Mutagenomics: A Rapid, High-Throughput Method to Identify Causal Mutations from a Genetic Screen

Charles Hodgens,a,b Nicole Chang,a G. Eric Schaller,c and Joseph J. Kieber,a,b,2,3

aDepartment of Biology, University of North Carolina at Chapel Hill, North Carolina 27599
bCurriculum in Genetics and Molecular Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599
cDepartment of Biological Sciences, Dartmouth College, Hanover, New Hampshire 03755

ORCID IDs: 0000-0002-0384-4458 (C.H.); 0000-0003-4032-2437 (G.E.S.); 0000-0002-5766-812X (J.J.K.).

Genetic screens are powerful tools to dissect complex biological processes, but a rate-limiting step is often the cloning of targeted genes. Here, we present a strategy, "mutagenomics," to identify causal mutations from a screen in a high throughput fashion in the absence of backcrossing. Mutagenomics is initiated by sequencing the genomes of the mutants identified, which are then subjected to a three-stage pipeline. The first stage identifies sequence changes in genes previously linked to the targeted pathway. The second stage uses heuristics derived from a simulation strategy to identify genes that are represented by multiple independent alleles more often than expected by chance. The third stage identifies candidate genes for the remaining lines by sequencing multiple lines of common descent. Our simulations indicate that sequencing as few as three to four sibling lines generally results in fewer than five candidate genes. We applied mutagenomics to a screen for Arabidopsis (Arabidopsis thaliana) mutants involved in the response to the phytohormone cytokinin. Mutagenomics identified likely causal genes for many of the mutant lines analyzed from this screen, including 13 alleles of the gene encoding the ARABIDOPSIS KINASE4 cytokinin receptor. The screen also identified 1-AMINOCYCLOPROPANE-1-CARBOXYLATE (ACC) SYNTHASE7, an ACC synthase homolog involved in ethylene biosynthesis, and ELONGATED HYPOCOTYL5 (HY5), a master transcriptional regulator of photomorphogenesis. HY5 was found to mediate a subset of the transcriptional response to cytokinin. Mutagenomics has the potential to accelerate the pace and utility of genetic screens in Arabidopsis.

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 (Jorgensen and Mango, 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 transfer DNA (T-DNA) or other insertion 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]), which induce point mutations, enable the facile generation of large mutagenized populations in nearly any genetic background and create 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 cosegregation 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 has 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 (Minevich et al., 2012), the SHOREmap pipeline (Schneeberger et al., 2012), 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. 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 (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 Arabidopsis (Arabidopsis thaliana) His kinase receptors (AHks; Inoue et al., 2001; Ueguchi et al., 2001; Yamada et al., 2001; Caesar et al., 2011; Wulfetange et al., 2011), which autophosphorylate on a His within the His 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 an Arabidopsis His phosphotransfer protein (AHPs; Miyata et al., 1998; Suzuki et al., 1999; Hutchison et al., 2006; Schaller et al., 2008), which in turn transfers it to 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 RR 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-ARRs (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 causative mutations in a parallel manner, which we call "mutagenomics." Genes identified in this screen include multiple alleles of the AHK4 receptor, the ethylene signaling element ETHYLENE-INSENSITIVE2 (EIN2), the ethylene biosynthesis gene 1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE7 (ACS7), the auxin efflux carrier AUXIN RESISTANT1 (AUX1), and the transcription factor ELONGATED HYPOCOTYL5 (HY5). 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 mutagenized an Arabidopsis line that is partially compromised for cytokinin signaling due to loss-of-function mutations in ahp2 and ahp3 with the alkylating agent EMS. This double ahp2,3 mutant has comparable cytokinin sensitivity to the wild type in a root elongation assay, as does the single ahp1 mutant (Fig. 1A; Hutchison et al., 2006). In contrast, the ahp1,2,3 triple mutant is partially resistant to cytokinin, suggesting it is genetically sensitized to modest perturbations of cytokinin responsiveness. Using a root elongation assay, we determined that 0.1 μM of 6-benzylaminopurine (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 (Fig. 1, A and B).

We mutagenized an ahp2,3 line and screened for altered sensitivity to cytokinin using these conditions. Approximately 75,000 Arabidopsis seeds were mutagenized with EMS, the resultant M1 plants were grown to maturity, and the M2 seeds were harvested in 111 pools, with a mean pool size of 140 fertile, viable plants per pool (Fig. 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 rescreening M3 seedlings. A representative mutant is shown in Figure 1B. We refer to these mutations as enhancer of AHPs (eh). 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., 1998a). To identify these ethylene-hyposensitive lines, we assessed the sensitivity of the ehh lines to ethylene using a triple response screen (Table 1;
Guzmán and Ecker, 1990). We removed most of the strongly ethylene-insensitive eah lines from further consideration, although we continued to analyze eight as a proof-of-concept for the subsequent analysis.

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 20× coverage. In addition, we sequenced the parental ahp2,3 plants used for mutagenesis to identify any background SNPs. 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.

