UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

BADDADAN

Mechanistic modelling of time-series gene module expression

Noordijk, Ben; Reinders, Marcel; van Dijk, Aalt D.J.; de Ridder, Dick

DOI
10.1017/qpb.2025.10017
Publication date
2025
Document Version
Final published version
Published in
Quantitative Plant Biology
License
CC BY-NC-SA

Link to publication

Citation for published version (APA):
Noordijk, B., Reinders, M., van Dijk, A. D. J., & de Ridder, D. (2025). BADDADAN: Mechanistic modelling of time-series gene module expression. *Quantitative Plant Biology*, *6*, Article e31. https://doi.org/10.1017/qpb.2025.10017

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:12 Sept 2026

Quantitative Plant Biology

cambridge.org/qpb

BADDADAN: Mechanistic modelling of time-series gene module expression

Ben Noordijk[1,2], Marcel Reinders[2,3], Aalt D.J. van Dijk[2,4] and Dick de Ridder[1,2]

[1]Bioinformatics Group, Wageningen University & Research, Wageningen, The Netherlands; [2]CropXR Institute, Utrecht, The Netherlands; [3]Pattern Recognition & Bioinformatics Group, Delft University of Technology, Delft, The Netherlands; [4]Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands

Original Research Article

Cite this article: B. Noordijk et al. BADDADAN: Mechanistic modelling of time-series gene module expression. *Quantitative Plant Biology*, **6**:e31, 1–14
https://dx.doi.org/10.1017/qpb.2025.10017

Received: 17 February 2025
Revised: 22 May 2025
Accepted: 16 June 2025

Keywords:
gene module; machine learning; mechanistic model; ordinary differential equations (ODEs); plant stress.

Corresponding author:
Ben Noordijk;
Email: ben.noordijk@wur.nl

Abstract

Plants respond to stresses like drought and heat through complex gene regulatory networks (GRNs). To improve resilience, understanding these is crucial, but large-scale GRNs (>100 genes) are difficult to model using ordinary differential equations (ODEs) due to the high number of parameters that have to be estimated. Here we solve this problem by introducing BADDADAN, which uses machine learning to identify gene modules—groups of co-expressed and/or co-regulated genes—and constructs an ODE model that predicts gene module dynamics under stress. By integrating time-series gene expression data with prior co-expression data it finds modules that are both coherent and interpretable. We demonstrate BADDADAN on heat and drought datasets of *A. thaliana*, modelling over 1,000 genes, recovering known mechanistic insights, and proposing new hypotheses. By combining machine learning with mechanistic modelling, BADDADAN deepens our understanding of stress-related GRNs in plants and potentially other organisms.

© The Author(s), 2025. Published by Cambridge University Press in association with John Innes Centre. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (https://creativecommons.org/licenses/by-nc-sa/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

1. Introduction

Climate change poses significant challenges to global food security due to its negative impact on crop growth and productivity (Suter & Widmer, 2013). With a growing world population, especially in climate-vulnerable areas (Dai, 2011), it becomes essential to develop crop varieties that are resilient to stresses, such as pests, heat and drought, particularly when these occur simultaneously or in rapid succession. To understand these plant environmental response mechanisms, as well as their potential trade-offs, we aim to understand the underlying gene regulatory networks (GRNs) (Quint et al., 2016), i.e., sets of genes which influence each other's expression. The temporal dynamics of these GRNs are commonly modelled using ordinary differential equations (ODEs) (Cao et al., 2012; de Luis Balaguer et al., 2017; Valentim et al., 2015) fit to time-series datasets. Importantly, such ODE models can reflect biological prior knowledge and enable causal insights through formal mathematical model analysis and human interpretation. They have successfully been used to describe the influence of a small group of genes on flowering time (Chew et al., 2022) or hypocotyl elongation (Nieto et al., 2022) in response to different temperature and light regimes, and enhanced our mechanistic understanding of these processes.

However, these models typically only include a small number of genes (<100) (Fröhlich et al., 2018) because they cannot (yet) be constructed for genome-scale plant stress response networks due to a lack of extensive, mechanistic, biological knowledge. On top of that, a large number of input genes makes parameter estimation prohibitively complex, as this requires significant amounts of data (Cao et al., 2012; Segal et al., 2003; Shaik & Ramakrishna, 2013) and results in exponentially more possible gene network topologies. To overcome these issues and enable ODE modelling of large-scale regulatory networks from time-series data, we focus on gene modules (also known as gene clusters or gene sets): groups of co-expressed and/or co-regulated genes (Saelens et al., 2018; Segal et al., 2003). By modelling temporal behaviour of gene modules instead of individual genes, we significantly reduce the number of interactions that require parameter estimates. Additionally, gene modules can help bridge the gap between single-gene expression

levels and high-level phenotypes by linking these modules to broader biological processes or well-studied plant hormones. Finally, modules improve our ability to convert novel knowledge from model plants, such as *A. thaliana* to agronomically relevant crops, since module-level insights are easier to translate between species than individual genes (Ficklin & Feltus, 2011; Julca et al., 2021; Shaik & Ramakrishna, 2013; Sreedasyam et al., 2023). Machine learning (ML) is a promising approach to extract gene modules from expression measurements, as it has demonstrated the ability to infer complex patterns from large amounts of (biological) data (Alber et al., 2019). Nonetheless, ML approaches do not elucidate interactions between biological mechanisms that underlie a plant's response to environmental influences, and therefore fall short in delivering the detailed understanding that is needed to improve plant stress resilience. Thus, recently there has been an increasing interest in combining mechanistic models with ML approaches (Alber et al., 2019; Baker et al., 2018; Noordijk et al., 2024). In our context, this involves combining ML-derived modules, which extract relevant patterns from large amounts of biological data, with mechanistic models to provide interpretability and causal insights.

For insightful ODE modelling, we want to find modules that are both coherent and interpretable. Here, coherent means that genes within the same module share an expression pattern in a given dataset, allowing us to summarise their expression as a single variable (e.g., the mean) in our ODE model. Interpretable means that they can be linked to a clear biological process, providing insight into how plants respond to stress and deepening our mechanistic understanding. There is a trade-off between these two desired characteristics, which is related to the way in which co-expression is calculated in the context of gene clustering. Firstly, 'local' co-expression, calculated from gene expression measured within only one experiment (i.e., over various time points), might yield clusters that are coherent but overfit the data (Mao et al., 2019), resulting in an ill-defined biological function. On the other hand, 'global' co-expression is calculated from a gene expression compendium which covers a wide range of different conditions (e.g., developmental stages, tissues or stresses). Global co-expression should reflect gene-gene relationships more generally, providing high biological functional similarity within modules, but potentially overlooking groups of genes that are only co-expressed in a particular context. In BADDADAN, we address this coherence-interpretability trade-off by employing a module finding approach that uses a combination of global and local co-expressions.

Module detection has been extensively researched and reviewed (Saelens et al., 2018), and has been used to infer functionally related groups of genes in various organisms from gene expression data. For example, modules were found by converting gene expression to an association network first, disregarding the temporal dimension (Forster et al., 2022; Langfelder & Horvath, 2008; Wang et al., 2008). Moreover, Janky et al. (2014) and Mao et al. (2019) have extended such methods to incorporate prior biological knowledge. However, none of these methods explicitly leverage temporal data. In contrast, methods that exploit the temporal nature of the data (Han et al., 2021; Heard et al., 2005; McDowell et al., 2018; Schulz et al., 2012) do not build upon prior biological data, and suffer from sparse, noisy measurements, common in biology. Additionally, data integration has been employed to find modules (Depuydt & Vandepoele, 2021; Forster et al., 2022; Heyndrickx & Vandepoele, 2012; Sanchez-Munoz et al., 2024), but these methods cannot easily be applied to new transcriptomics datasets of interest, hampering their ability to improve gene module coherence. Finally, Lu et al. (2011), Soltanalizadeh et al. (2020), and Wu et al. (2014) performed ODE modelling of gene modules in yeast, mice or human, but did not build upon existing data, potentially reducing the interpretability of their modules and failing to leverage prior knowledge available for these species. Thus, many distinct elements of module finding, biological data integration, and ODE modelling are available, but no method yet integrates all these steps.

Here, we present BADDADAN (a Bioinformatics Approach to Describe Dynamical Activations of a Dimension-reduced *A-priori*-informed Network) which integrates existing transcriptomics and transcription factor (TF) binding data with experiment-specific expression data to infer a dynamic network model of gene modules that are coherent and reflect clear biological functions. As a proof of concept, we show that we can quantitatively model gene module temporal expression through a system of ODEs with reasonable accuracy in two real-world datasets, and that these models recover known biological insight as well as provide novel hypotheses for drought or heat resilience mechanisms. Taken together, BADDADAN offers a quantitative and interpretable framework for studying temporal gene expression responses to external perturbations, applicable to plants and other organisms.

2. Results

To demonstrate BADDADAN (Figure 1), we apply it to two time-series transcriptomics datasets in *A. thaliana* Col-0: 1) a progressive drought stress dataset of five-week-old plants grown in increasingly dry soil over a period of two weeks, with transcriptomics of leaf 7 sampled each day (Bechtold et al., 2016) and 2) a dataset where six-week-old plants were subjected to elevated ambient temperature, with transcriptomic samples of rosette leaves collected unevenly over a 24-hour period (Caldana et al., 2011). We chose these datasets because they contain transcriptomics measurements over a large number of time points (>14), making them suitable for ODE modelling. Furthermore, they represent distinct stress response pathways, which allows us to test our approach on multiple types of stresses. Below, we will go over each step in our approach and its results on both datasets in more detail.

2.1. Module finding

As a first step of gene module finding, we filter our dataset for differentially expressed (DE) genes to reduce the chance of fitting modules to noise. DE filtering is performed using *limma* (Ritchie et al., 2015) (see Methods), and we find 2,363 and 4,444 DE genes (FDR $\leq$ 0.05) for drought and heat, respectively, with which we proceed for clustering. The set of DE genes is subsequently clustered. Hereto, we use either the (local) co-expression between the DE genes in the time-series data or (global) co-expression across a large compendium of datasets as compiled in ATTED (Obayashi et al., 2022), and also combine the local and global distances by summing them to express the similarities between the DE genes (see Methods). To investigate if clusters derived from the combined distance can balance coherence and interpretability, we compare them to clusters derived from either global or local distances alone. To assess coherence, we calculate the percentage of variance in gene expression over time explained by the first principal component for each module. To measure interpretability, we identify the enriched biological process gene ontology (GO) terms for each module (see Methods). Coherence is based on expression, which is used explicitly in the clustering process; in contrast, GO annotations are

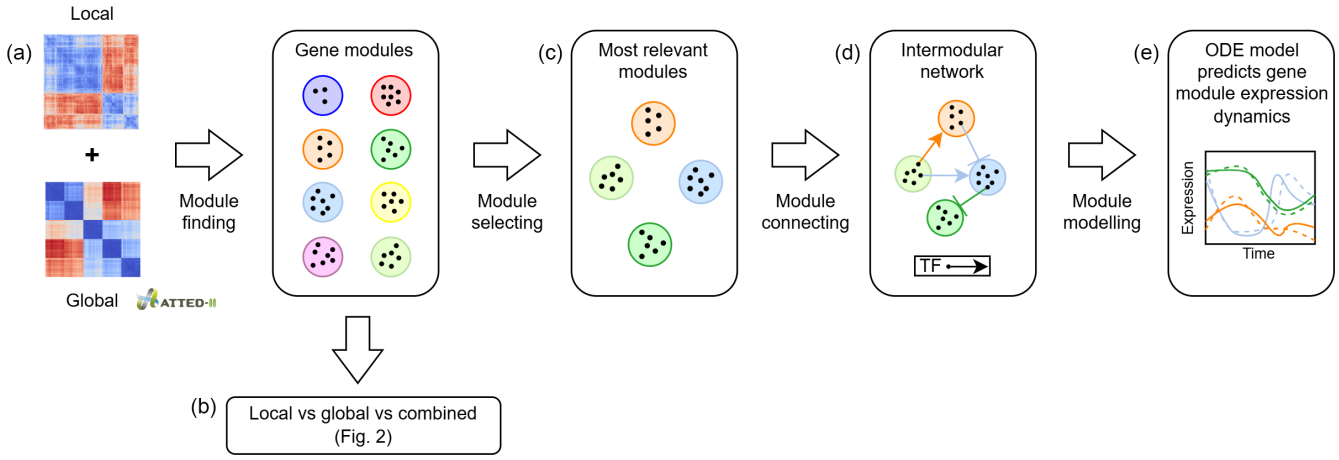

Figure 1. BADDADAN overview. (a) BADDADAN starts with candidate gene module finding based on a combination of experiment-specific (local) expression and compendium-wide (global) co-expression data (Obayashi et al., 2022). **(b)** To show this leads to improved coherence and interpretability, we compare these modules to modules created from either local or global data alone. **(c)** Our pipeline then selects a subset of the modules based on four criteria reflecting suitability for ODE modelling. **(d)** Next, it connects the modules using TFBS enrichment, and **(e)** creates an ODE model from this intermodular network which is fit to experimental (local) data and allows for biological insights.

