A Framework for Engineering Stress Resilient Plants Using Genetic Feedback Control and Regulatory Network Rewiring

Mathias Foo,†,‡,∥ Iulia Gherman,† Peijun Zhang,‡ Declan G. Bates,*† and Katherine J. Denby*‡,§

Warwick Integrative Synthetic Biology Centre, School of Engineering, University of Warwick, Coventry CV4 7AL, United Kingdom
Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, United Kingdom
Department of Biology and Centre for Novel Agricultural Products, University of York, York YO10 5DD, United Kingdom

Supporting Information

ABSTRACT: Crop disease leads to significant waste worldwide, both pre- and postharvest, with subsequent economic and sustainability consequences. Disease outcome is determined both by the plants’ response to the pathogen and by the ability of the pathogen to suppress defense responses and manipulate the plant to enhance colonization. The defense response of a plant is characterized by significant transcriptional reprogramming mediated by underlying gene regulatory networks, and components of these networks are often targeted by attacking pathogens. Here, using gene expression data from Botrytis cinerea-infected Arabidopsis plants, we develop a systematic approach for mitigating the effects of pathogen-induced network perturbations, using the tools of synthetic biology. We employ network inference and system identification techniques to build an accurate model of an Arabidopsis defense subnetwork that contains key genes determining susceptibility of the plant to the pathogen attack. Once validated against time-series data, we use this model to design and test perturbation mitigation strategies based on the use of genetic feedback control. We show how a synthetic feedback controller can be designed to attenuate the effect of external perturbations on the transcription factor CHE in our subnetwork. We investigate and compare two approaches for implementing such a controller biologically—direct implementation of the genetic feedback controller, and rewiring the regulatory regions of multiple genes—to achieve the network motif required to implement the controller. Our results highlight the potential of combining feedback control theory with synthetic biology for engineering plants with enhanced resilience to environmental stress.

KEYWORDS: plant synthetic biology, plant–pathogen interaction, synthetic gene circuits, feedback control, network rewiring, plant defense response

Unfavorable environmental conditions during the growth of crop plants can cause significant yield loss and reduction in quality. These conditions include abiotic stresses, such as drought and extreme temperature, as well as the biotic stresses of disease and herbivory. Climate change is driving increasingly unpredictable and variable weather, and bringing associated change in pathogen (and hence disease) prevalence and incidence. It is therefore important to develop crops that are resilient to varying conditions and able to maintain yield in suboptimal environments. The introduction and/or removal of single genes via genetic engineering has led to plants with enhanced tolerance to particular abiotic and biotic stresses, however, often such approaches have unintended consequences on other plant responses, and in the case of disease resistance they may not be durable. Recent increased understanding of how plant responses to different environmental conditions are controlled and integrated, together with the development of systems biology approaches, has opened up the possibility of designing stress resilient crops using engineering principles. In this work, we have focused on transcriptional regulation, as transcripational reprogramming is a significant component of plant stress responses and a point of cross-talk between responses to different stresses.

In this paper we focus on the regulation of the defense response induced in Arabidopsis by the fungal pathogen, Botrytis cinerea. When pathogens infect plants, disease is the result of dynamic interactions between the two organisms. Pathogens secrete a range of proteins, small RNAs and metabolites to disrupt host defense and manipulate the extra- and intracellular environment to aid colonization. This is thought to explain why some positive regulators of defense are downregulated during infection, for example expression of TGA3 decreases during B. cinerea infection of Arabidopsis, yet plants lacking TGA3 expression are more susceptible to this pathogen.

In this study, we use a control engineering approach to counteract such potentially pathogen-mediated perturbations of positive regu-
lators of defense. Constitutive overexpression of such positive regulators would be an obvious approach, but this brings significant drawbacks; the positive regulator of defense may have other roles in the plant which are disrupted due to constitutively high levels of expression, and constitutive activation of plant defense responses is known to often impact on growth.\textsuperscript{19} Our proposed approach, which seeks to dynamically respond to perturbations of expression over the time-course of infection, should overcome these drawbacks.

From the perspective of control engineering, this scenario can be naturally formulated mathematically as a disturbance attenuation problem. For such problems, control engineers have developed a variety of powerful theoretical tools and techniques that allow the design of feedback controllers that can attenuate the effects of external perturbations on the functioning of a system or network (see ref 16 and references therein). The application of these tools to the analysis and design of complex biological networks is now attracting significant interest within the synthetic biology community.\textsuperscript{17,18} To date, however, the potential usefulness of such approaches for engineering more resilient plants has not been investigated.

Here, we explore how combining control engineering design tools\textsuperscript{19−21} with synthetic biology techniques could be used to enhance resistance against \textit{B. cinerea} in \textit{Arabidopsis} by preventing downregulation of a positive regulator of defense during infection. We design and test our controller using a model of the \textit{Arabidopsis} gene regulatory subnetwork underlying the transcriptional response to \textit{B. cinerea} infection. This network model is formulated using ordinary differential equations (ODEs) and constructed from experimental data using network inference and system identification techniques. It is then validated against different time-series transcriptome data sets capturing the response of the plant’s regulatory network to pathogen attack. Simulation results show the capability of the proposed approach to significantly reduce the perturbation of a positive regulator of plant defense in response to infection. We propose a novel strategy for implementing the controller experimentally, which avoids the need for the incorporation of any exogenous synthetic control circuitry. This strategy is based on the insight that the network motif required for the controller can be implemented by rewiring the regulatory regions of existing genes in the plant’s stress-response network. We show how this can be done through the addition of gene coding sequences under the control of alternative regulatory regions.

