Title
Promoter-based integration in plant defense regulation

Permalink
https://escholarship.org/uc/item/1rw6c9cv

Journal
Plant Physiology, 166(4)

ISSN
0032-0889

Authors
Li, B
Gaudinier, A
Tang, M
et al.

Publication Date
2014-01-01

DOI
10.1104/pp.114.248716

Peer reviewed
A key unanswered question in plant biology is how a plant regulates metabolism to maximize performance across an array of biotic and abiotic environmental stresses. In this study, we addressed the potential breadth of transcriptional regulation that can alter accumulation of the defensive glucosinolate metabolites in Arabidopsis (Arabidopsis thaliana). A systematic yeast one-hybrid study was used to identify hundreds of unique potential regulatory interactions with a nearly complete complement of 21 promoters for the aliphatic glucosinolate pathway. Conducting high-throughput phenotypic validation, we showed that >75% of tested transcription factor (TF) mutants significantly altered the accumulation of the defensive glucosinolates. These glucosinolate phenotypes were conditional upon the environment and tissue type, suggesting that these TFs may allow the plant to tune its defenses to the local environment. Furthermore, the pattern of TF/promoter interactions could partially explain mutant phenotypes. This work shows that defense chemistry within Arabidopsis has a highly intricate transcriptional regulatory system that may allow for the optimization of defense metabolite accumulation across a broad array of environments.

An organism’s growth and fitness within its environment are largely determined by its ability to efficiently obtain and utilize energy and elements to create biomass to survive. Central to this process is primary metabolism, which determines the use of energy and chemicals from the environment to produce all of the necessary building blocks for cells and the resulting biomass. To optimize fitness, metabolism must be precisely tuned and coordinated to make the most efficient use of available resources. This basic supposition is central to a wide range of biological fields from the study of organismal growth to the study of interactions with the environment, and has led to strong interest in understanding how metabolism is regulated (Karban and Baldwin, 1997; Smith and Stitt, 2007).

The last decade has seen an upsurge in studies investigating the transcriptional control over metabolism. This has largely been driven by three key areas of focus: regulation of sugar metabolism, amino acid metabolism, and secondary metabolism (Desvergne et al., 2006). In most organisms, Glc levels cause transcriptional reprogramming of sugar metabolism and organismal development often by a mitogen-activated protein kinase cascade (Moore et al., 2003; Rolland et al., 2006). Transcriptional control over amino acid metabolism frequently involves General Control Nonderepressible4, whereas other transcriptional activators play a role in coordinating amino acid metabolism (Hope and Struhl, 1986; Neuwald and Lansman, 1997; Li et al., 2013). Within plants, there are also amino acid pathway-specific transcription factors (TFs) known for a couple of plant amino acid biosynthetic pathways (Celenza et al., 2005; Maruyama-Nakashita et al., 2006). In plants, the transcriptional regulation of secondary metabolites has also received significant attention with numerous myeloblastosis (MYB) TFs being identified (Dubos et al., 2010).

The above studies have largely focused on regulation via individual TFs or on a few specific genes within a pathway. Thus, there are a number of unanswered questions surrounding the transcriptional regulation of metabolism. For example, is a pathway a coordinately regulated unit or are there distinctively regulated modules within a pathway? Do converging signals, such as those from defense and growth, converge upstream of a key set of TFs to regulate a pathway or is there any capacity to integrate at the level of the promoters for the regulation of individual genes? Finally, how many TFs have the capacity to interact with a pathway and regulate the level of the metabolite(s) and under which conditions?

To begin asking these questions, we are using the model metabolic pathway of aliphatic glucosinolates (GLSs) within Arabidopsis (Arabidopsis thaliana; Fig. 1A). This pathway has been used to conduct systems
analyses on the function and regulation of metabolism in conjunction with the resulting ecological and evolutionary consequences (Kliebenstein, 2012). Aliphatic GLSs are sulfur-containing secondary metabolites that are the dominant resistance mechanism against chewing insects and a number of bacterial and fungal pathogens (Lambrix et al., 2001; Kliebenstein et al., 2002; Beekwilder et al., 2008; Hansen et al., 2008; Fan et al., 2011; Stotz et al., 2011). The aliphatic GLS pathway also contains genetic variation in enzymatic and regulatory loci that alters fitness at both the local and continental scales (Mauricio, 1998; Kliebenstein et al., 2001a, 2001b, 2001c; Hansen et al., 2007, 2008; Wentzell et al., 2007; Bidart-Bouzat and Kliebenstein, 2008; Chan et al., 2011; Züst et al., 2012). These fitness benefits are constrained by GLS accumulation altering growth and yield, suggesting that the pathway needs fine-tuning to optimize growth and defense (Mauricio, 1998; Lankau and Strauss, 2008; Paul-Victor et al., 2010; Kerwin et al., 2011; Züst et al., 2011). In addition, there is a nearly complete description of the genes and enzymatic steps within the GLS biosynthetic pathways, providing an additional practical advantage to studying these compounds (Fig. 1A; Wittstock and Halkier, 2002; Grubb and Abel, 2006; Halkier and Gershenzon, 2006). Finally, it is possible to cost-effectively, rapidly, and accurately quantify GLSs, further enabling large-scale systems studies (Wentzell et al., 2007; Chan et al., 2011, 2010, 2011).

The developing body of biosynthetic genetic tools has enabled the identification of TFs that regulate the aliphatic GLS pathway (Gigolashvili et al., 2007a, 2007b, 2008; Hirai et al., 2007; Sønderby et al., 2007, 2008; Schweizer et al., 2013). The current model, which is in general agreement with other defense pathway models, suggests that there are limited sets of TFs that can linearly account for most of the transcriptional regulation in the GLS pathway and that environmental signal integration occurs upstream of these TFs (Fig. 1B; Dombrecht et al., 2007; Beekwilder et al., 2008; Navarro et al., 2008; Hou et al., 2010; Lackman et al., 2011; Wild et al., 2012; Nakata et al., 2013; Xu et al., 2014). Although there are data supporting this model, specifically that removing two MYB (MYB28 and MYB29) and three MYC (MYC2, MYC3, and MYC4) TFs abolishes all aliphatic GLS accumulation, a closer investigation of published transcriptomic data suggests that this MYB/MYC model does not fully explain the biological observations (Beekwilder et al., 2008; Sønderby et al., 2010b; Schweizer et al., 2013). The expression of most genes in the aliphatic GLS pathway is not completely abolished in either of the polygenic mutants as the model suggests (Fig. 1C). Interestingly, it is only the transcripts for genes at the beginning of the pathway that were strongly altered in either polygenic mutant (Fig. 1C). This raises the potential for there to be additional unidentified TFs regulating the pathway.