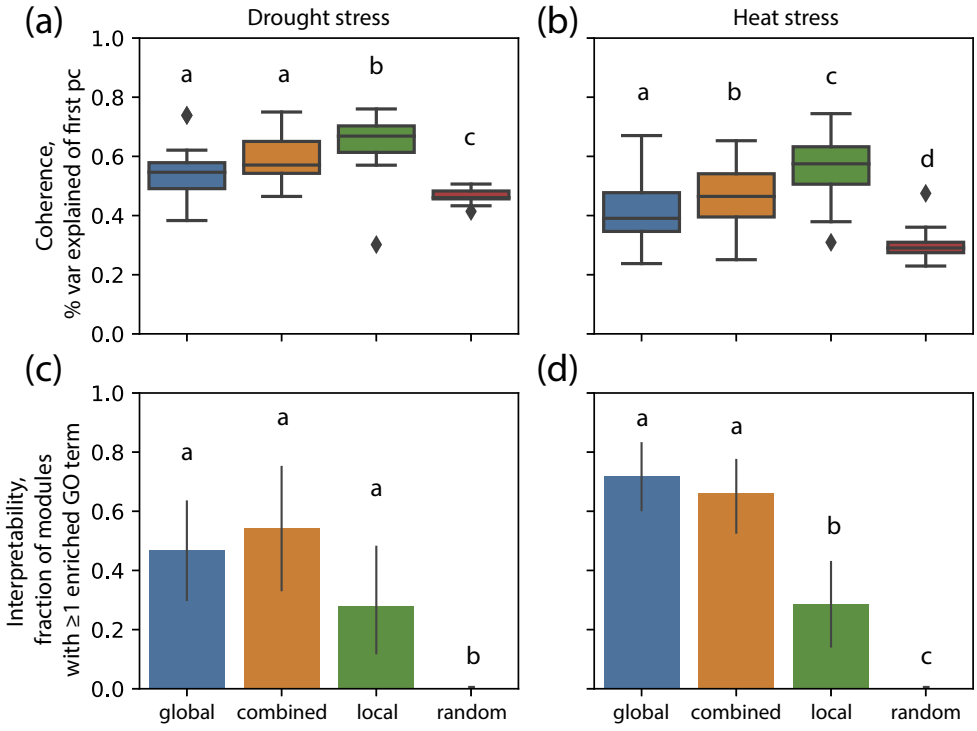

Figure 2. Module finding. (a and b) Per-module coherence, i.e., explained variance of the first principal component of modules in drought and heat stress datasets, respectively. **(c and d)** Fraction of gene modules with at least one enriched biological process GO term. The *x*-axis represents the distances used to form modules: global (derived from compendium (Obayashi et al., 2022)), local (based solely on the experiment of interest), combined (a sum of local and global) and random (modules formed by random gene assignment). Error bars indicate the 95% c.i. estimated by bootstrapping. Different letters indicate $p < 0.05$, based on a two-sided pairwise Mann-Whitney *U* test between distributions.

never explicitly used during the clustering, and thus provides an independent assessment of clustering performance. To prevent bias (e.g., modules of size 1 are 100% coherent), we also ensure that the number of modules and their sizes are approximately equal between different clusterings (see Table S1 in the Supplementary Material, Figure S1 in the Supplementary Material and Methods).

In both datasets, we see a similar pattern (Figure 2): local distance-based clusters are more coherent than global distance-based ones, though fewer are enriched for at least one GO term. We furthermore find that combined distance-based modules balance coherence and interpretability, supporting our hypothesis that combining them can address the trade-off between the two. However, most differences between the use of local, combined or global distances for the drought data are not significant (e.g., combined versus local GO enrichment, $p = 0.07$, Mann-Whitney *U* test). This could be due to the smaller number of modules in drought (≈ 25) compared to heat (≈ 45), which, in turn, reflects the fewer DE genes in the drought experiment. The total number of GO terms per module follows a trend similar to the fraction of modules with at least one annotated GO term (Figure S3 in the Supplementary Material).

We also assess GO term functional relatedness per module using mean pairwise semantic similarity (Wang et al., 2007) (Figure S4 in the Supplementary Material and Supp. Methods in the Supplementary Material), showing that clustering on combined distances increases interpretability without reducing functional relatedness.

2.2. Module selecting

Not all found candidate modules may be relevant to include in our dynamic model. We select modules most suitable for ODE modelling based on four characteristics, expressed as z-scores (see Methods): 1) activity, i.e., a high average expression over all samples; 2) coherence, as this facilitates representation of the expression pattern by a single variable; 3) relevance, i.e., differing in expression between control and treatment conditions; and 4) regulatory relevance, hypothesising that modules with a high number of enriched TF binding sites (TFBSs) are more likely to be involved in stress response GRNs (Heyndrickx et al., 2014). There is no strong correlation between these four scores (Figures S5 and S6 in the Supplementary Material), suggesting their complementary nature. We use a combined Stouffer z-score of these four characteristics to proceed with the top 50% of all candidate modules (see Methods), resulting in 12 modules for drought and 26 modules for the heat stress datasets. A breakdown of the scores per module shows that the highest scoring modules score well on coherence, relevance and activity, but not always on the regulatory relevance (Figure S7 in the Supplementary Material).

2.3. Module connecting

To incorporate a set of modules into an ODE model of intermodular regulation, we need to find regulatory interactions, i.e., determine which modules influence which other(s). Since such interactions cannot trivially be inferred from co-expression patterns and methods based on, for example, mutual information (Huynh-Thu & Geurts, 2018) do not always perform well (Aibar et al., 2017; Saelens et al., 2018), we choose to find these interactions based on existing knowledge of TFs and their TFBSs in *A. thaliana*. To this end, we used TF2Network (Kulkarni et al., 2018), which is based on extensive prior research (e.g., yeast one-hybrid and ChIP-Seq studies). However, connecting modules solely based on TFBSs can yield false positives, because TFs might not actually bind to a potential binding site due to, e.g., low chromatin accessibility (De Clercq et al., 2021; Khamis et al., 2018; Kulkarni et al., 2018). Thus, we additionally require that a module can only be connected to another module when the TF that has a potential binding site to the target module is part of the source module and that this TF has an absolute correlation > 0.75 with the target module expression average (Figure S8 in the Supplementary Material and Methods). The sign of the correlation is then used to infer whether the TF activates or inhibits the target module. Moreover, we only consider a TF when the correlation between its expression and the average expression of its source module is above 0.3 (Figure S9 in the Supplementary Material and Methods). Finally, we only keep modules that have at least one ingoing or outgoing connection to other modules. The resulting intermodular networks are shown in Figure 3b and d for the drought and the heat stress network, respectively.

2.4. Model fitting

Next, to evaluate whether BADDADAN is able to capture module expression dynamics, we construct ODEs based on the inferred module network, save them in SBML format (Hucka et al., 2003), and fit these using multi-start optimisation (see Methods). For the heat stress dataset, we explicitly model the circadian clock as input for the modules (see Methods) because the time points cover 24 hours and without the circadian clock as input our model was not able to attain an adequately accurate fit. For both datasets, the optimisation problem is hard; i.e., starting from 5,000 initial parameter sets, we rarely achieved the best objective value (Figure S10 in the Supplementary Material).

Nevertheless, the resulting ODE models match the dynamics of the experimental data, with a mean squared error of 0.15 and 0.05 for drought stress (Figure 3a) and heat stress (Figure 3c), respectively. However, they do not capture some of the early dynamics in the treatment condition (e.g., the peaks in modules 3, 4 and 6 for drought stress or the dip of module 27 for heat stress). In the drought stress experiment, the conditions are expected to be highly similar between control and treatment for the first few days, so what causes these peaks is not *a priori* clear. The fact that early peaks are not fitted well could be due to the way the external input is modelled. For drought, this is a linear increase from 0 to 1 over two weeks. Consequently, the drought condition does not have a strong signal initially in the model. For the heat stress dataset, the external input into our model is a constant 1, even though the influence of the heat stress on certain modules might change over time (Kappel et al., 2023). Hence, further exploring the representation of the external input might enhance these fits. Alternatively, the fit may be improved by including additional module interactions or by adopting a more flexible ODE formulation (Eq. (7) in Methods).

2.5. Mechanistic insights

We end up with ODEs for networks of eight and nine modules that represent 1,203 and 1,335 genes in total for drought and heat stress, respectively. Edges between modules are derived from one TF in most cases (12 edges for both drought and heat), with some edges representing two to six TFs (five and two edges for drought and heat, respectively). While comprehensive investigation of all interactions in these models is beyond the scope of this article, below we select some hypotheses from our models and relate these to earlier work (Datasets S1 and S2 in the Supplementary Material).

2.5.1. Drought. For the drought stress experiment, the original publication (Bechtold et al., 2016) found that many genes only become DE from day eight onwards, as this likely marks a shift from mild to severe drought stress. We see a similar pattern in our data (Figure 3a); many modules behave similarly in drought and control for approximately the first half of the experiment, and only diverge after that. Only two modules (25%) in our network could not be functionally annotated using GO terms (see Methods and Dataset S1 in the Supplementary Material), a far lower fraction than found in the original paper (Bechtold et al., 2016) where roughly 50% of the modules could not be annotated.