### RESULTS

**Inferring the Regulatory Subnetwork Containing a Positive Regulator of Defense.** We previously generated a high-resolution time series of the \textit{Arabidopsis} transcriptome during the first 48 h after inoculation by the pathogen \textit{B. cinerea}.\textsuperscript{10} Nearly 10,000 genes were identified as being differentially expressed in infected leaves compared to mock-inoculated leaves, including 883 TFs (Supplementary File S1a). We used the time-series transcriptome data for the differentially expressed TFs as input for network inference algorithms, to generate causal directed network models of the regulatory events underlying changes in expression of these TF genes. The algorithms chosen for this purpose (GENIE3,\textsuperscript{22} TIGRESS\textsuperscript{23} and Inferelator\textsuperscript{24}), were highly ranked in a recent assessment of network inference algorithms.\textsuperscript{25} GENIE3 approaches network inference as a tree-based regression problem and came first in the DREAM4 in silico multifactorial network inference challenge.\textsuperscript{25} Inferelator and TIGRESS both use feature selection and least angle regression to rank the potential regulators of a gene. The outputs from these three algorithms were used to generate a consensus network model, as a robust way of generating high confidence networks.\textsuperscript{26} A threshold (edges $\leq 10$ times the number of nodes) was applied to this consensus network to limit it to 8830 edges. Furthermore, only the top three regulators of each node were kept based on the highest probability score. From this final network, we looked for subnetworks surrounding positive regulators of defense against \textit{B. cinerea} that were downregulated during infection. This led us to focus on a 9-gene regulatory network, termed 9GRN (see Figure 2) containing the TF CHE, which includes predicted upstream regulators of CHE.

**CHE is a Positive Regulator of Defense against \textit{B. cinerea}**. Expression of the transcription factor (TF) CCA1 HIKING EXPEDITION (CHE) is downregulated during \textit{B. cinerea} infection (ref 10 and Figure 1a). Rhythmic expression of CHE is clear in the mock-inoculated samples (reflecting the role of CHE within the circadian clock\textsuperscript{27}) with downregulation due to infection beginning around 22 h post inoculation. A mutant with significantly reduced expression of CHE, che-1,\textsuperscript{27}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image.png}
\caption{Expression and role of CHE during infection with \textit{B. cinerea} (a) Expression of the TF CHE is downregulated during \textit{B. cinerea} infection of \textit{Arabidopsis} leaves. Leaves were drop-inoculated with \textit{B. cinerea} spores or mock-inoculated, and genome-wide gene expression determined every 2 h for both mock treatment (blue) and \textit{B. cinerea} infection (red). Open circles are the average of four biological repeats with bars representing standard deviation. This data is extracted from Windram et al.\textsuperscript{10} (b) CHE is a positive regulator of defense against \textit{B. cinerea}. Lesion size of \textit{Arabidopsis} leaves (n = 17) drop-inoculated with \textit{B. cinerea} spores were measured 36 and 72 h post infection. che-1 is an \textit{Arabidopsis} mutant with significantly reduced CHE expression. WT is the wildtype Col-0 \textit{Arabidopsis} accession. Error bars represent standard deviation, ** represents p $\leq 0.01$ and *** represents p $\leq 0.001$.}
\end{figure}
Validation of Edges in the 9GRN Model. To increase our confidence in the validity of the inferred 9GRN subnetwork model, we used yeast-1-hybrid (Y1H), a partial Arabidopsis cistrome map, and gene expression data from RAP2.6L overexpressors to test regulation predicted by the model. A set of pairwise Y1H had been carried out testing binding of 75 TFs to the promoter regions of 34 of the same TFs. Within this set, there were 4 edges in our model (RAP2.6L to ANAC055; ANAC055 to RAP2.6L, ANAC055 to ORA59 and at-ERF1 to ORA59) that had been tested. For two of these edges, strong binding was seen in the Y1H experiments; RAP2.6L could bind to the promoter of ANAC055 and at-ERF1 could bind to the promoter of ORA59 (Figure S2). In addition, the Y1H data suggested two additional edges that were missing from our model (RAP2.6L to ORA59, and ORA59 to ANAC055), however, expression data from RAP2.6L overexpressors and knockout mutant of ORA59 do not show any evidence for these regulatory edges. Additional interactions in the 9GRN were verified using data from an Arabidopsis cistrome map. The cistrome is the complete set of cis-elements or TF binding sites in an organism, and a partial map was generated by O’Malley et al. using DNA affinity purification sequencing (DAP-seq) to identify TF binding sites for 349 TFs (including CHE, ORA59, ANAC055 and MYB51 from our network). This analysis revealed that ANAC055 can bind to the promoters of ORA59 and RAP2.6L. Finally, the RAP2.6L overexpressing mutant showed increased expression of AT1G79150, providing evidence for this regulatory interaction. Edges with supporting experimental data are shown in green in Figure 2.