To begin searching for these additional TFs, we cloned the promoters from 22 of the known enzyme-encoding genes in the aliphatic GLS biosynthetic pathway. We then tested all of these promoters for interactions with 659 TFs in a high-throughput yeast one-hybrid (Y1H) platform (Gaudinier et al., 2011). This analysis found a large number of newly identified potential TFs that could interact with the biosynthetic genes and/or the known regulatory TFs. Among these TFs were ones known to be involved with growth, biotic, and abiotic signaling. Over 75% of the test mutant TFs found by Y1H altered the accumulation of the aliphatic GLSs similarly to single mutants in the known MYBs/MYCs (Beekwilder et al., 2008; Sønderby et al., 2010b; Schweizer et al., 2013). The phenotypic effects of the TF mutants were highly dependent upon the environment and tissue in which the accumulation was measured. This shows the importance of performing a gene-centered approach coupled with a nonbiased set of TFs to elucidate conditional-dependent regulatory TFs (Arda and Walhout, 2010). We could show that TFs that bound similar aliphatic GLS promoters also had similar phenotypic consequences using a nonlinear matrix analysis. This suggests that, at least for the aliphatic GLSs, the pathway appears to be regulated as a set of overlapping modules. These overlapping modules allow for the integration of multiple signals at the promoter level within the pathway. Future research will be needed to develop a cohesive model involving the dozens of newly identified TFs and the previously known TFs (Fig. 1).

RESULTS

cis-Regulatory Analysis of GLS Pathway Promoters

To develop a detailed graph of the potential transcriptional regulatory network mediating the expression of aliphatic GLS genes, we cloned approximately 2 kb upstream or to the nearest gene if this was less than 2 kb for the 22 known enzyme-encoding genes for this pathway (Figs. 1A and 2; Supplemental Table S1; Sønderby et al., 2010a; Gaudinier et al., 2011). These genes include all of the known genes for the pathway except for two genes (GGP and CYP79F2; Fig. 1A; Geu-Flores et al., 2009; Sønderby et al., 2010a). We also cloned the promoters from the known MYB regulatory factors (MYB28 and MYB29) that are critical for the aliphatic GLS pathway (Fig. 1B; Supplemental Table S1; Sønderby et al., 2007, 2010a, 2010b). In addition, we also included the promoter of MYB76, which is a homolog of MBY28 and MYB29 and also controls aliphatic GLS accumulations (Sønderby et al., 2007, 2010a, 2010b). As a comparison, we also included the promoters from six genes that had been peripherally associated with the GLS pathway either by being shared with the Val biosynthetic pathway or coexpressing with the pathway (IMD3, IPM2, IPM1, and PMSR1–PMSR3). Another comparison was provided by including two promoters from the indolic GLS pathway, which is largely discrete from the aliphatic biosynthetic pathway (UGT74B and GSTF9; Sønderby et al., 2010b).

Using the sequences of the cloned promoters, we queried the TF binding sites that may be enriched within these promoters using the Athena analysis suite
Figure 1. Proposed aliphatic GLS pathway regulatory model. A, Aliphatic GLS biosynthetic pathway with structures and enzymes. Enzymes in black are ones with promoters included in this study. Those in gray were not cloned. The use of n indicates the number of carbons introduced to the side chain from the elongation cycle and can vary from 1 to 7. The promoters are color coded based on their pathway membership as follows: indolic GLSs are pink, aliphatic regulatory genes are purple, and aliphatic peripheral genes are red. The aliphatic biosynthetic genes are parsed into the core pathway (green), elongation (blue), modifying (yellow), and seed specific (orange) to visualize regulation along the linear model of the pathway. B, A proposed model of aliphatic GLS pathway regulation. In this model, the pathway is regulated by the JA pathway via the bHLH’s MYC2/MYC3/MYC4 interacting with the MYB28/MYB29/MYB76 proteins and binding the respective promoters. C, Transcriptional analysis of the aliphatic GLS pathway in mutants missing the MYB28 and MYB29 TFs (blue) or MYC2/MYC3/MYC4 TFs (red). The value is shown as the relative value of the transcript in the mutant with the wild type set to 1 for each gene. The genes are ordered by their position in the biosynthetic pathway. The transcriptional data presented are from previous publications with full statistical analysis (Sonderby et al., 2010b; Schweizer et al., 2013).

WT, Wild type JA-ILE, isoleucyl jasmonic acid; JAZ, jasmonate ZIM domain proteins; Coi1, Coronatine Insensitive1; BCAT4, BRANCHCHAIN AMINOTRANSFERASE4; BAT5, BILE ACID TRANSPORTERS; IMD3, ISOPROPYL MALATE DEHYDROGENASE3; MAM1, METHYLIOALKYLMALATE SYNTHASE1; IPMI1, ISOPROPYL MALATE ISOMERASE1; CYP9F1/2, CYTOCHROME P450; GGP1, GAMMA GLUTAMYL PEPTIDASE1; GSOH, GLUCOSINOLATE HYDROXYLATION; GSOX1-5, GLUCOSINOLATE OXIDASE1-5; GSTF11, GLUTATHIONE-S-TRANSFERASE11; CSLyase, CYSTEINE-S-CONJUGATE β-LYASE; UGT, UDP-GLYCOSYLTRANSFERASE; SOT, SULFUR TRANSFERASE; AOP3, ALKENYLHYDROXYPROPYL3; BZO1, BENZOYLOXY1.
This showed that the aliphatic GLS promoters are enriched in binding sites for the two classes of TFs previously associated with regulating this pathway, the MYBs and basic helix-loop-helix (bHLH) MYCs (Fig. 2; Supplemental Table S2; Gigolashvili et al., 2007a, 2007b, 2008; Hirai et al., 2007; Sønderby et al., 2007, 2010a, 2010b; Schweizer et al., 2013). There was also a strong enrichment in Basic Leucine Zipper Domain (bZIP) and Homeodomain DNA binding elements and a weaker enrichment in binding elements for the MADs, Auxin Response Factor (ARF), and Apetala2 (AP2)/Ethylene-response factor (ERF) TF families (Supplemental Table S2). The Homeodomain, bZIP, MADs, ARFs, and AP2/ERF TF families have not been previously associated with regulating the aliphatic GLS pathway, suggesting that there are potentially a number of unknown TFs that regulate this pathway.

Y1H Survey of Aliphatic Biosynthetic Genes

We then assessed the interaction of each promoter with all 659 TFs of the published stele-expressed TF collection (Fig. 2; Gaudinier et al., 2011). The stele is the predominant tissue in which the aliphatic GLS biosynthetic pathway genes are expressed (Sønderby et al., 2010a; Moussaieff et al., 2013). We also added the MYB28, MYB29, and MYB76 TFs to this collection because they bind the majority of tested GLS promoters in Arabidopsis cell culture assays (Gigolashvili et al., 2007a, 2007b, 2008). Using our Y1H system, we identified three,
zero, and two interactions, respectively, between these TFs and aliphatic GLS biosynthetic promoters. Differences in interaction numbers in tissue cultures compared with Y1H are likely due to false positives and false negatives associated with each system (Yu et al., 2008; Walhout, 2011).

