INNOVATIONS IN CLONING GENES IDENTIFIED IN GENETIC SCREENS AND THEIR APPLICATION TO CYTOKININ SIGNALING IN *ARABIDOPSIS THALIANA*

Charles Hodgens

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum in Genetics and Molecular Biology in the School of Medicine.

Chapel Hill
2020

Approved by:

Joseph Kieber

Jason Reed

Sarah Grant

Jeremy Purvis

Zachary Nimchuk

ABSTRACT

Charles Hodgens: Innovations in cloning genes identified in genetic screens and their application to cytokinin signaling in Arabidopsis thaliana

(Under the direction of Joseph Kieber)

The plant hormone cytokinin is responsible for regulation of a diverse set of biological and developmental phenomena in plants such as leaf senescence, organ formation, nutrient distribution, abiotic stress response, and others. The core components of the cytokinin signaling pathway have been identified and characterized in *Arabidopsis thaliana* and other plant species, but questions remain about whether biological factors exist that may participate in or regulate the pathway.

In this work, I present the results of a sensitized genetic screen for *enhancers of ahp* (*eah*) mutants which are hyposensitive to exogenously supplied cytokinin signals. The *eah* screen identified several new alleles of *ahk4*, one of the cytokinin receptors in *Arabidopsis*, as well as implicating the light-signaling transcription factor *HY5*. As part of the screen, the "mutagenomics" strategy for analyzing uncharacterized mutants in a screen was developed. Mutagenomics involves high-throughput resequencing of a small number of mutant plants with the goal of detecting protein-altering single nucleotide polymorphisms (SNPs). The SNPs detected across the experiment are then assessed to determine if any genes were mutated more frequently than would be expected by chance alone. A second aspect of mutagenomics involves resequencing plant lines of common descent, referred to as siblings, to determine which mutations are homozygous in all related plants. The causative mutation should be contained in

the list of shared homozygous mutations. The mutagenomics strategies have the potential to accelerate the pace of genetic screens.

I also present a primer design tool, indCAPS, developed to aid the search for novel CRISPR/Cas9-derived mutant alleles. The indCAPS platform is intended to be used for Cleaved Amplified Polymorphic Sequence (CAPS) or derived CAPS (dCAPS) restriction-digest genotyping assays. Tools exist for designing dCAPS primers but they assume the researcher is genotyping a SNP allele. CRISPR/Cas9-derived alleles are typically insertion/deletion (indel) events that result in a frameshift of a protein, and the previously developed primer design tools often fail when given sequences with indels. The indCAPS platform was developed with the ability to distinguish indel alleles as well as SNP alleles.

ACKNOWLEDGEMENTS

This thesis is a long time coming, but it would not have been possible without the never-ending love and support of my family, friends, and coworkers.

To my parents – I could not have asked for a more supportive pair of people. You have been there for me every step of this long journey with unquestioning support. Every step I take in my life has brought new awareness of all the sacrifices you both made for me with no protest or bargaining, and I can never begin to repay your kindness.

To my mentors, past and present – Joe, Jason, Beth, Greg, Leena, – everything I am now professionally is the cumulative total of your patient instruction and advice. Thank you for your diligence in cultivating my talents; I definitely could not have done it on my own.

To my coworkers – too many to name – science is a game of standing on each other's shoulders. We share expertise, act as sounding boards, lend a hand for onerous experiments, and brighten each other's lives. At times, I was unwittingly my own worst enemy, and it was only with your sage advice that I was able to avoid making my life so much harder than it needed to be.

To my friends – again, too many to name – a life in science must be balanced with a life outside the lab, and you all kept me sane. Road trips, parties, tabletop game sessions, vacations, scavenger hunts, and just plain talking. Work in the lab is stressful, but the stress makes time with friends more refreshing. Time with friends recharges me to work harder in the lab. Both sides of life feed on each other to help me thrive as a person, and I am eternally grateful.

And finally, thank you, reader. The greatest gift you can give me as a scientist is your attention.

PREFACE

Chapter 2 was adapted from a manuscript submitted for review to *Plant Physiology* by Charles Hodgens, G. Eric Schaller, and Joseph J. Kieber. Chapter 4 has been adapted from (Hodgens et al., 2017) which has been made available under a Creative Commons Attribution License by PLoS ONE.

TABLE OF CONTENTS

LIST OF TABLES x

LIST OF FIGURES xii

ABBREVIATIONS xiii

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: MUTAGENOMICS: A HIGH-THROUGHPUT METHOD TO IDENTIFY CAUSATIVE MUTATIONS FROM A GENETIC SCREEN 10

Introduction 10

Results 13

Discussion 31

Material and Methods 36

CHAPTER 3: MUTAGENOMICS AND ITS APPLICATION TO ADDITIONAL MUTANT SCREENS 44

Further evidence of the usefulness of the mutagenomics strategies 44

Further analysis of the *eah* screen 45

The *acre* screen for ACC resistance demonstrates the enrichment of causative alleles in a screen 52

The *ghostbuster* screen demonstrates the utility of a sibling-focused sequencing strategy 59

The Heidelberg screen identified several enriched alleles 70

CHAPTER 4: INDCAPS: A TOOL FOR DESIGNING SCREENING PRIMERS FOR CRISPR/CAS9 MUTAGENESIS EVENTS 72

Introduction 72

Materials and methods 74

Results and discussion 75

CHAPTER 5: CONCLUSION 87

APPENDIX 1 92

REFERENCES 95

LIST OF TABLES

Table 2.1. *eah* mutant lines analyzed by mutagenomics. ……16

Table 3.1. Homozygous PA-SNPs observed in the *eah38* sibling set. ……48

Table 3.2. Candidate causative PA-SNPs in *eah49*. ……..…..……49

Table 3.3. Candidate causative PA-SNPs in *eah36*. ……49

Table 3.4. Candidate causative PA-SNPs in *eah40*. ……49

Table 3.5. *eah* lines not characterized by mutagenomic analysis. ……...…….50

Table 3.6. Summary of SNP densities in the *acre* screen mutants. ……..……54

Table 3.7. Mutations in *LHT1* observed in the *acre* screen. ……57

Table 3.8. EMS-consistent SNPs observed in the *gob* screen which are possible background mutations. ……..……..……61

Table 3.9. *gob2-23* and *gob2-24* form a sibling group. ……...……62

Table 3.10. *gob2-38*, *gob2-27*, and *gob2-14* form a sibling group, with 11 shared homozygous mutations between them. ……..……..……62

Table 3.11. *gob2-10* and *gob2-48* are a spurious sibling group. ……..……...…….62

Table 3.12. EMS-consistent mutations observed in the pool 2 set of *gob* mutants. ……...….63

Table 3.13. *gob1-18* and *gob1-15* form a sibling group. ……...…….……64

Table 3.14. *gob24-15* and *gob24-12* form a sibling group. ……..…...…….64

Table 3.15. *gob16-14*, *gob16-20*, *gob16-15*, and *gob16-31* form a sibling group, with 18 genes mutated in all four lines. ……..…….……65

Table 3.16. *gob9-1*, *gob9-3*, and *gob9-9* form a sibling group, with 19 genes mutated in all three lines. ……..……...….66

Table 3.17. *gob12-27*, *gob12-5*, and *gob12-30* form a sibling group, with 7 genes mutated in all three lines. ……..……...…….….66

Table 3.18. *gob21-63*, *gob21-47*, and *gob21-58* form a sibling group, with

10 genes mutated in all three lines. ……..…...….66

Table 3.19. *gob27-1*, *gob27-6*, and *gob27-9* form a sibling group, with 16 genes mutated in all three lines. ……...……..……...…….67

Table 3.20. *gob9-4* and *gob9-22II* form a sibling group, with 31 genes mutated in both lines. …….67

Table 3.21. Pool 16 mutants forming potential sibling group(s). ……..….…….……68

Table 3.22. Remaining mutants in the *gob* screen not part of a sibling group. …….…....69

Table 3.23. Simulation-derived probabilities of multiple alleles in the same gene arising by chance. ……..……...…….....….71

Table 3.24. Alleles identified in the Heidelberg screen indicated as enriched

by mutagenomics. ...…….…….71

Table 4.1. Number of productive primers generated for tested loci. ……...……..…77

LIST OF FIGURES

Figure 1.1. Cytokinin signaling in *Arabidopsis thaliana*. ……..…….2

Figure 2.1. Genetic screen for cytokinin hyposensitivity. ……...…….14

Figure 2.2. Overview of the mutagenomics process. ……...…….19

Figure 2.3. Disruptive and missense mutations were observed in AHK4. …..…….21

Figure 2.4. Inference of over-enriched mutations in independent mutant lines. ….…….24

Figure 2.5. Projected numbers of shared alleles between M3 siblings. …...…….26

Figure 2.6. Sibling analysis used to identify *hy5* in two sibling lines. …….28

Figure 3.1. Shared exact mutations in the *acre* screen mutants…….55

Figure 4.1. CAPS/dCAPS markers can distinguish alleles, but output of dCAPS Finder 2.0 can be flawed. …….……...…….73

Figure 4.2. Algorithm for generation of oligonucleotide primers useful for CAPS and dCAPS assays. ……...…….81

Figure 4.3. Homozygous editing events in AHK3 were identified. …….85

Figure A1.1. Full network diagram of PA-SNPs. ……..……..….92

Figure A1.2. Comparison of differentially expressed genes between BA and control-treated wild type plants and *hy5* control plants and wild type control plants. …..…..….93

Figure A1.3. Exact shared SNPs between *eah14-1* and *eah14-2*. ..…….….94

ABBREVIATIONS

SNPs – Single Nucleotide Polymorphisms

PA-SNPs – Protein-Altering Single Nucleotide Polymorphisms

CHAPTER 1: INTRODUCTION

If you examine the myriad sub-domains of biology, you can find groups of scientists so separated in subject matter, cross-referencing, collaboration, and typical methodology that an ecologist might define them as different species (if “ecology of science” was a thing, that is). However, all biologists share the same core goal: to study life. The form and technique may vary, but the ambition is the same. Many of the biological phenomena that allow a fruit fly or fish or worm to live are the same phenomena that keep rats and mice and our own bodies alive. Plant biology is no different, but if you will permit me a moment of romanticisim, plant biology has a certain importance in the role plants have in maintaining our lives. Life as we know it would not be possible without plants. Land plants and ocean algae are responsible for the oxygen we breathe; the food we and our animals eat; the material resources we use to build our homes, publish our books and newspapers, and power our engines; and the stabilization of the land against erosion and flooding. The study of plants is the study of the keystone holding human civilization upright.

Regrettably however, plants are unglamorous. To most people, they are static organisms that fill in the backdrops of their everyday lives. However, the same factors that make plants unremarkable when compared to animals also make them scientifically interesting and brought me to love them. No, plants are not capable of large movements. If a predator threatens a plant, it cannot run away or hide. Instead, it must change its biology, and that is truly fascinating. A plant reacts to its environment by changing its developmental programs, relying on a level of

biological plasticity rarely seen in animals. “Escape” for a plant requires the capability to recognize a hazard, an extant strategy to cope with the hazard, and the flexibility to change its own growth patterns to enact that strategy.

That flexibility in biological response is a central theme of the work described in this thesis. Specifically, we study the methods plants use to sense and respond to signals from within and without the plant. This requires recognition and integration of inputs, transport of molecular intermediaries of the signal through the cell, and the translation of the abundance of those intermediates into a low-level biological change that can affect the behavior of the cell. I will be discussing hormone signaling, and the hormone of interest to our research group is cytokinin. Cytokinins are a class of molecules capable of interacting with and activating the receptor proteins of the cytokinin signaling pathway (Figure 1.1) (Romanov et al., 2006; Stolz et al., 2011; Lomin et al., 2015). They regulate a broad, diverse set of biological processes, including, among others, organ formation, leaf expansion, senescence timing, control of apical dominance, nutrient distribution and uptake, cell proliferation, nodulation, environmental stress responses, plant pathogen responses, and

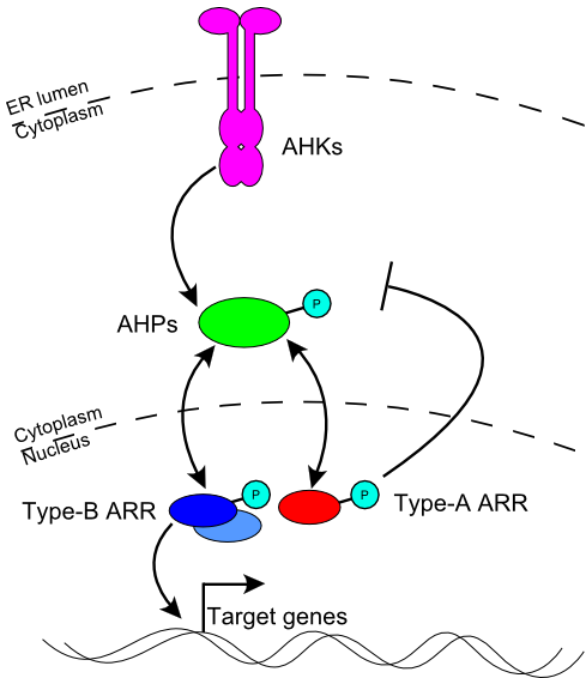

Figure 1.1. Cytokinin signaling in Arabidopsis thaliana. Cytokinin is perceived in the ER lumen, which causes the AHKs to undergo a phosphorylation event. The phosphoryl group is transferred to the AHPs, which then transfer the phosphoryl group to either a type-A or type-B ARR in the nucleus. Type-A ARRs act to negatively regulate the pathway through an unknown mechanism and are transcriptionally responsive to cytokinin signaling. Type-B ARRs act as transcription factors but are largely not responsive to cytokinin signaling.

chloroplast development (Mok and Mok, 2001; Argueso et al., 2010; Hwang et al., 2012; Kieber and Schaller, 2014).

Trans-zeatin biosynthesis in Arabidopsis begins with either ATP or ADP, which is converted to iPRTP/iPRDP with the addition of an isoprene side chain by one of several IPT enzymes (Takei et al., 2004). The side chain of isopentenyl adenine is hydroxylated by CYP735A to form trans-zeatin ribotide, which is then converted to the free base form by the LONELY GUY (LOG) family of enzymes (Kurakawa et al., 2007; Kuroha et al., 2009; Tokunaga et al., 2012). *Trans*-zeatin molecules typically are broken down by oxidation by CYTOKININ OXIDASE (CKX) proteins (Schmülling et al., 2003; Kowalska et al., 2010) or modified by glucosyltransferases for storage or degradation (Brzobohatý et al., 1993; Hou et al., 2004). Cytokinin is transported distally in the plant as ribosides (Matsumoto-Kitano et al., 2008; Kudo et al., 2010), which are then converted to their active form inside the cell by LOG proteins.

Cytokinin was first identified as an unknown compound in coconut milk capable of promoting growth in embryonic and somatic tissue in a variety of plant species (Overbeek et al., 1941; Ball, 1946; Caplin and Steward, 1948; Nickell, 1950; Morel and Wetmore, 1951; Steward and Caplin, 1951; Mauney et al., 1952; Miller C.O. et al., 1956). Similar growth-promoting activity was later identified in yeast extracts and autoclaved salmon sperm , leading to the identification of *kinetin*, a derivative of adenine (Miller C.O. et al., 1956). A class of chemicals, *kinins*, later re-named cytokinins, was also defined (Hall, 1873; Miller C.O. et al., 1956). Other cytokinins discovered since include zeatin (Letham, 1963), isopentenyl adenine (Dyson and Hall, 1972), and benzyladenine (BA) (Hamzi and Skoog, 1963). In *Arabidopsis*, the highest receptor affinity is to the *trans*-zeatin (the most abundant form) and isopentenyladenine (Stolz et al., 2011).

The cytokinin signaling pathway in *Arabidopsis* is similar to bacterial two-component signaling systems and was identified by comparisons of *Arabidopsis* proteins to bacterial TCS proteins (Perraud et al., 1999; Stock et al., 2000; West and Stock, 2001; Mizuno, 2005; Schaller et al., 2008). In a prototypical bacterial two-component signaling system, a sensor histidine kinase detects undergoes a phosphorylation event in response to stimulus, the kinase transfers its phosphoryl group through a His-to-Asp phosphorelay to a response regulator, and the response regulator mediates the signaling output of the pathway. The *Arabidopsis* incarnation of this pathway is an extended form that incorporates a phosphorelay step (Swanson et al., 1994; Appleby et al., 1996; Schaller et al., 2011). In this form, cytokinin is perceived by the ARABIDOPSIS HISTIDINE-CONTAINING KINASE (AHK) proteins. Three *AHK* genes have been shown to contribute to cytokinin signaling: CYTOKININ RESPONSE 1 (CRE1) / AHK2, AHK3, and WOODEN LEG (WOL) / AHK4 (Scheres et al., 1995; Mahonen et al., 2000; Inoue et al., 2001; Ueguchi et al., 2001a; Yamada et al., 2001). The AHKs consist of a Cyclases/Histidine Kinases Associated Sensor Extracellular domain (CHASE domain) linked to a histidine kinase domain and a receiver domain (Anantharaman and Aravind, 2001; Inoue et al., 2001; Heyl et al., 2007; Hothorn et al., 2011). Following binding by cytokinin in the lumen of the ER (Yamada et al., 2001; Romanov et al., 2006; Caesar et al., 2011; Lomin et al., 2011; Stolz et al., 2011; Wulfetange et al., 2011), the histidine kinase domain undergoes an autophosphorylation event, and then the phosphoryl group is transferred to a receiver domain at the C terminus of the protein (Hutchison and Kieber, 2002). The AHKs also possess phosphatase activity in the absence of cytokinin binding to the CHASE domain (Mähönen et al., 2006).

Phosphoryl groups from the AHKs are transferred to the ARABIDOPSIS HISTIDINE-CONTAINING PHOSPHOTRANSFER (AHP) proteins. Six genes have been identified in

Arabidopsis: *AHP1* through *AHP6*. The first five true *AHP*s were first identified by the presence of a histidine phosphotransfer (HPt) domain in five separate genes (Miyata et al., 1998; Suzuki et al., 1998; Initiative, 2000; Suzuki et al., 2000; Schaller et al., 2008) and a sixth, pseudo-*AHP* (*PHP*) incapable of being phosphorylated by the AHKs has also been identified (Mahonen et al., 2006). The AHPs physically interact with both the AHKs and the ARRs (Urao et al., 1998; Imamura et al., 1999; Tanaka et al., 2004; Dortay et al., 2006) and function by accepting phosphoryl groups from the AHKs and transferring the phosphoryl group to the ARRs (Suzuki et al., 1998; Hutchison et al., 2006; Deng et al., 2010; Cheng et al., 2013). The AHP6 protein lacks the conserved His site that serves as the phosphorylation carrier and thus cannot participate in phosphorelay-based cytokinin signaling (Mahonen et al., 2006). Genetically, it has been shown to inhibit cytokinin signaling and to play a role in the patterning of the vasculature of roots in *Arabidopsis* (Mahonen et al., 2006; Bishopp et al., 2011).

The phosphorylated AHPs interact with ARABIDOPSIS RESPONSE REGULATOR (ARR) proteins. These proteins come in two main classes, type-B and type-A ARRs. The type-A *ARR* genes are *ARR3* through *ARR9* and *ARR15* through *ARR17*, and the type-B *ARRs* are *ARR1*, *ARR2*, *ARR10* through *ARR14*, and *ARR18* through *ARR21*. The first class discussed here, the type-B ARRs, mediate the transcriptional output and downstream biological changes in response to cytokinin. Type-B ARRs were first identified by the presence of their RR domains (Brandstatter and Kieber, 1998; Imamura et al., 1998; Suzuki et al., 1998; Urao et al., 1998) and later identified as DNA-binding transcription factors (Sakai et al., 1998; Imamura et al., 1999; Lohrmann et al., 1999; Sakai et al., 2000; Hosoda et al., 2002; Imamura et al., 2003) and are primarily localized in the nucleus (Lohrmann et al., 1999; Imamura et al., 2001; Lohrmann et al., 2001; Hosoda et al., 2002). The principal type-B ARRs in *Arabidopsis* mediating cytokinin

signaling are *ARR1*, *ARR10*, and *ARR12* (Mason et al., 2005; Yokoyama et al., 2007; Argyros et al., 2008; Ishida et al., 2008).

The second class of ARRs is the type-A ARRs (D'Agostino et al., 2000; Schaller et al., 2008). Like the type-Bs, these proteins have a receiver domain and interact with the AHPs, but lack a DNA-binding domain (Imamura et al., 1998; D'Agostino and Kieber, 1999b; Hutchison and Kieber, 2002; Dortay et al., 2006; Schaller et al., 2008). Response regulators have been observed to interact with regulatory proteins in a phospho-dependent manner in other species for control of protein-protein interactions or sub-cellular localization (Jenal and Galperin, 2009), and have been observed in the same role in *Arabidopsis*, where the type-As are negative regulators of cytokinin signaling (Kiba et al., 2003; To et al., 2004; Leibfried et al., 2005; Lee et al., 2007a; To et al., 2007) and are transcriptionally upregulated following cytokinin signaling (Brandstatter and Kieber, 1998; Taniguchi et al., 1998; D'Agostino et al., 2000; Taniguchi et al., 2007; Xie et al., 2018). While the molecular mechanism for the negative regulation is unknown, several methods for negative regulation of cytokinin signaling by the type-A ARRs have been hypothesized. The most straightforward proposal is that they act as phosphate sinks, similar to some bacterial two-component systems such as CheA/CheY chemotaxis system in *E. coli* (Sourjik and Schmitt, 1998; Dortay et al., 2006). The type-As and type-Bs both undergo a phosphotransfer reaction with the AHPs, but only the type-As are upregulated by cytokinin treatment (Brandstatter and Kieber, 1998; Taniguchi et al., 1998; D'Agostino et al., 2000; Taniguchi et al., 2007). If more type-A ARRs are present than type-Bs, then more phosphoryl groups will be transferred to them rather than to the type-Bs, and thus there will be a reduction in the number of activated type-B ARRs. As a subset of type-A ARRs are stabilized by phosphorylation (To et al., 2007), the phosphorylation intercepted by the newly synthesized type-A ARRs would also further increase

their negative regulatory effect. However, a phosphomimic mutant type-A ARR can partially rescue a multiple type-A loss of function mutant (To et al., 2007). The phosphomimic protein replaces the conserved aspartic acid residue that receives the phosphoryl group from the AHPs with a glutamic acid (Kurepa et al., 2014). The glutamic acid is not phosphorylated, but has a chemical character similar enough to a phosphorylated aspartic acid that it can partially mimic a phosphorylated wild type ARR (Hass et al., 2004). If a phosphate sink mechanism were solely responsible, a phosphomimic protein would not rescue a loss-of-function mutant. However, phosphate sink activity may play at least some role in the negative regulation of type-A ARRs, as the relative abundance of the type-As does increase with cytokinin signaling increases. However, sink activity could be negligible. For example, if the binding affinity of phosphorylated AHPs with type-B ARRs is higher than for type-A ARRs, then the phosphate sink effect of increased type-A ARR prevalence would be muted. Sink activity alone may serve to extend flux through the cytokinin pathway rather than dampen it. In bacterial systems, phosphotransfer in two component systems has been shown to be reversible in some cases (Stewart, 1997; Zapf et al., 2000). If reversible phosphotransfer occurs in *Arabidopsis*, phosphoryl groups on Type-As could transfer back to the AHPs and then to the type-B ARRs. It has also been suggested that increased abundance of type-A ARRs could bind to and effectively sequester AHPs, preventing them from binding to other proteins such as the AHKs (Dortay et al., 2006). Another proposed mechanism similar to the phosphate sink concept is that type-A ARRs could dephosphorylate phosphorylated AHPs (Imamura et al., 1998; Suzuki et al., 1998; Dortay et al., 2006). This would be effectively a phosphate sink but by a different mechanism as the phosphates are "sunk" into the cytoplasm rather than onto the type-A ARRs. Three ARRs (ARR3, ARR4, and ARR22) were shown to have phosphohistidine phosphatase activity toward the AHPs (Imamura et al.,

1998; Kiba et al., 2004; Horák et al., 2008). However, the presence or extent of phosphatase activity in other ARRs has not been documented in the literature, so the relative contribution of this mechanism is unknown.

Other mechanisms for negative regulation of cytokinin signaling that do not rely on the type-As have been proposed or documented. For example, the *CKX* genes which degrade cytokinins are transcriptionally upregulated in response to exogenous cytokinin (Werner et al., 2003; Taniguchi et al., 2007; Bhargava et al., 2013). This transcriptional response results in an increase in the number of cytokinin-degrading enzymes in the cell, reducing the abundance of active cytokinin molecules faster than normal cytokinin movement and natural dephosphorylation would otherwise. Additionally, there may be phospho-dependent interaction with cytokinin signaling regulators.

There may be other regulators of cytokinin signaling. It has been proposed that there may be distinct phosphatases to inactivate type-B ARRs (Kiba et al., 2004; Horák et al., 2008). Auxiliary phosphatases have been observed in several bacterial two component systems (Perego et al., 1996; Silversmith, 2010) so there may be uncharacterized auxiliary phosphatases in the *Arabidopsis* genome as well. The type-B ARRs are regulated by the KISS ME DEADLY (KMD) proteins through targeting of E3 ligases (Kim et al., 2013b; Kim et al., 2013a; Zheng and Shabek, 2017), although others have suggested that KMD protein regulation of cytokinin signaling may be a downstream effect of regulation of phenylpropanoid synthesis resulting in altered auxin sensitivity (Kurepa et al., 2018). The Gα EXTRA-LARGE G PROTEIN (XLG) proteins have been shown to interact with E3 ligases to regulate cytokinin signaling through degradation of ARR10 (Wang et al., 2017). It is not known whether ubiquitination-mediated degradation is a factor in the regulation of any other cytokinin signaling proteins. Negative

regulation is also accomplished through nitrous oxide (NO). NO regulation of cytokinin signaling has been observed for AHP1 where NO molecules covalently modify a cysteine residue and disrupt phosphotransfer reactions (Feng et al., 2013).

My thesis studies focused on two areas of research. The first goal was to identify uncharacterized genes involved in cytokinin signaling. To that end, I performed a sensitized genetic screen for *enhancers of ahp* (*eah*) mutants which are hyposensitive to exogenously supplied cytokinin. This *eah* screen identified several new alleles of *ahk4*, one of the cytokinin receptors in *Arabidopsis*, as well as implicating the light-signaling transcription factor *hy5*. It also has identified a number of additional genes that have yet to be identified that are not in known targets.

The second goal was developed over the course of pursuing the first goal, and that was to develop tools to overcome obstacles I encountered in my own research that I anticipated would be useful resources for other scientists. As part of the *eah* screen, we developed the "mutagenomics" strategy for cloning uncharacterized mutants in a screen. Mutagenomics is discussed at length in Chapter 2. We believe the strategies developed during the completion of the screen have the potential to accelerate the pace of genetic screens in plant biology. I also developed a primer design tool, indCAPS, to aid the search for novel CRISPR/Cas9-derived mutant alleles. The indCAPS project is described in Chapter 4. It is designed to be used for Cleaved Amplified Polymorphic Sequence (CAPS) or derived CAPS (dCAPS) restriction-digest genotyping assays with special care given to compatibility with CRISPR/Cas9-derived insertion/deletion polymorphisms.

CHAPTER 2: MUTAGENOMICS: A HIGH-THROUGHPUT METHOD TO IDENTIFY CAUSATIVE MUTATIONS FROM A GENETIC SCREEN

Introduction

A wide variety of genetic screens have been used to dissect numerous biological processes, helping to define the elements involved as well as their roles in various pathways (Forsburg, 2002; Page and Grossniklaus, 2002; St Johnston, 2002; Candela and Hake, 2008). The most common approach utilized is to screen a population of randomly mutagenized lines for a phenotype of interest and to subsequently identify the genes corresponding to these mutations. Generating T-DNA or other insertional alleles simplifies subsequent cloning of the mutated genes, but limits the breadth of allelic diversity obtained, and it is tedious to generate such populations in desired genetic backgrounds. Chemical mutagens, such as ethyl methanesulfonate (EMS), that induce point mutations facilitate the creation of large mutagenized populations in nearly any genetic background and generate a large diversity of alleles; however, cloning the causative genes from such alleles is much more technically challenging. Diverse methods exist for these purposes, but most rely on marker-based mapping procedures. To this end, an outcross is performed between a mutant line of interest with a second line harboring numerous genomic variants, most commonly small nucleotide polymorphisms (SNPs). The F2 population from the cross is collected, and the co-segregation of the phenotype of interest with molecular markers is determined. With sufficient marker density, an interval defining the genomic region containing the targeted mutation can be deduced and the causative gene confirmed by genetic complementation or analysis of independent alleles.

The use of next generation high throughput sequencing technologies and experimental approaches that leverage bulked segregant mapping have greatly facilitated the process of cloning genes corresponding to point mutations (Michelmore et al., 1991). Variations of this include NGM (Austin et al., 2011), SIMPLE (Wachsman et al., 2017), CloudMap (Austin et al., 2011), the SHOREmap pipeline (Schneeberger et al., 2009), MutMap+ (Fekih et al., 2013), and NIKS (Nordström et al., 2013). Typically, these pipelines rely on segregation of SNPs (derived from sequence variations of a distinct parental ecotype or the numerous random SNPs induced by EMS mutagenesis) derived from an outcross. Comparisons of SNP frequencies between homozygous mutant lines and reference lines reveal regions of the genome linked to the causative mutation. Simulations for the effect of increasing the number of F2 segregants pooled from this analysis indicate that at least 40 segregants may be required to generate fewer than 20 candidate genes (Velikkakam James et al., 2013). While powerful, these methods are laborious and, consequently, only a small subset of lines are analyzed from a screen.