A Validated Dynamic Model of the CHE Regulatory Subnetwork. The network inference algorithms used to infer the large consensus network model are able to predict regulatory relationships between the genes in the 9GRN but the type of regulation (i.e., activating or inhibiting) cannot be determined. Since these are essential features of any model that can be used for controller design, we next determined the direction of the regulatory edges in the 9GRN using standard four-step system identification techniques: data collection, model structure selection, parameter estimation and model validation (see chapters 1 and 7 of ref 31). Previous studies that utilized this technique to identify regulation types in GRNs used linear models, and there is now strong evidence that the underlying dynamics of GRNs can be accurately described using such models (we define accurate as a model able to recapitulate experimental data within a single standard deviation of error). Moreover, as the model is subsequently to be used to design perturbation mitigation strategies, a linear model facilitates the use of linear control design techniques that are more established than their nonlinear counterparts. In system identification terminology, black box models refer to a set of ready-made models with no physical structure or biological interpretation. On the other hand, gray box models refer to models that are tailor-made given some prior information about the system. Since we have prior knowledge of the direction of regulation between the genes obtained from the inferred network above, we use a linear gray box model comprising nine ODEs (eq 4 in the Methods section) for the 9GRN, and thus only need to identify the regulation type and dynamics within the 9GRN. The values of the model parameters were estimated from the available mRNA time-series data using a nonlinear least-
squares algorithm\textsuperscript{37,38} (eq 5 in Methods section) and the estimated parameters are given in Table 3 in Methods section. As these mRNA time-series measurements are normalized using an intensity-dependent normalization method,\textsuperscript{39,40} the resulting measurements are dimensionless and are reported as relative expression. Figure 2 indicates the regulation types identified in the 9GRN subnetwork, where positive and negative values of production rate given in Table 3 in the Methods section denote transcriptional activators and inhibitors, respectively. In addition, all the estimated degradation rates had the expected negative sign, and had numerical values within the range expected.\textsuperscript{41} We validated the dynamic model by comparing its response against another mRNA time-series data set (see Methods section) that was not used in the parameter estimation process as well as two mutant behaviors. As shown in Figures 3 and S3, the identified model is able to accurately predict the expression behavior of the network. Additionally, the model shows good predictive capability against two mutant data sets (see Figure S4).

Design of a Feedback Controller for Perturbation Mitigation. As outlined above, our control objective is to employ feedback to prevent the reduction in CHE levels when the plant is subjected to pathogen attack. There are several frameworks available for designing genetic controllers.\textsuperscript{19,21,42} In refs 19 and 42, the authors proposed and extended a framework for implementing an integral controller using a negative feedback of a two-promoter gene network. In ref 21, the authors analyzed the dynamics of gene regulation using frequency domain tools from control theory and proposed the implementation of a genetic phase lag controller. Here, we based our design on the framework proposed in Harris et al.,\textsuperscript{21} where the proposed genetic controller is made up of a combination of genes and the regulatory relationships between them. In ref 21, these gene regulations are modeled using nonlinear Michaelis–Menten type functions and these functions are then linearized such that the controller design and analysis can be done using standard frequency domain methods. In this study, since we have used a linear model to describe the 9GRN, we also model the gene regulations in the controller using linear functions.

Figure 4a shows the genetic circuit diagram of the proposed feedback controller. The controller architecture is modified from the framework suggested in ref 21, whereby for the purposes of implementation in plants we replace the protease degradation component with a transcriptional inhibitor component. The modified circuit contains three genes and their associated proteins: genes X, Y and E giving proteins X, Y and E.

Let X denote an arbitrary gene that can be regulated by E, and its translated protein X denotes the TF that can regulate the output gene, Y, whose levels we ultimately want to control. E denotes the protein whose function is to regulate gene X and calculate the error signal. Here the error signal is the difference between the desired reference level and the output signal Y (see...
The ODE for the regulation of $X$ by $E$ is given by

$$\alpha \beta = - + $$

Here, $\alpha$, $\beta$, and $b_S$ represent production rate, degradation rate and basal expression level, respectively.

With $Y$ being the output of the process that we want to control, then the ODE describing the regulation of $Y$ by $X$ and $E$ can be written as

$$\alpha \alpha \beta = + - + $$

Taking Laplace Transforms of eqs 1 and 2, and after some algebraic manipulation, we obtain the following transfer function (see Section S1 of the Supporting Information):

$$Y(s) = \left( s + \beta X + (\alpha X, E \alpha Y, X, E) / \alpha Y, E \right) \left( s + \beta Y \right)$$

Eq 3 is the open-loop transfer function from $E$ to $Y$. In control theory, an open-loop transfer function is defined as the ratio of the output signal to the input signal in the absence of feedback and it is usually composed of the product of the transfer functions of the controller and the process. In a transfer function, the solutions making the numerator to zero are called the zeros of the system while the solutions making the denominator zero are called the poles of the system. Since $Y$ is the output of the process, its transfer function is given by $s^2 + \beta X / (s + \beta Y)$. Thus, the transfer function of the controller is then given by $s + \beta X + (\alpha X, E \alpha Y, X, E) / \alpha Y, E$, where the zeros and poles of the controller are $z = -((\beta X + (\alpha Y, X) / \alpha Y, E)) / \beta Y$, and $p = -\beta Y$, respectively. Since $|p| < |z|$, we obtain a phase lag controller. In control engineering, phase lag controllers are commonly used to improve disturbance
rejection and reduce steady-state error, and thus they are well suited to our control objective of achieving perturbation mitigation. Interestingly, based on the schematic diagram of the phase lag controller as shown in Figure 4a, we note that this controller structure is equivalent to a coherent feedforward loop type-I network motif, but with an added feedback loop. The role of this network motif in natural biological systems has been subjected to extensive studies and one of its key roles includes perturbation attenuation.