These results using known interacting TFs provide us an ability to develop an empirical threshold for calling TFs that likely interact with the pathway. Thus, as a compromise between false positives and false negatives, we decided to use three interactions as an empirically validated threshold to identify TFs that are likely to be biologically relevant. This threshold requires these TFs to have at least as many interactions as the validated MYB28 TF. Using this threshold, we are likely removing some biologically relevant promoter-TF interactions because the MYB76 TF, which does regulate the pathway but is not a critical regulator of the pathway, had only two interactions and would have been dropped from our analysis. Of the 144 TFs identified in the Y1H analysis, 52 interacted with three or more of the aliphatic biosynthetic gene promoters similar to the validated MYB28, suggesting that we have identified a large collection of TFs potentially regulating aliphatic GLS accumulation (Fig. 3A).

Our main hypothesis is that aliphatic biosynthetic gene expression is regulated partially independently of MYB28/MYB29 and MYC2/MYC3/MYC4. One mechanism by which this independence could occur is if some promoters have a larger suite of regulatory interactions. These other TFs would provide an added level of redundancy, leading to these promoters being less sensitive to knockouts in the MYBs or MYCs. To test whether this model of independence may be present, we compared the number of TFs that interact with a promoter versus the relative expression of that gene within MYB and MYC knockouts. Because the available transcriptomics studies are done in adult foliar tissues, we focused on only the 16 genes expressed in this tissue. This removed four genes that have embryo-specific expression (BZO1, AOP3, GSOX2 GLUCOSINOLATE, and GSOX4; Kliebenstein et al., 2001c, 2007; Hansen et al., 2007; Li et al., 2008). There was a significant positive correlation in the relative expression of the aliphatic biosynthetic genes in the myb28/myb29 double knockout and the number of identified TFs interacting with their promoters ($P < 0.001; R = 0.81; n = 16$; Sønderby et al., 2010a; Fig. 3B). There was also a significant positive relationship between the aliphatic gene expression within the myc2/myc3/myc4 triple mutant and the number of identified TFs interacting with their promoters ($P < 0.01; R = 0.46; n = 16$; Schweizer et al., 2013; Fig. 3C). Thus, the expression of 16 foliar aliphatic biosynthesis genes is regulated in a partially MYB- and MYC-independent manner, concordant with the number of TFs identified as binding to their promoters.

Figure 3. Relationship between the number of TFs binding a promoter and gene expression. A, Histogram of the number of TFs found to interact with different number of aliphatic biosynthetic gene promoters based on the Y1H analysis. B, Comparison of the relative expression of the aliphatic GLS genes within the myb28/myb29 knockout with the number of TFs found to interact with their promoters based on the Y1H analysis. The predicted linear trendline is shown. C, Comparison of the relative expression of the aliphatic GLS genes within the myc2/myc3/myc4 knockout with the number of TFs found to interact with their promoters based on the Y1H analysis. The predicted linear trendline is shown.

Y1H Survey of Aliphatic Regulatory and Peripheral Genes

Using the promoters for the three aliphatic-related MYBs (MYB28, MYB29, and MYB76), we identified 46 TFs that bound one or more of the promoters for these three MYBs, suggesting that they could act as putative upstream regulators (Supplemental Table S3). Thirty-six of these 46 TFs (78%) also interacted with an aliphatic biosynthetic gene promoter with a median of interacting with eight aliphatic biosynthetic gene promoters. This fraction of TFs interacting with both the aliphatic biosynthetic and aliphatic TF gene promoters...
is higher than expected by random chance when using the distribution of interactions detected using all 659 TF and 34 promoters to generate a random distribution (i.e. a null expectation) to empirically control for false positives and negative errors within this system ($\chi^2 = 17.64; P < 0.05$). Thus, over three-quarters of potential TFs that interact with the aliphatic MYB promoters also link to a significant fraction of biosynthetic gene promoters, suggesting a potential for the formation of feed-forward loops (Alon, 2006). Among these potential loops was a feedback loop identified by the interaction of the MYB28 TF with its own promoter. This is in agreement with previously observed nonlinear changes in gene expression in MYB28 transgenic plants (Sønderby et al., 2010b).

Interestingly, we found that 26 of the 44 TFs (59%) that interacted with promoters of genes peripherally related to the aliphatic GLS pathway could also interact with core aliphatic biosynthetic gene promoters. By contrast, less overlap was observed between TFs that interact with indolic GLS biosynthesis genes and those that interact with core aliphatic GLS biosynthetic genes (11 of 21 TFs with a median of 2 interactions). These data are supported by gene expression profiling experiments in which peripheral genes show a higher level of coexpression with genes in the aliphatic biosynthetic pathway, whereas the indolic GLS pathway shows much lower coexpression, suggesting that they are differentially regulated (Gigolashvili et al., 2007a, 2007b, 2008; Sønderby et al., 2007, 2010b, Wentzell et al., 2007).

**TF Binding Does Not Reflect Position within the GLS Biosynthetic Pathway**

Given the large number of potential candidate TFs identified in this survey, we wanted to query whether there were identifiable TF/promoter modules and whether these modules reflected position within the biosynthetic or regulatory pathways. We first conducted hierarchical clustering using Euclidean distance with the Ward clustering algorithm (Fig. 4). The promoters fit into four to six large groups dependent on the TFs with which they interacted (Fig. 4). Group I is a distinct cluster that contains no aliphatic biosynthetic gene promoters but instead contains MYB28 and peripheral and indolic promoters. This is in contrast with previously published data demonstrating that MYB28 most closely coexpresses with the aliphatic genes and not the indolic genes (Hirai et al., 2007) and indicates that TF binding behavior relative to coexpression is likely quite complex.

The other promoter groupings all contained promoters from the aliphatic biosynthetic genes and regulatory genes with no obvious linear relationship to the pathway. For example, cluster IVa included the promoter for MYB29, promoters for the core biosynthetic pathway genes CS-LYASE and SOT2, and the side chain modifying genes GSOH and GSOX3 (Fig. 4). These four genes do not work in a colinear fashion and instead are scattered throughout the pathway (Fig. 1A). This finding suggests that there may be modular TF interactions that do not require the targets to be linear within a biosynthetic pathway. A correlation analysis of TF interactions with respect to gene position within the biosynthetic pathway found no significant correlations between any TF and promoter position within the pathway (data not shown). A spring-embedded clustering approach showed a similar graph in which there was no observable relationship between pathway position and TF-promoter binding behavior for aliphatic GLS biosynthetic pathway genes (Fig. 5). Thus, our analyses demonstrate that TFs are likely to interact with promoters in nonlinear modules of the pathway (Figs. 4 and 5).