The plant hormone cytokinin regulates a diverse set of biological processes in plants, including shoot formation, meristem activity, nutrient uptake, various abiotic and biotic interactions, and multiple developmental pathways (Kieber and Schaller, 2014). Cytokinin binds to histidine kinase receptors (AHKs) (Inoue et al., 2001; Ueguchi et al., 2001b; Yamada et al., 2001; Caesar et al., 2011; Wulfetange et al., 2011) which autophosphorylate on a His within the histidine kinase domain and then shuttle the phosphate to an Asp residue within the fused receiver domain of the AHK. This phosphate is subsequently passed to a histidine phosphotransfer protein (AHPs) (Miyata et al., 1998; Suzuki et al., 1998; Imamura et al., 1999; Hutchison et al., 2006; Schaller et al., 2008), which in turn transfer it to a type A- or type B-Response Regulators (ARRs) (To et al., 2004; Mason et al., 2005; To et al., 2007).

Phosphorylation of the type-B RRs activates these transcription factors, which bind to their genomic targets to regulate the primary wave of cytokinin-regulated transcriptional changes (Mason et al., 2005; Argyros et al., 2008; Zubo et al., 2017), which includes the type A-*ARR*s (Brandstatter and Kieber, 1998; Taniguchi et al., 1998; D'Agostino et al., 2000; Taniguchi et al., 2007). Type-A ARRs negatively regulate the signaling pathway, acting as feedback regulators to dampen the response to cytokinin (Kiba et al., 2003; To et al., 2004; Leibfried et al., 2005; Lee et al., 2007a; To et al., 2007).

Here, we performed a genetic screen on an Arabidopsis line genetically sensitized to perturbation of cytokinin responsiveness to search for novel elements involved in the response to cytokinin. To analyze the output from this screen in an unbiased way, we developed a pipeline to identify multiple causative mutations in a parallel manner, which we call "mutagenomics". In additional to multiple alleles of the *AHK4* receptor, the ethylene signaling element *EIN2*, and *AUX1*, we identified *HY5* as an important mediator of cytokinin responsiveness. Further analysis revealed that *HY5* was necessary for a subset of the transcriptional response to cytokinin. The mutagenomics strategy described here is a powerful tool for researchers performing mutant screens. It is species-agnostic, although it is especially useful in self-fertilizing model organisms. Mutagenomics has the potential to accelerate the pace of discovery of novel genes by greatly reducing the time and effort required to identify causative mutations.

Results

Identification of mutants affected in the response to cytokinin using a sensitized genetic screen.

While prior analysis has defined a cytokinin response pathway from perception of cytokinin by the AHK receptors through phosphorylation and activation of transcription factors, it is likely that additional elements involved in the response to this phytohormone remain to be identified. One challenge with identifying mutations in genetic screens is that the effects of these alterations on the process being analyzed may be subtle, due to genetic redundancy, or alternatively, null alleles may be lethal for some genes due to pleiotropy and thus only hypomorphic alleles may be recoverable. This is particularly true for a well-characterized pathway such as cytokinin signaling, in which the most straightforward genes have already been identified (Kieber and Schaller, 2014). To exhaustively screen for genes involved in the response to cytokinin, we EMS mutagenized an Arabidopsis line that is partially compromised for cytokinin signaling due to loss-of-function mutations in *ahp2* and *ahp3*. This double *ahp2,3* mutant has comparable cytokinin sensitivity to the wild type in a root elongation assay, as does the single *ahp1* mutant (Figure 2.1A) (Hutchison et al., 2006). In contrast, the *ahp1,2,3* triple mutant is partially resistant to cytokinin, suggesting the double mutantis a genetically sensitized to modest perturbations of cytokinin responsiveness. Using a root elongation assay, we determined that 0.1 µM 6-benzyleaminopurine (BA), a synthetic cytokinin, resulted in a substantial, easily scored difference in root length between the *ahp2,3* and *ahp1,2,3* lines, but was not saturating for the response, indicating that this is a suitable condition to detect subtle changes in the response to cytokinin (Figures 2.1A and 2.1B).

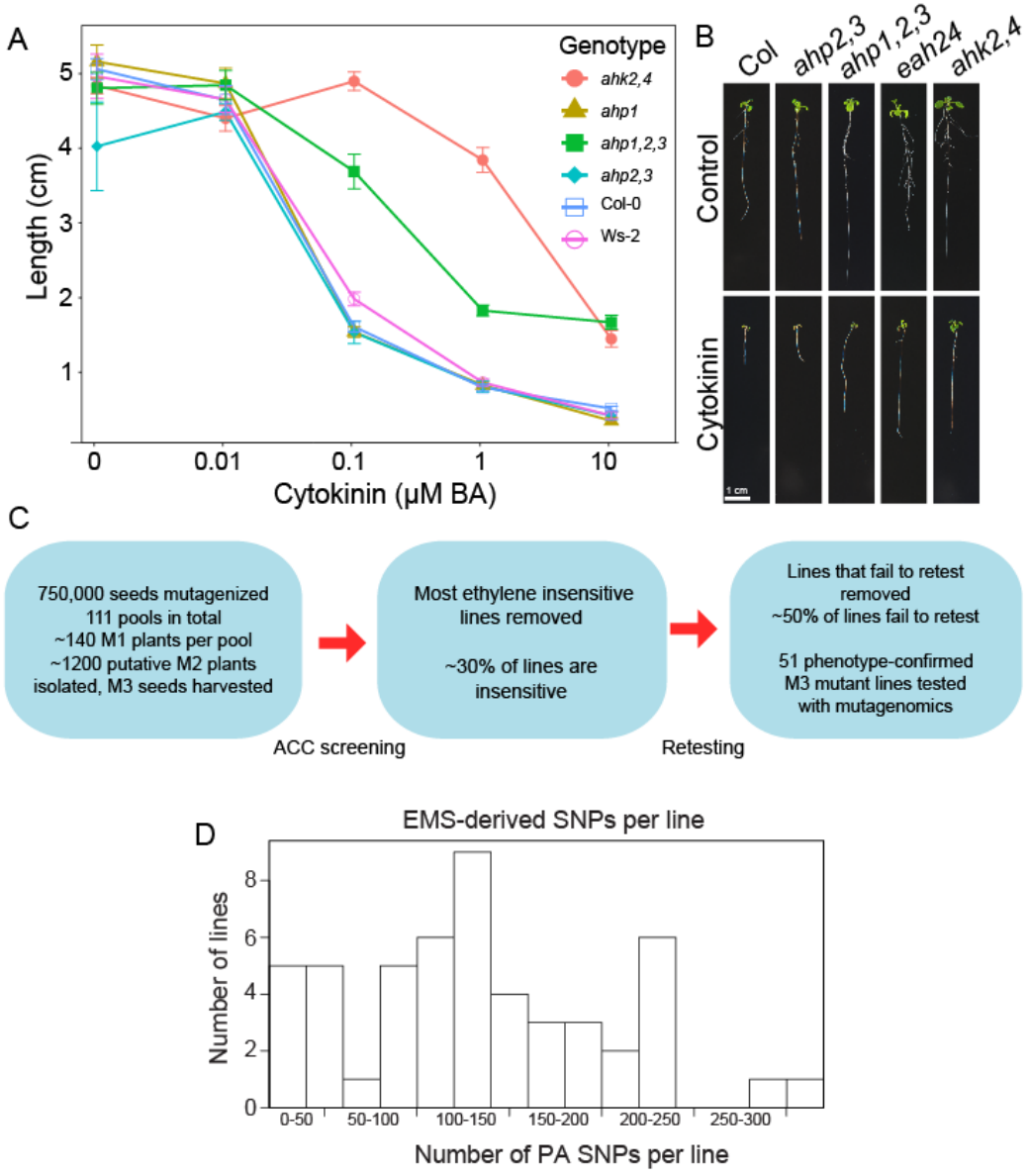

Figure 2.1. Genetic screen for cytokinin hyposensitivity. *We employed an ahp2,3 double mutant as a sensitized parent for a genetic screen of mutants altered in the response to cytokinin. A) Root elongation assay to determine the optimal concentration of benzyl adenine (BA) to use for the screen. Seedlings of the indicated genotypes were grown for three days on control media and then moved to media containing the indicated level of BA. The growth between 3 and 8 days after germination (DAG) was quantified. B) Representative images of wild type (Col) and the indicated mutants, including a representative eah line (eah24), grown as in (A). C) Flow chart describing the details of the genetic screen. D) Histogram of number of protein-altering (PA) SNP densities in sequenced mutant lines. PA-SNPs are inferred using SnpEff and consist of all mutations causing missense, nonsense, or splice site mutations*

We mutagenized an *ahp2,3* line and screened for altered sensitivity to cytokinin using these conditions. Approximately 750,000 Arabidopsis seeds were mutagenized with EMS, the resultant M1 plants grown to maturity and M2 seeds harvested in 111 pools, with a mean pool

size of 140 plants per pool (Figure 2.1C). An initial 1,200 putative mutants were identified from this population, which were refined to 272 strongly hyposensitive mutants from 78 independent pools after re-screening M3 seedlings. A representative mutant is shown in Figure 2.1B. We refer to these mutations as *enhancer of AHPs* (*eah*). Mutants disrupted in ethylene signaling would also be identified in this screen as exogenous cytokinin inhibits root elongation via elevated ethylene biosynthesis (Vogel et al., 1998). To identify these ethylene-hyposensitive lines, we assessed the sensitivity of the *eah* lines to ethylene using a triple response screen (Table 2.1) (Guzman and Ecker, 1990). We removed most of the ethylene-insensitive *eah* lines from further consideration, though we continued to analyze nine as a proof of concept for the subsequent mutagenomic analysis.

Table 2.1. *eah* mutant lines analyzed by mutagenomics.

Mutant	All EMS SNPs	All PA SNPs[3]	Hom. PA SNPs	Gene mutated	Mutation	ACC resp.[4]
eah1	400	112	99	*AHK4*	Splice acceptor variant c.1708-1G>A	S
eah12[5]	653	172	86	*AHK4*	Thr1034Ile	S
eah14-1[1,5]	153	20	14	*AHK4*	Gln928*	S
eah14-2[1]	125	27	11	*AHK4*	Gln928*	S
eah15[5]	565	137	79	*AHK4*	Asp1036Asn	w
eah17[5]	200	34	26	*AHK4*	Ser142Asn	S
eah25[5]	586	171	109	*AHK4*	Gly139Arg	w
eah32	427	106	55	*AHK4*	Ser142Asn	S
eah34	168	46	42	*AHK4*	Trp753*	nd
eah45	779	223	118	*AHK4*	Thr486Ile	S
eah46	142	30	24	*AHK4*	Gln927*	I
eah41	662	171	94	*AHK4* *ARR13*	Asp996Asn Asp66Asn	S
eah13	479	139	75	*AHK4*	Trp134*	S
eah22-1[2]	129	21	10	*HY5*	c.439-1G>A	S
eah22-2[2]	130	25	16	*HY5*	c.439-1G>A	S
eah2	529	137	107	*EIN2*	Cys1141Tyr	I
eah3[5]	364	98	55	*AHK4* *EIN2*	Arg124Gln Pro1185Leu	w
eah4	385	89	47	*EIN2*	Pro460Leu	I
eah6	335	70	41	*EIN2*	Gly247Arg	I
eah10	307	68	45	*EIN2*	Leu375Phe	I
eah18	531	132	35	*EIN2*	Trp1012*	I
eah21	429	103	76	*EIN2*	Trp1158*	I
eah47	401	105	51	*EIN2*	Gly882Glu	S
eah16	409	94	44	*AUX1*	Thr245Ile	S
eah20	318	59	43	*AUX1*	c.363+1G>A Splice donor variant	w
eah30	511	140	59	*CKI1*	Pro289Ser	S
eah31	304	86	50	*ARR3*	Ala140Val	S
eah39	372	83	62	*ARR12*	Pro538Ser	S
eah24	268	63	33	*AHK3*	Trp801*	nd
eah5	563	163	99			w
eah7	300	72	34			I
eah28	697	173	112			I
eahX	224-748	39-201	17-125	15 independent *eah* lines with no change in Table A1.S1 genes		S
eahY	115-615	24-182	12-107	6 independent *eah* lines with no change in Table A1.S1 genes		nd

[1]Sibling group for *eah14*.

[2]Sibling group for *eah22*.

[3]Indicates all PA-SNPs called by GATK and SnpEff, regardless of zygosity.

[4]ACC response: S; Sensitive, I; Insensitive, w; weakly insensitive, nd; not determined

[5]F1 failure to complement in cross with *cre1-12*.

To identify the causative mutations in the *eah* lines, we developed the mutagenomics pipeline, summarized in Figure 2. For this approach, the genomic sequences of a large subset of the mutant lines were determined, with the goal of identifying all the predicted protein-altering (PA) SNPs in each line. We identified M3 progeny that were homozygous for the *eah* phenotype and determined the sequence of their genomes using high throughput Illumina sequencing. We sequenced each mutant line to a target read depth of at least 20x coverage. In addition, we sequenced non-mutagenized *ahp2,3* plants from the same seed stock used for mutagenesis. Any background SNPs present in the pre-mutagenesis stock would be identified when mutants are sequenced, but they are not causative for the novel mutant phenotypes. Cataloguing the background mutations permits filtering those SNPs from final SNP calls for each mutant, restricting the SNPs in consideration to only those introduced by the mutagenesis process. A total of 1,825 background SNPS (relative to TAIR10) were observed in the *ahp2,3* background. In addition to background SNPs, we found that there was an average of 144 PA-SNPs per mutant line (both homozygous and heterozygous), ranging from as few as 20 to as many as 223 (Table 2.1; Figure 2.1D). When one considers only homozygous PA-SNPs, there are on average 57 per line, ranging from 11-125 per line. This is consistent with previous analyses of EMS-mutagenized Arabidopsis that found 100-200 PA-SNPs per line (Thole et al., 2014). This sequence information was used to sequentially analyze the lines with a three-step pipeline (Figure 2.2). In the first step, lines with PA-SNPs in genes known to be involved in the process of interest are identified. For example, in the case of the *eah* screen, we screened for genes previously identified that might, in combination with *ahp2,3*, result in an elongated root in the presence of exogenous cytokinin, such as *AHK4* (Inoue et al. 2001; Higuchi et al 2004). Lines that do not harbor mutations in the screened known targets then progress to step two of

mutagenomics. In this second step, genes that are found in multiple lines with PA-SNPs are identified. The idea is that if a gene is found mutated in multiple independent lines, then it is likely to be causative for the phenotype being screened for. In step three, to identify the causative mutations in the remaining lines, multiple lines from the same pool (likely siblings) are analyzed for overlapping homozygous PA-SNPs to narrow the candidate gene list down to a small number. As described below, we successfully applied mutagenomics to the *eah* screen as a proof of concept for this methodology.

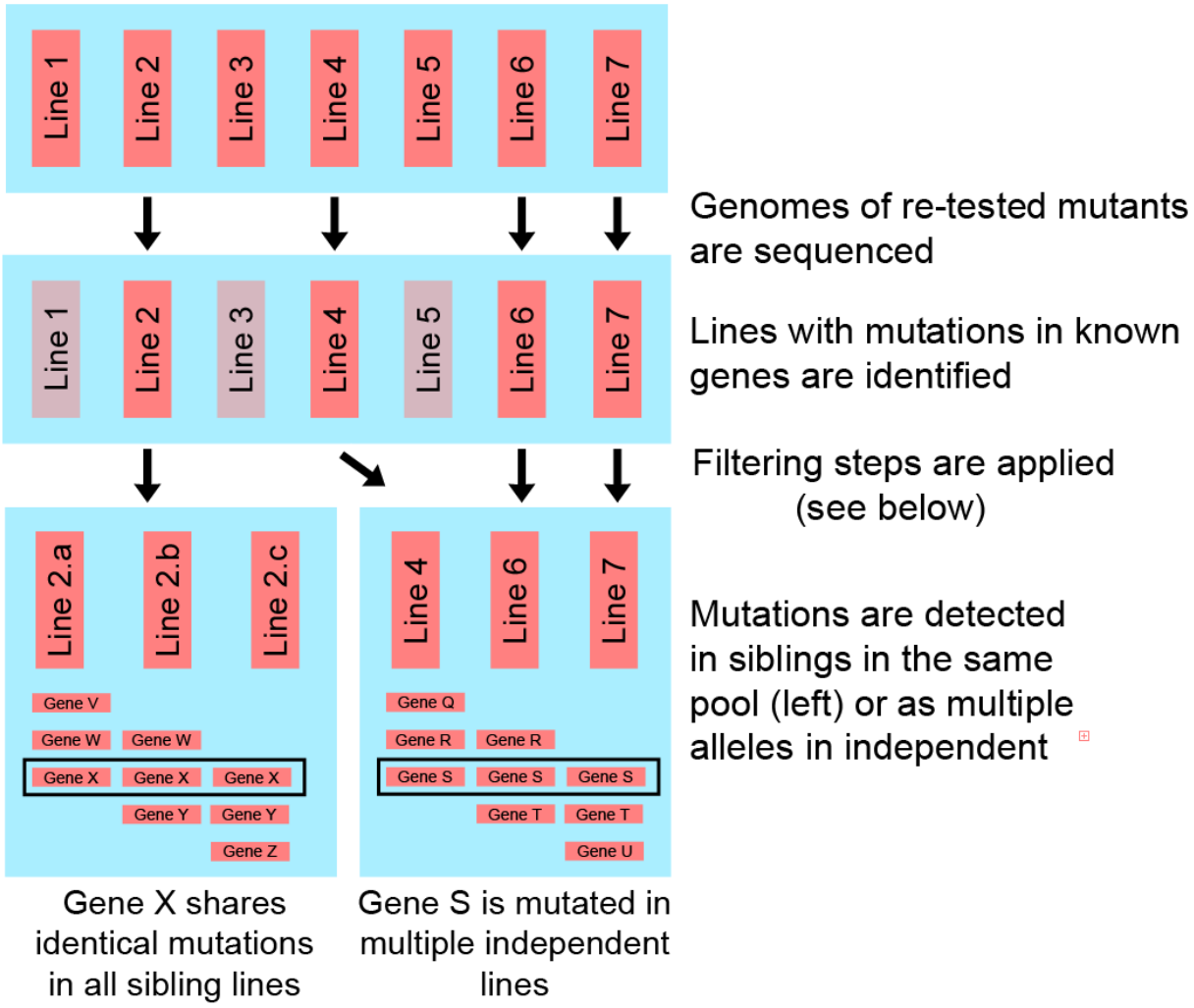

Figure 2.2 Overview of the mutagenomics process. *Top) Implementing mutagenomics is a multi-stop process that starts with a collection of mutant lines with unknown causative mutations. Middle) A number of lines are re-sequenced to detect mutations in known genes. Lines with no mutations in known genes can be prioritized in later stages. Bottom Right) Independent lines are examined to determine whether any genes are mutated in multiple independent lines more frequently than would be expected by chance alone. Bottom Left) Additional mutants from the same pool as previously-sequenced mutants can be re-sequenced to find lines of common descent (sibling lines). Sibling lines will share a subset of mutations between them, and this subset is examined to determine which genes are mutated in every related line. The causative mutation for the phenotype of interest will be in that set.*

Mutagenomics applied to the *eah* screen

Step 1: Screening for mutations in genes known to affect cytokinin responsiveness.

After re-testing and removing most ethylene-insensitive lines, we distilled our *eah* screen to 53 lines that we subjected to mutagenomics (Table 2.1). In the first step, we sought to identify lines harboring mutations in genes that have been previously linked to cytokinin sensitivity (Table S1). We screened the genomic sequences of the *eah* lines for PA-SNPs in these genes. We identified 13 alleles of *ahk4* in the *eah* screen (Table 2.1). *AHK4* encodes a cytokinin receptor and is the predominant member acting in the root response to cytokinin of the three Arabidopsis cytokinin receptors (Higuchi et al., 2004; Nishimura et al., 2004; Riefler et al., 2006). AHK4 has a cytoplasmic N-terminal domain, an ER-localized cytokinin-binding CHASE domain, and cytosolic kinase and receiver domains. We identified eight missense alleles in *ahk4*, four nonsense alleles, and one allele predicted to alter a splice site (Figure 2.3A). We confirmed that the causative mutations in a subset of these lines were indeed *ahk4* alleles by complementation tests with the *cre1-4* allele of *AHK4* (Table 2.1). Four of the *eah* missense mutations were predicted to alter residues within the transmembrane domains of AHK4. Remarkably, this includes two identical missense mutations (S142N) that were isolated from independent pools. These lines (*eah17* and *eah32*) were confirmed to be independent alleles as they did not share any SNPs throughout the genome other than S142N. Two missense mutations were predicted to alter residues within the His kinase domain, consistent with the observation that histidine kinase activity is required for cytokinin signaling (Mähönen et al., 2006). Finally, three of the mutations are predicted to alter residues within the receiver domain (Figure 2.3A). Two of these (D996N and T1034I) are annotated as part of the active site (Lin et al., 1999).. The Asp at position 996 is, in fact, the target Asp that accepts the phosphate group from the HK domain and has been

previously demonstrated to be essential for AHK4 function (Inoue et al., 2001; Mähönen et al., 2006). A third missense mutation, D1036N, is two residues away from? an active site residue on the AHK4 primary amino acid sequence. To examine the predicted effect of these missense alleles, the AHK4 receiver domain was modeled using the hhpred software suite (Zimmermann et al., 2018) and modeled in pymol (Figure 2.3B). Examination of the predicted 3D structure shows the D1036 residue is physically near the active site residues, similar to the D996 and T1034 residues, but further toward the outer surface of the domain.

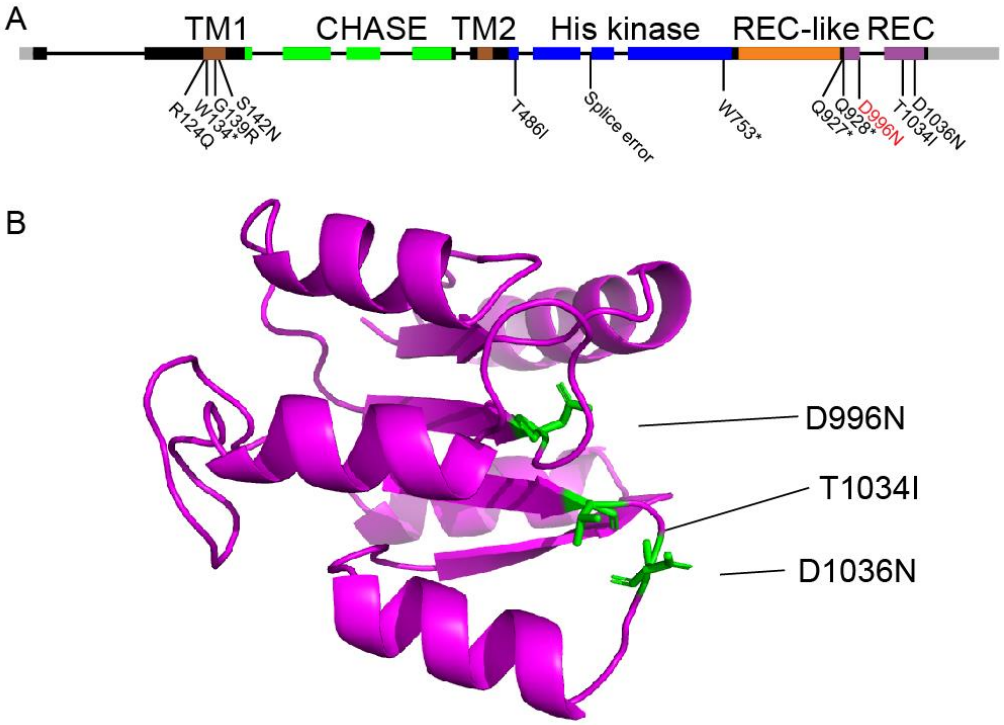

Figure 2.3. Disruptive and missense mutations were observed in AHK4. *A) Gene model of AHK4 showing the position and character of mutations in AHK4 identified in this screen. TM1 and TM2 (brown) are the transmembrane domains; CHASE domain (green); His kinase domain (blue); REC-like (orange) and REC (purple) are the receiver-like and receiver domains; UTRs (gray). The red D996N mutation is the second D in the conserved DDK motif (Mahonen et al., 2006) that is the conserved phosphorylation site for the receiver domain. B) Computationally predicted model of the AHK4 receiver domain. The three mutated residues in this domain are shown in green.*

We also identified mutations in other two-component signaling elements in the *eah* screen. The *eah24* mutant harbors a nonsense allele in the coding region of *AHK3*, which

encodes a second cytokinin receptor that plays a more minor role in cytokinin responses in the root (Higuchi et al., 2004; Nishimura et al., 2004). Single *ahk2* mutations do not significantly affect the response to exogenous cytokinin in root elongations assays (Higuchi et al., 2004; Nishimura et al., 2004) and so it is likely that the cytokinin hyposensitivity in this *eah30* line reflects the sensitized nature of the *ahp2,3* parental line. We also identified mutations in *CKI1*, which encodes a histidine kinase that lacks the cytokinin-binding CHASE domain (Kakimoto, 1996). Various lines of evidence suggest that CKI1 can activate downstream cytokinin signaling to regulate vascular and gametophytic development, though it lacks a CHASE domain and thus its activity is not regulated by cytokinin binding (Kakimoto, 1996; Hwang and Sheen, 2001; Pischke et al., 2002; Hejátko et al., 2009; Yuan et al., 2016). Null alleles of *cki1* are lethal (Pischke et al., 2002). It is possible that this *eah30* allele of *cki1* decreases input into cytokinin signaling sufficiently in an *ahp2,3* background to confer cytokinin hyposensitivity, though it is also plausible that this is not the causative mutation for the cytokinin response phenotype in this line. Finally, we identified two lines with mutations in type-B ARRs. In *eah39*, there is a mutation in *ARR12*; single *arr12* mutants display a modest hyposensitivity to cytokinin in root elongations assays (Mason et al., 2005), and this may act additively with the *ahp2,3* mutations. In line *eah41* there is a mutation in *ARR13*, as well as a second mutation in *AHK4*; it is likely that the *ahk4* mutation is the primary driver of cytokinin hyposensitivity in this line.

Of the *eah* lines that we analyzed, eight of them displayed a strong ethylene-insensitive phenotype as measured by a triple response assay (Table 2.1). Six of these are in the previously identified *EIN2* gene. One line with an *ein2* mutation (*eah47*) was fully sensitive to ethylene in a triple response assay, suggesting either that it is not the causative mutation for the cytokinin hyposensitive phenotype, or that this allele of *ein2* affects root elongation in the light, but not the

triple response in etiolated seedlings. Conversely, one line (*eah3*) with an *ein2* mutation had a weak ethylene-insensitive phenotype; *eah3* also had an *ahk4* mutation that was confirmed by a complementation test with *cre1-4*. Intriguingly, three *eah* lines that displayed an ethylene-insensitive phenotype did not display PA-SNPs in any of the screened ethylene-related candidate genes (Table S1; note that the *eah46* also had an *ahk4* mutation), suggesting that they may affect novel elements in the response to ethylene.

Mutagenomics step 2: Screening for potential causative mutations identified by multiple independent alleles.

The second step in the mutagenomics pipeline involves identifying genes that are found to be mutated in multiple independent lines, above what one would predict by chance. To ascertain the probability of finding any gene mutated in multiple lines randomly, we performed simulation modeling to define the probability of finding independently derived mutations with an increasing number of lines analyzed (Figure 2.4A). The simulations were performed using the average density of homozygous protein-altering SNPs observed in the *eah* screen and used previously reported estimates of crossover frequencies with the zero-crossover and one-crossover categories merged (Salomé et al., 2012). Statistically, the frequency of finding multiple independent alleles of a gene in a random population of mutant lines is fairly high (Figure 2.4A). For example, if one is analyzing 50 independent mutant lines, the probability of finding at least one gene that was mutated by chance four times is >90%, and five times ~25%. The actual rate at which any gene is mutated is dependent on multiple factors, including target size (including both gene size and the number of critical residues) and its chromatin environment.