Due to the random sampling procedure of whole-genome–library preparation, some regions of the genome may have low or no coverage. The size of low-coverage regions will shrink as average genome coverage rises. Empirically, at 20× coverage in our libraries, ~10% of the genome had coverage of <10×. Nearly all libraries had at least 5× coverage at the 2.5% percentile level. Thus, while it is formally possible that a causative SNP might fall in unsequenced regions of the genome in this experiment, it is highly unlikely.

A total of 1,825 background SNPs (relative to The Arabidopsis Information Resource 10) 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 (Fig. 1D; Table 1). When one considers only homozygous PA-SNPs, there are on average 57 per line, ranging from 11 to 125 per line. This is consistent with previous analyses of EMS-mutagenized Arabidopsis that found 100 to 200 PA-SNPs per line (Thole et al., 2014). In the first step, lines with PA-SNPs in genes known to be involved in the process of interest and likely to be causative, are identified. For example, in the case of the

Figure 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 BA to use for the screen. Seedlings of the indicated genotypes were grown for 3 d on control media and then moved to media containing the indicated level of BA. The growth between 3 and 8 d after germination (DAG) was quantified. Bars are se. n = 11 for all groups except ahp2,3 control (7), 0.01 (6), and 10 μM (4) of BA. B, Representative images of wild type (Col) and the indicated mutants, including a representative eah line (eah22), grown as in A. Scale bar = 1 cm. C, Flow chart describing the details of the genetic screen. D, Histogram of number of PA-SNP densities in sequenced mutant lines. PA-SNPs are inferred using the program SnpEff and consist of all mutations causing missense, nonsense, or splice-site mutations.
eha screen, we screened for genes previously identified that might 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 beyond what is expected by chance, then it is likely to be causative for the phenotype being screened for. In the third step, 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 to a small number. As described below, we successfully applied mutagenomics to the eha screen as a proof-of-concept for this methodology. A user guide to performing mutagenomics is presented in Supplemental Data S1.

### Mutagenomics Applied to the eha Screen

#### Step 1: Screening for Mutations in Genes Known to Affect Cytokinin Responsiveness

After retesting and removing most ethylene-insensitive lines, we distilled our eha screen to 51 lines that we subjected to mutagenomics (Table 1). These lines were sequenced to a target coverage of 20× using

Table 1. eha mutant lines analyzed by mutagenomics

| Mutant | All EMS SNPs | All PA-SNPs | Homozygous PA-SNPs | Candidate Gene | Mutation | ACC Response |
|--------|--------------|-------------|--------------------|---------------|----------|--------------|
| eah1   | 400          | 112         | 99                 | AHK4          | Splice acceptor variant c.1708-1G > A | S |
| eah14-1 | 153          | 20          | 14                 | AHK4          | Thr-1034Ile | S |
| eah14-2 | 125          | 27          | 11                 | AHK4          | Gln-928*   | S |
| eah15   | 565          | 137         | 79                 | AHK4          | Asp-1036Gln | w |
| eah17   | 200          | 34          | 26                 | AHK4          | Ser-142Asn | S |
| eah25   | 586          | 171         | 109                | AHK4          | Gly-139Arg | w |
| eah32   | 427          | 106         | 55                 | AHK4          | Ser-142Asn | S |
| eah34   | 168          | 46          | 42                 | AHK4          | Thr-486Idel | S |
| eah45   | 779          | 223         | 118                | AHK4          | Gln-927*   | w |
| eah46   | 142          | 30          | 24                 | AHK4          | Asp-996Gln | S |
| eah41   | 662          | 171         | 94                 | ARR13         | Asp-66Gln | S |
| eah13   | 479          | 139         | 75                 | AHK4          | Thr-134*   | S |
| eah22-1 | 129          | 21          | 10                 | HY5           | c.439-1G > A | S |
| eah22-2 | 130          | 25          | 16                 | HY5           | c.439-1G > A | S |
| eah2    | 529          | 137         | 107                | EIN2          | Cys-1141Tyr | l |
| eah3    | 364          | 98          | 55                 | EIN2          | Arg-124Gln | w |
| eah4    | 385          | 89          | 47                 | EIN2          | Pro-1185Leu | l |
| eah6    | 335          | 70          | 41                 | EIN2          | Pro-460Leu | l |
| eah10   | 307          | 68          | 45                 | EIN2          | Leu-375Phe | l |
| eah18   | 531          | 132         | 35                 | EIN2          | Thr-1012*  | l |
| eah21   | 429          | 103         | 76                 | EIN2          | Thr-1158*  | l |
| eah47   | 401          | 105         | 51                 | EIN2          | Gly-882Glu | w |
| eah16   | 409          | 94          | 44                 | AUX1          | Thr-245Idel | S |
| eah20   | 318          | 59          | 43                 | AUX1          | 363+1G > A Splice donor variant | w |
| eah30   | 511          | 140         | 59                 | CK11          | Pro-289Ser | S |
| eah39   | 372          | 83          | 62                 | ARK12         | Pro-538Ser | S |
| eah24   | 268          | 63          | 33                 | AHK3          | Thr-801*   | nd |
| eah9    | 526          | 120         | 71                 | ACS7          | Thr-417*   | S |
| eah44   | 448          | 117         | 68                 | ACS7          | Ser-249Phe | nd |
| eah48   | 297          | 90          | 57                 | ACS7          | Thr-83*    | S |
| eah5    | 563          | 163         | 99                 | —             | —          | w |
| eah7    | 300          | 72          | 34                 | —             | —          | l |
| eah28   | 697          | 173         | 112                | —             | —          | l |
| eahX    | 224–748      | 39–201      | 17–125             | Fourteen independent eah lines with no change | S |
| eahY    | 115–615      | 24–182      | 12–107             | Five independent eah lines with no change in | nd |