The interpretability of our approach helps generating hypotheses. For example, module 12, the only one with decreased expression levels in drought compared to control, is enriched for GO terms related to organismal-level ion homeostasis, iron ion starvation and nutrient availability, suggesting that in the late phases of drought stress these can play a significant role, which has been found before (Mukarram et al., 2021; Mulet et al., 2020, 2023). According to BADDADAN, genes related to this homeostasis are inhibited by module 4 through the GBF3 (AT2G46270) TF, a bZIP G-box binding protein that has been found to be

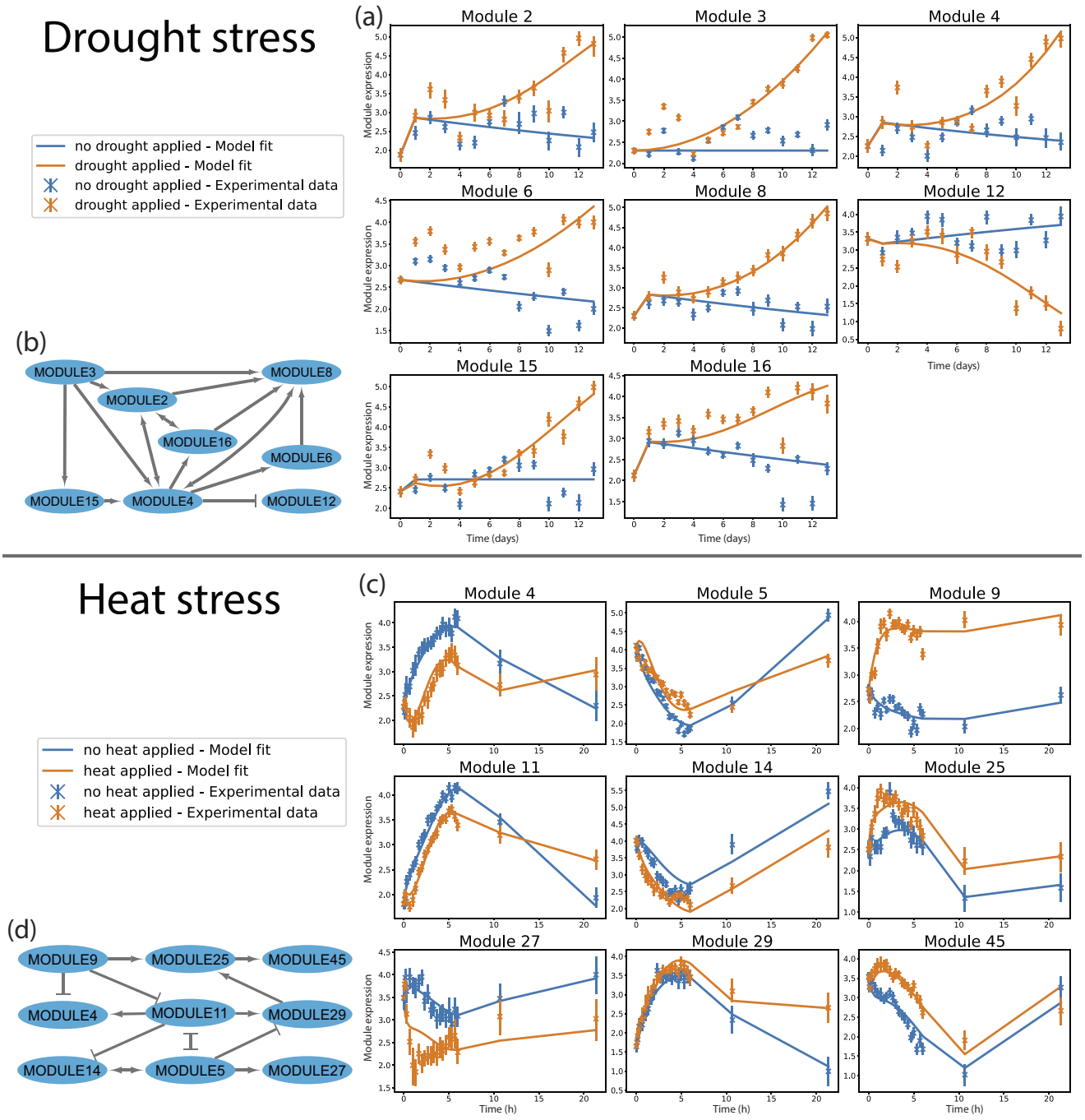

Figure 3. Intermodular networks and ODE fits. (a and c) Best ODE fit for the drought and heat stress experiments, respectively. Error bars indicate the 95% confidence interval of the mean module expression. **(b and d)** Intermodular network for drought and heat stress dataset, respectively. Arrows indicate activation, 'T'-shaped ends indicate inhibition. Each edge can represent more than one TF. Module numbering is discontinuous due to the module selection step.

induced by water deprivation and was originally found in a finger millet drought transcriptome analysis (Ramegowda et al., 2017). Additionally, overexpression of this gene in *A. thaliana* has been shown to increase drought resistance (Ramegowda et al., 2017). As a potential mechanism, the authors suggest altered ABA signalling, but BADDADAN raises the possibility that GBF3 expression could be associated with ion homeostasis in drought resistance, which would be interesting to validate experimentally.

Module 4 and module 15 are enriched for terms related to oxidative stress and salicylic acid (SA) production or response, which are known to be involved in drought stress (Ortega et al., 2024). Analysis of the original paper's data confirms that SA abundance correlates with the average expression of these modules in our ODE, with Pearson correlations of 0.61 and 0.66 for modules 4 and 15, respectively (Figure S14 in the Supplementary Material). This demonstrates how modules can be successfully linked to higher-level phenotypic traits, such as plant hormone levels. SA has been shown to induce stomatal closure (Okuma et al., 2014), and there seems to be a negative correlation between stomatal conductance (Bechtold et al., 2016) and the expression of modules 4 and 15.

Unfortunately, given the absence of the original raw measurement data, this is hard to quantify. Module 4 is also enriched for terms associated with biotic stress, in line with a role for stomatal closure in preventing water loss as well as bacterial entry (Okuma et al., 2014), and SA plays important roles in both biotic and abiotic stress (Ortega et al., 2024). Okuma et al. (2014) notes that stomatal closure is light dependent, which matches our model in that module 4 is regulated by module 2, which contains GO terms associated with light response. Module 4 has the highest closeness centrality of 1 in the network, which suggests it is more likely to play a key role in the overall response mechanism. Finally, module 16 is regulated by both modules 2 and 4 and enriched for starch, glucan and maltose metabolism. This suggests that these metabolic processes could be controlled through, e.g., circadian rhythm and light—all from module 2—but also stress-related processes from module 4, all of which have been reported (Lu et al., 2005; MacNeill et al., 2017; Skryhan et al., 2018). Knocking out *in silico* (see Methods) the CBF3 (AT4G25480) TF that connects module 2 to module 4 results in altered expression of, e.g., module 16 (Figure S15 in the Supplementary Material), and in literature, it is indeed reported that changed CBF3 expression can impact sugar metabolism (Gilmour et al., 2000). All in all, these findings show that BADDADAN recovers known mechanisms and TFs, and can link module expression to other measurements such as plant hormone abundance.

Note that the expression profiles of some modules appear relatively similar (e.g., those of modules 2 and 4). Recall that the modules are created by clustering both the local distances as well as the global distances that came from a compendium of data. The latter reflects that the two modules will have different biological functions. This is indeed true as the GO enrichments of modules 2 and 4 are different; module 2 is related to light response, circadian rhythm and organic hydroxy compound biosynthetic process, whereas module 4 is related to biotic and abiotic stress. This shows the potential benefit of clustering based on a combination of local and global expression measurements.

2.5.2. Heat. The heat stress dataset shows a large variety in up- and down-expression of modules over time, which is reflected in a large number of edges representing inhibition in the intermodular network (six inhibiting edges and eight activating ones versus one inhibiting and 16 activating in drought).

Module 4 displays interesting dynamics, initially lower in heat than control but higher at the last time point. Module 4 is linked to heat and temperature stress responses as well as protein folding. Potentially, module 4 might be involved in longer-term thermotolerance, as found recently for one of its constituent genes (HSFA2) (Pan et al., 2024). BADDADAN suggests this is potentially through protein-folding-related processes, evidence for which can be found (Wang et al., 2024). The expression pattern of module 4 seems to result from the type-2 coherent feed-forward loop (Alon, 2007) from module 9 through module 11. Module 9 inhibits module 4 through the LCL1 (AT5G02840) TF, which has been found to be involved in heat shock response (Kidokoro et al., 2023). Module 11 stimulates modules 4, 5, 14, and 29 through the ABF1 (AT1G49720) TF, which is linked to ABA response as well as heat (Song et al., 2017), highlighting it as a potential key regulator. Unfortunately, ABA was not measured in this experiment, so we cannot study how its abundance correlates to module expression. Nonetheless, knocking out ABF1 *in silico* (see Methods) reveals a stronger influence on modules 4 and 29 compared to modules 5 and 14 (Figure S15 in the Supplementary Material). Specifically, the parameters $\beta_{11,4} \approx 195$ and $\beta_{11,29} \approx 20$ are much higher than $\beta_{11,5} \approx 7 \cdot 10^{-4}$ and $\beta_{11,14} \approx 4 \cdot 10^{-4}$, highlighting that not all inferred intermodular edges represent equally strong relationships. This underscores the value of the interpretable parameters provided by our approach.

Module 5 is initially downregulated in both conditions, with a slightly higher expression in heat in the first 10 hours, after which we observe a higher expression in the control condition. This module is associated with defence response, stomatal movement and response to biotic (bacterial) stimulus. As we also found in the drought dataset, the biotic stimulus is likely related to the stomatal movement (Okuma et al., 2014). Module 5 is also enriched for programmed cell death, glucosinalate catabolism and callose localisation, which have been linked to (a)biotic stress responses (Janda et al., 2019; Hopkins et al., 2009). Overall, this shows how BADDADAN leverages prior knowledge to identify modules potentially related to combined stresses, and suggests that module 5 is an interesting target for studies on interactions between them.

Finally, Module 29 exhibits a similar response in control and heat during the initial phase of the heat shock, but at later time points is more active in heat. This could again be through a type-2 coherent feed-forward loop (Alon, 2007), this time from module 5. Module 29 is related to cellular homeostasis, glucosinolate transport and sulphur compound transport, all related to heat stress (Ihsan et al., 2019; Sugio et al., 2009). An increase in catabolism of glucosinalate (which contains sulphur) due to module 5 will lead to a need for transport. Module 5 inhibits module 29 through the LHY1 (AT1G01060) TF, linked to the circadian clock; BADDADAN suggests that it could also be involved in controlling sulphur transport, through module 29. This has been observed in *Medicago truncatula* (Achom et al., 2022), and LHY1 has been linked to sulphur in *A. thaliana* (Van De Mortel et al., 2008). As the underlying mechanism is still unknown, this is a potentially interesting direction for future research.

Of the modules that initially display a large difference between the control and heat conditions (4, 9, 25, 27 and 45), only three could be annotated for function (4, 9 and 25). Module 9 has interesting annotations such as fluid transport—which suggests the early phase response of heat is causing the plant to transport more water—and hydrogen peroxide transmembrane transport, but because it has no ingoing connections, we cannot infer what processes or TFs are involved in regulating this. This shows that BADDADAN only recovers a part of the intermodular network and thus does not necessarily provide insight into the full network; however, any model based on a subset of genes will be similarly limited.

3. Discussion

Here, we introduce BADDADAN, a framework for constructing an ODE model of gene modules encompassing more than 1,000 genes in total. We apply it to two real-world datasets covering different stresses and time scales, demonstrating its advancements in gene module expression modelling. First, by combining TFBS enrichment with expression data, it identifies directional, signed relationships between modules. Second, by estimating parameters from data that capture these relationships, it quantifies interaction strengths. Third, by combining experiment-specific expression data with compendium-wide co-expression data, it yields coherent and interpretable modules. Together, these features enable accurate predictions of gene module expression over time and *in silico*

predictions of responses to perturbations, such as knockouts, that would not be apparent from network structure alone. Moreover, BADDADAN uncovers known biological processes and their interactions while also generating novel hypotheses that can inform the design of future experiments.

BADDADAN enables out-of-the-box integration of experiment-specific data with prior experimental data (ATTED (Obayashi et al., 2022)), addressing a limitation of the widely-used WGCNA (Langfelder & Horvath, 2008), which does not directly support such integration and assumes a scale-free network topology. In contrast to recent black-box ML approaches (Baker et al., 2018; Sapoval et al., 2022; Shen et al., 2021), BADDADAN is composed of intermediate steps that are inspectable; e.g., module selection provides z-scores, connections between modules can be visualised as a directed graph, and the ODE model contains human-interpretable parameters (e.g., β can be linked to interaction strength). Moreover, existing ODE modelling approaches comparable to BADDADAN (Lu et al., 2011; Wu et al., 2014) infer connections between modules solely in a data-driven way—similar to LASSO regression—which has been shown to underperform for GRN inference (Saelens et al., 2018). However, as potential future research, such connections could be integrated with our TFBS-based ones, a hybrid approach similar to Aibar et al. (2017) where connections are inferred if they are important for predicting dynamics and/or underpinned by the presence of a TFBS. Also, incorporation of additional data, e.g., in the form of chromatin accessibility, could further inform finding of regulatory connections (Staut et al., 2024).