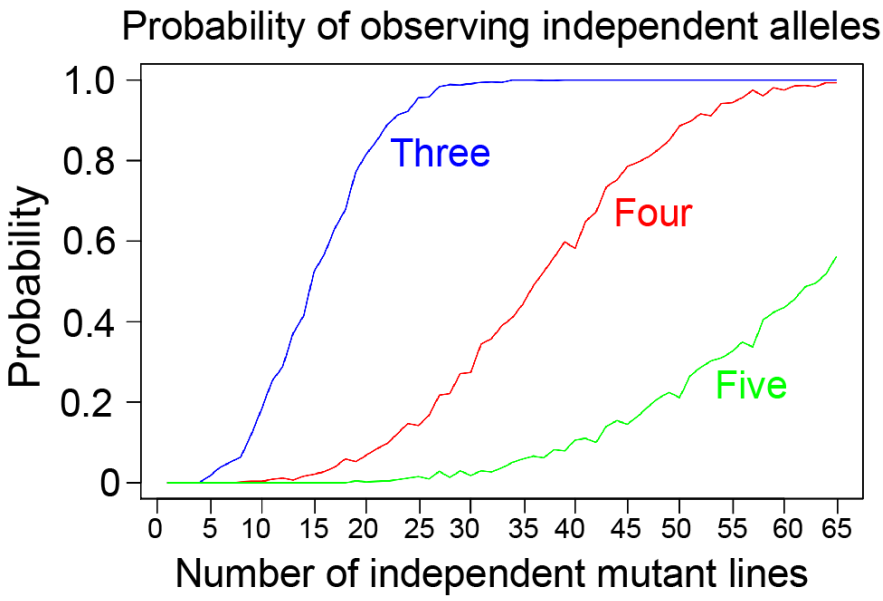

Figure 2.4. Inference of over-enriched mutations in independent mutant lines. *Simulation-derived probability of at least one occurrence of m independent alleles in any gene in n independent mutant lines. Different curves represent different numbers of independent alleles. The blue line shows 3 independent alleles; the red line 4 alleles; and the green line 5 alleles.*

Consistent with our simulation studies, we find many genes that are independently mutated at least two to four times from our pool of 53 *eah* lines (Figure S1). As there is a reasonable probability (~25%) that even five independent alleles would be recovered by chance (Figure 4), we searched for genes identified as being mutated in at least six independent lines as these are highly likely to be causative mutations. The only two genes identified with at least six lines harboring homozygous PA-SNPs in the same gene (i.e. independent alleles) are *ein2* (8 alleles) and *ahk4* (13 alleles). In screens for ethylene-insensitivity, *ein2* is the most frequently identified mutation, presumably reflecting its large size (encoded protein is 1,294 amino acids), non-redundant nature, and strong phenotype (Alonso et al., 1999). Likewise, *AHK4* is a large gene (encoded protein is 1,080 amino acids) and *ahk4* loss-of-function mutations are known to confer a strong cytokinin-hyposensitive phenotype in root elongations assays (Higuchi et al., 2004; Nishimura et al., 2004; Riefler et al., 2006). As noted, the *eah3* line harbors both *ein2* and

ahk4 mutations. As this line both has a weak ethylene-insensitive phenotype and fails to complement *cre1-4*, it is likely a double loss-of-function *ein2 ahk4* mutant.

Mutagenomics step 3: Analysis of lines of common descent to narrow down candidate genes.

In the first two steps of mutagenomics, we identified genes that were previously implicated in cytokinin signaling and/or which are represented by more than six independent alleles in our mutant pool. Twenty-seven lines remain (out the original 53 analyzed) for which a causative gene was yet not identified. For these, we have developed a third step in the mutagenomics pipeline to substantially narrow down the number of potential candidate genes. This step involves sequencing additional lines that are derived from the same pool as the targeted mutant; these will generally be sibling mutants harboring the same causative mutation (if the pool size is sufficiently small, < 100 lines). The basic idea is that, as you sequence increasing numbers of homozygous sibling lines, you continually enrich for the homozygous causative mutation as the other non-relevant PA-SNPs randomly segregate.

To further develop the theoretical underpinnings of this step, we performed a series of simulation experiments to estimate the number of overlapping homozygous PA-SNPs expected in comparisons of two to six M3 siblings. The M3 generation was simulated rather than the M2 in order to emulate a typical screen in which M2 individuals are identified and retested in the M3 generation to confirm the phenotype of interest. For this analysis, we used PA-SNP densities ranging from 10 to 285 per line, consistent with the observed numbers per line in our screen (Figure 2.1D). The PA-SNPs were randomly allocated across the genome and allowed to segregate independently. If a mutation was identified as homozygous in all lines, this was added to the number of candidate genes for that trial. This process was repeated for 10,000 trials to

estimate the degree of overlap observed between siblings. The simulation experiments are summarized in Figure 5. A striking prediction from this analysis is that even for a line with 285 PA-SNPs, sequence analysis of as few as four additional siblings reduces the number of candidate genes to fewer than five, which is a very manageable number for subsequent analysis. With lines containing fewer PA-SNPs, this is reduced to even fewer candidates.

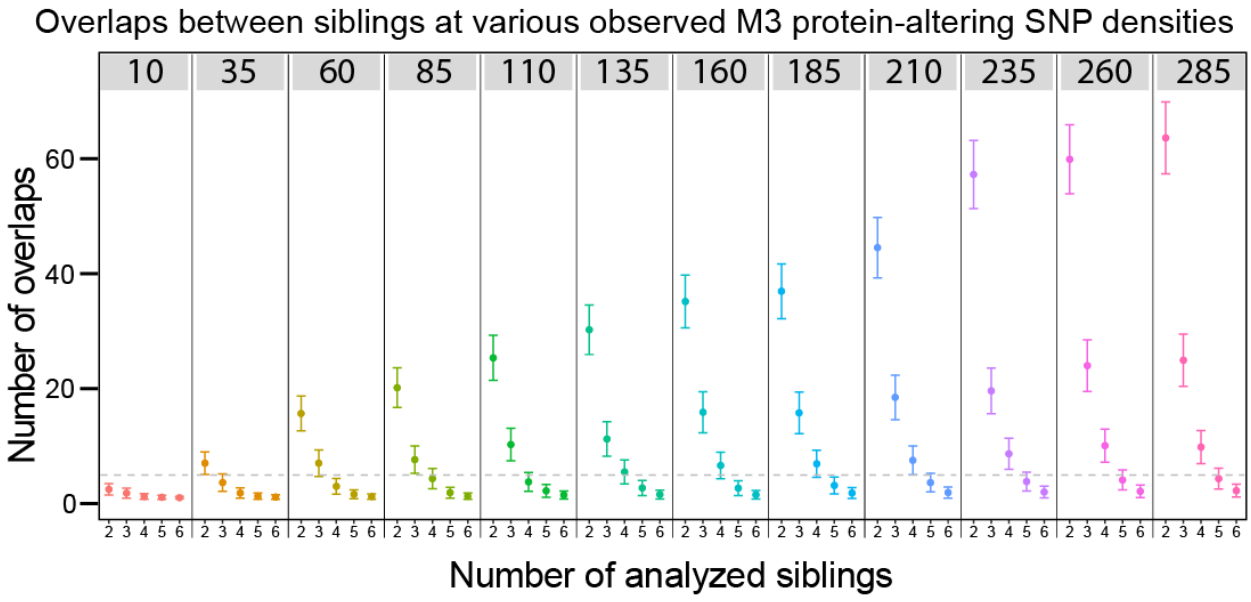

Figure 2.5. Projected numbers of shared alleles between M3 siblings. *Bars indicate the mean and standard deviation of simulation-derived distributions of mutations shared between sibling lines. The gray bar indicates 5 genes. The range of M3 PA-SNP densities is representative of the number of PA-SNPs observed in the* eah *screen.*

As a proof of concept, we sequenced a pair of siblings from the *eah14* mutant line (*eah14-1* and *eah14-2*). Note that *AHK4* was already identified as a causative mutation in the line from both step one and two of the mutagenomics analysis. This pair of siblings had a small mean number of PA-SNPs ($\bar{X}$ = 23.5 if include both homo- and heterozygous PA-SNPs; $\bar{X}$ = 12.5 homozygous; Table 2.1). When the second sibling line was sequenced, it was found to only share a single homozygous PA-SNP, that being in *AHK4* (Figure 2.6F), which is consistent with this

being the causative mutation. Given the observed number of PA-SNPs, this fits with what our simulation mode would predict (Figure 2.5). To confirm that these two lines were true siblings, the network of shared identical SNPs, including those not predicted to affect protein coding, was examined; the large number of shared mutations confirms that these two lines are indeed sibling (Figure A1.3).

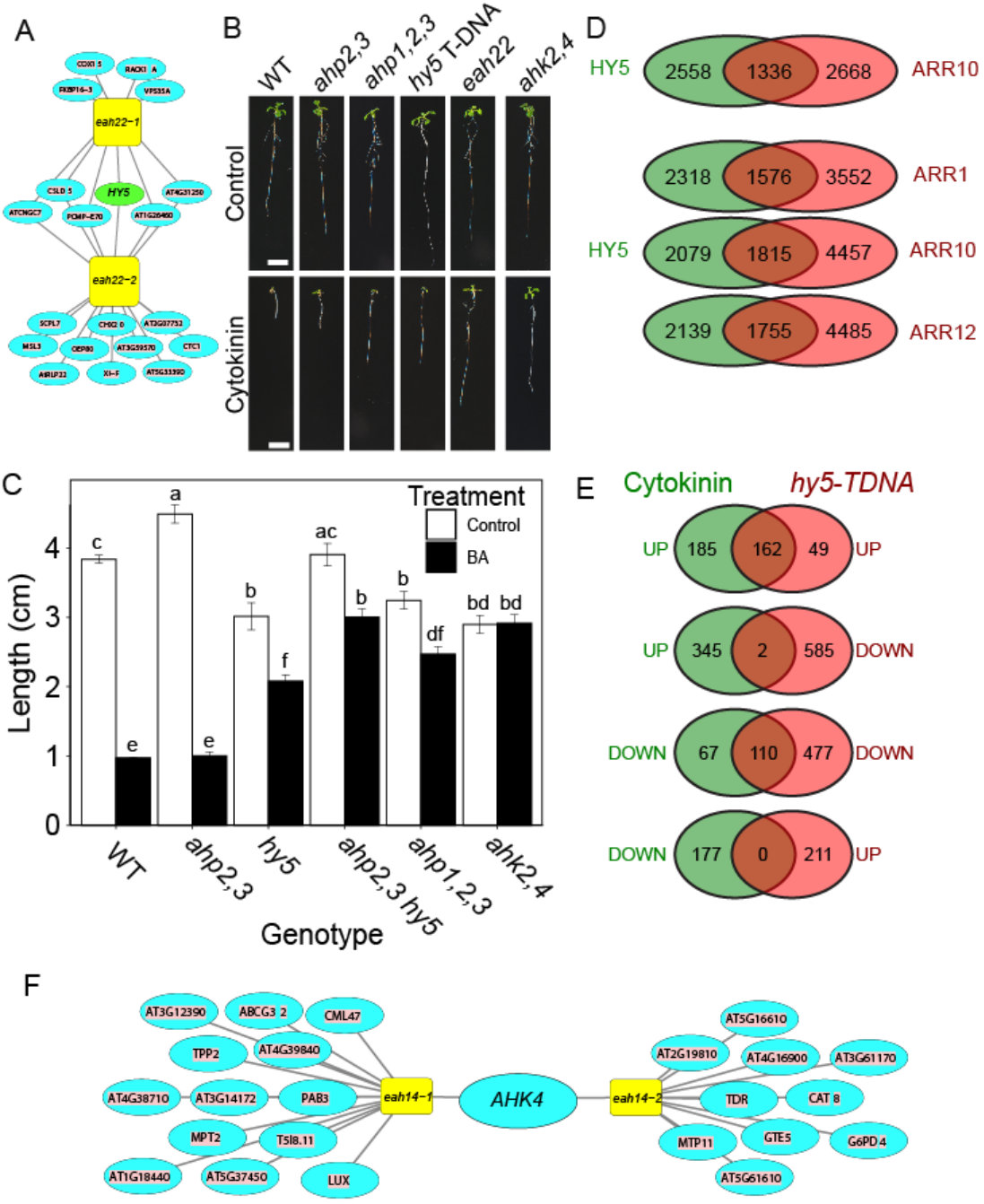

Figure 2.6. Sibling analysis used to identify* hy5 *in two sibling lines. *A) Network of shared and unshared PA-SNPs in* eah22-1 *and* eah22-2*, sibling lines containing a mutation in* hy5*. B,C) Representative images (B) of root elongation at 8 days after germination (DAG) for Col-0,* ahp2,3, ahp1,2,3, hy5-SALK_096651C, eah22, *and* ahk2,4*, grown on control and BA-treated media. Quantified root elongation data (C) from 3 to 8 DAG. Error bars are standard error. Letters indicate statistical significance following a Tukey post-hoc test. D) Comparison of HY5 binding sites with BA-treated type-B binding sites. Type-B binding sites for ARR10 (Zubo et al., 2017), and separately for ARR1, ARR10, and ARR12 (Xie et al., 2018) were compared with binding sites for HY5 (Lee et al., 2007b). E) Comparison of differentially expressed genes between BA and control-treated wild type plants and between BA and control-treated* hy5*. Differentially expressed genes are at least 0.5-fold different (on a log2 scale) with BH-adjusted p-value < 0.02. F) Network diagram of PA SNPs of eah14.*

We also analyzed the *eah22* line, which had a comparatively small mean number (13) of homozygous PA-SNPs (Table 2.1). We sequenced a second *eah* line (*eah22-2*) derived from the same M1 pool that *eah22-1* was identified from. There were only six genes that shared homozygous PA-SNPs in these two lines (Figure 2.6A). One of these, *HY5*, has in fact been suggested to affect cytokinin sensitivity (Cluis et al., 2004), but was not including in our original set of "known" genes due to its predominant role in light signaling. Consistent with the overlap of *hy5* PA-SNPs, both *eah22* siblings displayed characteristic *hy5* phenotypes, such as an elongated hypocotyls and elongated, horizontal lateral roots (Oyama et al., 1997). To confirm that *hy5* was the mutation responsible for cytokinin hyposensitivity in this line, we examined an independent *hy5* mutant line. Similar to *eah22*, a *hy5* T-DNA allele also conferred cytokinin hyposensitivity as measured using a root elongation assay (Figure 2.6B and 2.6C). We conclude that *hy5* is the causative mutation in *eah22*, and further characterized the role of HY5 in cytokinin responsiveness.

Mutations in *HY5* alter the transcriptional response to cytokinin

HY5 is a transcription factor involved in regulating the output of blue-light-responsive cryptochromes and their downstream processes, such as photomorphogenesis and anthocyanin production (Wang et al., 2001; Yang et al., 2001). In the dark, HY5 is degraded by COP1, an E3 ubiquitin ligase (Ma et al., 2002; Bauer et al., 2004). When plant cells are exposed to light, COP1 production declines and HY5 accumulates. We hypothesized that HY5 may play a role in the transcriptional response to cytokinin. To test this, we first determined if the HY5 binding sites in the genome correlated to the binding sites for the type-B ARRs. We compared the ChIP-chip (Lee et al., 2007b) peaks determined for HY5 with the peaks derived from a ChIP-seq analysis of several type-B ARRs (ARR1, ARR10, and ARR12) (Zubo et al., 2017; Xie et al.,

2018) (Figure 2.6D). There was highly significant overlap between genes with adjacent type-B ARR and HY5 binding sites across the genome (p-values for the overlap of HY5 with ARR10 (Zubo et al., 2017), ARR1, ARR10, and ARR12 (Xie et al., 2018) by a hypergeometric test are 6.81e-257, 2.4e-267, 5.6e-283, and 1.9e-252, respectively). For example, 45% of the genes identified as having adjacent HY5 binding sites also are targeted by ARR12. This suggests that these transcription factors share many common targets.

We next examined the role of HY5 in modulating cytokinin-mediated changes in gene expression. Wild-type and *hy5* (T-DNA allele) mutant roots were treated with a vehicle control or 0.1 µM BA and RNA transcript levels determined by RNA-Seq (Fig. 2.6E). A total of 347 genes were upregulated after cytokinin treatment in wild type in these conditions (FDR of 0.02 and $\log_2$ fold-change ≥ 0.5) (Table S2). Of those, 53% (185 genes) were not upregulated in the *hy5* mutant (Figure 2.6E and Table S3). Likewise, 177 genes were down-regulated in response to cytokinin in wild-type roots, and 38% of these were not down-regulated in the *hy5* mutant (Figure 2.6E and Table S3). Interestingly, there were a substantial number of genes up-regulated (49) and down-regulated (477) specifically in the *hy5* mutant (Figure 2.6E).

We also examined if changes in basal gene expression in the *hy5* mutant affected cytokinin-regulated genes. There were 1,277 genes that showed increased and 954 decreased expression in the *hy5* mutant as compared to wild-type roots (Table S4). These genes overlapped significantly with cytokinin DEGs in the wild-type roots (Figure S2), suggesting that disruption of *HY5* alters the expression of a subset of cytokinin regulated genes in the absence of exogenous cytokinin. Together with the ChIP data, this data is consistent with HY5 playing an important role in regulating the transcriptional response to cytokinin.

Discussion

Genetic screens are remarkably powerful tools to dissect biological processes, but the identification of causal genes can be a challenging task in Arabidopsis and other higher eukaryotic organisms. Genetic screens can result in dozens to hundreds of mutant lines depending on the design of the screen and the biological system in question. Current methods, such as mapping by sequencing, have greatly reduced the time required to map any single mutation, but these methods remain difficult to parallelize. The mutagenomics pipeline described here provides a means to identify causative mutations in a relatively rapid, parallel manner, facilitating the cloning of genes corresponding to many identified mutants. Mutagenomics has the capacity to accelerate the analysis of mutant screens, enabling exhaustive screening in Arabidopsis and other model organisms.

We utilized the mutagenomics pipeline to preliminarily analyze 53 independent lines from an enhancer screen for cytokinin response mutants. This analysis identified the causative mutations in approximately half of these lines, demonstrating the utility of the approach. Further analysis of siblings can be applied to identify the causative mutations for the remaining lines. Alternatively, one could also use mapping by sequencing to analyze the remaining lines as a complement to mutagenomics.

The foundational data for the mutagenomics pipeline is determining the genomic sequencing of the identified mutants as well as the un-mutagenized background genome for the screen. While, in theory, this could be done using M2 plants, in general it is highly advantageous to confirm the phenotypes of the isolated mutants after a re-screen of M3 seedlings, and this has the important advantage of identifying lines homozygous for the mutant phenotype. While recessive mutations will be necessarily homozygous in the M2 plants, dominant mutations will

be homozygous in the M2 generation only ~33% of the time, though if semi-dominant there would likely be a bias towards homozygous lines as they would have a stronger phenotype. To include dominant lines in mutagenomics, one could either identify a homozygous M2 sibling from the same pool (as evidence by lack of segregation of wild-type phenotypes in its progeny), or screen the progeny of various M3 plants for homozygosity. A target genome coverage of at least 15-20x provides adequate coverage for SNP calls. For *Arabidopsis*, a paired end 150 base pair sequencing protocol provided enough sequence depth to reach our target coverage and sequence 30 libraries per lane on a HiSeq4000 or HiSeq X10 platform. Attention should be paid to the expected number of clusters per lane for a given sequencing platform as this determines the number of individual fragments which will be contributing sequence information to the experiment. Paired end protocols will increase the sequencing depth for each library but will also increase the cost of the experiment. If a small number of libraries are to be sequenced and cost is a concern, a shorter read length protocol can be used or the experiment changed from paired end to single end.

The first step in the mutagenomics pipeline is the identification of lines with mutations in known genes. In the case of a well-characterized pathway such as cytokinin signaling, there can be many potential known targets, but in more naïve screens, there may be few or no presumptive candidate genes, rendering this step unnecessary. A critical piece of information for this step is knowledge of the SNPs already present in the genetic background of the plants mutagenized for the screen. Novel genetic variation can accumulate as a stock is maintained for many generations, and if not filtered out of the variants observed in the mutant lines those background variants may obscure new mutations in important genes. In the cytokinin screen analyzed here, we included 23 genes related to cytokinin (including *aux1*, which is known to give cytokinin-

insensitivity) and 18 previously identified in ethylene-insensitive screens that we anticipated might be identified in the *eah* screen (Table S1). As anticipated, mutations in several of these were indeed identified in the screen. However, identifying a mutation in a known gene does not guarantee that mutation is causative for the altered phenotype, so care must be taken to examine the specific mutation and to assess the plausibility of that mutation deleteriously altering the protein. Complementation tests can also be performed to confirm the causative nature of mutations in known genes. In any case, this step would help prioritize which lines to focus on in a screen for novel elements in a pathway as one would likely not pursue lines with mutations in known genes, whether complementation was tested or not.

The second stage of mutagenomics is the comparison of independent mutations across all mutant lines to identify any genes mutated more times than expected by chance. Our simulation modeling indicates that the same gene can be found mutated in multiple lines randomly, which is not necessarily an intuitive prediction. In our case, in which we analyzed 53 mutant lines, there was a high likelihood (>90%) that a random gene will be mutated in up to four independent lines, and a reasonable probability in up to five lines (~25%). In our simulation, we mutate genes proportionally to their coding region length, which is a simplification as other factors such as chromatin state can affect mutability (Fahmy and Fahmy, 1971). Further, different genes are likely more or less sensitive to mutations affecting their function due to differences in the number of critical residues and some genes that give a weak phenotype even when harboring a null mutation would be under-represented in the identified mutant pools. These differences likely underlie the low number of times some known genes were identified in the *eah* screen. For example, *AHP1* was surprisingly not found in our screen, even though it results in strong cytokinin-insensitivity when combined with *ahp2,3* (Hutchison et al., 2006); this may reflect its

small size (154 amino acids) and/or low mutability. An *ahk3* mutation was only found in a single line and *ahk2* not at all, which may reflect their relative minor contributions to cytokinin signaling in the root, necessitating a strong/null allele to yield a phenotype that could be identified in the primary screen.

The third stage of the mutagenomics pipeline relies on the ability to identify mutant lines of common descent (i.e. siblings). The ease with which this is accomplished depends to a large extent on the size of the M1 pools generated. At one extreme, an M1 pool size of one would greatly facilitate the identification of siblings, but would make the primary screen more tedious. At the other extreme, very large pool sizes would simplify the primary screen, but would make isolation of siblings more challenging. For example, the original pool size in our *eah* screen was slightly larger (140 per pool) than what is optimal for rapid sibling identification. Confirming that two mutant lines are siblings is clear-cut from genome sequencing data because siblings will share many common mutations, including both PA-SNPs and non-PA-SNPs, but independent lines should share no or extremely few common mutations.

Highlighting the power of the mutagenomics process, we observed a line (*eah14*) for which the causative gene was identified in all three stages of the mutagenomics process (detection of known gene mutations, over-enriched mutations, and mutations shared by siblings).

The idea behind the sibling analysis is that in the mutagenesis process, the M1 plant acquires randomly allocated heterozygous mutations. Upon selfing, this will produce M2 offspring segregating for all mutations, including the mutation of interest. The screen selects M2 lines homozygous for the mutation of interest if recessive. Alternatively, homozygotes can be selected from M3 re-screens for dominant mutations. Thus, as one sequences additional sibling lines, they will be all be homozygous for the mutation of interest (as the phenotype selected for

this), with other random mutations segregating. Our simulation study suggests that sequencing just a few additional individuals rapidly reduces the candidate pool list down to a small, manageable number. For example, for mutant lines with fewer than 60 M1 homozygous SNPs (the average that we found in our screen) comparing only two additional sibling M2 plants (a total of 3 siblings) can result in an overlap list of close to five genes. This process can be carried out in the M3 generation as well, although the overlapping gene sets will be larger as heterozygous M2 SNPs have an opportunity to fix in the M3 generation. Additional siblings can be analyzed by whole genome resequencing, but if the number of candidate genes is small (<10), it may be more feasible to analyze remaining SNPs using a targeted PCR-based approach rather than generate a whole genome sequencing library. One complication in this analysis is that random mutations closely linked to the causative mutation will be difficult to separate if they are in *cis* to the causative mutation in the M1 plant, but this should represent a small minority of the overall PA-SNPs present. Subsequent tests (genetic complementation, isolation of T-DNA or CRISPR alleles) are required in any case to definitively identify the causative gene.

While mutagenomics has the considerable advantages of being rapid, fairly easy to use, and capable of parallel analysis of many mutant lines, it may not be straightforward to identify the causative gene for every mutant line identified. For those mutant lines remaining after the mutagenomics pipeline, a more traditional bulked segregant mapping population could be employed.

We used a screen for reduced cytokinin response in a root elongation assay to test the mutagenomic pipeline. The most frequently mutated gene to come out of this screen was *AHK4*, and this allelic series shed light on the residues essential for AHK4 function. We also further characterized the role of *HY5* in the response to cytokinin. HY5 has been implicated as a link

between cytokinin and cryptochrome pathways (Vandenbussche et al., 2007), and *hy5* mutants were found to be resistant to exogenous cytokinin in root and hypocotyl assays (Cluis et al., 2004). We found that HY5 binds to an overlapping set of genomic targets with type-B ARRs, which mediate the primary transcriptional response to cytokinin (Zubo et al., 2017; Xie et al., 2018). Consistent with this, *HY5* is required for a subset of cytokinin-regulated gene expression changes.

There are 26 lines from our *eah* screen that do not harbor mutations in any of the known targets we screened for. These should provide rich fodder for further application of sibling analysis and perhaps bulk segregant mapping to identify new elements involved in the response to cytokinin. Future genetic screens would be better served by employing slightly smaller pool sizes to facilitate the third step in the mutagenomics pipeline.

Conclusions

Mutagenomics is a powerful approach to facilitate parallel processing of multiple mutants identified from genetic screens. This should empower all manner of screens in diverse genetic backgrounds in Arabidopsis. Further, mutagenomics is species agnostic, and can be applied to most model organisms.

Material and Methods

Plant material and growth conditions

The *ahp2,3* double mutant, the *ahp1* control in Ws-0, the *ahp1,2,3* mutant have been previously described (Hutchison et al., 2006). The *hy5* allele is from the SALK T-DNA collection (SALK_096651C). The *ahk2,4* line has been previously described as *ahk2-7 cre1-12* (Inoue et al., 2001; Yamada et al., 2001; Cheng et al., 2013). Seed were sterilized by incubating

in 95% (v/v) ethanol (1 min) followed by a 15 min incubation in 50% (v/v) bleach (Chlorox, 8.25% sodium hypochlorite), 0.5% Tween 20 for 15 min. After sterilization, the seeds were washed 5 times with sterile water and then stratified for 3 days at 4° C. Root elongation assays were performed on half strength MS with 0.1 µM BA with 8 g/L Phytagel (Sigma Aldrich). Seedlings were germinated on retest media to match the initial screen conditions. Retest plates were grown vertically in constant light conditions at 21° C.

EMS mutagenesis and screen for cytokinin hyposensitivity

Ethyl methanesulfonate (EMS) mutagenesis was performed by incubating 1 g of dry *ahp2,3* seeds in 100 mM pH 7.5 potassium phosphate buffer. EMS was added to a concentration of 0.4% v/v and the seeds were mixed on a rocking shaker for eight hours. The seeds were washed 20 times with water and sown directly on soil. The M1 plants were grown to maturity and the M2 seeds harvested in pools of approximately 140 M1 plants.

For the primary assay, approximately 100 µL of M2 seed was sterilized and plated densely on media containing 0.1 µM BA. Pooled M2 seeds were plated in dense rows on BA plates and grown for five to eight days. Seedlings with roots longer than the surrounding seedlings were picked and transplanted to soil. Transplanted seedlings were grown to maturity, and M3 seed was harvested and stored at -20° C. Retesting of M3 seed was performed on BA media to verify cytokinin hyposensitivity. M3 seed was sterilized and plated on media containing 0.1 µM BA plate at low density. A wild-type control was used on each plate. Three days after germination (DAG), the position of the primary root tips for each seedling was marked. Roots were scanned at 8 DAG and the distance from the mark to the primary root tip was measured using ImageJ (Abràmoff, 2004). The control and treatment root elongation values were compared using a single-tailed Welch's t-test with α=0.975.

The lines were screened for ethylene responsiveness by plating on MS media containing 10 µM ACC. After three days growth in the dark, the lines were scored for their triple response morphology (Guzman and Ecker, 1990).

Library preparation for genome resequencing and RNA-seq

For each mutant line of interest, one to four M3 plants were grown on selective media and inflorescences were collected from each plant. A modified CTAB extraction protocol (Doyle and Doyle, 1987) was performed to extract genomic DNA for genome resequencing. Frozen tissue was ground using a SPEX SamplePrep 2010 Geno/Grinder (Fisher Scientific). Ground tissue was mixed with 2x CTAB buffer. The homogenate was washed twice with chloroform. DNA was precipitated with 0.5X volume 5M NaCl and 2 volumes 100% ethanol and washed with 70% ethanol with 10 mM ammonium acetate. The DNA was then resuspended in 250 µL water with 10 µg ml^{-1} RNase A. The DNA was re-precipitated and washed with 70% ethanol, then resuspended in 50 µL water.

Genomic libraries were then prepared using an Illumina TruSeq PCR-Free DNA library kit (Illumina # FC-121-3003) for the first two phases, and a Kapa HyperPrep DNA library kit for the third phase. Library concentrations were determined using a Kapa library quantification kit (Sigma/Roche #7960140001). Mutant re-sequencing was performed in three rounds, twice on a HiSeq 4000 (first and third phases) and once on a HiSeq X10 platform (second phase).

SNP Detection

Several processing steps are necessary after identifying SNPs in sequencing data. SNPs are filtered out of this set based on quality scores, presence in background mutants, and consistency with an EMS-derived mutation. After filtering, SNPs are analyzed using SnpEff

(Cingolani et al., 2012) to determine whether they are protein-altering mutations. The SnpEff output is then parsed to create final output tables and figures. Currently, these functions are performed by a collection of batch scripts calling R scripts and SnpEff. At publication, the collection of scripts will be packaged together in a GitHub repository such that advanced users could run analyses locally.

Sequencing

Genome resequencing was performed in 3 rounds. The first round of sequencing was performed with 33 libraries on a single lane of a HiSeq 4000 instrument (2x150 bp) by the High Throughput Sequencing Facility (HTSF) at UNC Chapel Hill. The second round was performed on a HiSeq X10 platform (2x150 bp) by BGI (Shenzhen, China). The third round was sequenced on a HiSeq 4000 (2x75 bp) by the HTSF at UNC Chapel Hill. The RNA-seq experiment was sequenced on a HiSeq 4000 in high output mode (2x75 bp) by the HTSF at UNC Chapel Hill.