Supplemental Table S1 genes

| Supplemenatal Table S1 genes |

*Indicates all PA-SNPs called by GATK and SnpEff, regardless of zygosity. ** Likely caused by the nature of the mutation and role of the gene.  † F1 failure to complement in cross with cre1-12.  ‡ Sibling group for eah14.  § Sibling group for eah22.
did not share any SNPs throughout the genome other than S142N. Two mis-sense mutations were predicted to alter residues within the His kinase domain, consistent with the observation that His 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 (Fig. 3A). Two of these (D996N and T1034I) are annotated as part of the active site (Lin et al., 1999). A third mis-sense mutation, D1036N, is two residues adjacent to an active site residue on the AHK4 primary amino acid sequence. 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). To examine the predicted effect of these mis-sense alleles, the AHK4 receiver domain was modeled using the hprep software suite (Zimmermann et al., 2018) and modeled in the program PyMol (https://pymol.org/2/; Fig. 3B). Examination of the predicted 3D structure shows that 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.

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 ahk3 mutations do not significantly affect the response to exogenous cytokinin in root elongation assays (Higuchi et al., 2004; Nishimura et al., 2004), and so it is likely that the cytokinin hyposensitivity in this eah24 line reflects the sensitized nature of the ahp2,3 parental line. We also identified mutations in CYTOKININ INDEPENDENT1 (CKI1), which encodes a His 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, although 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, although 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 ARRI2; single arri2 mutants display a modest hyposensitivity to cytokinin in root elongation assays (Mason et al., 2005), and this may act additively with the ahp2,3 mutations. In line eah41 there is a mutation in ARRI3, 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.

Figure 2. Overview of the mutagenomics process. Top, Implementing mutagenomics is a multistep process that starts with a collection of mutant lines with unknown causative mutations. Middle, A number of lines are resequenced 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 resequenced 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.
Of the eah lines that we analyzed, eight of them displayed a strong EIN phenotype as measured by a triple response assay (Table 1). Six of these have mutations in the previously identified EIN2 gene. Two lines (eah3 and eah47) with an ein2 mutation had a weak EIN phenotype; eah3 also had an ahk4 mutation that was confirmed as a loss of function by a complementation test with the cre1-4 mutations. Intriguingly, two of the eah lines that displayed a strong EIN phenotype did not display PA-SNPs in any of the screened ethylene-related candidate genes (Supplemental Table S1), suggesting that they may affect novel elements in the response to ethylene. However, we cannot rule out the possibility that mutations in known genes were missed due to low sequence coverage, although, as described above, it is unlikely at this genome-sequencing depth.

**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 (Fig. 4). The simulations were performed using the average density of homozygous PA-SNPs observed in the eah screen.

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. To isolate the effect of gene size on the likelihood of finding independent mutations, we analyzed subsets of genes in the simulations defined by the quartiles of Arabidopsis gene sizes (Fig. 4). The probability of finding a gene mutated multiple times is a function of both its size and the number of lines examined. For the smallest quartile of genes (defined as having a coding length of <590 bp), the probability of observing three independent alleles by chance alone in 50 independent lines is <0.1 (Fig. 4A), while for the largest quartile (defined as having a coding length of >1,555 bp), the probability is ~95% (Fig. 4D). For three-quarters of all genes in the Arabidopsis genome, the probability of observing four independent alleles by chance in 50 independent lines is <0.1.

Consistent with our simulation studies, we find many genes that are independently mutated at least two to four times from our pool of 51 eah lines (Supplemental Fig. S1). As there is a reasonable probability (~25%) that a large gene would be independently mutated five times by chance (Fig. 4), we initially 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 (eight 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), nonredundant 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 elongation 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 has both a weak EIN phenotype and fails to complement the cre1-4 mutation, it is likely a double loss-of-function ein2 ahk4 mutant.