We illustrate here in simulation the use of the genetic phase lag controller in mitigating the perturbation affecting CHE in the 9GRN. The configuration for perturbation mitigation using the genetic phase lag controller is shown in Figure 4b. In 9GRN, the output gene Y is CHE and the feedback is delivered by CHE’s transcriptional repressor activity on gene E. As with standard perturbation mitigation strategies in feedback control theory, when a perturbation causes the output level to deviate from its desired level, the controller upon detecting this deviation will react in order to restore the output to its desired level. As CHE is a circadian gene, its expression level is not constant but oscillatory (ATML1 is also light regulated). In the absence of perturbations, the CHE expression levels oscillate around the relative expression value of 12.44 (black line in Figure 4c). In our simulations, the perturbation (B. cinerea inoculation) is introduced at time 120 h. Upon infection by B. cinerea, the average expression level of CHE drops from 12.44 to 9.77 as indicated by the yellow solid line in Figure 4c. The phase lag controller upon detecting this drop in the expression level of CHE should exert an appropriate control action to restore the level of CHE to its original level. When the phase lag controller is implemented (blue solid line), the controller almost completely attenuates the effect of the perturbation with the level of CHE oscillating around 12.33. Moreover, this control strategy is shown to be robust against variation in model and controller parameters through a Monte Carlo simulation (see Methods section), where we randomly varied the parameters within 20% of their nominal values.

It is known from control theory that to exactly restore the output to the desired reference level after a step disturbance requires an integral-type controller. In terms of the controller transfer function, an integral-type controller has a pole at $s = 0$. The transfer function of the phase lag controller given in eq 3 has a pole at $s = -\beta/\lambda$, and therefore the slower the degradation rate for X (which corresponds to a longer mRNA half-life), the more closely the controller will implement an integral-type control action that exactly restores the output to the desired reference level after a disturbance. In Arabidopsis, the longest half-life reported for mRNAs is approximately 26 h, which corresponds to a degradation rate of 0.026/hour (calculated using the standard equation for exponential decay, $\beta = \ln(2)/T$), and therefore, we have used this value in our simulations (blue solid line in Figure 4c). Full details of all the equations and parameter values underlying the simulations shown in Figure 4c can be found in Section S2.1 of the Supporting Information.

Controller Implementation Using Regulatory Network Rewiring. The direct implementation of the proposed controller in Arabidopsis presents a number of challenges, largely due to the choice of TFs for E and X and associated binding
sequences. In ref 21, the suggested genes for E and X are RhaS (E in Figure 4a) and XylS (X in Figure 4a). RhaS activates the production of XylS and CHE through a coherent feedforward loop, and XylS also acts as a regulator for the production of CHE. However, orthogonal TFs may not function in plants while using endogenous TFs is likely to have unintended consequences on other processes.

To get around these problems we propose an alternative approach for implementing the proposed controller, based on network rewiring. As shown in Figure 4a, (and Section S6 of the Supporting Information) the structure of a genetic phase lag controller is composed of a coherent feedforward loop type I motif with negative feedback. Thus, if we are able to realize this network motif through the rewiring of the 9GRN, we can obtain a genetic phase lag controller without the need to introduce new nonendogenous genetic circuitry.

For the 9GRN network shown in Figure 2, there are 46 potential rewiring combinations that can realize the network motif of a phase lag controller. However, not all genes within the 9GRN can be used in the rewiring exercise, due to functional constraints. Genes ATML1, LOL1 and AT1G79150 are not suitable for rewiring, as during B. cinerea infection, their expression levels decrease and to use them as part of the positive regulation of the network motif would lead to further decrease in the level of CHE. Another constraint is due to the gene at-ERF1, which is a negative regulator of plant defense (see Figure S1), and hence we would not wish to increase its expression further. Using at-ERF1 as part of the positive regulation of the network motif, however, would lead to an increase in its expression. In addition, the gene ORA59 is a positive regulator of defense, so decreasing its levels would negatively affect the defense response to B. cinerea. The gene RAP2.6L is highly responsive to stress hormones,30,49 and while its involvement in infection with B. cinerea has not been conclusively proven, we have also chosen to discard rewiring combinations that decrease its levels. Taking these constraints into account, we are left with 11 possible rewiring combinations (see Section S3 of the Supporting Information). Further analysis of these 11 rewiring combinations (see Section S3 of the Supporting Information) reveals that the rewiring strategy that requires the least amount of experimental modification involves the pathway from MYBS1 (E) to ORA59 (X) to CHE (Y). Note that we have included the equivalent function of the genetic phase lag controller in brackets.

Figure 5a shows the rewiring configuration using the pathway from MYBS1 to ORA59 to CHE. To realize the required network motif, CHE must inhibit expression of MYBS1, and MYBS1 and ORA59 must activate CHE expression. Implementing this in simulation, with the perturbation introduced at time 120 h, we notice only a small recovery in the expression level of CHE from around 9.77 to 10.31 after the perturbation (Figure 5a). Why is the increase in the level of CHE small given that we have implemented a phase lag controller through network rewiring? From eq 3, we note that the pole of the phase lag controller is given by the degradation rate of X, and in this network motif, this corresponds to the degradation rate of ORA59. From Table 3, the value of the degradation rate of ORA59 is 38.0062, which corresponds to placing the pole at \( s = -38.0062 \). From our previous discussion, it is desirable to have the pole of the controller to be as close to 0 in order for the controller to restore the output to its desired reference level. To move the pole associated with ORA59 closer to 0, we use positive autoregulation,15,21,50 i.e., we further rewire the network so that ORA59 activates itself. As expected, with the addition of autoactivation of ORA59, we observe that the expression level of CHE begins to show a significant increase at around 140 h. However, instead of returning to its original level, it increases by an extra 15% compared to its original value (Figure 5b).