Mutagenomics Data Processing

Sequenced libraries were aligned to the TAIR10 genome using Bowtie2 (Langmead and Salzberg, 2013). Variants were inferred using the Genome Analysis Tool Kit (GATK) (McKenna et al., 2010; DePristo et al., 2011; Van der Auwera et al., 2015). Detected variants were filtered based on their call qualities by GATK with the filterstring "QD<2.0||FS>60.0||MQ<40.0||MQRankSum<-12.5||ReadPosRankSum<-8.0". Variants passing all GATK filters were further filtered by the read depth of the detected variant and by the predicted effect as determined by SnpEff (Cingolani et al., 2012). Only protein-altering mutations were considered.

AHK4 receiver domain structure

The receiver domain structure was modeled using the HHpred and Modeller packages available at https://toolkit.tuebingen.mpg.de/ (Webb and Sali, 2016; Zimmermann et al., 2018). The response regulator domain (residues 946-1071) was submitted to the HHpred search and the top 100 matches were chosen to model the structure of the domain using Modeller.

RNA-Seq Analysis

The *hy5* gene expression experiment was performed using RNA extracted from flash-frozen tissue using an RNEasy kit (Qiagen, #74106). Samples of mRNA were isolated using Sera-Mag oligo dT beads (Thermo) in presence of RNase Out (Enzymatics) and heat-fragmented. First-strand synthesis was performed with Enzscript (Enzymatics), and second-strand synthesis with DNA PolI (Enzymatics) and RNase H (Enzymatics). End repair was performed with T4 DNA Polymerase (Enzymatics), Klenow polymerase (Enzymatics), and T4 PNK (Enzymatics). A-tailing was performed with Klenow exo$^-$ (Enzymatics). Adapter ligation was performed with T4 DNA ligase (Enzymatics). The libraries were PCR-amplified using KAPA HiFi HotStart ReadyMix (Kapa Biosystems). All wash steps were performed with AMPure XP beads (Beckman Coulter). Read alignment was performed using star (Dobin et al., 2013), and count data was normalized using Salmon (Patro et al., 2017). Differential gene expression analyzed using DESeq2 (Love et al., 2014) in R. The BH-adjusted p-value cutoff (Benjamini and Hochberg, 1995) used for determining significance was lowered to 0.02 to account for only having two replicates and a $\log_2$ fold-change cutoff of 0.5 was applied.

Independent Allele Simulations

The algorithm for the independent allele simulations is as follows. First, a decision is made on how many genomes to simulate and how many mutations to allocate. In this study, the allocated mutations are assumed to represent homozygous recessive mutations. Mutations are randomly allocated across the genome using gene coding length as a weighting factor with the assumption that no gene will be mutated more than once. Once all simulated genomes have had mutations allocated to them, a network of mutated genes is generated and analyzed to determine whether any genes were independently mutated *x* or more times. This simulation process is repeated until a population of simulated screens is collected. Then, the population of screens can be assessed to determine the proportion *n* of all simulated screens in which a gene was independently mutated *x* times. This process can be repeated for different numbers of genomes and different PA-SNP densities.

Sibling Allele Simulations

Empirical crossover frequencies were taken from (Salomé et al., 2012), although the frequencies of no-crossovers and single-crossovers were combined into a single value for single-crossovers. The algorithm for the simulation is as follows. For any given trial, a decision is made on how many siblings to simulate and how many protein-altering mutations to simulate. The chosen number of simulations are allocated to genes randomly in a heterozygous fashion. Mutations are allocated across the whole genome using individual gene coding region lengths as weighting factors. Two rounds of simulated meiosis are performed. In each, a number of crossovers to be performed is randomly sampled from the supplied empirical distribution. The positions of crossovers are randomly allocated, but assumed to always occur between genes. Sister chromatids are generated, and crossovers occur at the pre-determined positions between

randomly chosen non-sister chromatids. One of the four chromatids is randomly chosen to be either the pollen or ovule gamete. The recombination process is then repeated to generate the second gamete. The two simulated gametes are combined to form the M2 generation. The meiosis process described above is repeated to generate the M3 generation, and the whole process is repeated independently for each of the siblings in the trial. The number of genes in a homozygous state shared by all siblings in the trial is then recorded. This process can be repeated for different numbers of siblings and different PA-SNP densities.

User requirements to perform mutagenomics

Several programs are required for use of the Mutagenomics pipeline. The user should have access to a high-throughput computing platform to perform read alignment and variant calling. Once variant calling is completed, the user can move from a server to a personal computer. The scripts as written use Bowtie2 and the Genome Analysis Toolkit.

The pipeline has been set up to be compatible with either a Mac/Linux or Windows operating system. Different versions of the scripts are provided depending on the operating system in use. The user is required to have R installed, either through the base RGui or the RStudio project, and to have the Rscript front-end command available from the command line (e.g. set in the PATH variable on Windows). The user should also have a bash shell installed. Macs have one by default, and Windows users can use Git Bash (available at https://gitforwindows.org/). The scripts rely on several GNU utilities (sed, find) that are native on Macs and available on Windows through projects like Cygwin (available at https://www.cygwin.com/) or GOW (Gnu On Windows, available at https://github.com/bmatzelle/gow/wiki). These scripts were developed on a Windows platform

using Git Bash and GOW. The user should also have a copy of the SnpEff program downloaded, available at snpeff.sourceforge.net.

Code and Data Availability

Statistical and data analyses were performed using R and Python. Code is available at https://github.com/KieberLab/Mutagenomics. The RNA-seq GEO accession number is GSE149641. The Mutagenomics sequencing data access number is PRJNA631403.

Supplemental Data

For the supplemental data tables associated with this paper, please see (Hodgens, 2020).

Acknowledgements

The authors would like to thank Carly Sjogren for helpful discussion, Jamie Winshell for technical support, and Nicole Chang for her invaluable help in performing the screen. The authors also thank Yan Zubo for performing initial genetic analyses that confirmed a role for *HY5* in mediating the root response to cytokinin.

CHAPTER 3: MUTAGENOMICS AND ITS APPLICATION TO ADDITIONAL MUTANT SCREENS

Further evidence of the usefulness of the mutagenomics strategies

This chapter is intended to be an accounting of several successful applications of mutagenomics beyond the scope of Chapter 2. First, I return to the *eah* screen and discuss results of that screen not already described in the previous chapter. This includes an additional sibling not documented in Chapter 2 and a close analysis of the remaining mutations in several mutants with low protein-altering SNP (PA-SNP) densities.

Second, I will describe the use of mutagenomics by two colleagues and the results of those efforts. One of these is the *ACC resistance* (*acre*) screen performed by Asia Polko, a postdoctoral fellow in the Kieber lab, searching for mutations in *Arabidopsis thaliana* that confer resistance to the molecular precursor of ethylene, aminocyclopropane-1-carboxylic acid (ACC). One gene was observed to be mutated more often than by chance alone, and previous results strongly support that this is the causative gene. The second is the *ghostbuster* (*gob*) screen performed by Carly Sjogren, searching for suppressors of the *POLTERGEIST-LIKE1* (*PLL1*) gene. The sequenced lines in the *gob* screen were chosen to make maximal use of sibling analysis, and many sibling groups were identified. Both of these screens leverage different aspects of mutagenomics and considered together are a powerful proof of concept for the method.

Third, I will present a counterfactual exercise: if modern genome sequencing technology had been available to researchers in the 1970s and 1980s, how could mutagenomics have benefited mutant screens performed at the time? The example used is the Heidelberg screen, a classic genetic screen in *Drosophila melanogaster* which discovered many critical genes for embryonic development in flies. I show that if sequencing had been available, then several genes were identified that were mutated more frequently than would be expected by chance alone.

Further analysis of the *eah* screen

The previous chapter documented the *enhancer of AHPs* (*eah*) screen in *Arabidopsis thaliana* and the mutagenomics strategy developed during that screen. The results described therein were an abbreviated version of the full results of the *eah* screen. Several mutants with low SNP densities were documented but not discussed, one member of a sibling pair was omitted, and several experiments are needed to complete the analysis of the screen. It is my hope that future researchers can pick up these loose threads and follow them to new discoveries. In the interest of aiding those researchers, this chapter will complete the description of the *eah* screen.

Sequencing for the *eah* screen was carried out in three phases. The objective of the first phase was the detection of mutations in known genes in 29 *eah* mutants gathered during the screen. This phase focused on independent alleles, although through a book-keeping error, two closely related lines, *eah22* and *eah24*, were included. The observation of multiple mutations in the same genes between these two lines, and the realization that they were from the same pool and had similar root phenotypes, became the basis for sibling analysis.

The second phase was meant to be a mix of completely independent *eah* mutants and mutants suspected to be siblings to already sequenced lines. Of the lines suspected to be

members of sibling sets, three lines turned out to be new sibling sets. This brought the total number of sibling sets to three:

- *eah22-1* and *eah22-2*, identified as having *hy5* mutations
- *eah38-1* and *eah38-2*
- *eah14-1* and *eah14-2*, identified as having *ahk4* mutations

Due to bookkeeping errors, two sets of lines (*eah16* and *eah12*) were sequenced twice. While unfortunate, some new information can be inferred from these lines. From a naïve examination of the network of shared alleles, these lines appear as siblings, not exact copies of each other. There are genes mutated in each duplicate sample that are uniquely mutated. However, this is not an unexpected result. The goal of sibling analysis is to determine which mutations are homozygous across all sequenced M3 samples, assuming a recessive phenotype of interest. A non-causative mutation can have varying prevalence across the samples. If many M3 siblings are sequenced, then the experiment should have a comprehensive picture of the full set of mutations gained in the M1 plant. However, by only sampling a few M3 plants, there is the possibility that some genes are present in low levels or completely absent in the sampling population that contributed gDNA to the library. The set of genes which are homozygous in both of the two libraries will then be large, but not completely overlapping. However, it should still contain the potential causative genes of interest just as two true siblings would.

If no new siblings can be found either by screening collected mutants or by re-screening the pools of M2 seed, then generating a new sampling population of M3 plants from the same seed stock could be used to define a set of genes that must contain the causative gene just as sibling analysis does. This set of candidates will be much larger than can be achieved with true

sibling analysis, but will also be smaller than the set of candidate mutations derived from just one library.

The third phase of sequencing was an attempt purely to find siblings. However, many of the mutants sequenced in the third phase failed later phenotypic retests and have been removed from consideration. In the following sections, I will document the most promising candidates, summarize any relevant facts regarding their genome and detected SNPs, and give recommendations for follow-up.

eah38-1 and eah38-2 are a sibling pair with a small number of shared PA-SNPs

The *eah38* lines (stock numbers #213 and #222, number 15-2 and 15-4) form a sibling group with 15 shared PA-SNPs. These lines are ACC sensitive and have no mutations in or near known cytokinin or ethylene signaling components. *Eah38-1* has 83 total PA-SNPs and 47 homozygous PA-SNPs. *Eah38-2* has 99 total PA-SNPs and 20 homozygous PA-SNPs. The discrepancy in homozygous PA-SNPs (47 vs 20) is unusual compared to the other two sibling pairs (the *eah22* set, with 10 and 16 homozygous PA-SNPs, and the *eah14* set, with 14 and 11 homozygous PA-SNPs). However, the number of heterozygous SNPs in each (83 and 99) is comparable and much larger than for *eah22* and *eah14*, so the gap in homozygous PA-SNPs may be within the range of normal variation. The thresholds for calling SNP depth as heterozygous and homozygous are relaxed to account for a small degree of contamination (Hodgens, 2020), but if too much contamination is present then few confident SNP calls will be made. These two lines share 15 homozygous PA-SNPs which are list in Table 3.1. Identification of additional siblings or a backcross of one of the *eah38* siblings to *ahp2,3* may be useful, but 15 candidates is also a small enough list to manually screen. T-DNA lines for these genes have been ordered and testing of those lines is planned.

Table 3.1: Homozygous PA-SNPs observed in the *eah38* sibling set.

Locus	Info
AT5G06400	PPR superfamily protein
AT3G09700	Chaperone DnaJ-domain
AT3G25700	Aspartyl protease
AT3G19680	Hypothetical protein
AT3G22800	LRR protein
AT3G15130	TPR-like
AT3G58020	Chaperone DnaJ-domain
AT3G60370	FKBP20-2 immunophilin
AT3G23790	AMP-dependent synthetase/ligase
AT5G38520	CLD1
AT5G07440	Beta-subunit of glutamate dehydrogenase
AT1G49430	Long chain acyl-CoA synthetase
AT3G52860	Mediator subunit
AT3G11460	DYW PPR protein
AT2G02020	Major facilitator superfamily protein

Several lines may be amenable to manual assessment of fixed mutations

Three independent lines (*eah49*, *eah36*, and *eah40*) with a low number of PA-SNPs were identified that merit further investigation. These lines are not members of a sibling group and do not have mutated genes that rise to the level of enrichment in the experiment. *Eah49* was isolated as mutant 66-4, stock #658, with 40 total PA-SNPs and 27 homozygous PA-SNPs. *Eah 36* was isolated as mutant 110-7, stock #1157, with 24 total PA-SNPs and 12 homozygous PA-SNPs. *Eah40* was isolated as mutant 17-3, stock #445, with 31 total PA-SNPs and 17 homozygous PA-SNPs. We have screened the homozygous PA-SNPs for each line and narrowed the list down to a set of candidate mutations. The candidate mutations for *eah49*, *eah36*, and *eah40* are shown in Tables 3.2, 3.3, and 3.4, respectively.

Table 3.2. Candidate causative PA-SNPs in *eah49*.

Locus	Info
AT5G37720	Interacting with DNA-Binding Domain of ZN-Finger Parp 1 (DIP2)
AT3G08760	Stress Inducible Kinase
AT3G43670	Copper Amine Oxidase Gamma 2
AT1G77300	Early Flowering In Short Days (ESF)
AT3G15950	Similar to TSK-Associating Protein 1
AT2G26420	1-Phosphatidylinositol-4-phosphate 5-kinase 3
AT3G19190	Peroxisome Unusual Positioning 1/Autophagy 2
AT4G39940	APS-Kinase 2
AT4G27910	SET Domain Protein 16
AT2G17030	SKP1/ASK-interacting Protein 23
AT5G03670	Histone-lysine N-methyltransferase SETD1B-like protein
AT4G30020	PA-domain containing subtilase family protein

Table 3.3. Candidate causative PA-SNPs in *eah36*.

Locus	Info
AT5G13520	Peptidase M1 family protein
AT1G08270	Vacuolar protein sorting-associated protein
AT3G04030	Homeodomain-like superfamily protein
AT5G41610	Cation/H+ exchanger 18
AT2G19560	Enhanced ethylene reponse 5
AT3G17860	Jasmonate-zim-domain protein 3
AT5G60280	L-type Lectin Receptor Kinase I.8
AT5G08610	Pigment defective 340
AT5G10900	Protein phosphatase 7-like

Table 3.4. Candidate causative PA-SNPs in *eah40*.

Locus	Info
AT3G55410	Leucine-rich repeate protein kinase family protein
AT3G46340	2-oxoglutarate dehydrogenase, E1 component
AT5G57010	ATP-dependent caseinolytic protease/crotonase family protein
AT3G05980	Calmodulin-binding family protein
AT3G60510	Hypothetical protein
AT4G23270	Cystein-rich RLK 19
AT3G61600	Light-response BTB 2

Several uncharacterized promising mutants require mapping or additional siblings

A set of mutants were analyzed by mutagenomics but have not yet revealed a reasonable number of candidates genes to identify causative mutations. These lines (Table 3.5) are recommended for further study through a traditional mapping-by-sequencing approach (Velikkakam James et al., 2013) or by acquiring more siblings for each line. If one more sibling can be acquired for *eah23*, *eah33*, and possibly *eah26*, the set of shared mutations will likely fall below 20 genes and permit manual screening for candidate mutations. I recommend that a backcross be performed for *eah9*, *eah44*, and *eah19*, as two or more additional siblings may be required due to these lines' SNP densities.

Table 3.5. *eah* lines not characterized by mutagenomic analysis.

Mutant allele	Line designation	Stock #	All PA-SNPs	Homozygous PA-SNPs
eah9	89-1	932	120	71
eah33	85-2	964	89	53
eah26	110-3	1102	96	58
eah44	44-10	646	117	68
eah23	44-9	645	74	43
eah19	83-10	1105	201	125

***eah5* may have a mutation in an unknown ethylene signaling component**

This line, isolated as mutant 35-6, stock #528, has one of the largest EMS loads of the sequenced lines. Sequencing indicates it has 163 total PA-SNPs and 99 homozygous PA-SNPs, so it is not amenable to manual screening of fixed alleles. The curious feature of this line is that it was weakly insensitive to ACC in a triple response assay but did not have a mutation in or near any ethylene signaling components. Backcrossing or the acquisition of additional sibling lines is recommended for this line.

The causative mutation for *eah24* is likley be *ahk3*

This mutant, number 65-8, stock #874, with 63 total PA-SNPs and 33 homozygous PA-SNPS, possesses a stop mutation in *AHK3* at Trp801*. This stop is in the middle of the receiver-like domain in the protein, at residues 746-865 (Ueguchi et al., 2001a). This mutation is a plausible candidate for the causative mutation in *eah24*, but no complementation test has yet been performed. I have started growing plants to perform a complementation cross to *eah24*.

The *eah11* mutant may be an *arr13* mutant

This mutant, number 72-1, stock #809, has 97 total PA-SNPs and 65 homozygous PA-SNPs. This mutant was initially scored as a moderate mutant, but on subsequent retests was identified as an unambiguously having a strong cytokinin-insensitive phenotype. It is sensitive to ACC in a triple response assay. This line harbors the following homozygous mutations: an intron variant in *ARR13* (c.1665+65C>T), a mutation in the intergenic region downstream of *AHP1* (c.*3599G>A), and an intergenic mutation downstream of *EDF3* (c.*4961C>T), an ethylene responsive transcription factor (Alonso et al., 2003). The mutation near *ARR13*, a type-B ARR (Mason et al., 2004; Schaller et al., 2008) is unlikely to be causative as it is intergenic, not located in a coding or splice site region, and the gene is weakly expressed in roots (Winter et al., 2007; Klepikova et al., 2016). Mutations in *edf3* are associated with enhanced root growth in seedlings grown on ACC media (Alonso et al., 2003), which is consistent with *eah3*'s phenotype on BA media, but *eah11* is ACC sensitive by a triple response assay. The mutation in the region of *AHP1* represents one of two times a mutation was observed near the gene (the other mutant being *eah41*). *AHP1* was predicted to be mutated in this screen but no mutations within the *AHP1* locus were observed. If this mutation is causative for *eah11*, it must be in a regulatory region critical for *AHP1* expression. A complementation cross to *ahp1* could resolve this issue. .

There are too many fixed PA-SNPs to manually screen so analysis of further siblings or a backcross to *ahp2,3* could be pursued.

The *acre* screen for ACC resistance demonstrates the enrichment of causative alleles in a screen

The mutagenomics process was also applied to a screen performed by a postdoc in our lab, Asia Polko. Her screen was looking for mutants in pathways related the ethylene biosynthetic precursor molecule 1-aminocyclopropane-1-carboxylic acid (ACC). Previous research has established that a loss of function mutant in *ein2*, the primary receptor for ethylene disrupts biological responses to certain external stimuli. In etiolated wild type seedlings, ethylene treatment results in the triple response – apical hooking, inhibition of hypocotyl elongation, and hypocotyl swelling (Guzman and Ecker, 1990). Mutations in ethylene signaling components have only modest effects on plant growth and development and display an altered triple response in etiolated seedlings in response to exogenous ethylene.

Mutants disrupted for ACC synthesis are distinct from mutants disrupted in ethylene perception. In contrast to ethylene-insensitive mutants that have only subtle effects on growth and development, an octuple mutant in ACC synthase genes was embryo lethal (Tsuchisaka et al., 2009; Yoon and Kieber, 2013). This suggested that the ACC molecule itself may have a signaling role in plant development. Consistent with ACC acting as a signaling molecule independent of ethylene, inhibitors of ethylene biosynthesis (that is, affecting the ACC synthesis pathway) can affect root cell expansion phenotypes, but disruption of ethylene signaling cannot (Xu et al., 2008; Tsang et al., 2011; Yoon and Kieber, 2013).

The *acre* screen designed to identify mutants insensitive to ACC. An EMS-mutagenized population in an *ein2-5* background was screened for long roots in the presence of ACC. The number of EMS-consistent SNPs, all detected PA-SNPs, and homozygous PA-SNPs for the *acre* screen are summarized in Table 3.6. All the sequenced *acre* mutants are in theory from independent lines, but a naïve examination of the exact shared SNPs in the experiment suggests that two sibling sets exist: *acre26-2*/*acre51-1* and *acre61-3*/*acre56-1* (Figure 3.1A). The apparent *acre26-2*/*acre51-1* sibling group has a shared SNP set with a Jaccard index of 0.99 (SNP overlap shown in Figure 3.1B) and 13 shared homozygous PA-SNPs, and the apparent *acre61-3*/*acre56-1* sibling group has a shared SNP set with a Jaccard index of 0.97 (SNP overlap shown in Figure 3.1C) and no shared homozygous PA-SNPs. If the *acre* lines comprising the apparent sibling sets are removed, there is further SNP enrichment visible. *acre13-2* and *acre65-2* have their own distinct set of shared exact SNPs, and one SNP (Chromosome 5, position 10969685, G->A) is shared between 8 independent lines. It is possible that the 5-10969685-G-A mutation is a causative mutation and an indication the *acre* screen is very saturated, but I believe it to be unlikely. Exact shared SNPs between independent lines have been observed, but rarely in more than three lines. Also, that position is in AT5G28930, a transposable element gene. The 5-10969685-G-A mutation is likely a background mutation not captured in the libraries of pre-mutagenesis *ein2-5* plants.

Table 3.6. Summary of SNP densities in the *acre* screen mutants.

Line	All EMS SNPs	All detected PA-SNPs	Homozygous PA-SNPs	Notes
acre13-2	1049	304	116	shares SNPs with *acre65-2*
acre25-3	885	262	125	
acre28-2	491	123	72	
acre32-2	815	221	123	
acre39-5	485	137	99	
acre59-5	579	159	75	
acre61-3	698	187	0	shares SNPs with *acre56-1*
acre7-2	1002	304	238	
acre10-2	400	111	29	
acre14-1	762	225	173	
acre20-1	487	131	74	
acre21-1	635	154	79	
acre23-2	547	139	82	
acre26-2	824	234	27	shares SNPs with *acre51-1*
acre30-1	617	167	117	
acre43-3	466	106	73	
acre49-2	600	167	104	
acre51-1	425	122	83	shares SNPs with *acre26-2*
acre56-1	1030	298	146	shares SNPs with *acre61-3*
acre65-2	1191	351	63	shares SNPs with *acre13-2*

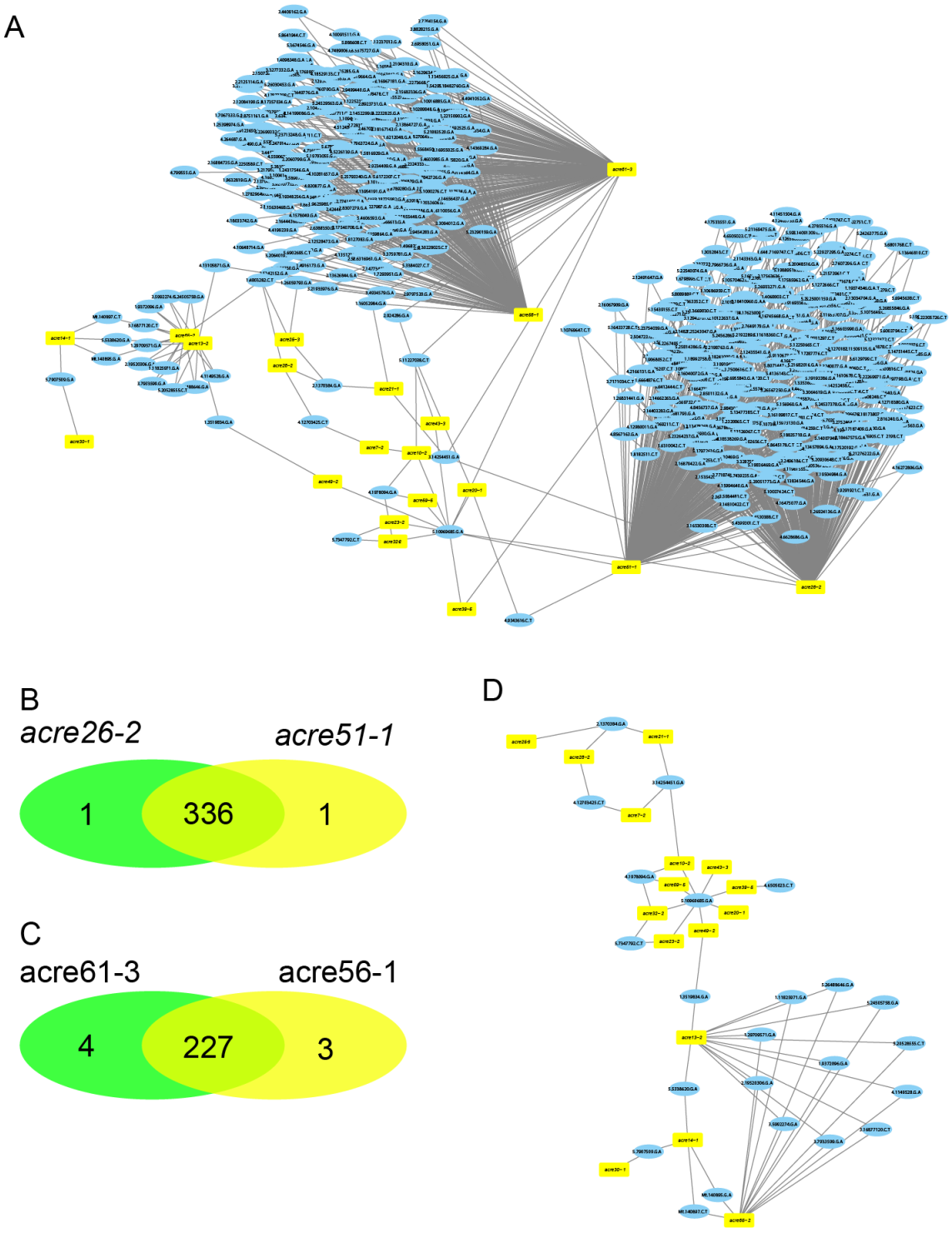

Figure 3.1. Shared exact mutations in the* acre *screen mutants. A. *Network diagram of EMS consistent SNPs (blue ellipses) shared by independent* acre *mutants (yellow rectangles).* ***B,C.*** *Overlapping and line-specific EMS mutations, both heterozygous and homozygous, for* acre26-2 *and acre51-1* ***(B)*** *and* acre61-3 *and* acre56-1 ***(C)***. ***D.*** *Subnetwork of* acre *lines with* acre26-2, acre51-1, acre61-3, *and* acre56-1 *removed.*

The shared exact SNPs in *acre26-2*, *acre51-1*, *acre61-3*, *acre56-1*, *acre13-2* and *acre65-2* are likely background mutations uncharacterized in the pre-mutagenesis libraries or real EMS mutations that appear shared by these independent libraries due to cross-contamination during tissue collection, library preparation, or seed harvest during the initial stages of the screen. If the pool of origin for one of each pair of mutants had been misrecorded, then the overlap in SNPs might represent legitimate common descent. I find this unlikely given that three different pairs of apparent siblings were observed. Without additional re-sequencing, there is no way to disentangle contaminant SNP calls from legitimate SNP calls. These lines will be removed from the current mutagenomic-oriented analysis.

If no sibling pairs are present, then two avenues of mutagenomic inference remain: independent allele analysis and manual analysis of mutant lines with low SNP density. Once the 6 suspect lines are removed, in the network of shared homozygous PA-SNPs, one gene (AT5G40780) was mutated four times and two genes (AT5G24740 and AT5G39000) were each mutated three times. In a network of 14 mutant lines, there is a near zero probability of a gene being independently mutated four times and a near 50% probability of a gene being mutated three times (Hodgens, 2020). The mutations in AT5G24740 and AT5G39000 may be false positives, but they may also be real signs of enrichment of a causative gene and should be examined.

AT5G40780 is *LYSINE HISTIDINE TRANSPORTER1* (*LHT1*) and has previously been characterized as an ACC transporter (Shin et al., 2015). LHT1 is transport protein involved in lysine and histidine uptake in *Arabidopsis* and is expressed in the root epidermis (Chen and Bush, 1997; Hirner et al., 2006; Svennerstam et al., 2007). ACC is an amino acid derivative synthesized from methionine (Yang and Hoffman, 1984; Kende, 1993; Zarembinski and

Theologis, 1994) and a loss of function allele in *lht1* was shown to have impaired uptake of radiolabeled ACC (Shin et al., 2015). The identification of four ACC resistant mutant lines with *lht1* mutations is consistent with our knowledge of its role in ACC transport. The ACC insensitivity in these lines is likely the result of an inability to take up the exogenously supplied ACC. The mutations in *acre25-3* and *acre10-2* are missense mutations disrupting one of the predicted transmembrane domains (Chen and Bush, 1997), the *acre20-1* mutation disrupts a splice donor site, and the *acre43-3* mutation is a missense mutation in a predicted extracellular domain. The *acre43-3* mutation, a leucine to phenylalanine change, is the most interesting, as its position in an extracellular domain suggests it may disrupt ligand binding or transport.

Table 3.7: Mutations in *LHT1* observed in the *acre* screen.