By addressing essential challenges in gene module network modelling—such as edge directionality determination, activation/inhibition inference and associated weights estimation within an ODE framework—BADDADAN opens up extensive possibilities for further analysis. For example, looking deeper into confidence interval estimation of parameters, applying an identifiability analysis to our model or performing topological sensitivity analysis (Babtie et al., 2014), where multiple different network topologies are tested for their ability to describe the data. This will tell us if we have found a network topology which would uniquely describe the resulting behaviour. Related to this, approaches for optimisation of a given network topology by adding or removing a small number of nodes (modules in our case) could be applied (Astola et al., 2014). Also, modelling choices (e.g., the value of the Hill coefficient, linear environmental effects or additive module interactions) can significantly affect the fitted parameter values—and possibly subsequent predictions of knockout effects—so care should be taken when interpreting them. Experimental validation of BADDADAN's hypotheses will help support or refine these modelling assumptions.

Settings of each individual step are now optimised separately in the pipeline, but it would be an interesting to research if gains can be made by optimising the pipeline end-to-end through differentiable optimisation methods developed for deep learning (AlQuraishi & Sorger, 2021). This would also allow the use of deep-learning-based embeddings for clustering, higher dimensional representations of module expressions and neural ODEs or biologically informed neural networks for ODE fitting. On top of that, the loss function could be extended to make modules adhere to, e.g., rate limits or other existing biological knowledge.

Furthermore, to aid in integration with existing or novel methodologies, BADDADAN builds on existing standards and software. For parametrisation of the ODE, we utilise standard formats and scripts for systems biology models (Schälte et al., 2023; Schmiester et al., 2021), and we save our models as SBML format files (Hucka et al., 2003), allowing interaction with the wide ecosystem that is already available for systems biology models. This would allow us, for example, to couple BADDADAN to high-quality ODE models for small GRNs that are already available (Chew et al., 2022; Nieto et al., 2022; Valentim et al., 2015). Given that both are composed of ODEs, integration should be possible. For example, we could use a BADDADAN module related to the production of a metabolite—similar to how we linked two modules to SA in the drought dataset—and use the abundance of that metabolite predicted by BADDADAN as input for an existing mechanistic model. Vice versa, mechanistic models that predict metabolite abundance could be used as input (cf. $c(t)$ or $u(t)$) to BADDADAN. Alternatively, a TF in the BADDADAN model could be taken to directly influence expression of a gene in the mechanistic model, improving its prediction accuracy.

BADDADAN may face computational challenges in parameter estimation, similar to those in GRN modelling, when applied to datasets requiring ODE modelling of large numbers of modules. However, this seems unlikely to pose a problem in practice, as we successfully apply BADDADAN here to real-world datasets representing complex biological systems. If needed, the number of modules requiring ODE modelling could be reduced by adjusting the deepsplit parameter or by being more stringent in the module selection step, at the cost of modelling accuracy, coherence and/or biological interpretability.

Both datasets contained microarray measurements, so BADDADAN's performance on RNA-seq data remains untested. However, existing studies suggest that clustering structures are similar between microarray and RNA-seq data (Sîrbu et al., 2012), which would suggest that BADDADAN's module finding would also work on RNA-seq. However, how the specific characteristics of RNA-seq measurements influence BADDADAN's ODE fitting of module dynamics is still unknown. Hence, the parameter limits and form of the ODEs might have to be re-evaluated. To further investigate this, we have concrete plans to apply BADDADAN to time-series RNA-seq data in the future.

Finally, in the future, BADDADAN could be applied to datasets of other (non-)plant species and/or stresses and could form the basis for translational science, using fundamental insights into model organism *A. thaliana* resilience mechanisms to breed relevant crops with higher resilience. Since our model can predict possible effects of knockouts under different stresses, it provides possibilities to study the effects of genetic variation on resilience. Also, a comparative analysis of modules between different individual or combined stresses could give insight into the temporal dynamics of processes important for multiple stresses specifically (Rizhsky et al., 2002; Verslues et al., 2023), which would further support targeted resilient crop development. For application to non-model species, alternatives to TF2Network and ATTED should be found. This could involve inferring regulatory interactions purely from expression data or applying the ATTED methodology to existing datasets available for the species of interest.

Summarising, BADDADAN provides a proof-of-concept approach for inferring gene module-based dynamic models from time-series transcriptomics data, to help unravel the intricate gene regulatory mechanisms that plants employ to deal with stresses. Increasing knowledge in this area might eventually be used to breed more climate resilient crops which can be used to provide food to a growing world population, even under the pressure of climate change.

4. Methods

4.1. Gene expression input data

Progressive drought microarray expression data (Bechtold et al., 2016) was extracted from Gene Expression Omnibus (GEO) (Barrett et al., 2013) accession GSE65046, and probe names were converted to TAIR gene IDs using the Python package GEOParse (v2.0.3) (Gumienny, 2024). For the heat dataset (Caldana et al., 2011), expression data was extracted from ArrayExpress accession E-MTAB-375 and again probe names were mapped to gene identifiers using the annotations available on GEO (Barrett et al., 2013). Both original datasets were already $\log_2$-transformed to make noise additive instead of multiplicative and to stabilise variance (Archer et al., 2004).

4.2. Differential gene expression

We used *limma* (v3.60.3) (Ritchie et al., 2015) in R (v4.4.0) for differential gene expression analysis, as it has been shown to perform well on time-course microarray datasets (Moradzadeh et al., 2019). For the drought stress dataset, contrasts were used to test if genes changed expression differently between subsequent time points in the drought group compared to the control group. However, such contrast testing only works if multiple biological replicates are available, which was not the case for the heat stress data. Thus, we ran *limma* on this dataset by fitting a natural cubic spline with five degrees of freedom through the time points using the R *splines* library. Next, we used *limma* to conduct an *F*-test between the conditions on the five parameters that correspond to interaction, i.e., difference between the control and heat groups. Finally, for both drought and control, only genes with a Benjamini-Hochberg false discovery rate (FDR) ≤ 0.05 were kept for further processing.

4.3. Module inference

To obtain global (compendium-wide) co-expression patterns, we downloaded pairwise *z*-scores from ATTED v11.1 (Obayashi et al., 2022), which were calculated based on 27,427 samples and corrected for biases due to differences in sampling conditions and sample sizes (Obayashi et al., 2022). The *z*-scores (denoted with W_{global}) were converted to a distance matrix D_{global}, which represents pairwise distances between genes, with each row and column corresponding to a gene:

$$D_{\text{global}} = \max(W_{\text{global}}) - W_{\text{global}}, \tag{1}$$

where $\max(M) = \max_{i,j} M_{ij}$ is the maximum value over all rows and columns. This is consistent throughout the manuscript and also holds for uses of $\min$.

Local distances were calculated within each dataset as follows:

$$D_{\text{local}} = 1 - C, \tag{2}$$

where C is the matrix of pairwise Pearson correlations between genes over the samples (taking the average expression level if biological replicates were available). Saelens et al. (2018) found that this 'unsigned' (Langfelder et al., 2011) metric yielded higher module detection accuracy compared to, e.g., an absolute Pearson correlation or Spearman correlation.

To obtain a clustering based on both global and local co-expression, we combined D_{global} and D_{local}. To avoid biases due to a difference in distributions, we first separately converted both distances to a *z*-score

$$Z = \frac{D - \mu_D}{\sigma_D}, \tag{3}$$

where μ_D is the average and σ_D is the standard deviation of all distances.

Next, we summed the resulting *z*-score matrices:

$$Z_{\text{combined}} = w \cdot Z_{\text{global}} + (1 - w) \cdot Z_{\text{local}}, \tag{4}$$

where w influences the balance between global and local co-expression; i.e., a lower w would result in higher coherence if that is more desired (Figure S2 in the Supplementary Material). We set $w = 0.5$ for all experiments as we want an equal balance between interpretability and coherence.

Finally, to ensure no distances were negative (needed for clustering), we converted them to a positive distance:

$$D_{\text{combined}} = Z_{\text{combined}} - \min(Z_{\text{combined}}). \tag{5}$$

To find modules, we use hierarchical clustering. This has been shown to obtain reasonable performance in gene module detection (Saelens et al., 2018), where it was benchmarked on simulated and experimental data with a known ground-truth. Average linkage hierarchical clustering was performed in R on D_{combined}, D_{global} and D_{local}. To extract modules from the resulting dendrogram, it could be cut at a specific height, but such a global approach has been shown to be suboptimal for gene module detection (Langfelder et al., 2008). Thus, we applied a Dynamic Tree Cut method (Langfelder et al., 2008) designed specifically for gene module detection. Also, since module size affects both coherence (e.g., modules of size 1 are always 100% coherent) and interpretability (e.g., a module containing all genes would lack enriched GO terms), we ensured that module sizes were roughly equal between the three clusterings. To achieve this, we tuned the deepsplit parameter (Figure S1 in the Supplementary Material). Increasing deepsplit changes the internal dynamic tree cut thresholds used to define clusters, promoting splits that yield lower within-cluster and higher between-cluster dissimilarity (Langfelder et al., 2008). This results in more, but smaller, clusters. For the drought stress data, we set deepsplit=1 (default value) for the global and combined distance matrices, but used deepsplit=2 for the local one, as this ensured cluster sizes similar to the other two clusterings (Figure S1a in the Supplementary Material). For the heat dataset, module sizes were comparable for all input distances at deepsplit=1 (default value) (Figure S1b in the Supplementary Material). All other Dynamic Tree Cut settings were kept at their default values.

4.4. Characteristics of clusterings

We compared our local, combined and global distance-based clusterings on coherence and interpretability (GO enrichment) of the resulting modules.

4.4.1. Coherence. To assess coherence, we performed a PCA on each module separately, where we first normalised the expression of each gene to have mean 0 and variance 1 (Eq. 3). Coherence was then calculated as the percentage of variance explained by the first principal component, found using the PCA function in Scikit-learn (v1.1.3) (Pedregosa et al., 2011). We plotted the distribution of these coherence scores over all modules using Seaborn (v0.11) (Waskom, 2021).

4.4.2. GO enrichment. GO enrichment was performed using GOAtools (v1.2.3) (Klopfenstein et al., 2018) with all DE genes as background. Only gene annotations of experimental

evidence codes were used to prevent data leakage from existing co-expression-based electronic annotations. We focused on biological process GO terms, as we were interested in biological interpretation of the gene modules. For all modules, only enriched GO terms with a Benjamini-Hochberg FDR ≤ 0.05 were used for further analysis. To study the variability in the fraction of modules that had at least one enriched GO term, we used Seaborn's bootstrapping (Waskom, 2021) to estimate the 95% confidence interval.

4.5. Statistical testing

Statistical tests were performed using the Python package statannotations v(0.5.0) (Charlier et al., 2022). Because we could not assume all data to be normally distributed, we compared coherence and interpretability using the non-parametric two-sided Mann-Whitney U test. As a null model, we measured the characteristics of random modules, where we ensured that the module sizes were distributed identically to those of the combined distances by randomly permuting these cluster assignments.