**Phenotypic Validation of GLS Regulatory Potential**

The Y1H analysis with the whole pathway identified a large number of candidate TFs that may regulate the aliphatic GLS pathway. However, there are currently no high-throughput chromatin immunoprecipitation assays that can rapidly screen through a large number of TFs with tissue-restricted expression. Thus, we decided to focus on the downstream phenotype of GLS metabolite accumulation in TF mutants as an approach to biologically validate these potential interactions. This approach has successfully identified GLS phenotypes in single regulatory TFs even when there are redundant genes providing a similar function (Sønderby et al., 2007, 2010a, 2010b). We focused on validating the potential regulatory activity of 29 TFs. These TFs were present in clusters that bind the most aliphatic GLS promoters or that clustered close to the known GLS TF MYB28 and had available homozygous transfer DNA (T-DNA) insertions (all TFs tested by T-DNAs are marked with a V in Fig. 4). In all possible cases, multiple alleles were identified for each gene (Supplemental Tables S4–S6). Obtaining mutants in both the dense and sparse clusters allow us to test whether overall levels in network connectivity are reflected in the magnitude of perturbations in aliphatic GLS quantities.

Using a randomized block design, we analyzed the accumulation of leaf and seed GLSs in the presence and absence of biotic stress for all TFs. Measuring GLS accumulation across multiple tissues and conditions allows us to sample a broader phenotypic spectrum to test whether TFs function only under specific conditions or developmental stages or more generally. To provide the biotic stress treatment, we utilized two growth chambers that have a similar abiotic environment, in which one was pest free (Controlled Environment Facility [CEF]) and one had an endogenous pest population (Life Sciences Addition [LSA]). We then utilized a three-way ANOVA with nesting to test whether insertions in a gene affected any aliphatic
Figure 4. Co-clustering of TFs and promoters via their interactions. Hierarchical clustering of promoter to TF interactions. Only promoters or TFs showing three or more interactions were included in the analysis, and the Ward algorithm was used for clustering. A V in front of the TF shows the TFs chosen to validate in mutant analysis. Within the plot, yellow shows a significant Y1H interaction between the promoter and TF, whereas red indicates no interaction. The promoters are color coded as shown in Figure 2 based on their pathway membership as follows: indolic GLSs are pink, aliphatic regulatory genes are purple, and aliphatic peripheral genes are red. The aliphatic biosynthetic genes are parsed into the core pathway (green), elongation (blue), modifying (yellow) and seed specific (orange) to visualize regulation along the linear model of the pathway. The roman numerals show potential promoter groupings.
or indolic GLSs across the experiments. The nesting term directly tested whether any GLS differences were due to individual insertions or if all insertions within a given TF had similar effects. Only results significant for the gene term (effect across all insertions) are discussed and insertion-specific events are ignored (Supplemental Tables S7 and S8).

The statistical analysis showed that all of the 29 TFs affected accumulation of at least one GLS with a median of nine GLSs affected across all TF insertion lines. To minimize the chance for false positive results, we established a baseline of a TF having to affect five GLS phenotypes because a random collection of gene mutations significantly affected levels of $3 \pm 1$ GLSs (Chan et al., 2011). This left 22 TFs that significantly altered the accumulation of at least five GLSs. Some of these TFs were previously characterized as having a role in abiotic stress responses, including COLD Responsive Binding Factor (CBF4; for full names and Arabidopsis Genome Initiative codes, see Supplemental Table S5; Haake et al., 2002; Zhou et al., 2011), Abscisic Acid-Responsive Element Binding Factor (ABF4; Choi et al., 2005; Xu et al., 2013), NAC Domain-Containing Protein102...
(NAC102; Christianson et al., 2009), RELATED TO AP2 (RAP2.6L; Che et al., 2006; Asahina et al., 2011; Krishnaswamy et al., 2011; Liu et al., 2012; Huang et al., 2013), and INDOLE-3-ACETIC ACID-LEUCINE RESISTANT3 (ILR3; Rampey et al., 2006; Long et al., 2010). Another set of TFs had previously been linked to growth and the cell cycle, including E2Fc (del Pozo et al., 2006; Berckmans et al., 2011), G-BOX BINDING FACTOR2 (GBF2; Schindler et al., 1992; Menkens and Cashmore, 1994; Terzaghi et al., 1997), GROWTH REGULATING FACTOR (GRF; (Hewezi et al., 2012; Casadevall et al., 2013), and AINTEGUMENTA (ANT; Elliott et al., 1996; Klucher et al., 1996; Krizek et al., 2000; Liu et al., 2000; Mizukami and Fischer, 2000; Horstman et al., 2014). Of these previously characterized TFs, only ERF9 had been previously associated with biotic stress and the resistance to necrotrophic fungi (Camehl and Oelmüller, 2010; Maruyama et al., 2013). This diversity of TF functions suggests that the aliphatic GLS biosynthetic gene promoters can function as a site of integration for a complex set of signals integrating across growth and defense to both abiotic and biotic stresses with respect to GLS levels.

If these TFs are central to integrating both diverse environmental and developmental inputs, then their effects should be conditional on the environment and developmental tissue in which we measured GLS levels. ANOVA analyses revealed that the majority of the statistically significant validations were specific to the growth conditions or tissue being tested (Table I). For example, ERF9 altered 10 GLSs in a tissue-dependent manner (Table I; Supplemental Tables S7–S9). By contrast, GBF2’s influence on GLS levels had an increased dependency on stress conditions (Table I; Supplemental Tables S7–S9). In total, 22 of the 29 TFs were able to significantly alter the accumulation of multiple GLS in a tissue- or stress-dependent manner arguing that promoters in biosynthetic pathways may allow for specificity to integrate signals.

Table I. Insertion mutant validation of putative TFs

Data are presented as the number of statistically significant tests for the gene term out of 21 different phenotypes. Gene shows the number of phenotypes that showed as significant difference between the wild type and the insertion mutants in the specific gene listed. This is after controlling for potential variation between the multiple insertion lines using nested ANOVA. KO indicates knockout and shows the number of independent T-DNA insertions tested per TF. Using the same exact model, Chamber:Gene shows the number of phenotypes with a significant interactions between the gene and chamber, and Tissue:Gene shows the number of phenotypes with a significant interaction between tissue and gene. The three-way term shows the number of phenotypes that were controlled by an interaction of chamber, tissue, and gene. The name column shows the published literature name for any gene with an identifiable name, and class shows the class of TF with which a gene is associated.