![Diagram of AHK4](https://plantphysiol.org)
There are 22 lines remaining that are not ethylene-insensitive (or were not tested), and which do not contain mutations in obvious cytokinin targets. We generated a network diagram consisting of these mutant lines connected to any genes with PA-SNPs in at least two lines (Fig. 5A). Three genes were found mutated in three independent lines; two of these are in genes that fall into the fourth quartile of Arabidopsis genes by the size of their coding region (CATION/H⁺ EXCHANGER2, 2.3 kb; emb1507, 6.5 kb). Finding three mutations by chance is not unlikely for genes in this size range (P > 0.9; Fig. 4). The third gene that is mutated in three lines (eah9, eahl44, and eahl48) is ACS7, which falls into the third quartile of genes (1.44-kb coding region). The probability of finding three hits in a gene of this quartile in 22 independent lines is <0.2, suggesting that it may be causative.

ACS7 encodes an ACC synthase protein that catalyzes the generally rate-limiting step in ethylene biosynthesis (Yang and Hoffman, 1984). Cytokinin inhibits root growth in Arabidopsis primarily by increasing the production of ethylene (Vogel et al., 1998a), which is why ein mutants are identified in the cehl screen. Previous genetic screens for cytokinin hyposensitivity using etiolated seedlings identified a different ACS gene, ACS5, as the target of cytokinin-induced ethylene production (Vogel et al., 1998b). There are eight genes in the Arabidopsis genome that encode active ACS enzymes (Yamagami et al., 2003), and these fall into three classes based on their C-terminal domains (Harpaz-Saad et al., 2012).

Cytokinin has been shown to regulate the stability of both class 1 and class 2 ACS proteins in etiolated seedlings, but did not affect ACS7 turnover, which is the sole class 3 ACS in Arabidopsis (Hansen et al., 2009; Lee et al., 2017). Further, cytokinin was found to not affect the level of ACS7 mRNA, and disruption of ACS7 did not reduce the level of ethylene made in response to exogenous cytokinin in etiolated seedlings (Lee et al., 2017). Thus, ACS7 was not included in the list of known targets for our screen.

The mutation in eahl44 is a mis-sense mutation in the ACS7 gene and is predicted to change Ser-249 to a Phe residue. This residue is adjacent to a residue (Tyr-248) strongly conserved in plant aminotransferases (Rottmann et al., 1991). If this acs7 mutation is causative, it may act by interfering with the active site required for the Asp aminotransferase activity of ACS7. The mutation in eahl9 is a nonsense allele causing a predicted premature translational stop at Trp-417. This mutation is near the end of the protein and is predicted to truncate a conserved domain of the ACS proteins (Yamagami et al., 2003). The mutation in eahl48 is a nonsense allele affecting an early codon (83).

To test if disruption of ACS7 does result in reduced sensitivity to cytokinin in a root elongation assay, we examined the response of the acs7-1 mutant, which harbors a T-DNA insertion mutation in the ACS7 gene (Tsuchisaka et al., 2009). The acs7-1 mutant displays strong insensitivity to cytokinin, similar to the response observed in lines cehl9 and cehl44 (Fig. 5, B and C). This confirms that the acs7 mutations are causative for
cytokinin hyposensitivity in the eah9, eah44, and eah48 lines. Consistent with observations from etiolated seedlings (Lee et al., 2017), our analysis indicates that the level of ACS7 mRNA is not significantly induced in response to cytokinin in Arabidopsis roots (see below, “RNA-Seq Analysis”: 760 versus 781 normalized ACS7 counts for vehicle control and 0.1 μM BA-treated roots, respectively), suggesting that the effect of cytokinin on ACS7 levels is post-transcriptional, similar to the stabilization of ACS5 and other type 1 and type 2 ACS proteins by cytokinin in etiolated seedlings (Hansen et al., 2009; Lee et al., 2017). In other studies, ACS7 protein levels were shown to be regulated by the XBAT32 RING E3 ligase (Lyzenga et al., 2012), and the N-terminal domain of ACS7 plays a role in the regulation of its turnover (Xiong et al., 2014). We conclude that, in contrast to etiolated seedlings, ACS7 is the primary target of cytokinin-induced ethylene biosynthesis in light-grown Arabidopsis roots, and this likely occurs via an increased stability of the ACS7 protein.

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-two lines remain (out of the original 51 analyzed) for which a causative gene was not yet 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
The range of M3 PA-SNP densities is representative of the number of PA-SNPs observed in a set of siblings. The average number of homozygous PA-SNPs observed in a set of siblings indicates five overlapping genes. The numbers above the graph indicate the number of mutations shared between sibling lines. The dashed gray line indicates siblings 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 6.** Projected numbers of shared alleles between M3 siblings. Error bars indicate the mean and SD of simulation-derived distributions of mutations shared between sibling lines. The dashed gray line indicates five overlapping genes. The numbers above the graph indicate the number of mutations shared between sibling lines. The dashed gray line indicates siblings 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.

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 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 (Fig. 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 6. 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.