4.6. Module selection

To select candidate modules for modelling, we assigned each module a score composed of four different components. 1) Activity: to obtain modules that are composed of genes that are active, we determined the average expression over all samples of all genes in a module. 2) Coherence: to measure the coherence of a module, we performed PCA on the normalised expression of its constituent genes over time in our dataset of interest and measured the percentage of variance explained by the first principal component, as described above. Coherence is desired because it facilitates representation of the module expression by a single value (e.g., average). 3) Relevance: to assess whether a module's expression differs between treatment and control, we first summarised the module by calculating the average expression of its genes over time for each condition. We then measured the difference between conditions using the mean squared error of these averaged expression profiles. 4) Regulatory relevance: we assumed modules with a high number of enriched TFBSs were more likely to be important for a plant's stress response; e.g., genes related to stimulus response and gene regulation in *A. thaliana* are known to have a higher number of regulators (Heyndrickx et al., 2014). Moreover, modules with many enriched TFBS are more likely to successfully be linked to other modules in our model. To score this regulatory relevance, we used the count of TFBS N_T that were enriched in a module according to TF2Network (Kulkarni et al., 2018). Since this distribution was highly skewed, we applied a transformation $\log(N_T + 1)$. We chose not to include interpretability (i.e., GO enrichment) in the scoring to ensure that it remains an independent criterion for evaluating the selected modules.

To ensure no single criterion dominated the selection, all scores were scaled to z-scores through Eq. (3). Subsequently, we combined these individual z-scores using Stouffer's method to obtain one combined z-score per module because this allowed for subsequent statistical interpretation and filtering:

$$Z = \frac{\sum_{i=1}^{k} Z_i}{\sqrt{k}}. \tag{6}$$

Here, k is the number of different characteristics (in our case, four). Only the activity was based on non-normalised gene expression, the other scores were calculated from gene expression where each gene was z-scored over all samples. We selected the top 50% (a user parameter, here chosen arbitrarily) of all modules based on their summary score for subsequent modelling. The module selection step results in an intermodular network that has a higher density compared to the network from modules that have not been filtered (Table S3 in the Supplementary Material).

4.7. Model structure

To infer connections between modules, we used the TFBS enrichment results of TF2Network (Kulkarni et al., 2018). We inferred regulatory interactions between modules based on three criteria: 1) The target module contained an enriched binding site for a TF that belonged to a source module. 2) The TF had a Pearson correlation of at least 0.3 with the average expression of its source module; any TFs with correlations to their source module lower than this were considered an outlier and excluded from regulatory interaction inference. 3) The absolute Pearson correlation between the expression of the TF and the average expression of its target module had to exceed a certain threshold. This step fulfilled two functions. Firstly, it reduced the number of false positives that could result from TFBS enrichment analysis, e.g., due to chromatin accessibility (De Clercq et al., 2021; Khamis et al., 2018; Kulkarni et al., 2018). Secondly, as we needed to choose a different form of the Hill equation for inhibition of activation (Eq. 7), we used the sign of the Pearson correlation as a heuristic to infer up- or down-regulation. If multiple TFs from the same source module regulated the same target module, we asserted they all consistently up- or down-regulated; if there was disagreement, no valid network would be output by BADDADAN. Finally, modules that were left with no interactions with any other modules were removed from the model.

To determine the correlation cutoff between TFs and their target modules, we inferred model structures as described above for cutoffs varying from 0 to 0.95 at intervals of 0.05, and kept track of the total number of modules, interactions and whether the model graph consisted of one single weakly connected component or multiple disjoint components (Figure S8 in the Supplementary Material). We then picked 0.75 as a cutoff for both the heat and drought stress datasets, as this led to the sparsest model not split into disjoint components. Intermodular network visualisation and analysis were performed in NetworkX (v2.8.8) (Hagberg et al., 2008) and Cytoscape (v3.10.3) (Shannon et al., 2003).

4.8. ODE model

For each module, we model its activity in an ODE representing the average gene expression levels over time after normalisation (Eq. 3) because related genes can have similar expression patterns but at different magnitudes (Wu et al., 2014). Since z-score normalisation of the gene expression could result in negative values, a constant of 3 (minimum value of module expression was around −2.5) was added to guarantee positive values, preventing biologically implausible scenarios where negative expressions would invert the decay term and falsely suggest gene production. For a module m that is inhibited by a set of modules I and activated by a set of modules A, we represent its rate of change as follows:

$$\frac{dy_m}{dt} = \gamma_m \cdot u(t) - \delta_m \cdot y_m + c_m(t) + \sum_{i \in I} \beta_{i,m} \frac{k_{i,m}^2}{k_{i,m}^2 + y_i^2} + \sum_{a \in A} \beta_{a,m} \frac{y_a^2}{k_{a,m}^2 + y_a^2}. \tag{7}$$

The expression of each module y_m is dictated by an external stimulus-dependent term weighted by γ_m, a decay term governed by the parameter δ_m, a circadian clock input $c_m(t)$ that we only used for the heat stress dataset (because this is sampled over a period of 24 hours) and activation/inhibition Hill equations of the second order where $\beta_{n,m}$ represents the maximum magnitude of the activation/inhibition of module n on module m, and $k_{n,m}$ the abundance of the regulator at which the module is 50% activated/inhibited. We chose these formulas as gene regulation is highly nonlinear and Hill kinetics are widely used in ODE modelling in systems biology (Frank, 2013; Polynikis et al., 2009; Valentim et al., 2015). To investigate the influence of the Hill coefficients, i.e., the strength of the nonlinearity, we also performed parameter optimisation (see below) with Hill coefficients of 1 and 3, and found that the ability to fit the data was not lost when the strength of nonlinearity was changed (Figure S11 in the Supplementary Material). However, resulting parameter values differed significantly (Figures S12 and S13 in the Supplementary Material), suggesting that the choice of Hill coefficient alters BADDADAN's prediction of underlying dynamics.

The function $u(t)$ captures the external input, which was different for the drought and heat experiment. For the former, we modelled the linear increase in drought stress over two weeks (Bechtold et al., 2016) as follows:

$$u(t) = \begin{cases} \frac{t}{13}, & 0 \le t \le 13, \quad \text{if drought,} \\ 0, & \text{control,} \end{cases} \tag{8}$$

to simulate a linear increase of drought where t is time in days.

For the latter, heat stress, we modelled the external input as follows:

$$u(t) = \begin{cases} 1, & 0 \le t \le 24, \quad \text{if heat,} \\ 0, & \text{control,} \end{cases} \tag{9}$$

as is also common in plant mechanistic modelling papers (Nieto et al., 2022). Here, we measured t in hours.

Initially, the model struggled to accurately capture the data in the heat experiment, likely due to the circadian clock having a strong influence on gene expression over a 24-hour period (which is not observed during daily sampling in the drought dataset). To address this, we explicitly included a circadian clock input $c_m(t)$ for each module:

$$c_m(t) = \begin{cases} a_m \sin\left(\frac{t}{24} 2\pi + \phi_m\right) + b_m, & 0 \le t \le 24, \quad \text{if heat,} \\ 0, & \text{if drought.} \end{cases} \tag{10}$$

Here, a_m quantifies the input strength of the circadian clock on the module expression change, ϕ_m determines the phase offset of the circadian clock influence (which we give a 24-hour period by definition) and b_m models the baseline influence of the circadian clock. Again, t is measured in hours. Finally, BADDADAN exported the full ODE models in the Systems Biology Markup Language (Hucka et al., 2003) to ensure reproducibility, consistency and interoperability.

4.9. Parameter optimisation

We specified the parameter optimisation problem per dataset (drought or heat) in PETAB (Schmiester et al., 2021) and optimised the parameters of the ODE model to match experimental measurements on both the control and treatment conditions using PyPESTO (Schälte et al., 2023) and AMICI (Fröhlich et al., 2021) for ODE integration. We started the ODE model with $t = 0$ being the measured initial module expressions and predicted the full time course data from there. Both conditions shared all parameters except for $u(t)$ and $c_m(t)$. We minimised the negative log-likelihood objective function using the L-BFGS optimisation algorithm implemented in SciPy (Virtanen et al., 2020) and AMICI's adjoint sensitivity method. For fitting, PyPESTO required the noise level for each module to calculate the negative log-likelihood. We set this value to 0.1, based on the 95% confidence interval estimated through bootstrapping of the mean per module. Parameter limits were chosen empirically to ensure the parameters of a best fit never hit their upper or lower bounds (Table S2 in the Supplementary Material), and all non-negative parameters were estimated on a $\log_{10}$ scale to aid exploring different orders of magnitude (Schälte et al., 2023). To maximally explore the optimisation landscape, models with random initial parameters were initialised 5,000 times in parallel, and waterfall plots were created to inspect if a global minimum was found (Figure S10 in the Supplementary Material). In the heat dataset, the last two data points ($t \approx 10$ and $t \approx 20$) were oversampled fourfold to mitigate the effects of uneven sampling and ensure a more balanced representation across time points. Finally, to predict TF knockout effects *in silico*, we set the parameter β of the associated edge(s) to 10^{-6}. ODE fits and resulting parameters were stored using MLFLOW (Chen et al., 2020). Optimisation took approximately four days per dataset on a server equipped with two Intel(R) Xeon(R) E5-2640 v3 CPUs.

Open peer review. To view the open peer review materials for this article, please visit http://doi.org/10.1017/qpb.2025.10016.

Acknowledgements

The authors thank Tijmen den Butselaar for helping find the public data that was used to test BADDADAN, Elena Del Pup and Francesca Giaume for helpful discussions regarding the visualisation and biological interpretation of the results and Jordi Alonso Esteve for input on the methodology.

Competing interests. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability statement. All source code is available at https://github.com/CropXR/BADDADAN.

Author contributions. B.N., A.D.J.vD. and D.dR. conceived the initial idea. B.N. carried out the research and wrote the initial manuscript. M.R., A.D.J.vD. and D.dR. provided critical feedback. All authors have read and approved the final manuscript.

Funding statement. This publication is part of the long-term program PlantXR: A new generation of breeding tools for extra-resilient crops (KICH3.LTP.20.005), which is financed by the Dutch Research Council (NWO), the Foundation for Food & Agriculture Research (FFAR), companies in the plant breeding and processing industry, and Dutch universities. These parties collaborate in the CropXR Institute (www.cropxr.org) that is founded through the National Growth Fund (NGF) of The Netherlands.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/qpb.2025.10017.