<table>
<thead>
<tr>
<th>Arabidopsis Genome Initiative Code</th>
<th>Name</th>
<th>Class</th>
<th>KO</th>
<th>Total</th>
<th>Gene</th>
<th>Chamber: Gene</th>
<th>Tissue: Gene</th>
<th>TGC</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT1G12630</td>
<td>ERF/AP2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>AT1G61730</td>
<td>Storekeeper</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>AT1G12610</td>
<td>DDF1</td>
<td>ERF/AP2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>AT2G30590</td>
<td>WRKY21</td>
<td>WRKY</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>AT5G15210</td>
<td>HB3</td>
<td>HB</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>AT4G37260</td>
<td>MYB73</td>
<td>MYB</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>AT2G44730</td>
<td>MYB-Like</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>AT3G33340</td>
<td>NF-YB1</td>
<td>NF-Y</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>AT1G68360</td>
<td>C2H2 ZF</td>
<td>1</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>AT3G1580</td>
<td>AP2/B3</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>AT1G76880</td>
<td>MYB-Like</td>
<td>2</td>
<td>8</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>AT3G19290</td>
<td>ABF4</td>
<td>hZIP</td>
<td>2</td>
<td>9</td>
<td>4</td>
<td>0</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>AT1G47870</td>
<td>E2FC</td>
<td>E2F</td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>AT5G44210</td>
<td>ERF9</td>
<td>ERF/AP2</td>
<td>3</td>
<td>11</td>
<td>3</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>AT5G63790</td>
<td>NAC102</td>
<td>NAC</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>AT3G28920</td>
<td>HB34</td>
<td>HB</td>
<td>3</td>
<td>9</td>
<td>6</td>
<td>5</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>AT5G13330</td>
<td>RAP2.6L</td>
<td>ERF/AP2</td>
<td>2</td>
<td>9</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>AT1G20910</td>
<td>Arid/Bright</td>
<td>2</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>AT2G28440</td>
<td>GRF1</td>
<td>C3H ZF</td>
<td>1</td>
<td>9</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>AT2G02540</td>
<td>HB21</td>
<td>HB</td>
<td>1</td>
<td>12</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>AT1G24625</td>
<td>ZFP7</td>
<td>C2H2 ZF</td>
<td>1</td>
<td>11</td>
<td>5</td>
<td>4</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>AT1G4880</td>
<td>Arid/Bright</td>
<td>2</td>
<td>12</td>
<td>10</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>AT4G01120</td>
<td>GBF2</td>
<td>hZIP</td>
<td>1</td>
<td>13</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>AT5G51990</td>
<td>CBF4</td>
<td>ERF/AP2</td>
<td>1</td>
<td>13</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>AT5G20200</td>
<td>ERF/AP2</td>
<td>1</td>
<td>13</td>
<td>4</td>
<td>8</td>
<td>3</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>AT5G61590</td>
<td>ERF/AP2</td>
<td>1</td>
<td>11</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>AT1G66140</td>
<td>ZFP4</td>
<td>C2H2 ZF</td>
<td>2</td>
<td>17</td>
<td>8</td>
<td>11</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>AT5G54680</td>
<td>ILR3</td>
<td>hHL</td>
<td>2</td>
<td>19</td>
<td>16</td>
<td>6</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>AT4G37750</td>
<td>ANT</td>
<td>AP2</td>
<td>1</td>
<td>21</td>
<td>15</td>
<td>16</td>
<td>18</td>
<td>9</td>
</tr>
</tbody>
</table>
Connectivity versus Phenotypic Validation

Systems biology studies like to focus validation efforts on hub genes, genes that show the highest level of connectivity, with the assumption that these genes are more likely to have phenotypic consequences. For instance, it was previously shown in Arabidopsis that genes bound by a relatively large number of TFs are significantly positively correlated with conferring a morphological phenotype (Brady et al., 2011). Thus, we proceeded to test this assumption using the phenotypes of our TF mutants compared with their connectivity. Within our data, there was no correlation between the number of promoters that a TF interacted with and the number of GLSs affected by in the corresponding mutant (n = 29; R²=0.06; P = 0.185). Thus, filtering Y1H data sets simply by the number of interactions will not provide an increase in the successful validation of new genes controlling a phenotype. Similarly, there was no pattern in the level of connectivity within the spring-embedded network diagram. Finally, the three TFs whose insertional mutants had no phenotype were among the most connected at the center of the diagram (Fig. 5). This has also been observed in yeast (Saccharomyces cerevisiae) pseudohyphal growth and in a Caenorhabditis elegans neuronal gene regulatory network, suggesting that it is a broader phenomenon than plant secondary metabolism (Bormann et al., 2006; Vermeirssen et al., 2007). In addition, this finding suggests that transcriptional regulation of secondary plant metabolism may be distinct from that of cell-type specification and differentiation and organ morphogenesis (Taylor-Teeples et al., 2014).

ANT and ILR3 Are Strong Regulators of GLS Biosynthesis

ANT, a TF important for controlling plant organ size, growth, and cell number, was the TF with the strongest effects both in terms of effect size and number of GLSs affected (Table I; Supplemental Table S9; Klucher et al., 1996; Mizukami and Fischer, 2000). However, we could only identify interactions of ANT with promoters of four aliphatic biosynthetic genes (Fig. 5). Although ANT affected aliphatic GLSs in all conditions, the direction and magnitude changed dependent on the condition (Fig. 6). ANT also influences leaf indole GLS accumulation even though we identified no promoter interactions with any indolic GLS-associated promoter, suggesting additional feedback between aliphatic and indolic GLS biosynthesis (Fig. 6). In contrast with ANT, loss of ILR3, a bHLH TF that has been connected to auxin conjugation and response to iron deprivation, had consistent effects on aliphatic GLS accumulation in the leaves but no effect in the seed (Rampey et al., 2006; Long et al., 2010; Fig. 6). This suggests that ILR3 functions as a tissue-specific repressor of aliphatic GLS biosynthesis (Fig. 6).

Phenotypic Network Analysis

To relate the phenotypic effects of the TF insertional mutations to the pattern of Y1H interactions we clustered the genes by their effect on aliphatic GLS accumulation compared with Col-0. We found that the TFs clustered into eight groupings (Fig. 7). Group C closely clustered with Col-0 and contained all of the TFs with the fewest significant phenotypic changes in either long-chain or short-chain aliphatic GLS accumulation (Figs. 7 and 8). The rest of the clusters all showed a significant effect on short-chain or long-chain aliphatic GLS accumulation of a magnitude similar to that seen in the single knockouts of the validated MYB28/MYB29 and MYB76 (Fig. 8). To our knowledge, although the MYB insertions always lead to lower GLS accumulation, most TF mutants in this study led to higher GLS in the absence of biotic stresses (Fig. 8, CEF chamber). However, this increase disappeared in the presence of pests (Fig. 8, LSA chamber), suggesting that they are TFs that repress aliphatic GLS accumulation in the absence of biotic or abiotic stress. Whether negative GLS regulators are more numerous than positive regulators, or the assay is more prone toward identifying them, is yet to be determined. However, these stele-expressed
TFs appear to act as general repressors of aliphatic GLS accumulation rather than activators.