A detailed look at the plot of MYBS1 reveals that the error computed by MYBS1 is higher than expected. The reason for the incorrect error computation is that there is unmodeled regulation affecting MYBS1 (see eq 4 in Methods section and Section S5 of Supporting Information). As a result, the controller “sees” a larger error than actually exists, and thus exerts a higher control action to mitigate this error, resulting in the observed further increase in the expression level of CHE.

To address this issue, a mechanism to negate the effect of unmodeled regulation on MYBS1 is required. This can be achieved by rewiring another gene, for example ANAC055, to regulate MYBS1 (see Section S4 in the Supporting Information). As the negation is independent of the process output, this is equivalent to using a feedforward controller. With the addition of autoregulation and feedforward control, the simulation results in Figure 5c show that the phase lag controller implemented via rewiring is now able to significantly attenuate the effect of the perturbation on CHE and return it to its original expression level. Additionally, the Monte Carlo simulations (see Methods section) show that the proposed strategy is robust against parameter variations. The details of the equations and parameter values underlying the simulations shown in Figure 5c can be found in Section S2.2 of the Supporting Information.

### DISCUSSION

We have presented a novel strategy, based on the use of feedback control, for mitigating the effects of pathogen attack on plant gene regulatory networks, and demonstrated via simulation the ability of this approach to restore the levels of CHE, a key defense gene in Arabidopsis, after infection by B. cinerea. The use of simple rewiring such as negative autoregulation of CHE and direct regulation from ANAC055 was found to be insufficient for restoring the level of CHE, therefore we employed a coherent feedforward type I motif with negative feedback. In order to develop the strategy, we employed system identification techniques to build and validate a new dynamical model of the infected gene regulatory subnetwork that accurately predicts the type of regulation between each node of the network. Then, using this model, we designed perturbation mitigation strategies using feedback control theory. In the proposed approach, we applied a combination of two positive and one negative regulatory interactions to implement genetic circuitry realizing a phase lag controller. Phase lag controllers are widely used in engineering systems to reduce the effects of disturbances on system performance, and have been proposed as a useful motif for implementing synthetic biological control systems.21 To date, however, practical strategies for implementing such controllers in vivo remain to be elucidated. Here, based on the observation that this control architecture resembles a coherent feedforward loop type-I with negative feedback, we propose a novel controller implementation strategy based on identifying groups of genes within the 9GRN whose regulation can be rewired to realize this network motif. Within the 9GRN, rewiring the pathway from MYBS1 to ORA59 to CHE was shown to provide the most straightforward implementation of the phase lag controller. When suitably augmented with rewired autoregulation and feedforward components, this implementation of the controller was shown to deliver almost perfect perturbation mitigation without the need for any nonendogenous synthetic circuitry.
The regulatory network rewiring described above can be carried out experimentally through the insertion of constructs expressing the desired TF from the appropriate promoter region or TF binding sites combined with a minimal promoter sequence. Given that there are multiple TF binding sites in a typical promoter sequence (see, e.g., ref 51), it is preferable to use specific TF binding regions. For the rewiring we propose for the 9GRN, regulation of MYBS1 by CHE, CHE by ORA59, and ORA59 regulation of its own expression, could be achieved using specific promoter regions that have been shown to confer the necessary regulation to drive expression of copies of the target TF coding sequence. CHE binds to the promoter of its target gene CCA1 at the sequence GGTCCACA. Both the region −363 to −192 bp of the CCA1 promoter encompassing this sequence and a trimer of the CHE binding sequence have been shown to be bound by CHE. ORA59 binds to two GCC boxes (GCCGCC and GCAGCCGCT) in the PDF1.2 promoter and a tetramer of one of these boxes is sufficient for ORA59 activation of expression. The other regulatory edges required for rewiring (MYBS1 activation of CHE expression and ANAC055 inhibition of MYBS1) would currently require using the full length promoter sequences and potentially fusion of transcriptional repression domains. Rewiring using full-length promoter sequences could be achieved relatively quickly (1–2 years) and methods to insert multiple gene constructs into Arabidopsis are available (for example, Golden gate cloning). However, the site of insertion of the necessary transgenic constructs (which is not controlled) may also influence resulting levels of expression and hence further optimization/selection of lines with appropriate levels will no doubt be necessary.

The main premise of this paper is to demonstrate potential application of the phase lag controller motif in preventing pathogen-induced perturbations of gene expression. Our 9GRN model is used to demonstrate how a phase lag controller could function. We have provided some evidence for edges in our network, but the presence of additional edges we have not modeled or false positive edges could have a significant impact on the performance of this controller in vivo. Strategies such as DAP-seq are making significant improvements in our knowledge of plant TF-promoter interactions but, particularly given the expansion of TF families in plants, greater mapping of plant gene regulatory networks under multiple environmental and developmental conditions will be necessary to drive successful plant synthetic biology strategies. Clearly, in all nonorthogonal rewiring strategies the new edges may have unintended consequences on plant physiology through changing regulatory interactions. In our in silico implementation, the controller is only triggered by a significant reduction in CHE levels (such as that driven by pathogen infection, not the daily circadian oscillations), and levels of CHE and MYBS1 quickly return to normal. The intention of the controller is to maintain oscillating expression levels of CHE given its key role in the circadian clock (regulating CCA1, a core transcriptional regulator). We also ensured that expression of positive regulators of defense was not compromised. However, the genes with new rewired links (MYBS1, ORA59 and ANAC055) are involved in response to other environmental conditions. For example, MYBS1 promotes expression of indolic glucosinolate biosynthetic genes in response to mechanical stimuli and ANAC055 is induced by drought, salt and abscisic acid stress. Changes in the expression of these genes due to other stimuli could prevent the controller from operating during B. cinerea infection (for example, induction of ANAC055 would lower levels of MYBS1 and indicate a lower level of error to the controller, leading to the controller not reacting properly). Our controller is not designed to handle more than one perturbation (environmentally induced shift in gene expression), and this limitation raises another key challenge in plant systems biology. The development of novel approaches to model and simulate dynamic networks of sufficient size to capture environmental stress cross-talk will significantly improve our ability to rationally engineer stress resilient plants.