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 steps one and two of the mutagenomics analysis. This pair of siblings had a small mean number of PA-SNPs ($X = 23.5$ if both homo- and heterozygous PA-SNPs are included; $X = 12.5$ homozygous; Table 1). When the second sibling line was sequenced, it was found to only share a single homozygous PA-SNP with the eah14-1 line, which was in the AHK4 gene (Fig. 7F), which is consistent with this being the causative mutation. Given the observed number of PA-SNPs, this fits with what our simulation model would predict (Fig. 6). 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 other shared mutations confirms that these two lines are indeed siblings (Supplemental Fig. S3).

We also analyzed the eah22 line, which had a comparatively small mean number (13) of homozygous PA-SNPs (Table 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 (Fig. 7A). One of these, HY5, has in fact been suggested to affect cytokinin sensitivity (Cluis et al., 2004), but was not included 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 sibling lines displayed characteristic hy5 phenotypes, such as 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 (Fig. 7, B and C). 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 CONSTITUTIVELY PHOTOMORPHOGENIC1, an E3 ubiquitin ligase (Ma et al., 2002; Bauer et al., 2004). When plant cells are exposed to light, CONSTITUTIVELY PHOTOMORPHOGENIC1 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 chromatin immunoprecipitation (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; Fig. 7D). There was highly significant overlap between genes with adjacent type-B ARR and HY5 binding sites across the genome (the $P$ values for the overlap of HY5 with ARR10 [Zubo et al., 2017] and 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 were determined by RNA sequencing (RNA-seq; Fig. 7E). A total of 347 genes were upregulated after cytokinin treatment in the wild type in these conditions (false discovery rate of 0.02 and log2 fold-change ≥ 0.5; Supplemental Table S3). Of those, 53% (185 genes) were not upregulated in the hy5 mutant (Fig. 7E; Supplemental Table S4). Likewise, 177 genes were downregulated in response to cytokinin in wild-type roots, and 38% of these were not downregulated in the hy5 mutant (Fig. 7E; Supplemental Table S4). Interestingly, there were a substantial number of genes upregulated (49) and downregulated (477) in response to cytokinin specifically in the hy5 mutant (Fig. 7E).

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 (Supplemental Table S5). These genes overlapped significantly with cytokinin differentially expressed genes (DEGs) in the wild-type roots (Supplemental Fig. 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, these data are 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 eukaryotic organisms. EMS and other mutagens that induce point mutations are extraordinarily useful in that a wide range of different alleles can be generated, including those that reduce but not do not eliminate the function of the encoded protein (i.e. hypomorphic alleles), thereby overcoming potential lethality or infertility of null alleles, or mutations that modify protein function that can overcome functional redundancy, such as dominant negative (antimorphic) or activating (neomorphic or hypermorphic) changes. Further, EMS can be used to rapidly develop mutagenized populations in any genetic background, allowing screens to be performed in sensitized backgrounds, enabling suppressor and enhancer screens. Such screens can be useful in many ways, including allowing the identification of genes with relatively weak effects on a pathway of interest, potentially overcoming partial functional redundancy, and/or identifying elements involved in signaling crosstalk. While extraordinarily useful, cloning the genes targeted by EMS mutagenesis remains challenging.

The advent of high-throughput sequencing has provided tools to facilitate the cloning of EMS-generated and other missense alleles. The most common approach in Arabidopsis and other sequenced organisms is to use “mapping-by-sequencing” (Schneeberger et al., 2009b). This approach has greatly facilitated cloning of genes in Arabidopsis and other model organisms. While these approaches are quite powerful and reduce the time and effort required for identifying causative mutations of interest, they still are focused on analyzing a single mutant line at a time and require multiple generations to develop a mapping population. 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.

The foundational data for the mutagenomics pipeline is determining the genomic sequencing of the identified mutants as well as the parental line. Although, 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 rescreen 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, although if semi-dominant there would likely be a bias toward 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 evidenced by lack of segregation of wild-type phenotypes in its progeny) or screen the progeny of various M3 plants for homozygosity.

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. In the cytokinin screen analyzed here, we included 23 genes related to cytokinin (including aux1, which is known to give cytokinin-insensitivity) and 77 previously identified in EIN screens (Merchante and Stepanova, 2017) that we anticipated might be identified in the eih screen (Supplemental Table S1). As anticipated, mutations in several of these were indeed identified in the eih 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 helps to 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 step 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 51 mutant lines, there was a high likelihood (>90%) that a gene will be mutated in up to four independent lines by chance when one includes genes of all sizes. As the frequency at which a gene is mutated is directly proportional to the length of its coding region (among other factors), we stratified our analysis by the size of the gene by considering quartiles of the genome binned by the size of the coding region. As expected, this analysis demonstrates that smaller genes are less likely to be mutated multiple times by chance. Further, the likelihood of multiple random mutations decreases with smaller numbers of mutant lines. We used these factors to reanalyze the unassigned eih lines, and this revealed three mutations in acs7 as the potentially causative mutation, which was subsequently confirmed. This suggests that ACS7 is the primary target by which cytokinin elevates ethylene biosynthesis in roots of light-grown Arabidopsis seedlings, in contrast to etiolated seedlings in which ACS5 is the primary target (Vogel et al., 1998b). Interestingly, ACS7 has also been linked to root gravitropism, lateral root formation, and ABA responsiveness (Prasad et al., 2010; Dong et al., 2011; Huang et al., 2013).