In contrast with this general pattern, mutants in TFs within cluster B (containing HOMEBOX PROTEIN21 [HB21], GBF1, and ERF/AP2s At5g52020 and At5g61590) were most similar to the effects seen in ANT, showing opposing effects with higher aliphatic GLS in the pathogen-free conditions and lower aliphatic GLS in response to pathogens (Fig. 8). Cluster B, however, did not display the seed effects of ANT, suggesting that these TFs function with different tissue specificity than ANT. TF group E (containing ILR3, HB30, C2H2 Zinc Finger At1g68360, and ARID/Bright At1g20910) showed the most consistent effects across conditions (Fig. 8). Thus, we can use Y1H to identify TFs that affect aliphatic GLS accumulation in a range of conditions with a high rate of validation.

**TF/Promoter Interactions Link to Metabolic Consequence in TF Mutants**

The above analysis found that T-DNA insertion mutants in 22 of the 29 tested TFs had significant effects on GLS accumulation. This shows that we can utilize the Y1H approach to identify TFs for the aliphatic GLS pathway with a success rate of 75% (Fig. 5). However, no clear pattern was observed with respect to TF-promoter interaction and specific mutant effects on GLS accumulation for the TFs (Figs. 4 and 7). Thus, we used a variety of linear modeling approaches (multiple regression and principal component) to test for a correlation between the TF-promoter interactions and their mutants’ associated metabolic changes. No such correlation was observed, suggesting nonlinear control of TF regulation of GLS biosynthesis (data not shown).

The above approaches to link mutant phenotype and promoter binding of specific TFs are limited because they function by focusing on linear analysis of the connections. To avoid the limitations of linear approaches, we tested whether it was possible to identify connections in the entire data set using a Mantel test (Mantel, 1967). The Mantel test allows us to conduct a correlation of all TF-promoter interactions against the GLS phenotypes of all of the TF mutants in a single nonlinear test. The Mantel test conducts a comparison of matrices. As such, we used the TF-promoter interactions and the relative GLS phenotypes of the TF mutants to generate two separate distance matrices. One distance matrix describes the similarity of how the TFs interact with the tested promoters, whereas the other distance...
also a significant correlation between TF-promoter interactions and indolic GLS phenotypes in the seeds independent of stress treatment (CEF, $r = 0.2607; P = 0.014$; and LSA, $R = 0.36; P = 0.001$). This significant influence of the TFs on indolic GLS phenotypes is in contrast with the fact that we found these TFs based on their binding to aliphatic GLS-related promoters, and further supports our conclusion that regulation of aliphatic GLS biosynthesis is interconnected with indolic GLS biosynthesis, likely in an indirect manner. Using the leaf GLS data, we only found a single significant relationship between the TF-promoter matrix under nonstress conditions with indolic GLS ($R = 0.26; P = 0.0480$). No significant interactions were found for aliphatic GLS under either treatment, or for indolic GLS under stress treatment.

Thus, the Mantel tests show that there is a significant similarity between the pattern of how the TFs interact with the aliphatic GLS promoters and the resulting GLS phenotype of the TF mutants in seeds. This relationship is strongest within the seeds potentially because the seeds contain GLS transported from the maternal plant and thus represent an integrative measurement of GLS regulation across the life span of the maternal plant (Nour-Eldin et al., 2012; Andersen et al., 2013). By contrast, the leaves represent a more temporal measurement specific to that time and condition and may require more spatial or temporal information to identify the linkage between promoter interaction and metabolic phenotype for the different TFs. In addition, the TF library is generated from stele-expressed TFs, which may indicate that this collection of TFs provides more information about loading of GLS into seeds than it does about foliar GLS accumulation (Gaudinier et al., 2011). It will require the use of a complete collection of Arabidopsis TFs to test whether we can generate a similarly significant relationship between TF-promoter interactions and foliar GLS accumulation.

cis-Element Motif Enrichment Largely Reflects TF Binding

To test whether there was a link between the cis-elements within a promoter and the Y1H identification of associated TF families, we assessed if specific TF families were enriched in their Y1H identification across the aliphatic GLS pathway promoters (Supplemental Table S4). This showed that the bZIP, AP2/ERF, ABSCISIC ACID INSENSITIVE3, GLABROUS ENHANCER BINDING PROTEIN, and Tribelix TF families interacted more often with the aliphatic GLS pathway gene promoters than expected by chance given their occurrence in the Y1H library. This agreed well with the enrichment of bZIP, AP2/ERF, and Tribelix binding motifs in these promoters (Supplemental Table S2). Our finding that nearly all of the TFs in these families that we tested affect aliphatic GLS accumulation suggests a previously unrecognized regulatory role for these TFs in Arabidopsis GLS biosynthesis.

By contrast, the enrichment in MYB, Homeodomain, and bHLH binding motifs in these promoters did not

---

**Figure 8.** Average GLS change in TF groups. The average change in short-chain and long-chain aliphatic GLSs of the TF groups A to G as defined by the phenotypic clustering are shown with 95% confidence intervals for the group change. The corresponding changes in ANT are also shown. For comparison, the average changes in the leaf and seed phenotypes in single insertion mutants in MYB28, MYB29, and MYB76 are shown. For the MYBs, the $r$ of the phenotype across published results in these same chambers is shown. Blue bars are values from leaves, with dark blue showing the CEF chamber and light blue showing the LSA chamber. Green bars are from seeds, with dark green showing the CEF chamber, and light green is from seeds in the LSA chamber. To estimate the average group changes for the TF clusters shown in Figure 7, we took the average change of each TF mutant in relation to wild-type Col-0. We then grouped the TFs into the appropriate groups based on Figure 7. We then estimated the average percent change and 95% confidence interval for each group of TFs. This was done for the total level of short-chain and long-chain aliphatic GLS. LC, Long chain; SC, short chain.
The TFs identified as aliphatic GLS regulators in this study include a number of known TFs controlling the responses to a variety of abiotic and biotic stresses as well as internal growth and development switches (Table I). This suggests that these TFs provide the ability to integrate a complex mixture of signals to regulate the pathway (Fig. 9). Furthermore, the fact that the TFs caused independent alterations in the amount and profile of aliphatic GLSs suggests that this integration can occur by interaction of the TFs with nonlinear components of the pathway. For example, comparing the predicted binding capacity of ABF4 and ANT shows that there is both overlap and independence in their promoter binding and the chemical profile of mutants in these genes (Figs. 7 and 9). This modularity involved some TFs such as ABF4 having the capacity to interact with the known MYB promoters allowing for feed-forward loops to regulate the pathway (Figs. 4 and 9). By contrast, other TFs such as ANT did not show any interaction with the MYBs but instead only connected directly to the biosynthetic gene promoters (Figs. 4 and 9). Thus, it is likely that this metabolic pathway has a mixture of direct regulatory processes that bypass the major MYC/MYB regulators as well as feed-forward loops that incorporate the MYC/MYB regulators.