METHODS

**Transgenic Arabidopsis Line.** The CHE T-DNA insertion line, SALK_143403c, was obtained from the SALK collection and confirmed to be homozygous. Expression of CHE in this line is significantly reduced compared to wildtype (Col-0). The coding region of at-ERF1 was cloned into the pB7WG2 vector and stable transgenic Arabidopsis lines generated in a Col-4 background. The coding region of at-ERF1 was cloned into the pB7WG2 vector and stable transgenic Arabidopsis lines generated in a Col-4 background.

**Infection Assay.** Arabidopsis plants were grown and B. cinerea strain pepper35,66 cultivated as described in ref 10. Leaves from 5-week-old plants were detached and placed on 0.8% agar in propagator trays. Each leaf was inoculated with a single 10 μL droplet of B. cinerea inoculum, or a 10 μL droplet of sterile grape juice diluted in a 1:1 ratio with sterile water. Each tray contained 9 control leaves and 81 infected leaves, with control and infected leaves in each row. The trays were covered with lids and kept in a growth cabinet under a 16:8 h light:dark cycle at 22 °C, with 90% humidity. Lesion area was assessed from photographs using ImageJ. Mean lesion area of leaves from WT and T-DNA insertion lines were compared using a Student’s two-tailed t-test, which assumed equal variance.

**Yeast-1-Hybrid Assay.** Yeast-1-Hybrid assays were performed as previously described. Three overlapping promoter regions (of approximately 400 bp) spanning 800 to 1200 bp upstream of the transcription start site were used as bait for transcription factors fused to a GAL4 activation domain in pDEST22 (Invitrogen). Yeast strain AH109 (Clontech) was transformed with these individual TF clones. The promoter fragments were amplified using two-step PCR and cloned into a pDonrZeo vector (Invitrogen) using Gateway cloning. Yeast strain Y187 (Clontech) was transformed with the individual vectors to create the bait strain. The promoter strain was spotted onto YPDA (yeast, peptone, dextrose, adenine) plates, overlaid with the TF strain, and incubated for 24 h at 30 °C. The diploid cells were replica plated onto selective plates and incubated overnight. This was followed by replica-cleaning and incubation for 4 days, after which growth was scored. Each interaction was tested twice. Primer sequences for the promoter fragments are given in Table 1.

**Accession Numbers.** Arabidopsis gene names and AGI locus codes referred to in this article are shown in Table 2.

**Generating the TF network.** An Arabidopsis TF list was generated by combining lists from ThaleMine, DATF, ref 62, and homology searches using DNA binding domains, followed by manual curation of genes only identified in one list. The final list of 2534 genes is given in Supplementary File S1b. The list of Arabidopsis genes differentially expressed during B. cinerea infection was obtained from ref 10, which included 883 differentially expressed TFs (Supplementary File S1a).

**Generating a Dynamic Model of the CHE Regulatory Subnetwork.** Traditionally, a model of a gene regulatory
network comprises both transcription and translation mechanisms. However, in our case, given that only mRNA accumulation time-series data are available, the following two assumptions are made in building the 9GRN model. First, the translation of the protein from mRNA follows a linear relationship and second, the behavior of the translated protein follows its mRNA closely. With these two assumptions, we can group together the protein translation rate with the mRNA transcription rate resulting in the entire 9GRN being modeled using only mRNA data. Based on the above assumptions, the model of the 9GRN shown in Figure 2 can be described by the following ODEs:

\[
\begin{align*}
\frac{dN_\text{ORA}}{dt} &= \alpha_\text{ORA,1} N_\text{MYB} + \alpha_\text{ORA,2} N_\text{ANA} + \alpha_\text{ORA,3} N_\text{ERF} \\
&\quad + \beta_\text{ORA} N_\text{ORA} + b_{S,\text{ORA}} \\
\frac{dN_\text{MYB}}{dt} &= \beta_\text{MYB} N_\text{MYB} + b_{S,\text{MYB}} + c_{\text{MYB}} W \\
\frac{dN_\text{LOL}}{dt} &= \beta_\text{LOL} N_\text{LOL} + b_{S,\text{LOL}} + c_{\text{LOL}} W \\
\frac{dN_\text{AT1}}{dt} &= \alpha_{\text{AT1,1}} N_\text{CHE} + \alpha_{\text{AT1,2}} N_\text{RAP} + \beta_{\text{AT1}} N_\text{AT1} + b_{S,\text{AT1}} \\
&\quad + c_{\text{AT1}} W \\
\frac{dN_\text{ANA}}{dt} &= \alpha_{\text{ANA,1}} N_\text{RAP} + \beta_{\text{ANA}} N_\text{ANA} + b_{S,\text{ANA}} + c_{\text{ANA}} W \\
\frac{dN_\text{ERF}}{dt} &= \beta_{\text{ERF}} N_\text{ERF} + b_{S,\text{ERF}} + c_{\text{ERF}} W \\
\frac{dN_\text{ATM}}{dt} &= \alpha_{\text{ATM,1}} N_\text{RAP} + \beta_{\text{ATM}} N_\text{ATM} + b_{S,\text{ATM}} + c_{\text{ATM}} W \\
&\quad + \gamma_{\text{ATM}} W \\
\frac{dN_\text{CHE}}{dt} &= \alpha_{\text{CHE,1}} N_\text{LOL} + \alpha_{\text{CHE,2}} N_\text{AT1} + \alpha_{\text{CHE,3}} N_\text{ATM} \\
&\quad + \beta_{\text{CHE}} N_\text{CHE} + b_{S,\text{CHE}} + c_{\text{CHE}} W \\
\frac{dN_\text{RAP}}{dt} &= \alpha_{\text{RAP,1}} N_\text{ANA} + \alpha_{\text{RAP,2}} N_\text{RAP} + b_{S,\text{RAP}} + c_{\text{RAP}} W
\end{align*}
\]

where \(\alpha_{ij} \in (−\infty, +\infty), \beta_i > 0, \gamma_{\text{CHE}} > 0, b_{S,j} \in (−\infty, +\infty), \) and \(c_i \in (−\infty, +\infty)\) are the unknown parameters that represent the production rate, degradation rate, scaled light effect, basal level and effect of the unmodeled regulation, respectively, with \(i\) and \(j\) denoting the appropriate indices describing the parameters given in eq 4. \(N_i\) represents the effect of the unmodeled regulation (e.g., direct regulation as a result of

<table>
<thead>
<tr>
<th>gene name</th>
<th>AGI</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORA59</td>
<td>AT1G06160</td>
</tr>
<tr>
<td>MYB51</td>
<td>AT1G18570</td>
</tr>
<tr>
<td>LOL1</td>
<td>AT1G32540</td>
</tr>
<tr>
<td>AT1G79150</td>
<td>AT1G79150</td>
</tr>
<tr>
<td>ANAC055</td>
<td>AT3G15500</td>
</tr>
<tr>
<td>at-ERF1</td>
<td>AT4G17500</td>
</tr>
<tr>
<td>ATML1</td>
<td>AT4G21750</td>
</tr>
<tr>
<td>CHE</td>
<td>AT5G08330</td>
</tr>
<tr>
<td>RAP2.6L</td>
<td>AT5G13330</td>
</tr>
</tbody>
</table>

Table 2. Associated AGI to Arabidopsis Gene Names
B. cinerea infection, noise and other regulations not identified by the network inference algorithms), where $W = 0$ (respectively $W = 1$) is used when the effect is absent (respectively present). In the experiments from which our data were generated, the time-series data from the control and infected experiments are treated as a continuous data set where the infection starts at the halfway point, i.e., time 48 h. Thus, the transition of W from 0 to 1 is not modeled as an instantaneous change but as a gradual increase. L represents the effect of light and CHE follows a sinusoidal rhythm as known.27 For details on the mathematical representation for W and L see Section S5 of the Supporting Information.

The values of the model parameters were estimated from the available mRNA time-series data using a nonlinear least-squares algorithm and the estimated parameters are given in Table 3.

### Table 3. Estimated Parameters of the Linear Model

<table>
<thead>
<tr>
<th>gene name</th>
<th>values</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORAS9</td>
<td>$α_{ORAS9} = 14.3800, β_{ORAS9} = -0.7359, γ_{ORAS9} = 41.5714, β_{ORA} = -38.0062, b_{ORA} = 15.2355$</td>
</tr>
<tr>
<td>MYBS1</td>
<td>$β_{MYBS1} = -0.6668, b_{MYBS1} = 5.6277, s_{MYBS1} = 1.1890$</td>
</tr>
<tr>
<td>LOL1</td>
<td>$β_{LOL1} = -0.0485, b_{LOL1} = 0.4874, c_{LOL} = -0.1241$</td>
</tr>
<tr>
<td>ATIG79150</td>
<td>$α_{ATIG79150} = 0.7577, β_{ATIG79150} = -0.7408, b_{ATIG79150} = 23.8635, c_{ATIG79150} = 0.091809$</td>
</tr>
<tr>
<td>ANAC055</td>
<td>$α_{ANAC055} = 25.6935, β_{ANAC055} = -28.4685, b_{ANAC055} = 0.0517, c_{ANAC055} = 82.5415$</td>
</tr>
<tr>
<td>at-ERF1</td>
<td>$β_{at-ERF1} = -0.2051, b_{at-ERF1} = 1.8699, s_{at-ERF1} = 0.8735$</td>
</tr>
<tr>
<td>ATML1</td>
<td>$α_{ATML1} = -0.7945, β_{ATML1} = -1.1142, b_{ATML1} = 19.3684, c_{ATML1} = 0.0040, c_{AT} = 0.5000$</td>
</tr>
<tr>
<td>CHE</td>
<td>$β_{CHE} = 38.4024, β_{CHE} = 3.3801, α_{CHE} = 17.6771, β_{CHE} = -40.1258, b_{CHE} = 3.7167, T_{CHE} = 16.3001$</td>
</tr>
<tr>
<td>RAP2.6L</td>
<td>$α_{RAP2.6L} = 0.4186, β_{RAP2.6L} = -0.7933, b_{RAP2.6L} = 3.7046, s_{RAP2.6L} = 0.0045$</td>
</tr>
</tbody>
</table>

All the simulations of the ODE models, phase genetic controller and network rewiring are done using MATLAB built-in solver ode45, and the initial condition for each gene to solve the ODE is the first data point of the mRNA time-series for each respective gene. For the simulation using the genetic phase lag controller (eq S2.1), the initial conditions for solving the ODEs for X and E are set to 0.