Line	Mutation	Note[1]
acre25-3	Pro387Ser	Helical domain (residues 379-401)
acre10-2	Ala271Val	Helical domain (residues 269-289)
acre20-1	581+1G>A	Splice donor variant
acre43-3	Leu148Phe	Extracellular (residues137-157)

[1]Domain predictions by TMHMM (Sonnhammer et al., 1998; Krogh et al., 2001)

Examination of the next most enriched genes demonstrates a subtle aspect of mutagenomics. It is not sufficient merely to look at the network of mutations in your experiment once. Rather, it must be an ongoing process where you consider the network in the light of new information. Two genes in the *acre* screen appear enriched, but on further examination the enrichment may be spurious. Three lines, *acre20-1*, *acre23-2*, and *acre49-2*, contained mutations in AT5G24740. This gene encodes *SHRUBBY* (*SHBY*), a vacuolar sorting protein (Gallagher and Koizumi, 2013). However, one of the lines with a *shby* mutation (*acre20-1*) also has an *lht1* mutation. The second case is similar. Three lines, *acre32-2*, *acre25-3*, and *acre7-2*, contained mutations in AT5G39000. This gene encodes *MEDOS2* (*MDS2*), a *CARANTHUS ROSEUS* RECEPTOR-LIKE KINASE 1-LIKE (CrRLK1L) protein involved in adaptation to metal ion

stress (Gouget et al., 2006; Richter et al., 2018). However, like *shby*, one of the lines with a *mds2* mutation also harbors an *lht1* mutation (*acre25-3*).

It is possible, though highly unlikely, that two causative mutations for the *acre* phenotype could be observed in the same mutant line. This may have been the case for *eah3*, which has mutations in *ein2* and *ahk4*, is cytokinin hyposensitive, and is weakly insensitive to ACC (Hodgens, 2020). However, two contributing mutations in the same mutant line is an unlikely scenario. For both the *shby* and *mds2* mutations, the fact that one of the lines containing them also contains a plausible causative mutation in *lht1* means that the mutations in *shby* and *mds2* are probably not causative for that line. The mutations in those genes may be false positives for those lines. That then reduces the number of informative mutations for each of those genes from three to two. For a 14 mutant screen, there is a 50% probability of at least one gene being mutated three times by chance alone, but a much higher probability of a gene being mutated twice.

Only one mutant line in the remaining 14 has a low enough density of homozygous PA-SNPs to make manual discrimination of the mutations feasible, and that is *acre10-2* with 29 homozygous PA-SNPs. As noted earlier, *acre10-2* has already been identified as a likely *lht1* mutant. The next lowest density line is *acre28-2*, with 72 homozygous PA-SNPs. Manual examination of the homozygous mutations is possible, but not feasible at this scale. This line would be an excellent candidate for sibling analysis, however. If two more sibling lines can be identified, the set of shared mutations would likely be reduced to 5-10 loci (Hodgens, 2020).

The *acre* screen was designed for a traditional mapping-by-sequencing approach, and the mutants available reflect that goal. Single, strong mutants were selected from each pool to ensure that all analyzed mutants were independent. A screen that intends to apply mutagenomics

benefits from a comprehensive harvest of mutants, ideally collecting multiple strong, phenotypically similar mutants from each pool. Because of this, one half of the mutagenomics strategy is inapplicable to the *acre* screen. However, the *acre* screen still represents a successful application of the mutagenomics strategy. The mutations in *lht1* were implicated as causative and independently implicated by a mapping-by-sequencing experiments.

The *ghostbuster* screen demonstrates the utility of a sibling-focused sequencing strategy

A second screen was performed and analyzed using a mutagenomics process by a colleague Carly Sjogren. This screen, titled the *ghostbuster* (*gob*) screen, was performed in a mutant with a loss of function mutation in *POLTERGEIST* (*POL*) and a hypomorphic allele of *N-MYRISTOYL TRANSFERASE 1* (*NMT1*). POLTERGEIST and POLTERGEIST-LIKE 1 (PLL1) are protein phosphatase type 2C proteins and regulate signaling through the clavata signaling pathway (Pogany et al., 1998; Yu et al., 2000; Song and Clark, 2005). Acylation by NMT1 is required for correct POL and PLL1 plasma membrane localization, a prerequisite for their regulation of clavata signaling (Song and Clark, 2005; Gagne and Clark, 2010). The hypomorphic allele of *nmt1* partially impairs the remaining PLL1 protein, resulting in failure of carpel organ development in approximately 50% of flowers. Mutations in genes affecting PLL1 function result in partial restoration of carpel organ development. All *gob* mutants sequenced for this analysis had 75% or more of flowers with wild type carpel development. This hypomorph approach is necessary because a *pol pll1* double mutant is seedling lethal (Song and Clark, 2005).

The *gob* screen consists of 81 lines from 19 pools. The priority in choosing lines was to identify sibling groups. Sibling groups of two, three, and four siblings were identified. I intend to briefly summarize the results of the *gob* screen from the point of view of assessing whether this

application of mutagenomics was successful in identifying shortlists of plausible candidates in sibling sets, enriched alleles, and mutants with low PA-SNP densities. I will briefly discuss the sibling sets identified in the experiment and the size of the sets of mutations shared by all siblings, but I will not be parsing the list of mutations for plausible candidate genes. Several genes show potential enrichment in the *gob* screen and I describe their identity, but I do not evaluate their biological plausibility as causative mutations. And finally, I identify several lines with homozygous PA-SNP densities low enough to make manual assessment of the mutations feasible.

The screen background libraries do not capture all variation in the mutated seed stock

On first examination, it is clear that background mutations are present in the experiment that were not accounted for in the pre-mutagenesis background libraries. An additional 40 mutations (Table 3.8) were detected that were present in five or more independent lines and may be background mutations rather than true EMS-derived variants. The largest sibling group in the *gob* screen is four, so any mutation present in more than four lines must be either a background SNP not identified by the background libraries or a coincidental exact SNP. Pools 20 and 21 contained a large portion of the background variants, accounting for 21 of the 40 SNPs. It is possible some of these SNPs are coincidences. If one non-background SNP was present in all four members of a sibling group (and thus one initial M1 plant) and in another singular line from another pool, then it would appear to be a background SNP when it is truly a novel EMS SNP. This might be the case for the 5.19173591.C.T mutation observed once in pool 14 and four times in pool 21, or for the 2.19465836.G mutation observed twice in pool 17 and three times in pool 16. However, pool 21 contains many instances of a SNP being present in four pool 21 lines and one pool 20 line. A few SNPs could have been mutated by chance in both pool 21 and pool 20,

but for 21 SNPs to be shared is unlikely. There may have been a rare lineage of background *pol stingy* plants only represented in the M0 seeds which were mutagenized to form pools 20, 21, and 14.

Table 3.8. EMS-consistent SNPs observed in the *gob* screen which are possible background mutations.

Mutation	Line Presence	Mutation	Line Presence
1.13082075.C.T	27 (4), 4	2.19697585.G.A	16, 17, 20, 28, 9
1.14855741.C.T	13, 24, 25 (4)	2.4923157.C.T	1, 21 (2), 2 (3), 24 (3), 25 (3), 26, 27, 28, 4 (2), 9
1.15041793.C.T	1, 24, 27, 28 (2), 4	2.5213550.C.T	13, 2, 28, 21, 9
1.15424317.C.T	9, 28, 4, 2, 21	3.10062955.G.A	20, 21 (4)
1.15435612.C.T	12, 20 (2), 2, 24, 25, 26	3.11029273.C.T	20, 21 (4)
1.15435619.C.T	2, 12, 20, 24, 26	3.14204677.C.T	1 (2), 26, 27 (2), 28, 9
1.17479888.C.T	20, 21 (4)	3.14254451.G.A	1, 12, 14, 18, 2 (2), 24, 25 (2), 27 (2), 4, 6, 9, 20, 21
1.17661315.C.T	20, 21 (4)	3.1705782.C.T	20, 21 (4)
1.19077605.C.T	20, 21 (4)	3.21485063.C.T	12 (2), 13, 17, 25
1.19452909.C.T	11, 25, 27, 9, 2	3.6055128.C.T	20, 21 (4)
1.20275608.C.T	20, 21 (4)	3.6349104.C.T	20, 21 (4)
1.20514812.C.T	20, 21 (4)	3.8901306.G.A	1, 2, 26, 27 (2), 4 (2), 9, 21(3)
1.23395332.G.A	20, 21 (4)	3.9911373.C.T	20, 21 (4)
1.28735948.C.T	20, 21 (4)	4.11477826.C.T	20, 21 (4)
1.3699161.G.A	20, 21 (4)	4.18043997.C.T	20, 21 (4)
1.9424452.G.A	20, 21 (4)	4.9044667.C.T	1, 12 (2), 20 (2), 2
2.11532776.G.A	20, 21 (4)	5.19173591.C.T	14, 21 (4)
2.15793293.G.A	20, 21 (4)	5.5283379.C.T	20, 21 (4)
2.17946228.G.A	17, 2, 21 (3), 25 (2), 26, 27, 28, 4, 9 (2)	5.8088953.C.T	20, 21 (4)
2.19465836.G.A	16 (3), 17 (2)	5.8201672.G.A	1 (3), 13, 14, 17, 21, 24, 25 (2), 26, 27, 28, 4, 9(6), 2, 21 (3)

The *ghostbuster* screen identified candidate causative mutations for several sibling sets

The *gob* pool 2 mutants contain three apparent sibling groups: one group of *gob2-23* and *gob2-24* (SNP information summarized in Table 3.9), a group of *gob2-38*, *gob2-14*, and *gob2-27*

(Table 3.10), and a group of *gob2-10* and *gob2-48* (Table 3.11). There is also a set of nine exact shared SNPs between individuals in the three groups (Table 3.12) which may be background SNPs.

Table 3.9. *gob2-23* and *gob2-24* form a sibling group.

Mutant Line	All EMS SNPs	All PA-SNPs	Homozygous PA-SNPs
gob2-23	451	117	94
gob2-24	492	136	97

Table 3.10. *gob2-38*, *gob2-27*, and *gob2-14* form a sibling group, with 11 shared homozygous mutations between them.

Mutant Line	All EMS SNPs	All PA-SNPs	Homozygous PA-SNPs
gob2-38	201	46	24
gob2-14	229	50	29
gob2-27	233	61	34

Table 3.11. *gob2-10* and *gob2-48* are a spurious sibling group. The number of EMS-consistent SNPs between the two lines varies by 279 yet these lines only share four SNPs not distinguishable as background SNPs.

Mutant Line	All EMS SNPs	All PA-SNPs	Homozygous PA-SNPs
gob2-10	732	186	103
gob2-48	453	101	73

The sibling group consisting of *gob2-23* and *gob2-24* (Table 3.9) shares 84 homozygous PA-SNPs. Few inferences can be made from this overlap set. The acquisition of additional siblings to shrink the set of shared mutations is recommended. The triple sibling group of *gob2-14*, *gob2-27*, and *gob2-38* contains only two genes with homozygous PA-SNPs. These two genes are AT3G49900 and AT3G03580. AT3G49900 is described as a phototropic-responsive NPH3 family protein but it does not have the NPH3 domain involved in blue light phototropism that other members of its family possess (Gingerich et al., 2005). AT3G03580 is *MITOCHONDRIAL*

EDITING FACTOR 26 (*MEF26*), a gene involved in transcript editing for several mitochondrial genes (Arenas-M. et al., 2014).

Some mutations identified in the pool 2 set of *gob* mutants have been classified as likely background contaminant SNPs. These SNPs are documented in Table 3.12. These SNPs are shared between at least two sets of the three apparent sibling groups in pool 2.

Table 3.12. EMS-consistent mutations observed in the pool 2 set of *gob* mutants. The "Line Presence" column indicates which of the *gob* pool 2 mutants contains the given mutation. The mutation format is [Chromosome].[Mutation position].[Reference base].[Mutant base].

Mutation	Line Presence
5.21712990.G.A	*gob2-24, gob2-48, gob2-14, gob2-10, gob2-27*
5.5987533.G.A	*gob2-48, gob2-14, gob2-10, gob2-27, gob2-38*
1.926694.C.T	*gob2-24, gob2-14, gob2-10, gob2-38, gob2-23*
5.6702396.G.A	*gob2-48, gob2-10, gob2-38*
5.8228650.G.A	*gob2-14, gob2-10, gob2-38*
5.284841.G.A	*gob2-24, gob2-48, gob2-14, gob2-10, gob2-27, gob2-38, gob2-23*
1.22743467.G.A	*gob2-24, gob2-48, gob2-14, gob2-10, gob2-27, gob2-38, gob2-23*
5.6025779.G.A	*gob2-48, gob2-14, gob2-10, gob2-27, gob2-38*
1.15645910.G.A	*gob2-24, gob2-48, gob2-14, gob2-27*

These SNPs were removed from consideration for further analysis of the pool 2 sibling sets. After removal of those SNPs, four SNPs remain shared between *gob2-48* and *gob2-10* and two of those are PA-SNPs at two loci: AT5G21326 and AT5G50920. AT5G50920 is *CLP PROTEASE C HOMOLOGUE 1* (*CLPC1*), a chloroplast chaperone protein (Adam and Clarke, 2002; Hengge and Bukau, 2003; Constan et al., 2004). A missense mutation at Gly534Glu was observed; this position overlaps a predicted UVR domain from residues 511-546 (Sigrist et al., 2013). AT5G21326 is *CBL-INTERACTING SERINE-THREONINE PROTEIN KINASE 26* (*CIPK26*) (Drerup et al., 2013). A missense mutation at Thr221Met was observed; this position is inside the protein kinase domain (Sigrist et al., 2013). However, there is also a possibility that the remaining SNPs are either coincidental or additional background mutations. *Gob2-48* has

453 EMS SNPs and *gob2-10* has 732 SNPs, almost twice as many. Siblings typically have different numbers of SNPs but the difference between their SNP densities is usually low, on the order of 50-100 SNPs. Also, only sharing 4 SNPs between siblings is highly unusual given the number of EMS-consistent SNPs the lines have; the *gob2-23* and *gob2-24* lines have about the same number of SNPs as *gob2-48* yet share 577 EMS-consistent SNPs. It is thus likely that the *gob2-48* and *gob2-10* sibling set is a spurious one and neither the *clpc1* nor the *cipk26* mutations are causative. Whether the shared EMS-consistent SNPs between these lines are background mutations or coincidence cannot be determined from this data.

The *gob1-18* and *gob1-15* sibling set (SNP data summarized in Table 3.13) have 51 homozygous PA-SNPs in common. The identification of additional siblings is recommended.

Table 3.13. *gob1-18* and *gob1-15* form a sibling group.

Mutant Line	All EMS SNPs	All PA-SNPs	Homozygous PA-SNPs
gob1-18	361	88	69
gob1-15	395	93	59

The sibling set consisting of *gob24-15* and *gob24-12* (SNP data summarized in Table 3.14) have 18 SNPs in common. A thorough accounting of the mutations observed in this sibling set is beyond the scope of this report, but the number of genes is well within the range of feasibility for manual screening for plausible candidates.

Table 3.14. *gob24-15* and *gob24-12* form a sibling group.

Mutant Line	All EMS SNPs	All PA-SNPs	Homozygous PA-SNPs
gob24-15	412	109	38
gob24-12	329	85	50

The pool 16 group of *gob* mutants illustrates one of the worst-case scenarios for sibling analysis (SNP densities summarized in Table 3.15). On average, including additional siblings in a sibling set should result in a set of shared mutations that is a smaller subset of the original sibling set. However, it is possible that the new sibling identified contains the exact set of mutations already observed. This could result in a case where moving from, for example, 3 siblings to 4 does not reduce the set of shared mutations. This was the case for the *gob16* set. *gob16-15* was the fourth mutant sequenced from pool 16 and it was suspected to be a sibling of *gob16-14, gob16-20,* and *gob16-31*, which already shared mutations in 18 genes. *Gob16-15* contained mutations in all 18 of those genes as well as a few mutations shared uniquely with two or three of the other mutant lines.

Table 3.15. *gob16-14, gob16-20, gob16-15,* and *gob16-31* form a sibling group, with 18 genes mutated in all four lines.

Mutant Line	All EMS SNPs	All PA-SNPs	Homozygous PA-SNPs
gob16-14	504	134	100
gob16-20	511	119	73
gob16-15	661	187	81
gob16-31	555	143	87

A sibling set is formed by *gob9-1*, *gob9-3*, and *gob9-9* (SNP data summarized in Table 3.16) with 19 PA-SNPs shared by all three siblings. A thorough accounting of the PA-SNPs in this sibling set is beyond the scope of this report, but 19 genes is within the range of feasibility for manual screening for plausible candidates.

Table 3.16. *gob9-1*, *gob9-3*, and *gob9-9* form a sibling group, with 19 genes mutated in all three lines.

Mutant Line	All EMS SNPs	All PA-SNPs	Homozygous PA-SNPs
gob9-1	618	178	128
gob9-3	788	210	111
gob9-9	604	159	140

A sibling set is formed by *gob12-27*, *gob12-5*, and *gob12-30* (SNP data summarized in Table 3.17) with 7 homozygous PA-SNPs shared by all three siblings. A thorough accounting of the PA-SNPs in this sibling set is beyond the scope of this report, but 7 genes is within the range of feasibility for manual screening for plausible candidates.

Table 3.17. *gob12-27*, *gob12-5*, and *gob12-30* form a sibling group, with 7 genes mutated in all three lines.

Mutant Line	All EMS SNPs	All PA-SNPs	Homozygous PA-SNPs
gob12-27	473	126	32
gob12-5	524	138	79
gob12-30	517	133	68

A sibling set is formed by *gob21-63*, *gob21-47*, and *gob21-58* (SNP data summarized in Table 3.18) with 10 homozygous PA-SNPs shared by all three siblings. A thorough accounting of the PA-SNPs in this sibling set is beyond the scope of this report, but 10 genes is within the range of feasibility for manual screening for plausible candidates.

Table 3.18. *gob21-63*, *gob21-47*, and *gob21-58* form a sibling group, with 10 genes mutated in all three lines.

Mutant Line	All EMS SNPs	All PA-SNPs	Homozygous PA-SNPs
gob21-63	404	100	62
gob21-47	382	100	78
gob21-58	496	116	46

A sibling set is formed by *gob27-1*, *gob27-6*, and *gob27-9* (SNP data summarized in Table 3.19), with 16 homozyous PA-SNPs shared by all three siblings. A thorough accounting of the PA-SNPs in this sibling set is beyond the scope of this report, but 16 genes is within the range of feasibility for manual screening for plausible candidates.

Table 3.19. *gob27-1, gob27-6,* and *gob27-9* form a sibling group, with 16 genes mutated in all three lines.

Mutant Line	All EMS SNPs	All PA-SNPs	Homozygous PA-SNPs
gob27-1	515	111	45
gob27-6	512	110	53
gob27-9	480	101	54

The sibling set formed by *gob9-4* and *gob-22II* (SNP data summarized in Table 3.20) contains 31 PA-SNPs shared between the two lines. An additional sibling would be beneficial in shrinking the set of shared PA-SNPs in this sibling set. It is possible to manually screen 31 genes for plausible candidates but the number of knockout lines required to test all candidates may approach infeasibility, especially if candidates from other sibling sets are being assessed at the same time.

Table 3.20. *gob9-4* and *gob9-22II* form a sibling group, with 31 genes mutated in both lines.

Mutant Line	All EMS SNPs	All PA-SNPs	Homozygous PA-SNPs
gob9-4	553	122	91
gob9-22II	467	117	74

Two lines, *gob17-42* and *gob17-44*, form a sibling group with 16 genes mutated in common, and a third line (*gob17-18*) may form a sibling group with *gob17-44* with two genes mutated in common. No PA-SNPs are shared by *gob17-18* and *gob17-42*. The SNP data for these three lines is summarized in Table 3.17. The putative sibling status of *gob17-18* may be

spurious. While all three lines have similar numbers of EMS-consistent SNPs and total PA-SNPs, *gob17-18* has more than twice as many homozygous PA-SNPs as *gob17-44* and *gob17-42* (Table 3.21). A thorough accounting of the PA-SNPs in this sibling set is beyond the scope of this report, but 16 genes is within the range of feasibility for manual screening for plausible candidates.

Table 3.21. Pool 16 mutants forming potential sibling group(s).

Mutant Line	All EMS SNPs	All PA-SNPs	Homozygous PA-SNPs
gob17-44	662	175	68
gob17-42	726	191	55
gob17-18	766	202	135

The remaining prong of mutagenomics is the examination of low-SNP-density mutants with no siblings or enriched mutations. The SNP density data for the remaining mutant lines is summarized in Table 3.22. Several lines stand out as good candidates for manual screening: *gob12-1* (19 homozygous PA-SNPs), *gob20-15* (16), *gob20-68I* (19), *gob21-4* (40), *gob21-55* (29), *gob21-6* (31), *gob26-11* (10), *gob26-14* (12), and *gob28-4* (11).

Table 3.22. Remaining mutants in the *gob* screen not part of a sibling group.

Mutant line	All EMS SNPs	All PA-SNPs	Homozygous PA-SNPs	Mutant line	All EMS SNPs	All PA-SNPs	Homozygous PA-SNPs
gob1-7	462	115	60	*gob25-22*	679	200	101
gob1-9	916	240	64	*gob25-3*	529	126	54
gob11-5	600	169	121	*gob25-4*	693	186	95
gob11-9	309	75	48	*gob25-7*	501	126	72
gob12-1	176	40	19	*gob25-9*	870	248	103
gob12-29	565	155	112	*gob26-10*	957	266	82
gob13-35	581	152	96	*gob26-11*	170	28	10
gob14-18	588	160	80	*gob26-14*	142	18	12
gob16-27	307	69	53	*gob26-15*	607	153	52
gob16-46	469	119	57	*gob26-16*	389	103	49
gob18-14	422	100	51	*gob27-10*	441	90	43
gob2-10	732	186	103	*gob27-4*	498	108	72
gob2-48	453	101	73	*gob28-1*	419	101	42
gob20-68II	376	77	50	*gob28-4*	757	223	11
gob20-15	429	94	16	*gob28-5*	346	81	44
gob20-68I	441	93	19	*gob28-6*	553	115	63
gob21-4	297	73	40	*gob28-7*	352	88	83
gob21-3	845	235	134	*gob4-1*	606	157	104
gob21-55	337	90	29	*gob4-13*	388	105	59
gob21-6	367	84	31	*gob4-18*	692	175	54
gob21-60	428	92	46	*gob4-38*	712	168	101
gob21-7	531	128	80	*gob6-17*	371	88	49
gob24-11	498	120	43	*gob9-21*	654	201	111
gob24-24	451	100	55	*gob9-21*	577	171	97
gob25-17	774	197	65	*gob9-23*	512	155	109

No genes harbored more mutations than expected by chance in the full 81-library network. If sibling groups are removed, reducing the network only to 50 libraries with no remaining sibling information, several genes have enrichment which is possible but not guaranteed with a network of the given size. Five mutations were observed in *COP1-INTERACTIVE PROTEIN1* (*CIP1*, AT5G41790), a protein involved in mediating regulation of COP1 protein (Matsui and Deng, 1995). Four mutations were observed in each of OCTOPUS-LIKE1 (*OPL1*, AT5G01170) and *ESSENTIAL FOR POTEXVIRUS ACCUMULATION1* (*EXA1*, AT5G42950). *OPL1* is a homolog of a regulator of protophloem development expressed in mature xylem tissue (Nagawa et al., 2006; Ruiz Sola et al., 2017). *EXA1 is* a protein involved in

the response to infection by a plantago asiatica mosaic virus (PIAMV) (Hashimoto et al., 2016). These three genes may not be causative in the *gob* screen but are still good candidates for further study.

Mutagenomics as applied to the *gob* screen was able to produce experimentally tractable lists of candidate mutations for 9 pools and lists which could be reduced to tractable status in two other pools. Additionally, 9 independent mutants were identified with tractable sets of homozygous mutations and three genes were tentatively identified. These results make the *gob* screen the most successful application of mutagenomics to date and an exemplary model for future screens.

The Heidelberg screen identified several enriched alleles

The *eah*, *acre*, and *gob* screens all demonstrate the efficacy of the mutagenomics strategies. However, the *acre* and *gob* screens are incomplete, and while some of the mutants in the *eah* screen have been characterized, there remain many un- or partially-characterized lines. Further demonstration of the viability of the mutagenomic strategy may help convince any who remain skeptical. To that end, I examined a classic genetic screen the results of which are long verified.

Specifically, I returned to the Heidelberg Screen in *Drosophila melanogaster* (Nüsslein-volhard and Wieschaus, 1980). In that screen, 600 mutants were identified describing 120 loci, with 100 loci having more than one allele represented. The *D. melanogaster* genome contains 13,957 protein coding genes and a 142 Mbp genome (Gelbart et al., 1996), fairly comparable to the Arabidopsis genome. The dosage used to mutagenize the flies mutated approximately 14 genes per sperm (Wieschaus and Nüsslein-Volhard, 2016). A series of 10,000 simulated

mutagenesis experiments was performed using the *D. melanogaster* genome statistics and the above mutation rate to estimate the null distribution of shared alleles in the Heidelberg Screen (Table 3.23) using the same simulation methods described in the previous chapter (Hodgens, 2020). If high-throughput sequencing had been available at the time, then within the mutagenomics framework, *wingless*, *patch*, *paired*, *Krüppel*, and *knirps* could have been quickly identified as plausible candidates for causative mutations without needing to map the genes (Table 3.24).

Table 3.23. Simulation-derived probabilities of multiple alleles in the same gene arising by chance.

Number of independent alleles	Probability by random chance
2	0.1358
3	0.02779
4	0.004167
5	5.03E-04
6	4.88E-05
7	4.33E-06
8	2.70E-07
9	2.00E-08

Table 3.24. Alleles identified in the Heidelberg screen indicated as enriched by mutagenomics.

Class	Locus	Number of alleles
Segment-polarity	*wingless*	6*
	gooseberry	1
	hedgehog	2
	patch	8*
	paired	3*
	even-skipped	2
Pair-rule	*odd-skipped*	2
	barrel	2
	runt	1
	engrailed	6
	Kruppel	6*
	knirps	5*
Gap	*hunchback*	1

* Gene which passes 0.05 probability threshold

CHAPTER 4: INDCAPS: A TOOL FOR DESIGNING SCREENING PRIMERS FOR CRISPR/CAS9 MUTAGENESIS EVENTS

Introduction

It is often necessary to genotype biological samples to select individuals from a large population with a desired genetic variant. Genetic variants generated by mutagenesis or natural variation can take the form of single nucleotide polymorphisms (SNPs) or insertions/deletions (indels). Sufficiently large indels can be distinguished using PCR followed by agarose or polyacrylamide gel electrophoresis (PAGE), but differences of one or two base pairs can be difficult to distinguish reliably even with PAGE, and SNP alleles are refractory to size-based genotyping. Diagnostic tools for genotyping samples with SNPs or small indels include PCR-based cleaved amplified polymorphic sequences (CAPS) or derived CAPS markers (dCAPS) (Neff et al., 1998). A typical CAPS assay consists of a short amplicon centered on a restriction site present in only one genotype. The CAPS assay identifies the genotype of the individual based on whether or not the PCR product is cleaved by the differential restriction enzyme (Fig 1A). A dCAPS assay can be used if there are no restriction sites differentially present in the wild-type and mutant genomic sequences. The dCAPS assay introduces or disrupts a restriction enzyme motif near the mutation by amplifying the target sequences using an oligonucleotide primer that includes one or more mismatches relative to the template (Fig 1B). The mismatches are chosen so that following amplification, a restriction site is introduced into either the wild-

type or the mutant amplified fragment, which can then be distinguished by restriction enzyme digestion followed by agarose gel electrophoresis.

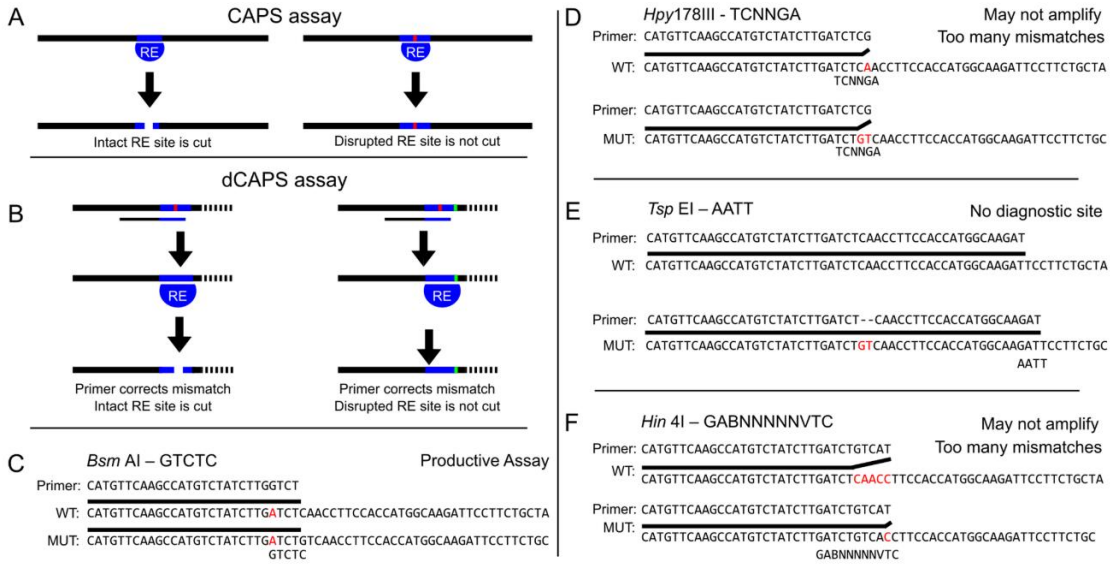

Figure 4.1. CAPS/dCAPS markers can distinguish alleles, but output of dCAPS Finder 2.0 can be flawed. *(A) Diagram of CAPS technique. An amplicon centered on a restriction site (blue bar) disrupted by a SNP or indel (red bar) is differentially cleaved by a restriction enzyme (RE) in the wild- type vs mutant. (B) Diagram of the dCAPS technique. A restriction site can be introduced into either the wild-type or mutant target sequences using mismatched oligonucleotide primers to discriminate two sequences. The mutation (green bar) disrupts the introduced restriction site such that it is not cleaved by the restriction enzyme (RE). Gel electrophoresis can be used to identify the size difference between the wild-type and mutant fragments in both the CAPS and dCAPS methods. (C-F) A sequence with a two base pair deletion at a CRISPR cut site, chosen using CRISPR-Plant (Xie et al., 2014), was supplied to dCAPS Finder 2.0 with a mismatch allowance of 1 base pair. A minority of proposed assays are viable (C), but others possess too many mismatches for successful amplification by PCR or do not introduce diagnostic restriction sites (D-F).*

As CRISPR/Cas9-generated mutant alleles become more prevalent, there is a growing need for a facile method for screening and genotyping indel alleles (Housden and Perrimon, 2016; Farboud, 2017; Karkute et al., 2017; Tandon et al., 2017). Assays based on dCAPS markers are ideal for this as they are simple, robust, inexpensive, and relatively high throughput. However, designing productive primers for allele-specific dCAPS assays can be cumbersome.