In addition to the size of the coding region, different genes are likely more or less sensitive to mutations affecting their function due to differences in the number of critical residues. Further, some genes will result in a weak phenotype even with a complete loss-of-function
as they, for example, may only modify a pathway rather than being a central component, and these 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 the ahp2,3 mutations (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 the ahk2 mutation was not found 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.

In cases where the number of mutations does not show clear enrichment, the exact nature of observed mutations can be considered when deciding whether a gene is worth further study. However, caution must be taken. In the eah screen, 12 alleles of ahk4 were identified (Fig. 3; Table 1), but only four nonsense alleles (three in the last third of the gene) and one splice mutation (about halfway through the gene) were observed. Stop mutations and splice errors are easy to classify as deleterious, mis-sense mutations less so. The most careful approach would be that the nature of observed mutations should be treated as reasons to include a gene which otherwise might be excluded, rather than to exclude any gene from further analysis.

The third step 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.

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 resccreens for dominant mutations. Thus, as one sequences additional sibling lines, they will 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 three 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 clustered regularly interspaced short palindromic repeats 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. Highlighting the power of the mutagenomics process, we observed a line (eail14) 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 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 22 lines from our eail 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. Mutagenomics should enable the parallel processing of mutant screens in Arabidopsis and other model organisms and should
further enhance the awesome power of genetics to dissect biological processes.

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.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

The Arabidopsis (Arabidopsis thaliana) ahp2,3 double mutant, the ahp1 mutant in the Ws-0 background, and the ahp1,2,3 triple mutant have been described in Hutchison et al. (2006). The hsp5 allele is from the SALK T-DNA collection (SALK_096651C). The ahk2-ahk4 line has been described as ahk2-7 crel1-12 in the literature (Insua et al., 2001; Yamada et al., 2001; Cheng et al., 2013). The aces7 line has been described as aces7-1 in Yamagami et al. (2003). Seed were sterilized by incubating in 95% (v/v) ethanol (1 min) followed by a 15-min incubation in 50% (v/v) bleach (8.25% [w/v] sodium hypochlorite; Cloroxy) and 0.5% (v/v) Tween 20. After sterilization, the seeds were washed five times with sterile water and then stratified for 3 d at 4°C. Root elongation assays were performed on a single plate of Murashige and Skoog medium with 0.1 μM of BA with 8 g L−1 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

EMS mutagenesis was performed by incubating 1.5 g of dry ahp2,3 seeds in 100 mM of potassium phosphate buffer at pH 7.5. EMS was added to a concentration of 0.4% (v/v) and the seeds were mixed on a rocking shaker for 8 h. The seeds were washed 20 times with water and sown directly on soil. The M1 plants were grown to maturity and the M2 seeds were harvested in pools of ~140 M1 plants.

For the primary assay, ~100 μL of M2 seed was sterilized and plated densely on media containing 0.1 μM of BA. Pooled M2 seeds were plated in dense rows on BA plates and grown for 5 to 8 d. 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 of BA at low density. A wild-type control was used on each plate. Three 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 the software ImageJ (Abramoff, 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 Murashige and Skoog medium containing 10 μM of ACC. After 3-d growth in the dark, the lines were scored for their triple response morphology (Guzmán and Ecker, 1990).

Library Preparation for Genome Resequencing

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 cetytrimethylammonium bromide extraction protocol (Doyle and Doyle, 1987) was performed to extract genomic DNA for genome resequencing. Frozen tissue was ground using a SPEX SamplePrep 10 Geno/Grinder (Thermo Fisher Scientific). Ground tissue was mixed with 2× cetytrimethylammonium bromide buffer. The homogenate was washed twice with chloroform. DNA was precipitated with 0.5× volume 5 M of NaCl and 2 volumes of 100% (v/v) ethanol and washed with 70% (v/v) ethanol with 10 μl of ammonium acetate. The DNA was then resuspended in 250 μl of water with 10 μg of mL−1 RNase A. The DNA was reprecipitated and washed with 70% (v/v) ethanol, then resuspended in 50 μl of water.

Genomic libraries were then prepared using a TrueSeq PCR-Free DNA Library Kit (cat. no. FC-121-3003; Illumina) for the first two phases, and a KAPA HyperPrep DNA Library Kit (Roche) for the third phase. Library concentrations were determined using a KAPA Library Quantification Kit (cat. no. 790140001; Sigma-Aldrich/Roche). Mutant resequencing was performed in three rounds—twice on a HiSeq 4000 (first and third phases; Illumina) and once on a HiSeq X10 platform (second phase; Illumina).

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 the program SnpEff (Cingolani et al., 2012) to determine whether they are PA 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.