### Parameter Estimation

For the 9GRN linear model, the values of the unknown parameters are estimated from the available mRNA time-series using nonlinear least-squares, given by

$$\hat{θ} = \arg \min_θ \frac{1}{N_L} \sum_{i \in ψ} \sum_{t=1}^{N_L} [N(t) - \hat{N}(t, θ)]^2$$

(5)

where $θ = [α_i, β_i, b_{SP}, c_i]$ with $i \in ψ = [ORA, MYB, LOL1, AT1, ANA, ERF, ATM, CHE, RAP2.6L]$, $N_L$ is the length of the time-series data, $\hat{N}$ is the simulated data from eq 4 and $N$ is the experimental data, which are the mRNA time-series taken from ref 10. There are four sets of mRNA time-series and we use the average mRNA expression from the first three sets for parameter estimation and use the fourth data set as an independent data set for validating the ODE model. Eq 5 is solved using MATLAB function fminsearch which uses the Nelder-Mead simplex algorithm.

As a quantitative measure of the model performance, we compute the Mean Square Error (MSE) for each gene between the experimental data and the model given by eq 4. The MSE for each gene is computed as follows:

$$MSE = \frac{1}{N_L} \sum_{i \in ψ} \sum_{t=1}^{N_L} [N(t) - \hat{N}(t, θ)]^2$$

(6)

The total MSE, $MSE_T$ is computed by summing the MSE for all nine genes in the 9GRN. Table 4 shows the MSE values for both the training and validation data sets.

### Table 4. MSE for Both Training and Validation Data Sets

<table>
<thead>
<tr>
<th>gene name</th>
<th>MSE (training)</th>
<th>MSE (validation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORAS9</td>
<td>0.7730</td>
<td>1.9828</td>
</tr>
<tr>
<td>MYBS1</td>
<td>0.3910</td>
<td>0.6949</td>
</tr>
<tr>
<td>LOL1</td>
<td>0.3703</td>
<td>0.6582</td>
</tr>
<tr>
<td>ATIG79150</td>
<td>0.1829</td>
<td>0.3587</td>
</tr>
<tr>
<td>ANAC055</td>
<td>0.9889</td>
<td>2.3849</td>
</tr>
<tr>
<td>at-ERF1</td>
<td>0.4394</td>
<td>1.0583</td>
</tr>
<tr>
<td>ATML1</td>
<td>0.3746</td>
<td>0.6682</td>
</tr>
<tr>
<td>CHE</td>
<td>0.8759</td>
<td>1.0819</td>
</tr>
<tr>
<td>RAP2.6L</td>
<td>0.3452</td>
<td>0.6366</td>
</tr>
<tr>
<td>MSET</td>
<td>4.7410</td>
<td>9.5245</td>
</tr>
</tbody>
</table>

### Performance and Robustness Analysis

To analyze the performance and robustness of the proposed strategies, we perform a Monte Carlo simulation where we randomly draw all the parameters from a uniform distribution. Then, we vary the parameters within ranges of 20%, around their nominal values. Mathematically, we have $p(1 + Δp(x))$, where $p$ denotes the model and the controller parameters, $p(x)$ is the probability distribution and $Δ = 0.2$. Using the Chernoff bound and associated guidelines for Monte Carlo simulation, a total number of 1060 simulations is required to achieve an accuracy level of 0.05 with a confidence level of 99%.63,64

### ASSOCIATED CONTENT

Supplementary Information
The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.8b00037.

Supplementary text and figures (PDF)
List of Arabidopsis TF genes and their associated biological functions (XLSX)
List of TF genes that are differentially expressed during B. cinerea infection (XLSX)

### AUTHOR INFORMATION

Corresponding Authors
*E-mail: d.bates@warwick.ac.uk.
*E-mail: katherine.denby@york.ac.uk.

ORCID
Mathias Foo: 0000-0003-1400-2659

Present Address
School of Mechanical, Aerospace and Automotive Engineering, Coventry University, CV1 5FB, United Kingdom.

Author Contributions
MF and IG are joint first authors. KJD and DGB conceived the study. MF and IG developed the models and simulations. Experimental data was generated by PZ and IG. MF, IG, DGB, and KJD analyzed the data. All authors reviewed and wrote the manuscript.

Notes
The authors declare no competing financial interest.
ACKNOWLEDGMENTS

This work was supported by the Biotechnological and Biological Sciences Research Council (BBSRC)/Engineering and Physical Sciences Research Council (EPSRC) Warwick Integrative Synthetic Biology Centre (WISB) via research grant BB/M017982/1 (MF, KJD, DGB) and the EPSRC/BBSRC Oxford/Warwick/Bristol Centre for Doctoral Training in Synthetic Biology (SynBio CDT) via research grant EP/L016494/1 (IG). PZ and KJD were supported by the BBSRC/EPSRC-funded Plant Response to Environmental Stress Arabidopsis (BB/F005806/1).

REFERENCES