References

Achom, M., Roy, P., Lagunas, B., Picot, E., Richards, L., Bonyadi-Pour, R., Pardal, A. J., Baxter, L., Richmond, B. L., Aschauer, N., Fletcher, E. M., Rowson, M., Blackwell, J., Rich-Griffin, C., Mysore, K. S., Wen, J., Ott, S., Carré, I. A., & Gifford, M. L. (2022). Plant circadian clock control of Medicago truncatula nodulation via regulation of nodule cysteine-rich

peptides. *Journal of Experimental Botany*, **73**(7), 2142–2156. https://doi.org/10.1093/jxb/erab526

Aibar, S., **González-Blas, C. B.**, **Moerman, T.**, **Huynh-Thu, V. A.**, **Imrichova, H.**, **Hulselmans, G.**, **Rambow, F.**, **Marine, J.-C.**, **Geurts, P.**, **Aerts, J.**, **van den Oord, J.**, **Atak, Z. K.**, **Wouters, J.**, & **Aerts, S.** (2017). SCENIC: Single-cell regulatory network inference and clustering. *Nature Methods*, **14**(11), 1083–1086. https://doi.org/10.1038/nmeth.4463

Alber, M., **Buganza Tepole, A.**, **Cannon, W. R.**, **De, S.**, **Dura-Bernal, S.**, **Garikipati, K.**, **Karniadakis, G.**, **Lytton, W. W.**, **Perdikaris, P.**, **Petzold, L.**, & **Kuhl, E.** (2019). Integrating machine learning and multiscale modeling—Perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *Npj Digital Medicine*, **2**(1), 1–11. https://doi.org/10.1038/s41746-019-0193-y

Alon, U. (2007). Network motifs: Theory and experimental approaches. *Nature Reviews Genetics*, **8**(6), 450–461. https://doi.org/10.1038/nrg2102

AlQuraishi, M., & **Sorger, P. K.** (2021). Differentiable biology: Using deep learning for biophysics-based and data-driven modeling of molecular mechanisms. *Nature Methods*, **18**(10, 1169–1180. https://doi.org/10.1038/s41592-021-01283-4

Archer, K. J., **Dumur, C. I.**, & **Ramakrishnan, V.** (2004). Graphical technique for identifying a monotonic variance stabilizing transformation for absolute gene intensity signals. *BMC Bioinformatics*, **5**(1), 60. https://doi.org/10.1186/1471-2105-5-60

Astola, L., **Stigter, H.**, **van Dijk, A. D. J.**, **van Daelen, R.**, & **Molenaar, J.** (2014). Inferring the gene network underlying the branching of tomato inflorescence. *PloS One*, **9**(4), e89689. https://doi.org/10.1371/journal.pone.0089689

Babtie, A. C., **Kirk, P.**, & **Stumpf, M. P. H.** (2014). Topological sensitivity analysis for systems biology. *Proceedings of the National Academy of Sciences*, **111**(52), 18507–18512. https://doi.org/10.1073/pnas.1414026112

Baker, R. E., **Peña, J.-M.**, **Jayamohan, J.**, & **Jérusalem, A.** (2018). Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biology Letters*, **14**(5), 20170660. https://doi.org/10.1098/rsbl.2017.0660

Barrett, T., **Wilhite, S. E.**, **Ledoux, P.**, **Evangelista, C.**, **Kim, I. F.**, **Tomashevsky, M.**, **Marshall, K. A.**, **Phillippy, K. H.**, **Sherman, P. M.**, **Holko, M.**, **Yefanov, A.**, **Lee, H.**, **Zhang, N.**, **Robertson, C. L.**, **Serova, N.**, **Davis, S.**, & **Soboleva, A.** (2013). NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Research*, **41**(D1), D991–D995. https://doi.org/10.1093/nar/gks1193

Bechtold, U., **Penfold, C. A.**, **Jenkins, D. J.**, **Legaie, R.**, **Moore, J. D.**, **Lawson, T.**, **Matthews, J. S.**, **Vialet-Chabrand, S. R.**, **Baxter, L.**, **Subramaniam, S.**, **Hickman, R.**, **Florance, H.**, **Sambles, C.**, **Salmon, D. L.**, **Feil, R.**, **Bowden, L.**, **Hill, C.**, **Baker, N. R.**, **Lunn, J. E.**, . . . **Mullineaux, P. M.** (2016). Time-series transcriptomics reveals that AGAMOUS-LIKE22 affects primary metabolism and developmental processes in drought-stressed arabidopsis. *The Plant Cell*, **28**(2), 345–366. https://doi.org/10.1105/tpc.15.00910

Caldana, C., **Degenkolbe, T.**, **Cuadros-Inostroza, A.**, **Klie, S.**, **Sulpice, R.**, **Leisse, A.**, **Steinhauser, D.**, **Fernie, A. R.**, **Willmitzer, L.**, & **Hannah, M. A.** (2011). High-density kinetic analysis of the metabolomic and transcriptomic response of Arabidopsis to eight environmental conditions. *The Plant Journal*, **67**(5), 869–884. https://doi.org/10.1111/j.1365-313X.2011.04640.x

Cao, J., **Qi, X.**, & **Zhao, H.** (2012). Modeling gene regulation networks using ordinary differential equations. In **J. Wang**, **A. C. Tan**, & **T. Tian** (Eds.), *Next generation microarray bioinformatics: Methods and protocols* (Methods in Molecular Biology, pp. 185–197). Humana Press. https://doi.org/10.1007/978-1-61779-400-1_12

Charlier, F., **Weber, M.**, **Izak, D.**, **Harkin, E.**, **Magnus, M.**, **Lalli, J.**, **Fresnais, L.**, **Chan, M.**, **Markov, N.**, **Amsalem, O.**, **Proost, S.**, **Krasoulis, A.**, & **Getzze Repplinger, S.** (2022). Statannotations. Zenodo.

Chen, A., **Chow, A.**, **Davidson, A.**, **DCunha, A.**, **Ghodsi, A.**, **Hong, S. A.**, **Konwinski, A.**, **Mewald, C.**, **Murching, S.**, **Nykodym, T.**, **Ogilvie, P.**, **Parkhe, M.**, **Singh, A.**, **Xie, F.**, **Zaharia, M.**, **Zang, R.**, **Zheng, J.**, & **Zumar, C.** (2020). Developments in MLflow: A system to accelerate the machine learning lifecycle. In *Proceedings of the fourth international workshop on data management for end-to-end machine learning DEEM'20* (pp. 1–4). Association for Computing Machinery. https://doi.org/10.1145/3399579.3399867

Chew, Y. H., **Seaton, D. D.**, **Mengin, V.**, **Flis, A.**, **Mugford, S. T.**, **George, G. M.**, **Moulin, M.**, **Hume, A.**, **Zeeman, S. C.**, **Fitzpatrick, T. B.**, **Smith, A. M.**, **Stitt, M.**, & **Millar, A. J.** (2022). The Arabidopsis framework model version 2 predicts the organism-level effects of circadian clock gene mis-regulation. *Silico Plants*, **4**(2), diac010. https://doi.org/10.1093/insilicoplants/diac010

Dai, A. (2011). Drought under global warming: A review. *WIREs Climate Change*, **2**(1), 45–65. https://doi.org/10.1002/wcc.81

De Clercq, I., **Van de Velde, J.**, **Luo, X.**, **Liu, L.**, **Storme, V.**, **Van Bel, M.**, **Pottie, R.**, **Vaneechoutte, D.**, **Van Breusegem, F.**, & **Vandepoele, K.** (2021). Integrative inference of transcriptional networks in Arabidopsis yields novel ROS signalling regulators. *Nature Plants*, **7**(4), 500–513. https://doi.org/10.1038/s41477-021-00894-1

de Luis Balaguer, M. A., **Fisher, A. P.**, **Clark, N. M.**, **Fernandez-Espinosa, M. G.**, **Möller, B. K.**, **Weijers, D.**, **Lohmann, J. U.**, **Williams, C.**, **Lorenzo, O.**, & **Sozzani, R.** (2017). Predicting gene regulatory networks by combining spatial and temporal gene expression data in Arabidopsis root stem cells. *Proceedings of the National Academy of Sciences*, **114**(36), E7632–E7640. https://doi.org/10.1073/pnas.1707566114

Depuydt, T., & **Vandepoele, K.** (2021). Multi-omics network-based functional annotation of unknown Arabidopsis genes. *The Plant Journal*, **108**(4), 1193–1212. https://doi.org/10.1111/tpj.15507

Ficklin, S. P., & **Feltus, F. A.** (2011). Gene coexpression network alignment and conservation of gene modules between two grass species: Maize and rice. *Plant Physiology*, **156**(3), 1244–1256. https://doi.org/10.1104/pp.111.173047

Forster, D. T., **Li, S. C.**, **Yashiroda, Y.**, **Yoshimura, M.**, **Li, Z.**, **Isuhuaylas, L. A. V.**, **Itto-Nakama, K.**, **Yamanaka, D.**, **Ohya, Y.**, **Osada, H.**, **Wang, B.**, **Bader, G. D.**, & **Boone, C.** (2022). BIONIC: Biological network integration using convolutions. *Nature Methods*, **19**(10), 1250–1261. https://doi.org/10.1038/s41592-022-01616-x

Frank, S. A. (2013). Input-output relations in biological systems: Measurement, information and the Hill equation. *Biology Direct*, **8**(1), 31. https://doi.org/10.1186/1745-6150-8-31

Fröhlich, F., **Kessler, T.**, **Weindl, D.**, **Shadrin, A.**, **Schmiester, L.**, **Hache, H.**, **Muradyan, A.**, **Schütte, M.**, **Lim, J.-H.**, **Heinig, M.**, **Theis, F. J.**, **Lehrach, H.**, **Wierling, C.**, **Lange, B.**, & **Hasenauer, J.** (2018). Efficient parameter estimation enables the prediction of drug response using a mechanistic pan-cancer pathway model. *Cell Systems*, **7**(6), 567–579.e6. https://doi.org/10.1016/j.cels.2018.10.013

Fröhlich, F., **Weindl, D.**, **Schälte, Y.**, **Pathirana, D.**, **Paszkowski, Ł.**, **Lines, G. T.**, **Stapor, P.**, & **Hasenauer, J.** (2021). AMICI: High-performance sensitivity analysis for large ordinary differential equation models. *Bioinformatics*, **37**(20), 3676–3677. https://doi.org/10.1093/bioinformatics/btab227

Gilmour, S. J., **Sebolt, A. M.**, **Salazar, M. P.**, **Everard, J. D.**, & **Thomashow, M. F.** (2000). Overexpression of the arabidopsis CBF3 transcriptional activator mimics multiple biochemical changes associated with cold acclimation. *Plant Physiology*, **124**(4), 1854–1865

Gumienny, R. (2024). GEOparse. GitHub. https://github.com/guma44/GEOparse.

Hagberg, A. A., **Schult, D. A.**, and **Swart, P. J.** (2008). Exploring network structure, dynamics, and function using NetworkX. In **G. Varoquaux**, **T. Vaught**, & **J. Millman** (Eds.), *Proceedings of the 7th python in science conference* (pp. 11–15). Pasadena.

Han, M., **Liu, X.**, **Zhang, W.**, **Wang, M.**, **Bu, W.**, **Chang, C.**, **Yu, M.**, **Li, Y.**, **Tian, C.**, **Yang, X.**, **Zhu, Y.**, & **He, F.** (2021). TSMiner: A novel framework for generating time-specific gene regulatory networks from time-series expression profiles. *Nucleic Acids Research*, **49**(18), e108. https://doi.org/10.1093/nar/gkab629

Heard, N. A., **Holmes, C. C.**, **Stephens, D. A.**, **Hand, D. J.**, & **Dimopoulos, G.** (2005). Bayesian coclustering of Anopheles gene expression time series: Study of immune defense response to multiple experimental challenges. *Proceedings of the National Academy of Sciences*, **102**(47), 16939–16944. https://doi.org/10.1073/pnas.0408393102

Heyndrickx, K. S., & **Vandepoele, K.** (2012). Systematic identification of functional plant modules through the integration of complementary data sources. *Plant Physiology*, **159**(3), 884–901. https://doi.org/10.1104/pp.112.196725

Heyndrickx, K. S., **Van de Velde, J.**, **Wang, C.**, **Weigel, D.**, & **Vandepoele, K.** (2014). A functional and evolutionary perspective on transcription factor binding in Arabidopsis thaliana[C][W]. *The Plant Cell*, **26**(10), 3894–3910. https://doi.org/10.1105/tpc.114.130591

Hopkins, R. J., **van Dam, N. M.**, & **van Loon, J. J. A.** (2009). Role of glucosinolates in insect-plant relationships and multitrophic interactions. *Annual Review of Entomology*, **54**, 57–83. https://doi.org/10.1146/annurev.ento.54.110807.090623

Hucka, M., **Finney, A.**, **Sauro, H. M.**, **Bolouri, H.**, **Doyle, J. C.**, **Kitano, H.**, **Arkin, A. P.**, **Bornstein, B. J.**, **Bray, D.**, **Cornish-Bowden, A.**, **Cuellar, A. A.**, **Dronov, S.**, **Gilles, E. D.**, **Ginkel, M.**, **Gor, V.**, **Goryanin, I. I.**, **Hedley, W. J.**, **Hodgman, T. C.**, **Hofmeyr, J.-H.**, … **Wang, J.** (2003). The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**(4), 524–531. https://doi.org/10.1093/bioinformatics/btg015