Sequenceing

Genome resequencing was performed in three rounds. The first round of sequencing was performed with 33 libraries on a single lane of a HiSeq 4000 instrument (2×150 bp) by the High Throughput Sequencing Facility (HTSF) at University of North Carolina (UNC) at Chapel Hill. The second round was performed on a HiSeq ×10 platform (2×150 bp; BGII). The third round was sequenced on a HiSeq 4000 (2×75 bp) by the HTSF at UNC Chapel Hill. The RNA-seq experiment was sequenced on a HiSeq 4000 in high output mode (2×75 bp) by the HTSF at UNC Chapel Hill.

Mutagenomics Data Processing

Sequenced libraries were aligned to The Arabidopsis Information Resource 10 genome (https://www.arabidopsis.org/) using the software Bowtie2 (Langmead and Salzberg, 2012). Variants were inferred using the Genome Analysis Tool Kit (GATK; McKenna et al., 2010; DePristo et al., 2011; Van der Auwer et al., 2013). Detected variants were filtered based on their call qualities by GATK with the filtering “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 the program SnpEff (Cingolani et al., 2012). Only PA 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 sequence regulator domain (residues 946 to 1,071) was submitted to the HHpred search and the top 100 matches were chosen to model the structure of the domain using Modeller.

Library Preparation and Analysis for RNA-Seq

The hsp5 gene expression experiment was performed using RNA extracted from flash-frozen tissue using an RNaseasy kit (cat. no. 74106; Qiagen). Samples of mRNA were isolated using Sera-Mag oligo dT beads (Thermo Fisher Scientific) in the presence of RNase Out (Enzymics) and heat-fractionated. First-strand synthesis was performed with Enzscript (Enzymics), and second-strand synthesis with DNA Poll (Enzymics) and RNase H (Enzymics). End repair was performed with T4 DNA Polymerase (Enzymics), Klenow polymerase (Enzymics), and T4 PNK (Enzymics). A-tailing was performed with T4 DNA ligase (Enzymics). The libraries were PCR-amplified using KAPA HiFi Hot- Start ReadyMix (Kapa Biosystems). All wash steps were performed with AMPure XP beads (Beckman Coulter). Read alignment was performed using the tool Star (Dobin et al., 2013), and count data were normalized using the
program Salmon (Patro et al., 2017). Differential gene expression was analyzed using the program DESeq2 (Love et al., 2014) in R. The Benjamini-Hochberg-adjusted P-value cutoff (Benjamini and Hochberg, 1995) was used for determining significance was lowered to 0.02 to account for only having two replicates, and a log₂ 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. 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 of all simulated screens in which a gene was independently mutated. 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 PA 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, crossovers to be performed are 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 predetermined positions between randomly chosen nonsister 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 the software BowTie2 and the Genome Analysis Toolkit.

### Supplemental Data

The following supplemental materials are available.

- **Supplemental Table S1.** List of genes previously linked to cytokinin response in roots.
- **Supplemental Table S2.** Genes mutated in at least two sequenced libraries of the cah screen.
- **Supplemental Table S3.** Genes differentially expressed in response to BA in wild-type roots.
- **Supplemental Table S4.** Genes differentially expressed in response to BA in hy5 roots.
- **Supplemental Table S5.** Genes differentially expressed between wild-type and hy5 roots in the absence of BA.
- **Supplemental Data S1.** User guide to mutagenomics.

### ACKNOWLEDGMENTS

We thank Jamie Winsholl for technical support, Gyeong Mee Yoon for the acs7 line, and Caroline Spgren for helpful discussions. The authors also thank Yan Zhuo for performing initial genetic analyses that confirmed a role for HY5 in mediating the root response to cytokinin. Received May 12, 2020; accepted September 2, 2020; published September 4, 2020.

### LITERATURE CITED

Abrámová M (2004) Image processing with Imagej. Biophotonics Internat. 11: 36–42

Alonso JM, Hirayama T, Roman G, Nourizadeh S, Ecker JR (1999) EIN2, a bifunctional transducer of ethylene and stress responses in Arabidopsis. Science 284: 2148–2152

ArgyroS RD, Mathews DE, Chiang YH, 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, Stamatiu 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

Bauer D, Viczíán A, Kircher S, Nobis T, Nitschke R, Kunkel T, Panigrahi KC, Adám E, Fejes E, Schäfer 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

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

BrandsStad 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

### Accession Numbers

Locus data for the genes from this article can be found in the GenBank/EMBL data libraries or at https://www.arabidopsis.org/ under accession numbers: AHPI (AT3G21510); AHP2 (AT3G29350); AHP3 (AT5G39340); AHPK2 (AT5G39750); AHPK3 (AT1G27320); AHPK4 (AT1G21830); HY5 (AT5G11260); ACS7 (AT4G26200); ARR1 (AT3G16857); ARR10 (AT4G31920); ARR12 (AT2G25180); ARR13 (AT2G27070); EIN2 (AT5G03280); AUX1 (AT2G38120); and CK11 (AT2G47430). The RNA-seq Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) accession number is GSE149641. The Mutagenomics sequence read archive (https://www.ncbi.nlm.nih.gov/sra) data access number is PRJA631403.