This diverse set of independent and overlapping regulatory links between TFs and the aliphatic GLS pathway is potentially a way to optimize the ability of the aliphatic GLS to protect the plant at minimal cost in a highly diverse biotic environment. For example, a number of the TF mutants lead to a significant change in the blend of aliphatic GLS within the plant. This ability to modulate the blend may be a response to the fact that each aliphatic GLS is most effective as a defense against a specific set of biotic attackers (Lambrix et al., 2001; Kliebenstein et al., 2002; Kroymann and Mitchell-Olds, 2005; Hansen et al., 2008; Stotz et al., 2011; Züst et al., 2012). Thus, altering the blend may be a way to optimize the defense against a specific set of biotic pests and away from another (Wentzell and Kliebenstein, 2008; Burrow et al., 2009). By having a mixture of TFs that directly bypass and TFs that create feed-forward loops with the MYBs, this creates a system in which the plant may be able to precisely fine-tune the regulation of the biosynthetic genes to optimize the defense of the system using a myriad of inputs including biotic, abiotic, and developmental signals.

**Master Regulators in Development versus Defense Metabolism**

Studies routinely classify genes as master or key regulators controlling a process. In developmental biology,
this term reflects the ability of a single TF to be both necessary and sufficient to drive a developmental program or cell identity. MYB28 and MYB29 were proposed to act as a master regulator of this process because their absence results in a complete absence of GLS biosynthesis and because their transcript levels are strongly correlated with GLS abundance. However, the master regulator terminology may be a misleading interpretation because it assumes that the biological pathway associated with the trait being measured has a largely hierarchical regulatory architecture. In contrast with this assumption, it has long been known that development and defense metabolism should be differently regulated (Waddington, 1942). Developmental characters such as floral tissue specification must be tightly coordinated and likely involve irreversible regulatory switches. An organism gains no benefit from having stamens and pistils rapidly interconverting because their functions are defined by their end states. In plants, these irreversible regulatory switches comprised the homeotic floral regulators.

By contrast, metabolism is characterized by rapid interconversion between compounds that is likely critical to allow for specific fine-tuning of metabolism (Samal and Jain, 2008; Chandrasekaran and Price, 2010; Watson and Walhout, 2014). Thus, metabolism likely has very few if any irreversible regulatory switches and a greater potential for fine-tuning. Hence, a myb28/myb29 double mutant that abolishes a compound without abolishing the majority of pathway transcripts does not fit the concept of master regulator as defined by classical studies of developmental processes. The idea that metabolic and developmental pathways may have different regulatory architectures is supported by systems studies, which show that the molecular architecture of regulatory networks is highly dependent on the phenotype or output that is being controlled (Milo et al., 2002, 2004; Sullivan et al., 2014).

Molecular Validation

We used high-throughput metabolic profiling to validate the potential of the identified TFs to alter metabolite accumulation. The goal of this approach was to identify a small subset of candidates upon which to conduct further molecular validation studies. Instead, however, we were able to find an effect of 75% or higher of the tested TFs mutants upon aliphatic GLS accumulation. Although it is tempting to assume that these are indirect effects, the 75% rate of validation is dramatically higher than a random null sample of genes, and these genes were identified via a potential direct in yeast interaction with promoters for aliphatic GLS biosynthetic genes (Fig. 4; Chan et al., 2011). Current methods to validate direct regulatory interactions would require performing transgenic plant generation to independently develop lines containing tagged versions of each putative TF and then conducting a chromatin immuno-precipitation assay in each line coupled with high-throughput expression profiling. Furthermore, this would have to be conducted in multiple tissues and environments given the fact that the TFs were highly conditional in their effect. This illuminates a critical need to develop new methodologies to conduct high-throughput in vivo validation under a variety of abiotic and biotic stress treatments of potential TF-promoter interactions or other molecular interactions to fully develop the complete model for how this pathway is regulated.

One potential method to provide further evidence of direct regulatory interactions would be to use the regulatory network model to identify putative double mutants that would epistatically interact and create significantly larger or even opposing effects upon the pathway. For example, our model predicts that ANT directly regulates the aliphatic GLS pathway as a repressor, whereas MYB28/MYB29 activates the pathway. Because these two processes are independent, our model would predict that a triple myb28/myb29/ant mutant would potentially rescue aliphatic GLS accumulation. Similar work in yeast has suggested that double mutants between regulatory modules have epistatic effects in contrast with those within modules (Segrè et al., 2005). Thus, it may be possible to further refine the direct regulatory module by working to predict the phenotype of double or higher order mutants.

CONCLUSION

By utilizing greater than 90% of the promoters for aliphatic biosynthetic genes with a large, focused Y1H library, we were able to develop a better image of the potential scale of GLS pathway transcriptional regulation. Incorporating metabolic profiling, we were able to show that there are at least dozens of TFs that have the capacity to significantly affect this pathway’s function. Furthermore, the effects of these TFs effects are highly conditional, indicating that the aliphatic GLS pathway is regulated to optimize growth and defense in a multifaceted environment. Future work is required to fully map out the molecular connections in this pathway and to expand this analysis to the full Arabidopsis complement of TFs. This will then allow the development of a regulatory model that is more inclusive of multiple environmental and growth inputs then have currently been proposed.

MATERIALS AND METHODS

Promoter Cloning

PCR primers were designed to amplify about 2,000 bp of each of the aliphatic GLS promoters upstream of the predicted translational start codon or the promoter sequence to the start of the next gene. The promoters were amplified by Phusion High-Fidelity DNA Polymerase (M0530S; New England Biolabs) and cloned into pENTR 5’ TOPO TA Cloning vector (K991-20; Invitrogen). The promoters were then transferred into the pMW2 and pMW3 destination vectors by LR reaction (11791-019; Invitrogen) and confirmed by DNA sequencing (Gaudinier et al., 2011).
MYB TF Cloning

To include the known aliphatic GLS MYB TFs MYB28, MYB29, and MYB76 into the existing YIH library, we cloned their open reading frames from Col-0 complementary DNA of leaf material by Phusion High-Fidelity DNA Polymerase (M05085), and the PCR products were cloned into pENTR Directional TOPO Cloning vector (K2400-20; Invitrogen). Finally, the three MYB TFs were transferred into the pDESTAD-2a destination vector and confirmed by DNA sequencing (Gaudinier et al., 2011).

