primers</td>
<td>S26A_SDM_1</td>
<td>AGGATGAAATATGATGAGGAGAGGAGAG</td>
</tr>
<tr>
<td>Cloning</td>
<td>ABF2 Site Directed Mutagenesis Primers</td>
<td>S26A_SDM_1</td>
<td>AGGATGAAATATGATGAGGAGAGGAGAG</td>
</tr>
<tr>
<td>Cloning</td>
<td>GAL1 Promoter</td>
<td>GAL1 - PV2F</td>
<td>CTCGGCAAGATTCTGCAGATATAGTCAACAGA</td>
</tr>
<tr>
<td>Cloning</td>
<td>GAL1 Promoter</td>
<td>GAL1 - PG2R</td>
<td>ATGACAGAATCTGATGAGGAGATATTG</td>
</tr>
<tr>
<td>Cloning</td>
<td>GAL1 Promoter</td>
<td>GAL1 Promoter Subcloning Primers</td>
<td>GAL1-PV2F-Short</td>
</tr>
<tr>
<td>Cloning</td>
<td>GPD Promoter Subcloning Primers</td>
<td>CYC-PV2F-Short</td>
<td>GPD-PV2F-Short</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>---------------------------------</td>
<td>-------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Cloning ABF2 Site Directed Mutagenesis Primers</td>
<td>$\text{S86A}<em>{-}\text{SDM}</em>{-}1$</td>
<td>$\text{AGGGCGACTUCCTCAGCTGACGCT}$</td>
<td>$\text{AGGGCGACTUCCTCAGCTGACGCT}$</td>
</tr>
<tr>
<td>Cloning ABF2 Site Directed Mutagenesis Primers</td>
<td>$\text{S86A}<em>{-}\text{SDM}</em>{-}2$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning ABF2 Site Directed Mutagenesis Primers</td>
<td>$\text{S86D}<em>{-}\text{SDM}</em>{-}1$</td>
<td>$\text{AGGGCGAGTCCCTCAGCTGACGCT}$</td>
<td>$\text{AGGGCGAGTCCCTCAGCTGACGCT}$</td>
</tr>
<tr>
<td>Cloning ABF2 Site Directed Mutagenesis Primers</td>
<td>$\text{S86D}<em>{-}\text{SDM}</em>{-}2$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning ABF2 Site Directed Mutagenesis Primers</td>
<td>$\text{S94A}<em>{-}\text{SDM}</em>{-}1$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning ABF2 Site Directed Mutagenesis Primers</td>
<td>$\text{S94A}<em>{-}\text{SDM}</em>{-}2$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning ABF2 Site Directed Mutagenesis Primers</td>
<td>$\text{S94D}<em>{-}\text{SDM}</em>{-}1$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning ABF2 Site Directed Mutagenesis Primers</td>
<td>$\text{S94D}<em>{-}\text{SDM}</em>{-}2$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning ABF2 Site Directed Mutagenesis Primers</td>
<td>$\text{S135A}<em>{-}\text{SDM}</em>{-}1$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning ABF2 Site Directed Mutagenesis Primers</td>
<td>$\text{S135A}<em>{-}\text{SDM}</em>{-}2$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning ABF2 Site Directed Mutagenesis Primers</td>
<td>$\text{S135D}<em>{-}\text{SDM}</em>{-}1$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning ABF2 Site Directed Mutagenesis Primers</td>
<td>$\text{S135D}<em>{-}\text{SDM}</em>{-}2$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning Subcloning Primers for FLAG3X Fusions</td>
<td>$\text{pADH}_{-}\text{Cyc}$</td>
<td>$\text{AGGGCGACTUCCTCAGCTGACGCT}$</td>
<td>$\text{AGGGCGACTUCCTCAGCTGACGCT}$</td>
</tr>
<tr>
<td>Cloning Subcloning Primers for FLAG3X Fusions</td>
<td>$\text{pGAL1}_{-}\text{tADH}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning Subcloning Primers for FLAG3X Fusions</td>
<td>$\text{3XFLAG}_{-}\text{Cyc}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning Subcloning Primers for FLAG3X Fusions</td>
<td>$\text{3XFLAG}_{-}\text{tADH1}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning Subcloning Primers for FLAG3X Fusions</td>
<td>$\text{pADH}_{-}\text{Cyc} _2$</td>
<td>$\text{AGGGCGACTUCCTCAGCTGACGCT}$</td>
<td>$\text{AGGGCGACTUCCTCAGCTGACGCT}$</td>
</tr>
<tr>
<td>Cloning GFP Tag Primer</td>
<td>$\text{Cterm}<em>{-}\text{GFP}</em>{-}\text{Fusion}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning GAL1 Subcloning for NLS Tests</td>
<td>$\text{pGAL1}_{-}\text{Cyc}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning ABF2 Subcloning primers for untagged ABF2</td>