Here, we present the development of a new web-based tool to design dCAPS primers for indels that should be of general utility for analysis of CRISPR/Cas9-generated mutant alleles in any species. We demonstrate the utility of this tool using CRISPR/Cas9 targeting of the AHK3

locus in *Arabidopsis thaliana*. AHK2, AHK3, and AHK4 are the three receptors present in *Arabidopsis* that are involved in the perception of cytokinin, a plant hormone regulating a diverse set of biological functions in plants (Kieber and Schaller, 2014). Previous studies have identified null alleles of *ahk2* and *ahk4*, but the previously identified *ahk3-3* allele is hypomorphic rather than a null allele, as residual full-length *AHK3* transcript was found to be present in *ahk3-3* seedlings (Cheng et al., 2013). Primers generated by the indCAPS tool were successfully used to identify editing events at the *AHK3* locus, and viable triple null mutant lines for the cytokinin histidine kinase receptors were identified. The indCAPS tool has the potential to be an important resource for investigators seeking to find new CRISPR alleles or design genotyping primers for known alleles.

Materials and methods

Software

The indCAPS package was written in Python (version 3.5.2) and is implemented as a webapplication using the flask framework (0.12), the bleach package (1.5.0) for input scrubbing, and the gunicorn WSGI HTTP server (19.7.1). It is provided through an OpenShift application platform available from UNC-Chapel Hill. The website is available at http://indcaps.kieber.cloudapps.unc.edu. The source code is available at https://github.com/KieberLab/indCAPS.

Plant growth and transformation

Plants were grown at 21˚C in long days (16 h light). The *ahk2-7 cre1-12* double mutant was transformed with pCH59, a pCUT series binary expression vector containing *AHK3*-targeting gRNA sequences and expressing plant codon optimized Cas9 (Peterson et al., 2016), by the floral dip method (Clough and Bent, 1998). Putative transformants were selected on Murashige and Skoog media containing 50 μg/ml hygromycin and then transferred to soil and allowed to set seed. T2 seeds were plated on Murashige and Skoog plant growth media (2400 mg/L MS salts, 250 mg/L MES buffer) (Murashige and Skoog, 1962), and then seedlings were transplanted to soil and genotyped for editing at the AHK3 locus.

Detection of ahk3 mutations using dCAPS

Oligonucleotide primers were designed to detect editing at the *AHK*3 locus using the indCAPS tool. Amplification of the *AHK3* locus was performed with primers AHK3.dC.F and AHK3.dC.rc (data in S1 Text) followed by digestion of the amplicon with *Bsa* BI. Digests were analyzed using gel electrophoresis with a 3% agarose gel. Lines lacking any wild-type digestion pattern were selected for analysis. Sanger sequencing was used to characterize editing events as single base-pair indels and to confirm homozygosity.

Results and discussion

dCAPS Finder 2.0 has poor compatibility with indel alleles

While a web tool for the design ofdCAPS primers has been described (Neff et al., 2002), it was designed primarily to detect SNP alleles. Primers generated with the tool for small indels often will not actually amplify either the wild-type or mutant sequences by PCR, or in some cases will not actually distinguish between the wild-type and mutant sequences. For example, an

analysis of potential dCAPS primers generated by the existing dCAPS program (http://helix.wustl.edu/dcaps/dcaps.html) for indels in several genes in *A. thaliana* demonstrated that as few as 15% of the suggested primers are capable of distinguishing the provided alleles (AHK3, 50% capable; CENH3, 14%; AGO1, 17%; MED20, 13%; RB1, 15%, see Table 2.1 for further details). Primers from dCAPS Finder 2.0 were examined by constructing the amplicons which the primer would generate in both the wild-type and an edited sequence and examining the amplicons for utility in a CAPS or dCAPS assay. Many of the primer pairs are non-functional and either do not generate a diagnostic restriction site or likely would not amplify the target DNA due to alignment gaps or extensive 3' mismatches between the primer and one of the template sequences (Fig 4.1C–4.1F). There are workaround methods for dCAPS Finder 2.0 in which the user supplies two sequences in which the terminal base is the indel, rather than placing the indel in the middle of the provided sequences. This approach will ignore any potential assays in which a restriction motif may overlap the indel site, as the program has no information about bases on the other side of the indel. The reason dCAPS Finder 2.0 falters on indel alleles is not clear as the source code is no longer available.

Table 4.1. Number of productive primers generated for tested loci. Simulated amplicons were made using generated primers. Non-productive primers did not amplify sequences capable of being distinguished with a restriction digest. Problematic primers amplify sequences capable of being distinguished, but the reaction would likely not amplify DNA due to primer defects such as 3' mismatches or gaps in alignment to provided sequence. Productive primers are expected to successfully amplify DNA capable of being distinguished by a restriction digestion.

Gene	Locus	Target Sequence	Productive primers	Problematic primers	Non-productive primers
AHK3	AT1G27320	GGTTGAGATCAAGATAGACA	8	3	5
CENH3	AT1G01370	TCACAACCTCGGAATCAAAC	3	3	22
AGO1	AT1G48410	GAGCCTTCACCTCCTTCAGA	2	5	29
MED20	AT2G28230	GGCTGCTTACTGTTGATCCT	0	3	20
RB1	AT3G12280	CCCATTTGGTTCAATGGGCG	5	1	35

A web-based tool for design of primers to detect indels: indCAPS

A new software package, indCAPS, was developed to facilitate the design of dCAPS primers for indels. This software has also been adapted for the design of CAPS and dCAPS oligonucleotide primers used in PCR amplification of target sequences in order to screen for editing events following CRISPR/Cas9-mediated mutagenesis (Cong et al., 2013). The tool is available at http://indcaps.kieber.cloudapps.unc.edu.

The interface presents two dialog boxes to the user. The first box is for the generation of dCAPS primers for known alleles. The second box is for the generation of dCAPS primers for detecting unknown alleles. The first box requires the user to submit two sequences. No assumptions are made about either sequence being a wild-type or mutant allele, so order does not matter. Ideally, each sequence is centered on the mutation of interest. The two sequences do not need to be the same length, but should have homology arms of at least 20 bases flanking the mutation of interest. The user is also asked to submit a maximum number of mismatches in the primer. The default value is 1 mismatch. Increasing the maximum mismatch value should result

in more enzymes being reported, but as with any dCAPS assay, this may result in primers which are less likely to successfully amplify DNA.

Several advanced options are also available. Amplicon length can be specified. This parameter dictates how far downstream the tool examines the sequence when looking for exact matches for each restriction site. A restriction enzyme is rejected if there is a cleavage site in the shared downstream region, which would complicate the analysis of the diagnostic cleavage at the site of interest. Depending on the size of the submitted sequence, the amplicon length may be longer than the sequence available to the program. In this case, the entire submitted sequence downstream of the primer is considered. If the user intends to use a paired primer lying outside the sequence supplied to the program, the user should check that either no exact restriction digests are present in the region not shown to the program or any exact restriction sites will still permit discrimination between diagnostic bands when analyzed with gel electrophoresis. Primers can be chosen based on a strict primer length or by a target melting temperature. A target size may be desired if the user wishes to ensure that a sufficiently large fragment will be cleaved. This may be useful in GC rich areas where a primer designed to match a target melting temperature would be short, resulting in small shifts in band sizes after cleavage and electrophoresis. Melting temperature calculations are performed using the Nearest Neighbor method (Breslauer et al., 1986) using thermodynamic parameters published by Sugimoto et al. (Sugimoto et al., 1996), as implemented in the Oligo Calc tool (Kibbe, 2007). It is necessary to assume certain information about the primer concentration and sodium ion concentration in the PCR reaction to calculate the melting temperature. Default values have been provided, but those parameters can be modified as necessary by the user through the web interface. Also, primers which contain terminal 3’ mis-matches are rejected by default. Some researchers have reported

that certain terminal 3’ mis-matches are compatible with PCR (Newton et al., 1989; Kwok et al., 1990; Simsek and Adnan, 2000; Inoue et al., 2001), but due to inconsistencies in the literature, the default assumption for this tool is that 3’ mismatches will not amplify. If the user wishes to allow certain 3’ mismatches, the option is available. If enabled, 3’ G/T mismatches are ignored.

The second box presented by the tool to the user permits screening for mutagenesis events in a CRISPR-mediated mutagenesis experiment. The box requires the user to submit the wild-type genomic sequence the user intends to target. The sequence should contain at least twenty bases on each side of the cleavage site. The user should also include the CRISPR target site oriented 5’ to 3’, not including the PAM. The CRISPR target sequence is not required to be in the same orientation as the wild-type sequence. The tool assumes that the last base of the provided sequence immediately precedes the PAM if aligned to the wild-type sequence and that cleavage occurs at the -3 position. The mismatch max parameter behaves as it does in the first box. The final major parameter is the acceptable loss threshold, which is the percent of editing events the user is willing to miss with their screening. Lower values mean the user wants to detect more editing events. Higher values mean the user is willing to accept missing certain editing events. Missed editing events, in this context, are most likely to occur if an insertion event occurs relative to an enzyme with degenerate bases in its recognition motif. The advanced options are the same as for those in the first box.

An additional application is facilitated by the first box, the known-alleles tool. CRISPR-mediated mutagenesis events create random mutations at the target locus. It may be desired in some cases to generate an isogenic mutation in a novel biological context, such as a different ecotype or genetic background. This is especially useful in cases where multiple mutant loci must be maintained and introducing an isogenic mutation would prove easier than screening

multiple segregating loci as a result of a cross. The output of the known-alleles tool indicates which of the two supplied sequences is cleaved by the restriction digest. If the user supplies the wild-type sequence and the specific mutant sequence they wish to find and choose an assay where the mutant sequence is cleaved, then a CAPS/dCAPS assay can be used to screen a pool of CRISPR-generated mutants for a specific mutation. This could be feasibly accomplished by a two-step process, where primers generated by the unknown allele indCAPS application are used to screen for lines showing any evidence of CRISPR-mediated editing events, and then a second primer set is used to screen for a specific mutation within that population.

After analysis is complete, the user may be presented with several candidate primers. The user is required to choose their own downstream primer. Currently, the primers are evaluated only by their length or melting temperature, depending on the user's specification.. The user can choose any of the reported primers, but may prefer certain assays over others due to cost or availability of the enzyme, compatibility of the enzyme with their PCR conditions, or personal preference in enzyme choice.

Technical details of the indCAPS package

A general outline of the algorithm that was developed is illustrated in Fig 4.2. The user-supplied sequences (based on the mutagenesis target) are compared by defining shared and unshared regions in each sequence. In the case of a SNP allele, each sequence will have an unshared region of one base. For indel alleles, the sequence with the deletion relative to the other will have an unshared region of 0 bases and the sequence with the relative insertion will have an unshared region equal to the number of inserted bases.

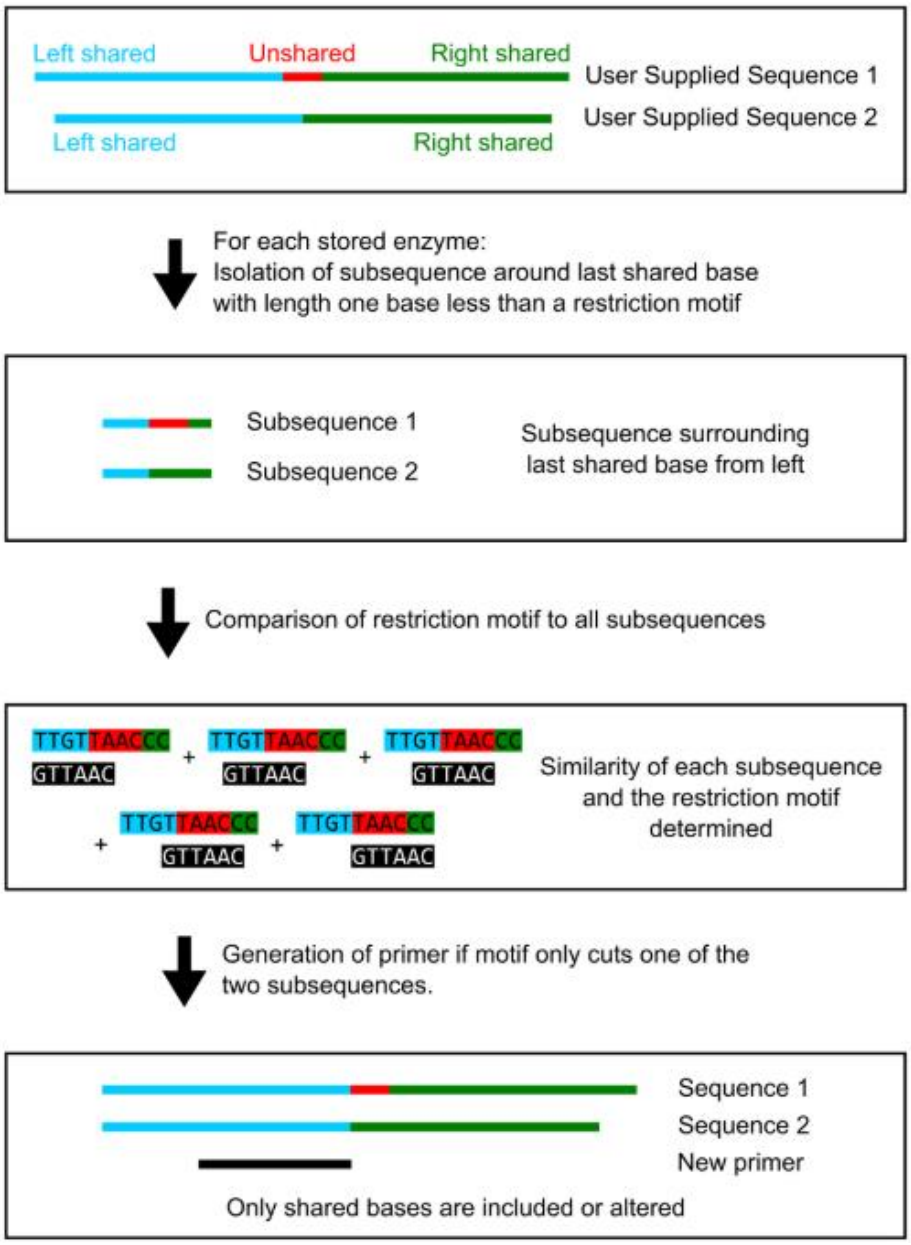

Figure 4.2. Algorithm for generation of oligonucleotide primers useful for CAPS and dCAPS assays. *Two user- supplied sequences are analyzed, with one end near the predicted mutation site. Shared and unshared regions are identified in each sequence. A sub-sequence near the last shared base from each direction is isolated and compared to a library of restriction enzyme recognition motifs (https://www.neb.com/tools-and-resources/selection-charts/alphabetized- list-of-recognition-specificities). If a diagnostic site is detected, determined by an exact or close motif match in only one sequence, a primer is generated. The primer disrupts any exact matches present in the shared regions and is checked to ensure the mismatch number is less than the specified maximum.*

Two core assumptions are made when designing diagnostic assays: 1) designed primers must be wholly contained in the shared region; and 2) putative restriction sites must have at least one base pair of overlap with both the shared and unshared regions. A library of restriction

enzyme recognition sites (https://www.neb.com/tools-and-resources/selection-charts/alphabetized-list-of-recognition-specificities) (excluding nicking or double-cutting enzymes) is compared to both shared regions and to a subsequence near the last shared base. Sequences are compared using a modified Hamming distance metric, where the number of non-similar bases is counted, allowing for degeneracy. For example, the sequences GCAT and GGTT have a distance of 2, and the sequences GCAT and GYTT have a distance of 1. Each recognition site is iterated across the sub-sequence and the distance at each position is calculated. If the distance for a comparison is below the user-specified mismatch threshold, that position is stored as a potential assay. An enzyme is rejected if it cuts in the downstream shared region, and exact matches in the primer are disrupted with mismatches during the primer design stage. Each set of sequences is analyzed twice, once as supplied, and again using the reverse complement of each sequence. An enzyme that is rejected because it cuts in the shared, downstream region of both sequences from one direction may be suitable if the primer is aligned to the reverse complement of the two sequences. For efficiency purposes, primers are designed only if a potential CAPS/dCAPS assay is detected in comparisons of the input sequences; the number of comparisons made is reduced if primers are designed for specific assays rather than designing all possible primers and assessing them for compatibility with CAPS/dCAPS assays. Primers are designed with their 3’ end at the last shared base of the two sequences. The primer length is then extended in the 5’ direction to generate candidate output primers. Each candidate primer is assessed by its length or melting temperature, according to the user’s choice, and the primer best matching the chosen criterion is reported.

For the purpose of CRISPR/Cas9 screening, a profile of possible editing events is used to simulate editing events in the wild-type sequence. Currently, the default editing events are single

base pair insertion and deletion events. Future versions of the software will allow the user to supply a custom profile of events. The user is required to provide the specific target sequence used in CRISPR/Cas9 mutagenesis and the cut site is assumed to be at the -3 position of the provided 20-bp target (Cong et al., 2013). All possible sequence variants are created and compared to the wild-type sequence to identify the last shared base. The last shared base is taken to be the last base in the wild-type sequence shared with all sequence variants. A simplifying assumption is made when designing primers for unknown alleles: cleavage will occur in the wild-type sequence and will be disrupted in mutant sequences.

Use of indCAPS to find mutants in cytokinin signaling

Cytokinins, a class of adenine-derived signaling molecules are involved in regulating a diverse set of biological processes. Cytokinins are perceived by Arabidopsis Histidine Kinase (AHK) proteins (Inoue et al., 2001; Yamada et al., 2001) in the endoplasmic reticulum (Caesar et al., 2011; Wulfetange et al., 2011), which undergo autophosphorylation on a His residue. This phosphate is ultimately transferred to either type-A or type-B Response Regulator proteins (ARRs) via the Histidine Phosphotransfer (AHP) proteins (Hutchison et al., 2006; Punwani et al., 2010; Punwani and Kieber, 2010). Type-B ARRs are transcription factors activated by phosphorylation (Mason et al., 2005). Type-A ARRs lack a DNA-binding domain, are cytokinin-inducible, and negatively regulate cytokinin signaling (To et al., 2004; To et al., 2007).

The cytokinin AHK receptors are encoded by *AHK2*, *AHK3*, and *AHK4/CRE1* in Arabidopsis. Multiple mutant lines with various loss-of-function T-DNA insertion alleles of each gene have been identified. The most severely affected triple mutant line, *ahk2-7 ahk3-3 cre1-12*, harbors null alleles for *ahk2* and *ahk4*, but still contains residual full-length wild-type transcript for *AHK3* (Cheng et al., 2013). We sought to identify a CRISPR-induced null allele of *ahk3* in an

ahk2 ahk4 background by introducing a frameshift mutation in *AHK3*. This would reveal the effect of complete disruption of the AHK cytokinin receptors in *Arabidopsis*.

The indCAPS tool was tested by designing primers for a CRISPR mutagenesis experiment targeting the *AHK3* gene (Fig 4.3). A CRISPR/Cas9 target was designed to target the first exon of *AHK3* using the CRISPR-Plant resource (Xie et al., 2014). The target site chosen is before the Cyclases/Histidine kinases Associated Sensory Extracellular (CHASE) domain, the cytokinin binding domain of the AHKs. The AHK3 targeting plasmid, pCH59, was stably transformed into an *ahk2-7 cre1-12* mutant line (Inoue et al., 2001; Yamada et al., 2001; Cheng and Kieber, 2013), referred to as *ahk2,4* hereafter. The pCH59 vector was constructed by cloning a commercially synthesized gRNA fragment into a pCUT binary vector system expressing plant codon optimized Cas9 (Peterson et al., 2016). T1 transformed seedlings were identified by hygromycin selection and grown to maturity. The T2 progeny were screened for *AHK3* editing events using primers generated by indCAPS. Viable seedlings with homozygous, single base pair insertions causing frameshift mutations disrupting the *AHK3* coding region were identified. These alleles, denoted *ahk3-9* and *ahk3-10*, are single base pair insertions of A and C, respectively. The frameshift produces an early stop codon 25 residues after the edit location. The resulting predicted protein retains two transmembrane domains, but no functional CHASE domain or cytosolic histidine kinase or receiver domains. These triple *cre1-12 ahk2-7 ahk3* mutants were viable and resembled the *cre1-12 ahk2-7 ahk3-3* (Argyros et al., 2008). These results demonstrate that the complete disruption of all three AHK cytokinin receptors does not result in embryo lethality. As these are the only CHASE-domain containing proteins in *Arabidopsis*, this suggests that either cytokinin is not essential for early development, or that there are other as yet unidentified cytokinin receptors.

In addition to analysis of CRISPR-induced *ahk4* alleles, indCAPS tool has been successfully use by multiple other members of the authors' research groups to successfully target at least five additional genes.

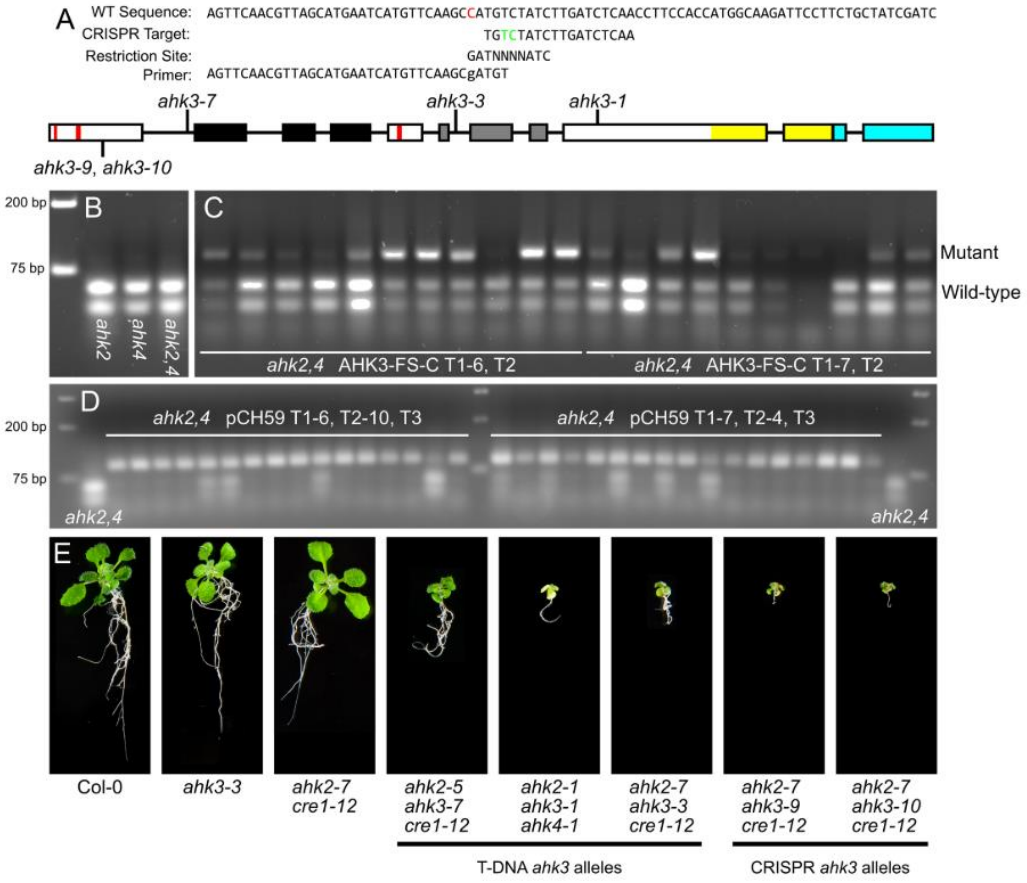

Figure 4.3. Homozygous editing events in AHK3were identified. *(A) The indCAPS package was used to generate a primer recognizing a Bsa BI site spanning the CRISPR cut site (between the green bases). A single mismatch was required in the primer (indicated in red). The genomic locus for AHK3 is shown. Boxes indicate exons, red bars—transmembrane domains, black region—CHASE domain, grey region—histidine kinase domain, yellow region—receiver domain, blue region– 3' UTR. Locations of T-DNA insertion sites (*ahk3-1, ahk3-3, *and* ahk3-7*) and targeted editing site are indicated. (B,C) The assay was used to screen A. thaliana plants stably transformed with a pCUT binary vector system (Peterson et al., 2016) and sgRNA constructs targeting* AHK3. *(B) Wild-type controls at edit location. (C) T2 plants from two representative independent transformation events are shown. The uncut amplicon is 90 bp and the wild-type allele is cleaved to produce 36 bp and 54 bp fragments. (D) Progeny from two T2 plants heterozygous for editing events were selected and analyzed for editing. (E) A. thaliana seedlings imaged at 2.5 weeks of growth. Shown are Col-0;* ahk3-3*;* ahk2,4*;* ahk2-5 ahk3-7 cre1-12; ahk2-1/+ ahk3-1 ahk4-1; ahk2-7 ahk3-3/+ cre1-12; ahk2-7 ahk3-9 cre1-12; ahk2-7 ahk3-10 cre1-12. *All plants at same scale.*

Conclusions

The indCAPS package provides a useful tool for researchers using CRISPR-mediated mutagenesis as it facilitates the screening of individuals in which editing of the target has occurred. It also provides replacement for existing tools for the design of primers for dCAPS analysis capable of distinguishing known indel alleles. We employed this tool to successfully design diagnostic primers to identify CRISPR-induced *ahk3* null alleles, the subsequent analysis of which showed that the cytokinin AHK receptors are not essential for embryo development.

Author Contributions

Conceptualization: Charles Hodgens, Zachary L. Nimchuk, Joseph J. Kieber.

Funding acquisition: Joseph J. Kieber.

Investigation: Charles Hodgens.

Methodology: Charles Hodgens

Project administration: Zachary L. Nimchuk, Joseph J. Kieber.

Software: Charles Hodgens.

Writing – original draft: Charles Hodgens, Zachary L. Nimchuk, Joseph J. Kieber.

Writing – review & editing: Charles Hodgens, Zachary L. Nimchuk, Joseph J. Kieber.

We would like to thank G. Eric Schaller, Brian Nalley, Christian Burr, Amala John, Joanna Polko, Lauren Ross, and Ariel Aldrette for helping with the visual design ofthe website and the figures for this paper and Jamie Winshell for technical support. This work was supported by grants from the National Science Foundation (IOS-1238051 and IOS-1455607).

CHAPTER 5: CONCLUSION

The scientist as a craftsman

If you are reading this work and made it to this conclusion, you are almost certainly a scientist. And if you are reading a biologist's dissertation, you are likely yourself a biologist, or may know a few. For a moment, please pretend you are not. Imagine that you are a child, or an adult in nearly any other profession, and you are on a tour of a laboratory. Your guide walks you past some desks, some computers, some objects that your guide calls microscopes but that look nothing like the tools you used in your science classes, and into an open area full of counters. The counters are covered with the worst mess you have seen in your life, and you feel sorry for anyone who must use them. Your guide refers to these counters as "benches." This name strikes you as odd. These are waist-high counters with shelves and drawers, not a chair you might find at a park or bus stop. And after a moment, the epiphany: your guide means a workbench, not a sitting bench.

I bring this to your attention, reader, because I believe that the workbench is a perfect simile for the contents of this dissertation (also because it's a moment of confusion even I have from time to time, and I believe the occasional brain fart is part of what makes us human). Some time ago, I came across the Youtube channel of a woodworker named Rex Krueger. One of the first videos I watched dealt with his fascination with English joiners' benches. Those joiners' benches were sturdy, purpose-built tools crafted in every detail to aid the work of the joiner. He proceeded to build one himself, describing his thought process at each step and how he used the

scrap wood from the project to further enhance the bench. He had a goal. He attempted to reach it. He ran into problems, looked at his available resources, thought carefully, and addressed the problems. His bench was a scientific bench, even though it was made of 2x8 lumber.

Similarly, this dissertation is a bench. I did not begin my graduate career with the intention of developing tools for other scientists. My goal was to discover new facts about hormone signaling in plants. In the pursuit of that goal, I discovered problems – problems of time, of complexity, and of absent tools. Giving up was not an option, so I looked at my available resources and found a way forward. I solved my problems and incorporated the solutions into my work. After many years, what I am left with is a joiner's bench built of words rather than wood.