YIH

The root vascular-expressed TF collection used for the YIH screening is as previously described (Gaudinier et al., 2011). The MYB28, MYB29, and MYB76 proteins were added to this collection. All of the GLS bait promoters were screened against the stele-expressed TF collection using the YIH protocol as previously described (Gaudinier et al., 2011; Taylor-Teeples et al., 2011).

Plant Material and Experimental Conditions

For the GLS analysis of the T-DNA insertion lines, Arabidopsis (Arabidopsis thaliana) plants were grown in two different controlled-environment chambers with 16-h light at 100- to 120-mE light intensity with two complete independent replicates of the whole experiment. One growth chamber (CEF) is routinely cleaned and sterilized to maintain it in a pest-free format and used exclusively for the Arabidopsis study. The other walk-in growth chamber (LSA) is used for multiple plant species, including Arabidopsis, tobacco (Nicotiana tabacum), and tomato (Solanum lycopersicum), and has an endogenous pest population composed of a mix of aphids and fungus gnats. This chamber is a model of a stressful pest and pathogen environment. In each growth chamber, each line was planted in 6-fold replication with four other mutants and wild-type controls within a complete block design. Wild-type Col-0 controls were included in each flat to minimize any spatial aspects of the growth chamber. This experiment was replicated in both growth chambers providing 12 biological repeats in total for most of the mutant genotype.

T-DNA Lines

T-DNA lines for the selected TFs were ordered from the Arabidopsis Biological Resource Center (Sussman et al., 2000; Alonso et al., 2003), and lines with homozygous insertions were identified by PCR-based genotyping (Supplemental Tables S5 and S6). Multiple independent insertions within each gene were obtained and phenotyped where possible (Table I).

GLS Analysis

Four weeks after sowing, we harvested a leaf from each plant for GLS analysis. One leaf from the first fully mature leaf pair of each plant was removed and stored in 90% (v/v) methanol to inhibit enzymatic breakdown of chemical compounds and begin the extraction. The plants were then allowed to continue to develop and produce seeds. We collected seeds from each plant individually when survived and extracted GLSs from 40 seeds. GLSs from both tissues were extracted and analyzed by HPLC according to previously described methods (Kliebenstein et al., 2001b).

Statistical Analysis

To test the effect of each TF upon GLS accumulation, all GLSs were analyzed via ANOVA using a general linear model within R software (R Development Core Team, 2014). This was done using a nested model to account for potential differences between independent insertions when comparing the wild type with the mutants. The following model was used to test for differences in GLS accumulation in the mutants within each specific gene to the respective wild-type Col-0. We then grouped the TFs into the appropriate groups based on the tissue series or the biostress series as the control matrix. The root vascular-expressed TF collection used for the YIH screening is as previously described (Gaudinier et al., 2011; Taylor-Teeples et al., 2011).

Maternal and paternal interactions were transformed into distances matrices in R. To generate a TF/promoter interaction matrix, the TF-promoter interactions were recorded in two states (0 = no interaction observed; 1 = interaction observed), and the distance matrix was computed using binary correlation. Binary correlation was utilized because the data are in a binary format. Only TFs tested by TF-TDNA lines were included in creating the matrix. To test the effect of TF insertions on glucosinolate accumulation, all GLSs were analyzed via ANOVA using a general linear model within R software (R Development Core Team, 2014). All presented P values are false discovery rates adjusted to 0.2 and least square means are presented (Supplemental Tables S7 and S8). The model was run independently for each TF.

Matrix Comparisons

To test for a link between the matrix of interactions between TFs and GLS promoters and the matrix of GLS phenotypes in the TF knockout mutants, we utilized the vegan R package for running Mantel and partial Mantel tests (Dixon, 2003; Oksanen et al., 2013; R Development Core Team, 2014). Mantel and partial Mantel tests were implemented using the Pearson’s coefficient with 1,000 permutations to test significance. The Mantel test calculates the correlation between two distance matrices. The TF-promoter interaction matrix and the relative mutant TF-GLS phenotype matrix (Supplemental Table S2) were transformed into distances matrices in R. To generate a TF/promoter interaction matrix, the TF-promoter interactions were recorded in two states (0 = no interaction observed; 1 = interaction observed), and the distance matrix was computed using binary correlation. Binary correlation was utilized because the data are in a binary format. Only TFs tested by TF-TDNA lines were included in creating the matrix. To test the effect of TF insertions on glucosinolate accumulation, all GLSs were analyzed via ANOVA using a general linear model within R software (R Development Core Team, 2014). All presented P values are false discovery rates adjusted to 0.2 and least square means are presented (Supplemental Tables S7 and S8). The model was run independently for each TF.

Group Changes

To test the effect of each TF upon GLS accumulation, all GLSs were analyzed via ANOVA using a general linear model within R software (R Development Core Team, 2014). This was done using a nested model to account for potential differences between independent insertions when comparing the wild type with the mutants. The following model was used to test for differences in GLS accumulation in the mutants within each specific gene to the respective wild-type Col-0. We then grouped the TFs into the appropriate groups based on the tissue series or the biostress series as the control matrix. This was done independently for the aliphatic and indolic GLSs, creating eight total TF/metabolite matrices.

We also generated a third set of distance matrices to describe the TF relationships using the gene expression levels of the same TFs across stress and developmental conditions. Average gene expression values for the TFs were obtained from the Arabidopsis Expression Browser (Bio-Analytic Resource, http://bar.utoronto.ca), specifically the tissue and biostress series within this database (Patel et al., 2012). The average expression of the TFs across the data sets were then used to create coexpression matrices with Euclidean distance (TF/expression matrices). We then tested whether these TF/expression matrices altered the relationship between the TF/promoter and TF/metabolite matrices using partial Mantel tests. This was conducted separately using either the tissue series or the biostress series as the control matrix.

Supplemental Data

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

Supplemental Table S1. Promoters cloned for YIH analysis.

Supplemental Table S2. Promoter motif enrichment in the tested promoters.

Supplemental Table S3. List of all promoter TF interactions identified.

Supplemental Table S4. TF family enrichment in interactions with aliphatic glucosinolate promoters.

Supplemental Table S5. T-DNAs used to validate specific transcription factor capacity to affect aliphatic glucosinolate accumulation.

Supplemental Table S6. Primers used for validating T-DNAs and cloning MYB promoters.

Supplemental Table S7. P values of significance for models testing the effect of TF insertions on glucosinolate.

Supplemental Table S8. Type III sums-of-squares for models testing the effect of TF insertions on glucosinolate.

Supplemental Table S9. Least square means and standard error for glucosinolate accumulation in the TF insertions and the comparator Col-0 samples.


R Development Core Team (2014) R: A Language and Environment for Statistical Computing, R Project for Statistical Computing, Vienna