<td>$\text{ABF2}<em>{-}\text{Cterm}</em>{-}\text{del}_{-}1$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning ABF2 Subcloning primers for untagged ABF2</td>
<td>$\text{ABF2}<em>{-}\text{Cterm}</em>{-}\text{del}_{-}2$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning ABF2 Coding Sequence</td>
<td>$\text{ABI5}<em>{-}\text{USER}</em>{-}\text{overhang}_{-}1$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning ABI5 Coding Sequence</td>
<td>$\text{ABI5}<em>{-}\text{USER}</em>{-}\text{overhang}_{-}2$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning ABI5 Subcloning Primers</td>
<td>$\text{ABI5}<em>{-}\text{USER}</em>{-}\text{PV2F}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Cloning ABI5 Subcloning Primers</td>
<td>$\text{ABI5}<em>{-}\text{USER}</em>{-}\text{Ov2R}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
<td>$\text{AACCGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Sequencing Internal Sequence Verification Primers</td>
<td>$\text{ABN}_{-}\text{FWD}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Sequencing Internal Sequence Verification Primers</td>
<td>$\text{ABN}_{-}\text{REV}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Sequencing Internal Sequence Verification Primers</td>
<td>$\text{ABN}_{-}\text{FWD}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Sequencing Internal Sequence Verification Primers</td>
<td>$\text{ABN}_{-}\text{REV}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Sequencing Internal Sequence Verification Primers</td>
<td>$\text{AREB1}_{-}\text{FWD}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Sequencing Internal Sequence Verification Primers</td>
<td>$\text{AREB1}_{-}\text{REV}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Sequencing Internal Sequence Verification Primers</td>
<td>$\text{PyR1}_{-}\text{FWD}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Sequencing Internal Sequence Verification Primers</td>
<td>$\text{PyR1}_{-}\text{REV}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Sequencing Internal Sequence Verification Primers</td>
<td>$\text{SnR2F}_{-}\text{FWD}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Sequencing Internal Sequence Verification Primers</td>
<td>$\text{SnR2F}_{-}\text{REV}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Sequencing Sequencing Primers (EasyClone2.0)</td>
<td>$\text{CYC1}<em>{-}\text{test}</em>{-}\text{rv}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
</tr>
<tr>
<td>Sequencing Sequencing Primers (EasyClone2.0)</td>
<td>$\text{ADH1}<em>{-}\text{test}</em>{-}\text{fw}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
<td>$\text{CACGGGCACTGGCTGACGCTG}$</td>
</tr>
</tbody>
</table>
Permissions

Figure 1.1A from Furihata et al. (2006) is used with permission (http://www.pnas.org/page/about/rights-permissions):

"Anyone may, without requesting permission, use original figures or tables published in PNAS for noncommercial and educational use (i.e., in a review article, in a book that is not for sale), provided that the full journal reference is cited and, for articles published in volumes 90–105 (1993–2008), "Copyright (copyright year) National Academy of Sciences." Commercial reuse of figures and tables (i.e., in promotional materials, in a textbook for sale) requires permission from PNAS."

Figure 1.3 yeast panel from Fehrenbacher et al. (2005) is used with permission (https://www.molbiolcell.org/permissions):

"Under the License and Publishing Agreement, authors grant to the general public, effective two months after publication of (i.e., the appearance of the edited manuscript in an online issue of MBoC), the nonexclusive right to copy, distribute, or display the manuscript subject to the terms of the Creative Commons–Noncommercial–Share Alike 3.0 Unported license (http://creativecommons.org/licenses/by-nc-sa/3.0)."