Huynh-Thu, V. A., & **Geurts, P.** (2018). dynGENIE3: Dynamical GENIE3 for the inference of gene networks from time series expression data. *Scientific Reports*, **8**(1), 3384. https://doi.org/10.1038/s41598-018-21715-0

Ihsan, M. Z., **Daur, I.**, **Alghabari, F.**, **Alzamanan, S.**, **Rizwan, S.**, **Ahmad, M.**, **Waqas, M.**, & **Shafqat, W.** (2019). Heat stress and plant development: Role of sulphur metabolites and management strategies. *Acta Agriculturae Scandinavica, Section B—Soil & Plant Science*, **69**(4), 332–342. https://doi.org/10.1080/09064710.2019.1569715

Janda, M., **Lamparová, L.**, **Zubíková, A.**, **Burketová, L.**, **Martinec, J.**, & **Krčková, Z.** (2019). Temporary heat stress suppresses PAMP-triggered immunity and resistance to bacteria in Arabidopsis thaliana. *Molecular Plant Pathology*, **20**(7), 1005–1012. https://doi.org/10.1111/mpp.12799

Janky, R., **Verfaillie, A.**, **Imrichová, H.**, **de Sande, B. V.**, **Standaert, L.**, **Christiaens, V.**, **Hulselmans, G.**, **Herten, K.**, **Sanchez, M. N.**, **Potier, D.**, **Svetlichnyy, D.**, **Atak, Z. K.**, **Fiers, M.**, **Marine, J.-C.**, & **Aerts, S.** (2014). iRegulon: From a gene list to a gene regulatory network using large motif and track collections. *PLoS Computational Biology*, **10**(7), e1003731. https://doi.org/10.1371/journal.pcbi.1003731

Julca, I., **Ferrari, C.**, **Flores-Tornero, M.**, **Proost, S.**, **Lindner, A.-C.**, **Hackenberg, D.**, **Steinbachová, L.**, **Michaelidis, C.**, **Gomes Pereira, S.**, **Misra, C. S.**, **Kawashima, T.**, **Borg, M.**, **Berger, F.**, **Goldberg, J.**, **Johnson, M.**, **Honys, D.**, **Twell, D.**, **Sprunck, S.**, **Dresselhaus, T.**, … **Mutwil, M.** (2021). Comparative transcriptomic analysis reveals conserved programmes underpinning organogenesis and reproduction in land plants. *Nature Plants*, **7**(8), 1143–1159. https://doi.org/10.1038/s41477-021-00958-2

Kappel, C., **Friedrich, T.**, **Oberkofler, V.**, **Jiang, L.**, **Crawford, T.**, **Lenhard, M.**, & **Bäurle, I.** (2023). Genomic and epigenomic determinants of heat stress-induced transcriptional memory in Arabidopsis. *Genome Biology*, **24**(1), 129. https://doi.org/10.1186/s13059-023-02970-5

Khamis, A. M., **Motwalli, O.**, **Oliva, R.**, **Jankovic, B. R.**, **Medvedeva, Y. A.**, **Ashoor, H.**, **Essack, M.**, **Gao, X.**, & **Bajic, V. B.** (2018). A novel method for improved accuracy of transcription factor binding site prediction. *Nucleic Acids Research*, **46**(12), e72. https://doi.org/10.1093/nar/gky237

Kidokoro, S., **Konoura, I.**, **Soma, F.**, **Suzuki, T.**, **Miyakawa, T.**, **Tanokura, M.**, **Shinozaki, K.**, & **Yamaguchi-Shinozaki, K.** (2023). Clock-regulated coactivators selectively control gene expression in response to different temperature stress conditions in Arabidopsis. *Proceedings of the National Academy of Sciences*, **120**(16), e2216183120. https://doi.org/10.1073/pnas.2216183120

Klopfenstein, D. V., **Zhang, L.**, **Pedersen, B. S.**, **Ramírez, F.**, **Warwick Vesztrocy, A.**, **Naldi, A.**, **Mungall, C. J.**, **Yunes, J. M.**, **Botvinnik, O.**, **Weigel, M.**, **Dampier, W.**, **Dessimoz, C.**, **Flick, P.**, & **Tang, H.** (2018). GOATOOLS: A python library for gene ontology analyses. *Scientific Reports*, **8**(1), 10872. https://doi.org/10.1038/s41598-018-28948-z

Kulkarni, S. R., **Vaneechoutte, D.**, **Van de Velde, J.**, & **Vandepoele, K.** (2018). TF2Network: Predicting transcription factor regulators and gene regulatory networks in Arabidopsis using publicly available binding site information. *Nucleic Acids Research*, **46** (6), e31. https://doi.org/10.1093/nar/gkx1279

Langfelder, P., & **Horvath, S.** (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**(1), 559. https://doi.org/10.1186/1471-2105-9-559

Langfelder, P., **Zhang, B.**, & **Horvath, S.** (2008). Defining clusters from a hierarchical cluster tree: The dynamic tree cut package for R. *Bioinformatics*, **24**(5), 719–720. https://doi.org/10.1093/bioinformatics/btm563

Langfelder, P., **Luo, R.**, **Oldham, M. C.**, & **Horvath, S.** (2011). Is my network module preserved and reproducible? *PLoS Computational Biology*, **7**(1), e1001057. https://doi.org/10.1371/journal.pcbi.1001057

Lu, T., **Liang, H.**, **Li, H.**, & **Wu, H.** (2011). High dimensional ODEs coupled with mixed-effects modeling techniques for dynamic gene regulatory network identification. *Journal of the American Statistical Association*, **106**(496), 1242–1258. https://doi.org/10.1198/jasa.2011.ap10194

Lu, Y., **Gehan, J. P.**, & **Sharkey, T. D.** (2005). Daylength and circadian effects on starch degradation and maltose metabolism. *Plant Physiology*, **138**(4), 2280–2291. https://doi.org/10.1104/pp.105.061903

MacNeill, G. J., **Mehrpouyan, S.**, **Minow, M. A. A.**, **Patterson, J. A.**, **Tetlow, I. J.**, & **Emes, M. J.** (2017). Starch as a source, starch as a sink: The bifunctional role of starch in carbon allocation. *Journal of Experimental Botany*, **68**(16), 4433–4453. https://doi.org/10.1093/jxb/erx291

Mao, W., **Zaslavsky, E.**, **Hartmann, B. M.**, **Sealfon, S. C.**, & **Chikina, M.** (2019). Pathway-level information extractor (PLIER) for gene expression data. *Nature Methods*, **16**(7), 607–610. https://doi.org/10.1038/s41592-019-0456-1

McDowell, I. C., **Manandhar, D.**, **Vockley, C. M.**, **Schmid, A. K.**, **Reddy, T. E.**, & **Engelhardt, B. E.** (2018). Clustering gene expression time series data using an infinite Gaussian process mixture model. *PLoS Computational Biology*, **14**(1), e1005896. https://doi.org/10.1371/journal.pcbi.1005896

Moradzadeh, K., **Moein, S.**, **Nickaeen, N.**, & **Gheisari, Y.** (2019). Analysis of time-course microarray data: Comparison of common tools. *Genomics*, **111**(4), 636–641. https://doi.org/10.1016/j.ygeno.2018.03.021

Mukarram, M., **Choudhary, S.**, **Kurjak, D.**, **Petek, A.**, & **Khan, M. M. A.** (2021). Drought: Sensing, signalling, effects and tolerance in higher plants. *Physiologia Plantarum*, **172**(2), 1291–1300. https://doi.org/10.1111/ppl.13423

Mulet, J. M., **Campos, F.**, & **Yenush, L.** (2020). Editorial: Ion homeostasis in plant stress and development. *Frontiers in Plant Science*, **11**, 618273. https://doi.org/10.3389/fpls.2020.618273

Mulet, J. M., **Porcel, R.**, & **Yenush, L.** (2023). Modulation of potassium transport to increase abiotic stress tolerance in plants. *Journal of Experimental Botany*, **74**(19), 5989–6005. https://doi.org/10.1093/jxb/erad333

Nieto, C., **Catalán, P.**, **Luengo, L. M.**, **Legris, M.**, **López-Salmerón, V.**, **Davière, J. M.**, **Casal, J. J.**, **Ares, S.**, & **Prat, S.** (2022). COP1 dynamics integrate conflicting seasonal light and thermal cues in the control of Arabidopsis elongation. *Science Advances*, **8**(33), eabp8412. https://doi.org/10.1126/sciadv.abp8412

Noordijk, B., **Garcia Gomez, M. L.**, **ten Tusscher, K. H. W. J.**, **de Ridder, D.**, **van Dijk, A. D. J.**, & **Smith, R. W.** (2024). The rise of scientific machine learning: A perspective on combining mechanistic modelling with machine learning for systems biology. *Frontiers in Systems Biology*, **4**, 1407994. https://doi.org/10.3389/fsysb.2024.1407994

Obayashi, T., **Hibara, H.**, **Kagaya, Y.**, **Aoki, Y.**, & **Kinoshita, K.** (2022). ATTED-II v11: A plant gene coexpression database using a sample balancing technique by subagging of principal components. *Plant & Cell Physiology*, **63**(6), 869–881. https://doi.org/10.1093/pcp/pcac041

Okuma, E., **Nozawa, R.**, **Murata, Y.**, & **Miura, K.** (2014). Accumulation of endogenous salicylic acid confers drought tolerance to Arabidopsis. *Plant Signaling & Behavior*, **9**(3), e28085. https://doi.org/10.4161/psb.28085

Ortega, M. A., **Celoy, R. M.**, **Chacon, F.**, **Yuan, Y.**, **Xue, L.-J.**, **Pandey, S. P.**, **Drowns, M. R.**, **Kvitko, B. H.**, & **Tsai, C.-J.** (2024). Altering cold-regulated gene expression decouples the salicylic acid–growth trade-off in Arabidopsis. *The Plant Cell*, **36**(10), 4293–4308. https://doi.org/10.1093/plcell/koae210

Pan, Y., **Yu, B.**, **Wei, X.**, **Qiu, Y.**, **Mao, X.**, **Liu, Y.**, **Yan, W.**, **Linghu, Q.**, **Li, W.**, **Guo, H.**, & **Tang, Z.** (2024). Suppression of SMXL4 and SMXL5 confers enhanced thermotolerance through promoting HSFA2 transcription in Arabidopsis. *The Plant Cell*, **36**(10), 4557–4575. https://doi.org/10.1093/plcell/koae224

Pedregosa, F., **Varoquaux, G.**, **Gramfort, A.**, **Michel, V.**, **Thirion, B.**, **Grisel, O.**, **Blondel, M.**, **Prettenhofer, P.**, **Weiss, R.**, **Dubourg, V.**, **Vanderplas,**

J., **Passos, A.**, **Cournapeau, D.**, **Brucher, M.**, **Perrot, M.**, & **Duchesnay, É.** (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, **12**(85), 2825–2830.