### Code and Data Availability

Statistical and data analyses were performed using R and Python. Code is available at https://github.com/KieberLab/Mutagenomics.
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

Patio R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14: 417–419

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

Prasad ME, Schoof D, Lyzewski W, Liu H, Stone SI (2010) Arabidopsis RING E3 ligase XBAT32 regulates lateral root production through its role in ethylene biosynthesis. Plant Physiol 153: 1587–1596

Riedler M, Novak O, Strnad M, Schmülling 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

Rottmann WH, Peter GF, Oeller PW, Keller JA, Shen NF, Nagy BP, Taylor LP, Campbell AD, Theologis A (1991) 1-Aminocyclopropane-1-carboxylate synthase in tomato is encoded by a multigene family whose transcription is induced during fruit and floral senescence. J Mol Biol 222: 937–961

Salomé PA, Bomblies K, Fitz J, Laitinen RAE, Warthmann N, Yant L, Prasad ME, Schoof D, Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2008) Two-component signaling elements encode a cytokinin receptor that transduces cytokinin signals across the membrane. J Biol Chem 283: 16045–16054

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

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

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

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

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 S (2007) ARR1 directly activates cytokinin response genes that encode proteins with diverse regulatory functions. Plant Cell Physiol 48: 263–277

Thole JM, Reins E, Liu J, Venkova SV, Strader LC (2014) Abscisic acid regulates root elongation through the activities of auxin and ethylene in Arabidopsis thaliana. G3 (Bethesda) 4: 1259–1274

To JPC, Deruère 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 JPC, Haberer G, Ferreira FJ, 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

Tsusakura A, Yu G, Jin H, Alonso JM, Ecker JR, Zhang X, Gao S, Theologis A (2009) A combinatorial interplay among the 1-amino-cyclopropane-1-carboxylate isoforms regulates ethylene biosynthesis in Arabidopsis thaliana. Genetcis 183: 799–1003

Ueguchi C, Koizumi H, Suzuki T, Mizuno T (2001) Novel family of sensor histidine kinase genes in Arabidopsis thaliana. Plant Cell Physiol 42: 231–235

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43: 11.10.1–11.10.33

Vandenbussche F, Habricot Y, Condiff AS, Maldiney R, Van der Straeten D, Ahmad M (2007) HY5 is a point of convergence between cryptochrome and cytokinin signalling pathways in Arabidopsis thaliana. Plant J 49: 428–444

Vogel JP, Schuerman P, Woeste K, Brandstatter I, Kieber JJ (1998a) Isolation and characterization of Arabidopsis mutants defective in the induction of ethylene biosynthesis by cytokinin. Genetics 147: 417–427

Vogel JP, Woeste KE, Theologis A, Kieber JJ (1998b) Rec sessive and dominant mutations in the ethylene biosynthetic gene ACS5 of Arabidopsis confer cytokinin insensitivity and ethylene overproduction, respectively. Proc Natl Acad USA 95: 4766–4771

Wachsmann 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 cytokinome with COP1 in light control development. Science 294: 154–158

Webb B, Sali A (2016) Comparative protein structure modeling using MODELLER. Curr Protoc Protein Sci 2016: 2.9.1–2.9.37

Wulftange K, Lomin SN, Romanov GA, Stolz A, Heyl A, Schmülling T (2011) The cytokinin receptors of Arabidopsis are located mainly to the endoplasmic reticulum. Plant Physiol 156: 1808–1818

Xie M, Chen H, Huang L, O’Neill RC, Shokhirev MN, Ecker JR (2018) A B-ARR-mediated cytokinin transcriptional network directs hormone cross-regulation and shoot development. Nat Commun 9: 1604

Xiong L, Xiao D, Xu X, Guo Z, Wang NN (2014) The non-catalytic N-terminal domain of ACS7 is involved in the post-translational regulation of this gene in Arabidopsis. J Exp Bot 65: 4397–4408

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: 1017–1023

Yamagami T, Tsuchisaka A, Yamada K, Haddon WF, Harden LA, Theologis A (2003) Biochemical diversity among the 1-amino-cyclopropane-1-carboxylate synthase isoforms encoded by the Arabidopsis gene family. J Biol Chem 278: 49102–49112

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, Hofman NE (1984) Ethylene biosynthesis and its regulation in higher plants. Annu Rev Plant Physiol 35: 155–189

Yuan L, Liu Z, Song X, Johnson C, Yu X, Sundaresan V (2016) The CKII histidine kinase specifies the female gametic precursor of the endosperm. Dev Cell 37: 34–46

Zimmermann I, 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 HPPred server at its core. J Mol Biol 430: 2237–2243

Zubo YO, Blakley IC, Yamureenko MV, Worthen JM, Street HI, 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 USA 114: E5995–E6004