I discovered new facts about cytokinin signaling, and in that, I accomplished my original goal. In addition, I made improvements to my own toolset that I hope will be useful to many other researchers around the world. In the modern era, CRISPR/Cas9-derived mutant alleles are common, but there were problems with the screening tools that I and other plant molecular biologists used. PCR-based restriction digest genotyping is a quick and easy technique that should be part of every molecular biologist's basic tool kit, but the existing tools were not compatible with insert/deletion alleles without difficult workarounds or manual primer creation. A member of my committee, Zack Nimchuk, was aware that I possessed experience in programming and suggested that I update the tools. After a lot of trial and error, I developed an algorithm that would work and implemented both the code and web portal to make a tool that any scientist could use. There are many ways to genotype an organism or screen for new mutations, but the ease with which PCR can be parallelized makes it an attractive method. I hope that future scientists will find that the indCAPS tool makes their lives easier.

Similarly, mutant screens are nothing new. Screens are a foundational tool for a geneticist, but it can take tremendous time and effort to characterize a mutant obtained from a screen. I performed the *eah* screen in order to identify genes with an unknown role in cytokinin signaling. I came out with one gene with a little recognized role in cytokinin signaling (*HY5*) and several excellent candidates for further study that I hope will yield fertile ground for future scientists. In the process, I used high throughput sequencing and my knowledge of bioinformatics to develop strategies to identify the low-hanging fruit in the experiment and accelerate the pace of discovery of novel mutations, and I implemented those strategies in a set of scripts and pipelines that any end-user should be able to use to gain new insights about their screen.

In fact, mutagenomics has already yielded results for other scientists. The *acre* screen, performed within our own research group, had an *lht1* mutation confidently implicated as causative for several of the mutant lines in the experiment. The *gob* screen does not yet have final confirmed causative mutations, but the sheer number of sibling groups in the *gob* experiment should keep researchers busy for years to come and hopefully will provide many new avenues of research to the Nimchuk group.

Additional avenues of research for the work described in this thesis remain. For example, we have shown that *HY5* is involved in cytokinin signaling through genetic means, but the mechanism of interaction is unknown. HY5 proteins have binding sites near many of the same genes as type-B ARRs, but whether this is the result of HY5 and ARRs sharing a common set of target genes for regulation or whether this represents interaction as a dimer or members of a large protein complex is unknown. Additional experiments are required to explore further whether HY5 and any of the ARRs interact. Some negative data has been collected regarding ARR and

HY5 interaction via a bimolecular fluoresence complementation assay, but other experiments such as a co-immunoprecipitation or a fluorescence resonance energy transfer assay could be attempted. Complementation tests have only been performed for five of the lines observed to have *ahk4* mutations; the other putative *ahk4* lines should be crossed to an *ahk4* T-DNA to confirm the causative nature of the *ahk4* mutations.

There are also wholly uncharacterized lines. At least one sibling set and several promising independent lines have SNP densities low enough they can be manually screened for plausible candidate mutations, but several other lines remain uncharacterized and have high SNP densities. These lines should be pursued with both mutagenomic and traditional mapping approaches. If additional siblings for these lines can be identified, then they should be sequenced, and in the meantime while the search for siblings is underway, they should be back-crossed to *ahp2,3* plants for a mapping experiment.

Uncharacterized mutants in the *acre* and *gob* screens remain as well. The *acre* screen is likely to require traditional mapping-by-sequencing efforts, but if siblings can be identified for any of the *acre* lines, mutagenomics may be helpful. Also, many of the *gob* sibling sets and independent lines are amenable to manual screening for candidate mutations. It is my hope that the lines identified by mutagenomics in the *gob* screen are fruitful areas of research for future scientists.

Finally, the current software pipeline for mutagenomics is implemented as a set of scripts and programs the end user must download to their computer and run from the command line. I have provided written instructions, but it would be useful in the future to create a website where users could upload their files of mutant SNPs and receive network files automatically. This is likely to require funding source for a website to host it. While the computational work in the

mutagenomics pipeline is light enough that a laptop can perform it it is not an immediate or computationally cheap process and free hosting services may not be sufficient to run the pipeline.

What I am left after many years is a combination of new knowledge and new techniques. It may not be the prettiest collection of ideas, but it fits together snugly. The new facts are inextricably intertwined with the new methods, and to discuss one without the other would be to tell an incomplete story. My hope is that this work will continue to support the work of other scientists for years to come.

APPENDIX 1

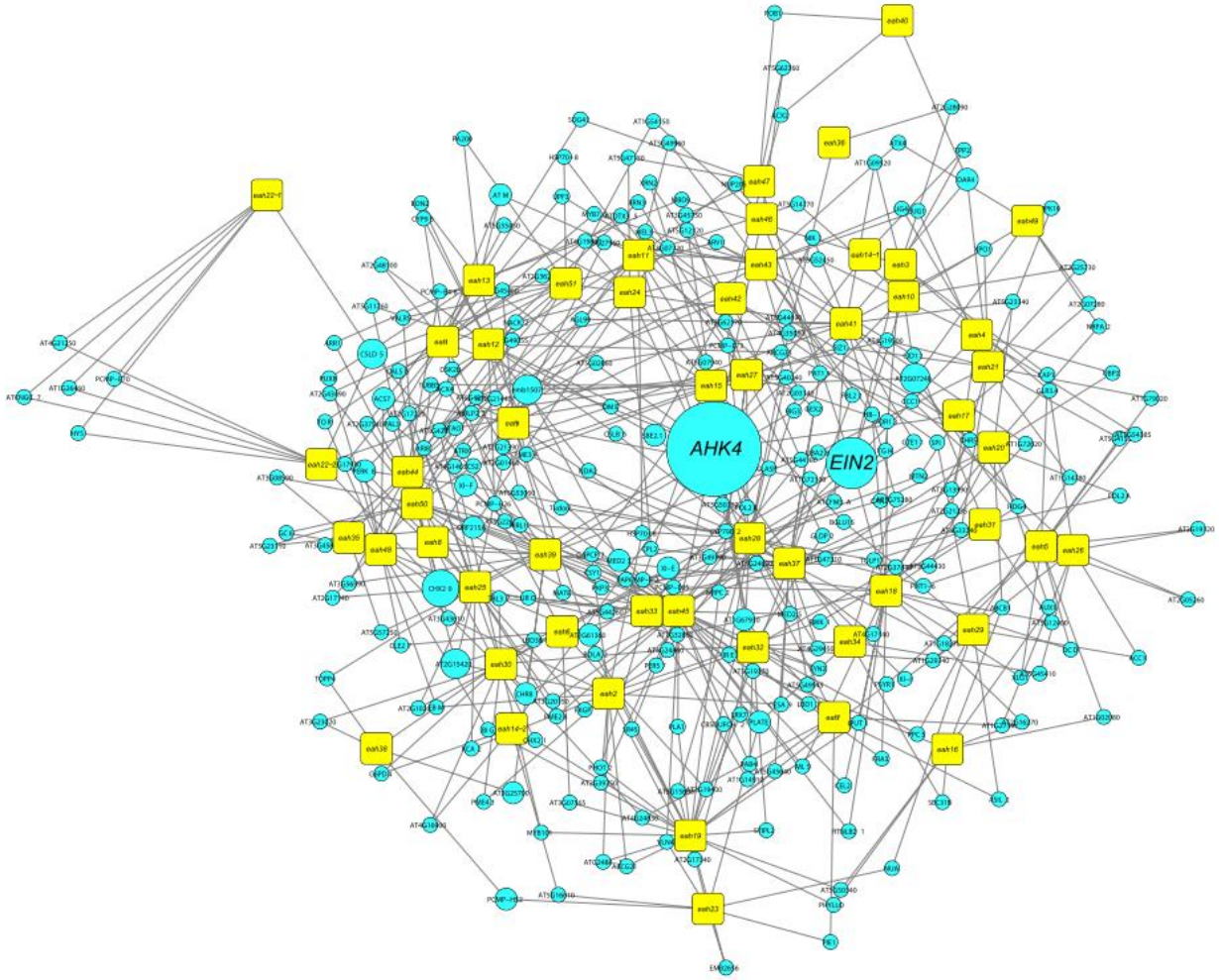

Figure A1.1. Full network diagram of PA-SNPs. *Yellow boxes represent individual mutant lines. Blue circles represent genes. Gene objects are scaled proportionally to the number of mutations observed in each gene. Lines are drawn from boxes to circles if a given mutant line harbored a PA-SNP in that gene. Blue circles are scaled by how many mutations in that gene were observed.*

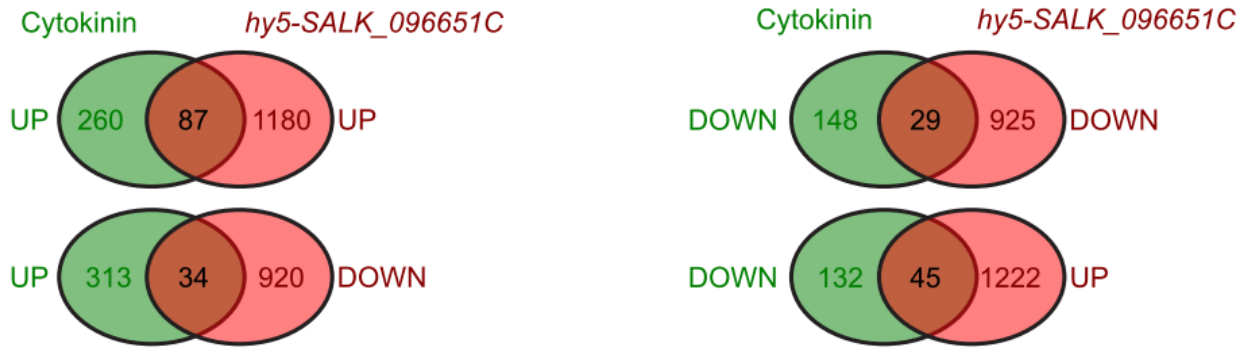

Figure A1.2. Comparison of differentially expressed genes between BA and control-treated wild type plants (green) and hy5 control plants and wild type control plants (red). *Differentially expressed genes are at least 0.5-fold different (on a log2 scale) with BH-adjusted p-value < 0.02. All overlaps are significant by a hypergeometric test. P-values are 8.1e-40 for the BA-up vs hy5-up comparison, 5.47e-8 for BA-up vs hy5-down, 3.1e-12 for BA-down vs hy5-down, and 1.97e-21 for BA-down vs hy5-up.*

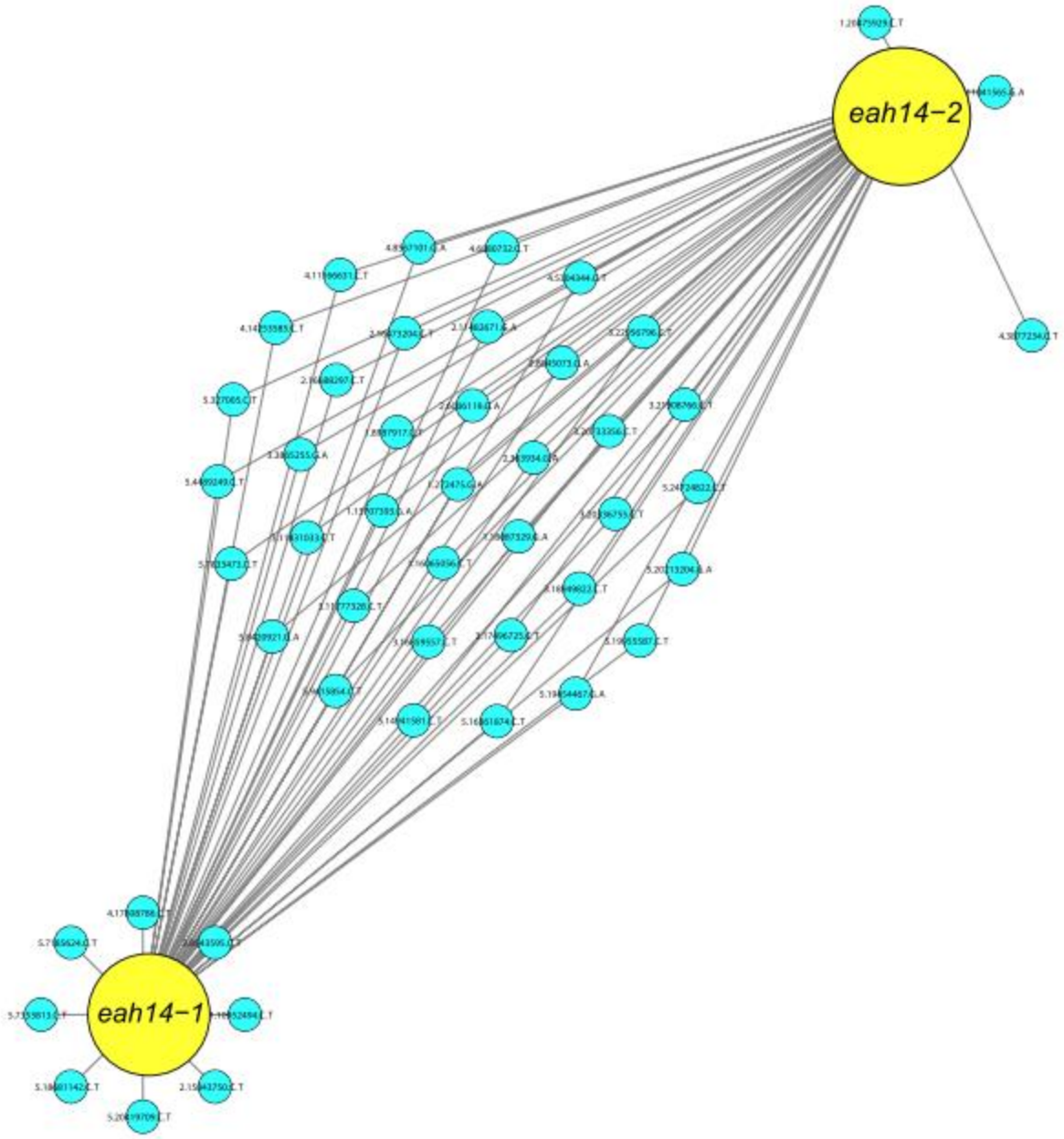

Figure A1.3. Exact shared SNPs between eah14-1 and eah14-2. *Although these siblings (yellow circles) share only one homozygous PA-SNP, examination of the shared EMS-derived SNPs (blue circles) confirms that they are sibling lines. Each circle represents a specific combination of chromosome, position, and mutation.*

REFERENCES

Abràmoff M (2004) Image processing with ImageJ. Biophotonics Int.

Adam Z, Clarke AK (2002) Cutting edge of chloroplast proteolysis. Trends Plant Sci **7**: 451–456

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science (80-) **301**: 653–657

Anantharaman V, Aravind L (2001) The CHASE domain: A predicted ligand-binding module in plant cytokinin receptors and other eukaryotic and bacterial receptors. Trends Biochem Sci **26**: 579–582

Appleby JL, Parkinson JS, Bourret RB (1996) Signal Transduction via the Multi-Step Phosphorelay: Not Necessarily a Road Less Traveled. Cell **86**: 845–848

Arenas-M. A, Zehrmann A, Moreno S, Takenaka M, Jordana X (2014) The pentatricopeptide repeat protein MEF26 participates in RNA editing in mitochondrial *cox3* and *nad4* transcripts. Mitochondrion **19**: 126–134

Argueso C, Raines T, Kieber J (2010) Cytokinin signaling and transcriptional networks. Curr Opin Plant Biol **13**: 533–539

Argyros RD, Mathews DE, Chiang Y, Palmer CM, Thibault DM, Etheridge N, Argyros DA, Mason MG, Kieber JJ, Schaller GE (2008) Type B Response Regulators of *Arabidopsis* Play Key Roles in Cytokinin Signaling and Plant Development. Plant Cell **20**: 2102–2116

Austin RS, Vidaurre D, Stamatiou G, Breit R, Provart NJ, Bonetta D, Zhang J, Fung P, Gong Y, Wang PW, et al (2011) Next-generation mapping of Arabidopsis genes. Plant J **67**: 715–725

Ball E (1946) Development in Sterile Culture of Stem Tips and Subjacent Regions of Tropaeolum Majus L. And of Lupinus Albus L. Am J Bot **33**: 301–318

Bauer D, Viczian A, Kircher S, Nobis T, Nitschke R, Kunkel T, Panigrahi KCS, Adam E, Fejes E, Schafer E, et al (2004) Constitutive Photomorphogenesis 1 and Multiple Photoreceptors Control Degradation of Phytochrome Interacting Factor 3, a Transcription Factor Required for Light Signaling in *Arabidopsis*. Plant Cell **16**: 1433–1445

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. J R Stat Soc **57**: 289–300

Bhargava A, Clabaugh I, To JP, Maxwell BB, Chiang Y-H, Schaller GE, Loraine A, Kieber JJ (2013) Identification of cytokinin-responsive genes using microarray meta-analysis and RNA-Seq in Arabidopsis. Plant Physiol **162**: 272–94

Bishopp A, Help H, El-Showk S, Weijers D, Scheres B, Friml J, Benková E, Mähönen AP, Helariutta Y (2011) A Mutually Inhibitory Interaction between Auxin and Cytokinin Specifies Vascular Pattern in Roots. Curr Biol **21**: 917–926

Brandstatter I, Kieber JJ (1998) Two genes with Similarity to Bacterial Response Regulators are Rapidly and Specifically Induced by Cytokinin in Arabidopsis. Plant Cell **10**: 1009–1019

Breslauer KJ, Frank R, Blöcker H, Marky LA (1986) Predicting DNA duplex stability from the base sequence. Proc Natl Acad Sci U S A **83**: 3746–50

Brzobohatý B, Moore I, Kristoffersen P, Bako L, Campos N, Schell J, Palme K (1993) Release of Active Cytokinin by a β-Glucosidase Localized to the Maize Root Meristem. Science (80-) **262**: 1051–1054

Caesar K, Thamm AMK, Witthöft J, Elgass K, Huppenberger P, Grefen C, Horak J, Harter K (2011) Evidence for the localization of the Arabidopsis cytokinin receptors AHK3 and AHK4 in the endoplasmic reticulum. J Exp Bot **62**: 5571–5580

Candela H, Hake S (2008) The art and design of genetic screens: Maize. Nat Rev Genet **9**: 192–203

Caplin SM, Steward FC (1948) Effect of Coconut Milk on the Growth of Explants From Carrot Root. Science (80-) **108**: 655–657

Chen L, Bush DR (1997) LHT1, a lysine- and histidine-specific amino acid transporter in Arabidopsis. Plant Physiol **115**: 1127–1134

Cheng CY, Kieber JJ (2013) The role of cytokinin in ovule development in Arabidopsis. Plant Signal Behav **8**: 1–4

Cheng CY, Mathews DE, Schaller GE, Kieber JJ (2013) Cytokinin-dependent specification of the functional megaspore in the Arabidopsis female gametophyte. Plant J **73**: 929–940

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; *iso-2*; *iso-3*. Fly (Austin) **6**: 80–92

Clough SJ, Bent a F (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J **16**: 735–43

Cluis CP, Mouchel CF, Hardtke CS (2004) The Arabidopsis transcription factor HY5 integrates light and hormone signaling pathways. Plant J **38**: 332–347

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al (2013) Multiplex Genome Engineering Using CRISPR/Cas Systems. Science (80-) **339**: 819–823

Constan D, Froehlich JE, Rangarajan S, Keegstra K (2004) A Stromal Hsp100 Protein Is Required for Normal Chloroplast Development and Function in Arabidopsis. Plant Physiol **136**: 3605–3615

D'Agostino IB, Deuere J, Kieber JJ (2000) Characterization of the Response of the Arabidopsis Response Regulator Gene Family to Cytokinin. Plant Physiol **124**: 1706–1717

D'Agostino IB, Kieber JJ (1999a) Phosphorelay signal transduction: The emerging family of plant response regulators. Trends Biochem Sci **24**: 452–456

D'Agostino IB, Kieber JJ (1999b) Molecular mechanisms of cytokinin action. Curr Opin Plant Biol **2**: 359–364

Deng Y, Dong H, Mu J, Ren B, Zheng B, Ji Z, Yang WC, Liang Y, Zuo J (2010) *Arabidopsis* Histidine Kinase CKI1 Acts Upstream of HISTIDINE PHOSPHOTRANSFER PROTEINS to Regulate Female Gametophyte Development and Vegetative Growth. Plant Cell **22**: 1232–1248

Dortay H, Mehnert N, Bürkle L, Schmülling T, Heyl A (2006) Analysis of protein interactions within the cytokinin-signaling pathway of *Arabidopsis thaliana*. FEBS J **273**: 4631–4644

Doyle JJ, Doyle JL (1987) A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue. Phytochem Bull **19**: 11–15

Drerup MM, Schlücking K, Hashimoto K, Manishankar P, Steinhorst L, Kuchitsu K, Kudla J (2013) The Calcineurin B-Like Calcium Sensors CBL1 and CBL9 Together with Their Interacting Protein Kinase CIPK26 Regulate the *Arabidopsis* NADPH Oxidase RBOHF. Mol Plant **6**: 559–569

Dyson WH, Hall RH (1972) N6-(d2-Isopentenyl)adenosine: Its Occurrence as a Free Nucleoside in an Autonomous Strain of Tobacco Tissue. Plant Physiol **50**: 616–621

Fahmy OG, Fahmy MJ (1971) Mutability at Specific Euchromatic and Heterochromatic Loci with Alkylating and Nitroso Compounds in *Drosophia Melanogaster*. Mutat Res **13**: 19–34

Farboud B (2017) Targeted genome editing in *Caenorhabditis elegans* using CRISPR/Cas9. Wiley Interdiscip Rev Dev Biol **6**: 14–16

Fekih R, Takagi H, Tamiru M, Abe A, Natsume S, Yaegashi H, Sharma S, Sharma S, Kanzaki H, Matsumura H, et al (2013) MutMap+: Genetic Mapping and Mutant Identification without Crossing in Rice. PLoS One **8**: 1–10

Feng J, Wang C, Chen Q, Chen H, Ren B, Li X, Zuo J (2013) S-nitrosylation of phosphotransfer proteins represses cytokinin signaling. Nat Commun **4**: 1529

Forsburg SL (2002) The art and design of genetic screens: Yeast. Nat Rev Genet **3**: 356–369

Gagne JM, Clark SE (2010) The Arabidopsis stem cell factor POLTERGEIST Is membrane localized and phospholipid stimulated. Plant Cell **22**: 729–743

Gallagher KL, Koizumi K (2013) Identification of SHRUBBY, a SHORT-ROOT and SCARECROW interacting protein that controls root growth and radial patterning. Dev **140**: 1292–1300

Gelbart WM, Rindone WP, Chillemi J, Russo S, Crosby M, Mathews B, Ashburner M, Drysdale RA, De Grey A, Whitfield EJ, et al (1996) FlyBase: The *Drosophila* database. Nucleic Acids Res **24**: 53–56

Gingerich DJ, Gagne JM, Salter DW, Hellmann H, Estelle M, Ma L, Vierstra RD (2005) Cullins 3a and 3b Assemble with Members of the Broad Complex/Tramtrack/Bric-a-Brac (BTB) Protein Family to Form Essential Ubiquitin-Protein Ligases (E3s) in *Arabidopsis*. J Biol Chem **280**: 18810–18821

Gouget A, Senchou V, Govers F, Sanson A, Barre A, Rougé P, Pont-Lezica R, Canut H (2006) Lectin Receptor Kinases Participate in Protein-Protein Interactions to Mediate Plasma Membrane-Cell Wall Adhesions in Arabidopsis. Plant Physiol **140**: 81–90

Guzman P, Ecker JR (1990) Exploiting the Triple Response of *Arabidopsis* to Identify Ethylene-Related Mutants. Plant Cell **2**: 513–523

Hall RH (1873) Cytokinins as a Probe of Developmental Processes. Annu Rev Plant Physiol **24**: 415–444

Hamzi HQ, Skoog F (1963) KINETIN-LIKE GROWTH-PROMOTING ACTIVITY OF 1-SUBSTITUTED ADENINES [1-BENZYL-6-AMINOPURINE AND 1-(GAMMA,GAMMA-DIMETHYLALLYL)-6-AMINOPURINE]. Proc Natl Acad Sci U S A **51**: 76–83

Hashimoto M, Neriya Y, Keima T, Iwabuchi N, Koinuma H, Hagiwara-Komoda Y, Ishikawa K, Himeno M, Maejima K, Yamaji Y, et al (2016) EXA1, a GYF domain protein, is responsible for loss-of-susceptibility to plantago asiatica mosaic virus in *Arabidopsis thaliana*. Plant J **88**: 120–131

Hass C, Lohrmann J, Albrecht V, Sweere U, Hummel F, Yoo SD, Hwang I, Zhu T, Schäfer E, Kudla J, et al (2004) The response regulator 2 mediates ethylene signalling and hormone signal integration in Arabidopsis. EMBO J **23**: 3290–3302

Hejátko J, Ryu H, Kim GT, Dobešová R, Choi S, Choi SM, Souček P, Horák J, Pekárová B, Palme K, et al (2009) The Histidine kinases CYTOKININ-INDEPENDENT1 and ARABIDOPSIS HISTIDINE KINASE2 and 3 Regulate Vascular Tissue Development in *Arabidopsis* Shoots. Plant Cell **21**: 2008–2021

Hengge R, Bukau B (2003) Proteolysis in prokaryotes: protein quality control and regulatory principles. Mol Microbiol **49**: 1451–1462

Heyl A, Wulfetange K, Pils B, Nielsen N, Romanov GA, Schmülling T (2007) Evolutionary proteomics identifies amino acids essential for ligand-binding of the cytokinin receptor CHASE domain. BMC Evol Biol **7**: 1–8

Higuchi M, Pischke MS, Mähönen AP, Miyawaki K, Hashimoto Y, Seki M, Kobayashi M, Shinozaki K, Kato T, Tabata S, et al (2004) *In planta* functions of the *Arabidopsis* cytokinin receptor family. Proc Natl Acad Sci U S A **101**: 8821–8826

Hirner A, Ladwig F, Stransky H, Okumoto S, Keinath M, Harms A, Frommer WB, Koch W (2006) Arabidopsis LHT1 is a High-Affinity Transporter for Cellular Amino Acid Uptake in Both Root Epidermis and Leaf Mesophyll. Plant Cell **18**: 1931–1946

Hodgens C, Nimchuk ZL, Kieber JJ (2017) indCAPS: A tool for designing screening primers for CRISPR/Cas9 mutagenesis events. PLoS One 1–11

Horák J, Grefen C, Berendzen KW, Hahn A, Stierhof YD, Stadelhofer B, Stahl M, Koncz C, Harter K (2008) The *Arabidopsis thaliana* response regulator ARR22 is a putative AHP phospho-histidine phosphatase expressed in the chalaza of developing seeds. BMC Plant Biol **8**: 1–18

Hosoda K, Imamura A, Katoh E, Hatta T, Tachiki M, Yamada H, Mizuno T, Yamazaki T (2002) Molecular Structure of the GARP Family of Plant Myb-Related DNA Binding Motifs of the Arabidopsis Response Regulators. Plant Cell **14**: 2015–2029

Hothorn M, Dabi T, Chory J (2011) Structural basis for cytokinin recognition by *Arabidopsis thaliana* histidine kinase 4. Nat Chem Biol **7**: 766–768

Hou B, Lim EK, Higgins GS, Bowles DJ (2004) N-glucosylation of cytokinins by glycosyltransferases of Arabidopsis thaliana. J Biol Chem **279**: 47822–47832

Housden BE, Perrimon N (2016) Cas9-mediated Genome Engineering in *Drosophila melanogaster*. Cold Spring Harb Protoc **2016**: 747–750

Hutchison CE, Kieber JJ (2002) Cytokinin Signaling in Arabidopsis. Plant Cell 269–289

Hutchison CE, Li J, Argueso C, Gonzalez M, Lee E, Lewis MW, Maxwell BB, Perdue TD, Schaller GE, Alonso JM, et al (2006) The Arabidopsis Histidine Phosphotransfer Proteins Are Redundant Positive Regulators of Cytokinin Signaling. Plant Cell **18**: 3073–87

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

Hwang I, Sheen J, Müller B (2012) Cytokinin signaling networks. Annu Rev Plant Biol **63**: 353–80

Imamura A, Hanaki N, Nakamura A, Suzuki T, Taniguchi M, Ueguchi C, Sugiyama T, Mizuno T (1999) Compilation and Characterization of *Arabiopsis thaliana* Response Regulators Implicated in His-Asp Phosphorelay Signal Transduction. Plant Cell Physiol **40**: 733–742

Imamura A, Hanaki N, Umeda H, Nakamura A, Suzuki T, Ueguchi C, Mizuno T (1998) Response regulators implicated in His-to-Asp phosphotransfer signaling in Arabidopsis. Proc Natl Acad Sci U S A **95**: 2691–6

Imamura A, Kiba T, Tajima Y, Yamashino T, Mizuno T (2003) In Vivo and In Vitro Characterization of the ARR11 Response Regulator Implicated in the His-to-Asp Phosphorelay Signal Transduction in *Arabidopsis thaliana*. Plant Cell Physiol **44**: 122–131

Imamura A, Yoshino Y, Mizuno T (2001) Cellular Localization of the Signaling Components of *Arabidopsis* His-to-Asp Phosphorelay. Biosci Biotechnol Biochem **65**: 2113–2117

Initiative TAG (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408**: 796–815