Polynikis, A., **Hogan, S. J.**, & **di Bernardo, M.** (2009). Comparing different ODE modelling approaches for gene regulatory networks. *Journal of Theoretical Biology*, **261**(4), 511–530. https://doi.org/10.1016/j.jtbi.2009.07.040

Quint, M., **Delker, C.**, **Franklin, K. A.**, **Wigge, P. A.**, **Halliday, K. J.**, & **van Zanten, M.** (2016). Molecular and genetic control of plant thermomorphogenesis. *Nature Plants*, **2**(1), 1–9. https://doi.org/10.1038/nplants.2015.190

Ramegowda, V., **Gill, U. S.**, **Sivalingam, P. N.**, **Gupta, A.**, **Gupta, C.**, **Govind, G.**, **Nataraja, K. N.**, **Pereira, A.**, **Udayakumar, M.**, **Mysore, K. S.**, & **Senthil-Kumar, M.** (2017). GBF3 transcription factor imparts drought tolerance in Arabidopsis thaliana. *Scientific Reports*, **7**(1), 9148. https://doi.org/10.1038/s41598-017-09542-1

Ritchie, M. E., **Phipson, B.**, **Wu, D.**, **Hu, Y.**, **Law, C. W.**, **Shi, W.**, & **Smyth, G. K.** (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, **43**(7), e47. https://doi.org/10.1093/nar/gkv007

Rizhsky, L., **Liang, H.**, & **Mittler, R.** (2002). The combined effect of drought stress and heat shock on gene expression in tobacco. *Plant Physiology*, **130**(3), 1143–1151. https://doi.org/10.1104/pp.006858

Saelens, W., **Cannoodt, R.**, & **Saeys, Y.** (2018). A comprehensive evaluation of module detection methods for gene expression data. *Nature Communications*, **9**(1), 1090. https://doi.org/10.1038/s41467-018-03424-4

Sanchez-Munoz, R., **Depaepe, T.**, **Samalova, M.**, **Hejatko, J.**, **Zaplana, I.**, & **Van Der Straeten, D.** (2025). Machine-learning metaanalysis reveals ethylene as a central component of the molecular core in abiotic stress responses in Arabidopsis. *Nature Communications*, **16**(1), 4778. https://doi.org/10.1038/s41467-025-59542-3

Sapoval, N., **Aghazadeh, A.**, **Nute, M. G.**, **Antunes, D. A.**, **Balaji, A.**, **Baraniuk, R.**, **Barberan, C. J.**, **Dannenfelser, R.**, **Dun, C.**, **Edrisi, M.**, **Elworth, R. A. L.**, **Kille, B.**, **Kyrillidis, A.**, **Nakhleh, L.**, **Wolfe, C. R.**, **Yan, Z.**, **Yao, V.**, & **Treangen, T. J.** (2022). Current progress and open challenges for applying deep learning across the biosciences. *Nature Communications*, **13**(1), 1728. https://doi.org/10.1038/s41467-022-29268-7

Schälte, Y., **Fröhlich, F.**, **Jost, P. J.**, **Vanhoefer, J.**, **Pathirana, D.**, **Stapor, P.**, **Lakrisenko, P.**, **Wang, D.**, **Raimúndez, E.**, **Merkt, S.**, **Schmiester, L.**, **Städter, P.**, **Grein, S.**, **Dudkin, E.**, **Doresic, D.**, **Weindl, D.**, & **Hasenauer, J.** (2023). pyPESTO: A modular and scalable tool for parameter estimation for dynamic models. *Bioinformatics*, **39**(11), btad711. https://doi.org/10.1093/bioinformatics/btad711

Schmiester, L., **Schälte, Y.**, **Bergmann, F. T.**, **Camba, T.**, **Dudkin, E.**, **Egert, J.**, **Fröhlich, F.**, **Fuhrmann, L.**, **Hauber, A. L.**, **Kemmer, S.**, **Lakrisenko, P.**, **Loos, C.**, **Merkt, S.**, **Müller, W.**, **Pathirana, D.**, **Raimúndez, E.**, **Refisch, L.**, **Rosenblatt, M.**, **Stapor, P. L.**, … **Weindl, D.** (2021). PEtab—Interoperable specification of parameter estimation problems in systems biology. *PLoS Computational Biology*, **17**(1), e1008646. https://doi.org/10.1371/journal.pcbi.1008646

Schulz, M. H., **Devanny, W. E.**, **Gitter, A.**, **Zhong, S.**, **Ernst, J.**, & **Bar-Joseph, Z.** (2012). DREM 2.0: Improved reconstruction of dynamic regulatory networks from time-series expression data. *BMC Systems Biology*, **6**(1), 104. https://doi.org/10.1186/1752-0509-6-104.

Segal, E., **Shapira, M.**, **Regev, A.**, **Pe'er, D.**, **Botstein, D.**, **Koller, D.**, & **Friedman, N.** (2003). Module networks: Identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature Genetics*, **34**(2), 166–176. https://doi.org/10.1038/ng1165

Shaik, R., & **Ramakrishna, W.** (2013). Genes and co-expression modules common to drought and bacterial stress responses in arabidopsis and rice. *PLoS One*, **8**(10), e77261. https://doi.org/10.1371/journal.pone.0077261

Shannon, P., **Markiel, A.**, **Ozier, O.**, **Baliga, N. S.**, **Wang, J. T.**, **Ramage, D.**, **Amin, N.**, **Schwikowski, B.**, & **Ideker, T.** (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, **13**(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Shen, J., **Liu, F.**, **Tu, Y.**, & **Tang, C.** (2021). Finding gene network topologies for given biological function with recurrent neural network. *Nature Communications*, **12**(1), 3125. https://doi.org/10.1038/s41467-021-23420-5

Sîrbu, A., **Kerr, G.**, **Crane, M.**, & **Ruskin, H. J.** (2012). RNA-seq vs dual- and single-channel microarray data: Sensitivity analysis for differential expression and clustering. *PLoS One*, **7**(12), e50986. https://doi.org/10.1371/journal.pone.0050986

Skryhan, K., **Gurrieri, L.**, **Sparla, F.**, **Trost, P.**, & **Blennow, A.** (2018). Redox regulation of starch metabolism. *Frontiers in Plant Science*, **9**, 01344. https://doi.org/10.3389/fpls.2018.01344

Soltanalizadeh, B., **Gonzalez Rodriguez, E.**, **Maroufy, V.**, **Zheng, W. J.**, & **Wu, H.** (2020). Modelling of hypoxia gene expression for three different cancer cell lines. *International Journal of Computational Biology and Drug Design*, **13**(1), 124–143. https://doi.org/10.1504/ijcbdd.2020.10026794

Song, C., **Kim, T.**, **Chung, W. S.**, & **Lim, C. O.** (2017). The *Arabidopsis* phytocystatin AtCYS5 enhances seed germination and seedling growth under heat stress conditions. *Molecules and Cells*, **40**(8), 577–586. https://doi.org/10.14348/molcells.2017.0075

Sreedasyam, A., **Plott, C.**, **Hossain, M. S.**, **Lovell, J. T.**, **Grimwood, J.**, **Jenkins, J. W.**, **Daum, C.**, **Barry, K.**, **Carlson, J.**, **Shu, S.**, **Phillips, J.**, **Amirebrahimi, M.**, **Zane, M.**, **Wang, M.**, **Goodstein, D.**, **Haas, F. B.**, **Hiss, M.**, **Perroud, P.-F.**, **Jawdy, S. S.**, … **Schmutz, J.** (2023). JGI Plant Gene Atlas: An updateable transcriptome resource to improve functional gene descriptions across the plant kingdom. *Nucleic Acids Research*, **51**(16), 8383–8401. https://doi.org/10.1093/nar/gkad616

Staut, J., **Manosalva Pérez, N. M.**, **Matres Ferrando, A.**, **Dissanayake, I.**, & **Vandepoele, K.** (2025). A map of integrated cis-regulatory elements enhances gene regulatory analysis in maize. *Plant Communications*, **6**(7), 101376. https://doi.org/10.1016/j.xplc.2025.101376

Sugio, A., **Dreos, R.**, **Aparicio, F.**, & **Maule, A. J.** (2009). The cytosolic protein response as a subcomponent of the wider heat shock response in arabidopsis. *The Plant Cell*, **21**(2), 642–654. https://doi.org/10.1105/tpc.108.062596

Suter, L., & **Widmer, A.** (2013). Phenotypic effects of salt and heat stress over three generations in Arabidopsis thaliana. *PLoS One*, **8**(11), e80819. https://doi.org/10.1371/journal.pone.0080819

Valentim, F. L., **van Mourik, S.**, **Posé, D.**, **Kim, M. C.**, **Schmid, M.**, **van Ham, R. C. H. J.**, **Busscher, M.**, **Sanchez-Perez, G. F.**, **Molenaar, J.**, **Angenent, G. C.**, **Immink, R. G. H.**, & **van Dijk, A. D. J.** (2015). A quantitative and dynamic model of the arabidopsis flowering time gene regulatory network. *PLoS One*, **10**(2), e0116973. https://doi.org/10.1371/journal.pone.0116973

Van De Mortel, J. E., **Schat, H.**, **Moerland, P. D.**, **Van Themaat, E. V. L.**, **Van Der Ent, S.**, **Blankestijn, H.**, **Ghandilyan, A.**, **Tsiatsiani, S.**, & **Aarts, M. G.** (2008). Expression differences for genes involved in lignin, glutathione and sulphate metabolism in response to cadmium in Arabidopsis thaliana and the related Zn/Cd-hyperaccumulator Thlaspi caerulescens. *Plant, Cell & Environment*, **31**(3), 301–324. https://doi.org/10.1111/j.1365-3040.2007.01764.x

Verslues, P. E., **Bailey-Serres, J.**, **Brodersen, C.**, **Buckley, T. N.**, **Conti, L.**, **Christmann, A.**, **Dinneny, J. R.**, **Grill, E.**, **Hayes, S.**, **Heckman, R. W.**, **Hsu, P.-K.**, **Juenger, T. E.**, **Mas, P.**, **Munnik, T.**, **Nelissen, H.**, **Sack, L.**, **Schroeder, J. I.**, **Testerink, C.**, **Tyerman, S. D.**, … **Wigge, P. A.** (2023). Burning questions for a warming and changing world: 15 unknowns in plant abiotic stress. *The Plant Cell*, **35**(1), 67–108. https://doi.org/10.1093/plcell/koac263.

Virtanen, P., **Gommers, R.**, **Oliphant, T. E.**, **Haberland, M.**, **Reddy, T.**, **Cournapeau, D.**, **Burovski, E.**, **Peterson, P.**, **Weckesser, W.**, **Bright, J.**, **van der Walt, S. J.**, **Brett, M.**, **Wilson, J.**, **Millman, K. J.**, **Mayorov, N.**, **Nelson, A. R. J.**, **Jones, E.**, **Kern, R.**, **Larson, E.**, … **van Mulbregt, P.** (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, **17**(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wang, J. Z., **Du, Z.**, **Payattakool, R.**, **Yu, P. S.**, & **Chen, C.-F.** (2007). A new method to measure the semantic similarity of GO terms. *Bioinformatics*, **23**(10), 1274–1281. https://doi.org/10.1093/bioinformatics/btm087

Wang, X., **Dalkic, E.**, **Wu, M.**, & **Chan, C.** (2008). Gene-module level analysis: Identification to networks and dynamics. *Current Opinion*

in Biotechnology, **19**(5), 482–491. https://doi.org/10.1016/j.copbio.2008.07.011

Wang, X., **Zhu, Y.**, **Tang, L.**, **Wang, Y.**, **Sun, R.**, & **Deng, X.** (2024). Arabidopsis HSFA9 acts as a regulator of heat response gene expression and the acquisition of thermotolerance and seed longevity. *Plant and Cell Physiology*, **65**(3), 372–389. https://doi.org/10.1093/pcp/pcad164

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, **6**(60), 3021. https://doi.org/10.21105/joss.03021

Wu, S., **Liu, Z.-P.**, **Qiu, X.**, & **Wu, H.** (2014). Modeling genome-wide dynamic regulatory network in mouse lungs with influenza infection using high-dimensional ordinary differential equations. *PLoS One*, **9**(5), e95276. https://doi.org/10.1371/journal.pone.0095276