Inoue T, Higuchi M, Hashimoto Y, Seki M, Kobayashi M, Kato T, Tabata S, Shinozaki K, Kakimoto T (2001) Identification of CRE1 as a cytokinin receptor from *Arabidopsis*. Nature **409**: 1060–1063

Ishida K, Yamashino T, Yokoyama A, Mizuno T (2008) Three Type-B Response Regulators, ARR1, ARR10 and ARR12, Play Essential but Redundant Roles in Cytokinin Signal Transduction Throughout the Life Cycle of *Arabidopsis thaliana*. Plant Cell Physiol **49**: 47–57

Jayson AP, Claire EH, Schaller GE, Joseph JK (2010) The subcellular distribution of the Arabidopsis histidine phosphotransfer proteins is independent of cytokinin signaling. Plant J **62**: 473–482

Jenal U, Galperin MY (2009) Single domain response regulators: molecular switches with emerging roles in cell organization and dynamics. Curr Opin Microbiol **12**: 152–160

Kakimoto T (1996) CKI1, a Histidine Kinase Homolog Implicated in Cytokinin Signal Transduction. Science (80-) **274**: 982–985

Kang NY, Cho C, Kim J (2013) Inducible Expression of *Arabidopsis Response Regulator 22* (*ARR22*), a Type-C *ARR*, in Transgenic *Arabidopsis* Enhances Drought and Freezing Tolerance. PLoS One. doi: 10.1371/journal.pone.0079248

Karkute SG, Singh AK, Gupta OP, Singh PM, Singh B (2017) CRISPR/Cas9 mediated genome engineering for improvement of horticultural crops. Front Plant Sci **8**: 1–6

Kende H (1993) Ethylene biosynthesis. Annu Rev Plant Physiol Plant Mol Biol **44**: 283–307

Kiba T, Aoki K, Sakakibara H, Mizuno T (2004) *Arabidopsis* Response Regulator, ARR22, Ectopic Expression of Which Results in Phenotypes Similar to the *wol* Cytokinin-Receptor Mutant. Plant Cell Physiol **45**: 1063–1077

Kiba T, Yamada H, Sato S, Kato T, Tabata S, Yamashino T, Mizuno T (2003) The type-A response regulator, ARR15, acts as a negative regulator in the cytokinin-mediated signal transduction in *Arabidopsis thaliana*. Plant Cell Physiol **44**: 868–874

Kibbe WA (2007) OligoCalc: An online oligonucleotide properties calculator. Nucleic Acids Res **35**: 43–46

Kieber JJ, Schaller GE (2014) Cytokinins. Arab. B. pp 1–36

Kim HJ, Chiang YH, Kieber JJ, Schaller GE (2013a) SCF(KMD) controls cytokinin signaling by regulating the degradation of type-B response regulators. Proc Natl Acad Sci U S A **110**: 10028–10033

Kim HJ, Kieber JJ, Schaller GE (2013b) The rice F-box protein KISS ME DEADLY2 functions as a negative regulator of cytokinin signaling. Plant Signal Behav **8**: 10–12

Klepikova A V., Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA (2016) A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. Plant J **88**: 1058–1070

Kowalska M, Galuszka P, Frébortová J, Šebela M, Béres T, Hluska T, Šmehilová M, Bilyeu KD, Frébort I (2010) Vacuolar and cytosolic cytokinin dehydrogenases of *Arabidopsis thaliana*: Heterologous expression, purification and properties. Phytochemistry **71**: 1970–1978

Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a Hidden Markov Model: Application to Complete Genomes. J Mol Biol **305**: 567–580

Kudo T, Kiba T, Sakakibara H (2010) Metabolism and long-distance translocation of cytokinins. J Integr Plant Biol **52**: 53–60

Kurakawa T, Ueda N, Maekawa M, Kobayashi K, Kojima M, Nagato Y, Sakakibara H, Kyozuka J (2007) Direct control of shoot meristem activity by a cytokinin-activating enzyme. Nature **445**: 652–655

Kurepa J, Li Y, Perry SE, Smalle JA (2014) Ectopic expression of the phosphomimic mutant version of Arabidopsis response regulator 1 promotes a constitutive cytokinin response phenotype. BMC Plant Biol. doi: 10.1186/1471-2229-14-28

Kurepa J, Shull TE, Karunadasa SS, Smalle JA (2018) Modulation of auxin and cytokinin responses by early steps of the phenylpropanoid pathway. BMC Plant Biol **18**: 1–15

Kuroha T, Tokunaga H, Kojima M, Ueda N, Ishida T, Nagawa S, Fukuda H, Sugimoto K, Sakakibara H (2009) Functional analyses of *LONELY GUY* Cytokinin-Activating Enzymes Reveal the Importance of the Direct Activation Pathway in *Arabidopsis*. Plant Cell **21**: 3152–3169

Kwok S, Kellogg DE, Mckinney N, Spasic D, Goda L, Levenson C, Sninsky JJ (1990) Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. Nucleic Acids Res **18**: 999–1005

Langmead B, Salzberg S (2013) Fast gapped-read alignment with Bowtie 2. Nat Methods **9**: 357–359

Lee DJ, Park JY, Ku SJ, Ha YM, Kim S, Kim MD, Oh MH, Kim J (2007a) Genome-wide expression profiling of *Arabidopsis Response Regulator 7*(*ARR7*) overexpression in cytokinin response. Mol Genet Genomics **277**: 115–137

Lee J, He K, Stolc V, Lee H, Figueroa P, Gao Y, Tongprasit W, Zhao H, Lee I, Deng XW (2007b) Analysis of Transcription Factor HY5 Genomic Binding Sites Revealed Its Hierarchical Role in Light Regulation of Development. Plant Cell **19**: 731–749

Leibfried A, To JPC, Busch W, Stehling S, Kehle A, Demar M, Kieber JJ, Lohmann JU (2005) WUSCHEL controls meristem function by direct regulation of cytokinin-inducible response regulators. Nature **438**: 1172–1175

Letham DS (1963) Zeatin, a factor inducing cell division isolated from *Zea mays*. Life Sci **2**: 569–573

Lin X, Kaul S, Rounsley S, Shea TP, Benito MI, Town CD, Fujii CY, Mason T, Bowman CL, Barnstead M, et al (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature **402**: 761–765

Lohrmann J, Buchholz G, Keitel C, Sweere U, Kircher S, Bäurle I, Kudla J, Schäfer E, Harter K (1999) Differential expression and nuclear localization of response regulator-like proteins from Arabidopsis thaliana. Plant Biol **1**: 495–505

Lohrmann J, Sweere U, Zabaleta E, Bäurle I, Keitel C, Kozma-Bognar L, Brennicke A, Schäfer E, Kudla J, Harter K (2001) The response regulator ARR2: A pollen-specific transcription factor involved in the expression of nuclear genes for components of mitochondrial complex I in Arabidopsis. Mol Gen Genet **265**: 2–13

Lomin SN, Krivosheev DM, Steklov MY, Arkhipov D V., Osolodkin DI, Schmülling T, Romanov GA (2015) Plant membrane assays with cytokinin receptors underpin the unique role of free cytokinin bases as biologically active ligands. J Exp Bot **66**: 1851–1863

Lomin SN, Yonekura-Sakakibara K, Romanov GA, Sakakibara H (2011) Ligand-binding properties and subcellular localization of maize cytokinin receptors. J Exp Bot **62**: 5149–5159

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15**: 1–21

Ma L, Gao Y, Qu L, Chen Z, Li J, Zhao H, Deng XW (2002) Genomic evidence for COP1 as a repressor of light-regulated gene expression and development in Arabidopsis. Plant Cell **14**: 2383–2398

Mahonen A, Bishopp A, Higuchi M, Nieminen K, al et (2006) Cytokinin Signaling and Its Inhibitor AHP6 Regulate Cell Fate During Vascular Development. Sci. (New York, NY) 311:

Mahonen AP, Bonke M, Kauppinen L, Riikonen M, Benfey PN, Helariutta Y (2000) A novel two-component hybrid molecule regulates vascular morphogenesis of the Arabidopsis root. Genes Dev **14**: 2938–2943

Mähönen AP, Higuchi M, Törmäkangas K, Miyawaki K, Pischke MS, Sussman MR, Helariutta Y, Kakimoto T (2006) Cytokinins Regulate a Bidirectional Phosphorelay Network in Arabidopsis. Curr Biol **16**: 1116–1122

Mason MG, Li J, Mathews DE, Kieber JJ, Schaller GE (2004) Type-B Response Regulators Display Overlapping Expression Patterns in Arabidopsis. Plant Physiol **135**: 927–937

Mason MG, Mathews DE, Argyros DA, Maxwell BB, Kieber JJ, Alonso JM, Ecker JR, Schaller GE (2005) Multiple Type-B Response Regulators Mediate Cytokinin Signal Transduction in *Arabidopsis*. Plant Cell **17**: 3007–3018

Matsui M, Deng X (1995) *Arabidopsis* COP1 protein specifically interacts *in vitro* with a cytoskeleton-associated protein, CIP1. Proc Natl Acad Sci U S A **92**: 4239–4243

Matsumoto-Kitano M, Kusumoto T, Tarkowski P, Kinoshita-Tsujimura K, Václavíková K, Miyawaki K, Kakimoto T (2008) Cytokinins are central regulators of cambial activity. Proc Natl Acad Sci U S A **105**: 20027–20031

Mauney JK, Hillman WS, Miller CO, Skoog F, Clayton RA, Strong FM (1952) Bioassay, Purification, and Properties of a Growth Factor from Coconut. Physiol Plant **5**: 485–497

Michelmore RW, Paran I, Kesseli R V (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci U S A **88**: 9828–32

Miller C.O., Skoog F, Okomura FS, Saltza MH, Strong FM (1956) Isolation, structure and synthesis of kinetin, a substrance promoting cell division. J Am Chem Soc 1375–1380

Miyata S, Urao T, Y KY, Shinozaki K (1998) Characterization of genes for two-component phosphorelay mediators with a single HPt domain in *Arabidopsis thaliana*. FEBS Lett **437**: 11–14

Mizuno T (2005) Two-Component Phosphorelay Signal Transduction Systems in Plants: Rrom Hormone responses to Circadian Rhythms. Biosci Biotechnol Biochem **69**: 2263–2276

Mok DW, Mok MC (2001) Cytokinin Metabolism and Action. Annu Rev Plant Physiol Plant Mol Biol **52**: 89–118

Morel G, Wetmore RH (1951) Tissue Culture of Monocotyledons. Am J Bot **38**: 138

Murashige T, Skoog F (1962) A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. Physiol Plant **15**: 473–497

Nagawa S, Sawa S, Sato S, Kato T, Tabata S, Fukuda H (2006) Gene Trapping in *Arabidopsis* Reveals Genes Involved in Vascular Development. Plant Cell Physiol **47**: 1394–1405

Neff MM, Neff JD, Chory J, Pepper a E (1998) dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in Arabidopsis thaliana genetics. Plant J **14**: 387–92

Neff MM, Turk E, Kalishman M (2002) Web-based primer design for single nucleotide polymorphism analysis. Trends Genet **18**: 613–615

Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res **17**: 2503–2516

Nickell LG (1950) Effect of Coconut Milk on the Growth In vitro of Plant Virus Tumor Tissue. Bot Gaz **112**: 225–228

Nishimura C, Ohashi Y, Sato S, Kato T, Tabata S, Ueguchi C (2004) Histidine Kinase Homologs That Act as Cytokinin Receptors Possess Cverlapping Functions in the Regulation of Shoot and Root Growth in *Arabidopsis*. Plant Cell **16**: 1365–1377

Nordström KJ V, Albani MC, James GV, Gutjahr C, Hartwig B, Turck F, Paszkowski U, Coupland G, Schneeberger K (2013) Mutation identification by direct comparison of whole-genome sequencing data from mutant and wild-type individuals using k-mers. Nat Biotechnol **31**: 325–30

Nüsslein-volhard C, Wieschaus E (1980) Mutations affecting segment number and polarity in drosophila. Nature **287**: 795–801

Overbeek J van, Conklin ME, Blakeslee AF (1941) FACTORS IN COCONUT MILK ESSENTIAL FOR GROWTH AND DEVELOPMENT OF VERY YOUNG DATURA EMBRYOS. Science (80-) **94**: 350–351

Oyama T, Shimura Y, Okada K (1997) The *Arabidopsis HY5* gene encodes a bZIP protein that regulates stimulus-induced development of root and hypocotyl. Genes Dev **11**: 2983–2995

Page DR, Grossniklaus U (2002) The art and design of genetic screens: *Arabidopsis thaliana*. Nat Rev Genet **3**: 124–136

Perego M, Glaser P, Hoch JA (1996) Aspartyl-phosphate phosphatases deactivate the response regulator components of the sporulation signal transduction system in *Bacillus subtilis*. Mol Microbiol **19**: 1151–1157

Perraud AL, Weiss V, Gross R (1999) Signalling pathways in two-component phosphorelay systems. Trends Microbiol **7**: 115–120

Peterson BA, Haak DC, Nishimura MT, Teixeira PJPL, James SR, Dangl JL, Nimchuk ZL (2016) Genome-wide assessment of efficiency and specificity in crispr/cas9 mediated multiple site targeting in arabidopsis. PLoS One **11**: 1–11

Pischke MS, Jones LG, Otsuga D, Fernandez DE, Drews GN, Sussman MR (2002) An *Arabidopsis* histidine kinase is essential for megagametogenesis. Proc Natl Acad Sci U S A **99**: 15800–15805

Pogany JA, Simon EJ, Katzman RB, De Guzman BM, Yu LP, Trotochaud AE, Clark SE (1998) Identifying novel regulators of shoot meristem development. J Plant Res **111**: 307–313

Punwani J a., Hutchison CE, Schaller GE, Kieber JJ (2010) The subcellular distribution of the Arabidopsis histidine phosphotransfer proteins is independent of cytokinin signaling. Plant J **62**: 473–782

Punwani JA, Kieber JJ (2010) Localization of the arabidopsis histidine phosphotransfer proteins is independent of cytokinin. Plant Signal Behav **5**: 896–898

Richter J, Watson JM, Stasnik P, Borowska M, Neuhold J, Berger M, Stolt-Bergner P, Schoft V, Hauser MT (2018) Multiplex mutagenesis of four clustered *CrRLK1L* with CRISPR/Cas9 exposes their growth regulatory roles in response to metal ions. Nat Sci Reports **8**: 1–14

Riefler M, Novak O, Strnad M, Schmu T (2006) *Arabidopsis* Cytokinin Receptor Mutants Reveal Functions in Shoot Growth, Leaf Senescence, Seed Size, Germination, Root Development, and Cytokinin Metabolism. Plant Cell **18**: 40–54

Romanov GA, Lomin SN, Schmülling T (2006) Biochemical characteristics and ligand-binding properties of Arabidopsis cytokinin receptor AHK3 compared to CRE1/AHK4 as revealed by a direct binding assay. J Exp Bot **57**: 4051–4058

Ruiz Sola MA, Coiro M, Crivelli S, Zeeman SC, Schmidt Kjølner Hansen S, Truernit E (2017) *OCTOPUS-LIKE 2*, a novel player in Arabidopsis root and vascular development, reveals a key role for *OCTOPUS* family genes in root metaphloem sieve tube differentiation. New Phytol **216**: 1191–1204

Sakai H, Aoyama T, Bono H, Oka A (1998) Two-Component Response Regulators from *Arabidopsis thaliana* Contain a Putative DNA-Binding Motif. Plant Cell Physiol **39**: 1232–1239

Sakai H, Aoyama T, Oka A (2000) *Arabidopsis* ARR1 and ARR2 response regulators operate as transcriptional activators. Plant J **24**: 703–711

Salomé PA, Bomblies K, Fitz J, Laitinen RAE, Warthmann N, Yant L, Weigel D (2012) The recombination landscape in *Arabidopsis thaliana* F2 populations. Heredity (Edinb) **108**: 447–455

Schaller GE, Kieber JJ, Shiu S (2008) Two-Component Signaling Elements and Histidyl-Aspartyl Phosphorelays. Arab B. doi: 10.1199/tab.0112

Schaller GE, Shiu SH, Armitage JP (2011) Two-Component Systems and Their Co-Option for Eukaryotic Signal Transduction. Curr Biol **21**: R320–R330

Scheres B, Di Laurenzio L, Willemsen V, Hauser MT, Janmaat K, Weisbeek P, Benfey PN (1995) Mutations affecting the radial organisation of the *Arabidopsis* root display specific defects throughout the embryonic axis. Development **121**: 53–62

Schmülling T, Werner T, Riefler M, Krupková E, Bartrina Y Manns I (2003) Structure and function of cytokinin oxidase/dehydrogenase genes of maize, rice, Arabidopsis and other species. J Plant Res **116**: 241–252

Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jørgensen JE, Weigel D, Andersen SU (2009) SHOREmap: Simultaneous mapping and mutation identification by deep sequencing. Nat Methods **6**: 550–551

Shin K, Lee S, Song WY, Lee RA, Lee I, Ha K, Koo JC, Park SK, Nam HG, Lee Y, et al (2015) Genetic identification of *ACC-RESISTANT2* reveals involvement of *LYSINE HISTIDINE TRANSPORTER1* in the uptake of 1-Aminocyclopropane-1-Carboxylic Acid in *Arabidopsis thaliana*. Plant Cell Physiol **56**: 572–582

Sigrist CJA, De Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I (2013) New and continuing developments at PROSITE. Nucleic Acids Res **41**: 344–347

Silversmith RE (2010) Auxiliary phosphatases in two-component signal transduction. Curr Opin Microbiol **13**: 177–183

Simsek M, Adnan H (2000) Effect of single mismatches at 3'-end of primers on polymerase chain reaction. J Sci Res Med Sci / Sultan Qaboos Univ **2**: 11–4

Song SK, Clark SE (2005) POL and related phosphatases are dosage-sensitive regulators of meristem and organ development in *Arabidopsis*. Dev Biol **285**: 272–284

Sonnhammer ELL, Heijne G von, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequence. Proc. Sixth Int. Conf. Intell. Syst. Mol. Biol. pp 175–182

Sourjik V, Schmitt R (1998) Phosphotransfer between CheA, CheY1, and CheY2 in the Chemotaxis Signal Transduction Chain of Rhizobium meliloti. Biochemistry **37**: 2327–2335

St Johnston D (2002) The art and design of genetic screens: Drosophila melanogaster. Nat Rev Genet **3**: 176–188

Steward FC, Caplin SM (1951) A Tissue Culture from Potato Tuber: The Synergistic Action of 2,4-D and of coconut milk. Science (80-) **113**: 518–520

Stewart RC (1997) Kinetic Characterization of Phosphotransfer between CheA and CheY in the Bacterial Chemotaxis Signal Transduction Pathway. Biochemistry **36**: 2030–2040

Stock AM, Robinson VL, Goudreau PN (2000) T Wo -C Omponent S Ignal T Ransduction. Reactions **69**: 183–215

Stolz A, Riefler M, Lomin SN, Achazi K, Romanov G a, Schmülling T (2011) The specificity of cytokinin signalling in </i>Arabidopsis thaliana</i> is mediated by differing ligand affinities and expression profiles of the receptors. Plant J **67**: 157–168

Sugimoto N, Nakano SI, Yoneyama M, Honda KI (1996) Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. Nucleic Acids Res **24**: 4501–4505

Suzuki T, Imamura A, Ueguchi C, Mizuno T (1998) Histidine-Containing Phosphotransfer (HPt) Signal Transducers Implicated in His-to-Asp Phosphorelay in *Arabidopsis*. Plant Cell Physiol **39**: 1258–1268

Suzuki T, Sakurai K, Imamura A, Nakamura A, Ueguchi C, Mizuno T (2000) Compilation and Characterization of Histidine-Containing Phosphotransmitters Implicated in His-to-Asp Phosphorelay in Plants: AHP Signal Transducers of *Arabidopsis thaliana*. Biosci Biotechnol Biochem **64**: 2486–2489

Svennerstam H, Ganeteg U, Bellini C, Näsholm T (2007) Comprehensive Screening of Arabidopsis Mutants Suggests the Lysine Histidine Transporter 1 to Be Involved in Plant Uptake of Amino Acids. Plant Physiol **143**: 1853–1860

Swanson R V., Alex LA, Simon MI (1994) Histidine and aspartate phosphorylation: two-component systems and the limits of homology. Trends Biochem Sci **19**: 485–490

Takei K, Ueda N, Aoki K, Kuromori T, Hirayama T, Shinozaki K, Yamaya T, Sakakibara H (2004) AtIPT3 is a key determinant of nitrate-dependent cytokinin biosynthesis in Arabidopsis. Plant Cell Physiol **45**: 1053–1062

Tanaka Y, Suzuki T, Yamashino T, Mizuno T (2004) Comparative Studies of the AHP Histidine-containing Phosphotransmitters Implicated in His-to-Asp Phosphorelay in Arabidopsis thaliana. Biosci Biotechnol Biochem **68**: 462–465

Tandon P, Conlon F, Furlow JD, Horb ME (2017) Expanding the genetic toolkit in Xenopus: Approaches and opportunities for human disease modeling. Dev Biol **426**: 325–335

Taniguchi M, Kiba T, Sakakibara H, Ueguchi C, Mizuno T, Sugiyama T (1998) Expression of *Arabidopsis* response regulator homologs is induced by cytokinins and nitrate. FEBS Lett **429**: 259–262

Taniguchi M, Sasaki N, Tsuge T, Aoyama T, Oka A (2007) ARR1 directly activates cytokinin response genes that encode proteins with diverse regulatory functions. Plant Cell Physiol **48**: 263–277

Thole JM, Beisner ER, Liu J, Venkova S V., Strader LC (2014) Abscisic acid regulates root elongation through the activities of auxin and ethylene in *Arabidopsis thaliana*. G3 Genes, Genomes, Genet **4**: 1259–1274

To JPC, Deruere J, Maxwell BB, Morris VF, Hutchison CE, Ferreira FJ, Schaller GE, Kieber JJ (2007) Cytokinin Regulates Type-A *Arabidopsis* Response Regulator Activity and Protein Stability via Two-Component Phosphorelay. Plant Cell **19**: 3901–3914

To JPCJ, Haberer G, Ferreira FFJ, Deruère J, Mason MG, Schaller GE, Alonso JM, Ecker JR, Kieber JJ (2004) Type-A Arabidopsis Response Regulators Are Partially Redundant Negative Regulators of Cytokinin Signaling. Plant Cell **16**: 658–671

Tokunaga H, Kojima M, Kuroha T, Ishida T, Sugimoto K, Kiba T, Sakakibara H (2012) Arabidopsis lonely guy (LOG) multiple mutants reveal a central role of the LOG-dependent pathway in cytokinin activation. Plant J **69**: 355–365

Tsang DL, Edmond C, Harrington JL, Nühse TS (2011) Cell wall integrity controls root elongation via a general 1-aminocyclopropane-1-carboxylic acid-dependent, ethylene-independent pathway. Plant Physiol **156**: 596–604

Tsuchisaka A, Yu G, Jin H, Alonso JM, Ecker JR, Zhang X, Gao S, Theologis A (2009) A combinatorial interplay among the 1-aminocyclopropane-1-carboxylate isoforms regulates ethylene biosynthesis in Arabidopsis thaliana. Genetics **183**: 979–1003

Ueguchi C, Koizumi H, Suzuki T, Mizuno T (2001a) Novel Family of Sensor Histidine Kinase Genes in Arabidopsis thaliana. Plant Cell Physiol **42**: 231–235

Ueguchi C, Koizumi H, Suzuki T, Mizuno T (2001b) Novel Family of Sensor Histidine Kinase Genes in Arabidopsis thaliana. Plant Cell Physiol **42**: 231–235

Urao T, Yakubov B, Yamaguchi-Shinozaki K, Shinozaki K (1998) Stress-responsive expression of genes for two-component response regulator-like proteins in *Arabidopsis thaliana*. FEBS Lett **427**: 175–178

Vandenbussche F, Habricot Y, Condiff AS, Maldiney R, Straeten DVD, Ahmad M (2007) HY5 is a point of convergence between cryptochrome and cytokinin signalling pathways in Arabidopsis thaliana. Plant J **49**: 428–441

Velikkakam James G, Patel V, Nordstrom KJ, Klasen JR, Salome P a, Weigel D, Schneeberger K, James GV, Nordström KJ, Salomé P a (2013) User guide for mapping-by-sequencing in *Arabidopsis*. Genome Biol **14**: R61

Vogel JP, Schuerman P, Woeste K, Brandstatter I, Kieber JJ (1998) Isolation and characterization of Arabidopsis mutants defective in the induction of ethylene biosynthesis by cytokinin. Genetics **149**: 417–427

Wachsman G, Modliszewski JL, Valdes M, Benfey PN (2017) A SIMPLE pipeline for mapping point mutations. Plant Physiol **174**: 1307–1313

Wang H, Ma LG, Li JM, Zhao HY, Xing Wang Deng (2001) Direct interaction of *Arabidopsis* cryptochromes with COP1 in light control development. Science (80-) **294**: 154–158

Wang Y, Wu Y, Yu B, Yin Z, Xia Y (2017) EXTRA-LARGE G PROTEINS Interact with E3 Ligases PUB4 and PUB2 and Function in Cytokinin and Developmental Processes. Plant Physiol **173**: 1235–1246

Webb B, Sali A (2016) Using MODELLER. Curr Protoc Protein Sci 1–37

Werner T, Motyka V, Laucou V, Smets R, Van Onckelen H, Schmülling T (2003) Cytokinin-Deficient Transgenic Arabidopsis Plants Show Multiple Developmental Alterations Indicating Opposite Functions of Cytokinins in the Regulation of Shoot and Root Meristem Activity. Plant Cell **15**: 2532–2550

West AH, Stock AM (2001) Histidine kinases and response regulator proteins in two-component signaling systems. Trends Biochem Sci **26**: 369–376

Wieschaus E, Nüsslein-Volhard C (2016) The Heidelberg Screen for Pattern Mutants of *Drosophila* : A Personal Account. Annu Rev Cell Dev Biol **32**: 1–46

Winter D, Vinegar B, Nahal H, Ammar R, Wilson G V., Provart NJ (2007) An "Electronic Fluorescent Pictograph" Browser for exploring and Analyzing Large-Scale Biological Data Sets. PLoS One **2**: 1–12

Wulfetange K, Lomin SN, Romanov GA, Stolz A, Heyl A, Schmu T (2011) The Cytokinin Receptors of Arabidopsis Are Located Mainly to the Endoplasmic Reticulum. Plant Physiol **156**: 1808–1818

Xie K, Zhang J, Yang Y (2014) Genome-wide prediction of highly specific guide RNA spacers for CRISPR-Cas9-mediated genome editing in model plants and major crops. Mol Plant **7**: 923–926

Xie M, Chen H, Huang L, O'Neil RC, Shokhirev MN, Ecker JR (2018) A B-ARR-mediated cytokinin transcriptional network directs hormone cross-regulation and shoot development. Nat Commun **9**: 1–13

Xu SL, Rahman A, Baskin TI, Kieber JJ (2008) Two leucine-rich repeat receptor kinases mediate signaling, linking cell wall biosynthesis and ACC synthase in arabidopsis. Plant Cell **20**: 3065–3079

Yamada H, Suzuki T, Terada K, Takei K, Ishikawa K, Miwa K, Yamashino T, Mizuno T (2001) The Arabidopsis AHK4 histidine kinase is a cytokinin-binding receptor that transduces cytokinin signals across the membrane. Plant Cell Physiol **42**: 1017–23

Yang HQ, Tang RH, Cashmore AR (2001) The signaling mechanism of Arabidopsis CRY1 involves direct interaction with COP1. Plant Cell **13**: 2573–2587

Yang SF, Hoffman NE (1984) Ethylene Biosynthesis and its Regulation in Higher Plants. Annu Rev Plant Physiol **35**: 155–189

Yokoyama A, Yamashino T, Amano Y-I, Tajima Y, Imamura A, Sakakibara H, Mizuno T (2007) Type-B ARR transcription factors, ARR10 and ARR12, are implicated in cytokinin-mediated regulation of protoxylem differentiation in roots of Arabidopsis thaliana. Plant Cell Physiol **48**: 84–96

Yoon GM, Kieber JJ (2013) 1-Aminocyclopropane-1-carboxylic acid as a signaling molecule in plants. AoB Plants **5**: 1–6

Yu LP, Simon EJ, Trotochaud AE, Clark SE (2000) *POLTERGEIST* functions to regulate meristem development downstream of the *CLAVATA* loci. Development **127**: 1661–1670

Yuan L, Liu Z, Song X, Johnson C, Yu X, Sundaresan V (2016) The CKI1 Histidine Kinase Specifies the Female Gametic Precursor of the Endosperm. Dev Cell **37**: 34–46

Zapf J, Sen U, Madhusudan, Hoch JA, Varughese KI (2000) A transient interaction between two phosphorelay proteins trapped in a crystal lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction. Structure **8**: 851–862

Zarembinski TI, Theologis A (1994) Ethylene biosynthesis and action: a case of conservation. Plant Mol Biol **26**: 1579–1597

Zheng N, Shabek N (2017) Ubiquitin Ligases: Structure, Function, and Regulation. Annu Rev Biochem **86**: 129–157

Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol **430**: 2237–2243

Zubo YO, Blakley IC, Yamburenko M V., Worthen JM, Street IH, Franco-Zorrilla JM, Zhang W, Hill K, Raines T, Solano R, et al (2017) Cytokinin induces genome-wide binding of the type-B response regulator ARR10 to regulate growth and development in *Arabidopsis*. Proc Natl Acad Sci U S A **114**: E5995–E6004