ANALYSIS OF THE DYNAMIC NITROGEN RESPONSE OF ROOT AND SHOOT TRANSCRIPTS IN *ARABIDOPSIS THALIANA* TO GAIN INSIGHTS ON LONG-DISTANCE NITROGEN SIGNALING INTERACTIONS

BY

SACHIN HEERAH

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Plant Biology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2020

Urbana, Illinois

Master's Committee:

Assistant Professor Amy Marshall-Colón, Chair
Adjunct Professor Elizabeth Ainsworth
Professor Donald Briskin
Assistant Professor Stéphane Guerrier, University of Geneva

ABSTRACT

Genetic advances account for less than 2% of global annual crop yield growth, with nitrogen (N) fertilizer use comprising the remaining 98% and its use expected to increase to 138Tg to meet crop demand by 2030. This overapplication of N fertilizer results in direct economic and environmental consequences as plants only take up 30-50% of the available soil N, with the rest being lost to the environment. To combat this, many studies have focused on improving the N use efficiency of plants through understanding the mechanisms and pathways involved in the N-responsive long-distance signaling pathways between roots and shoots. These studies, however, often fall short of integrating data across time and space due to various biological constraints, while others attempt to use time series models not designed for biological systems. Here, I propose a new time series model that is suitable for biological systems, accounting for these constraints. This model was applied to unevenly spaced, multivariate time-series data from root and shoot tissue in *Arabidopsis thaliana* in response to a N signal. From 2,173 shoot and 568 root differentially expressed genes, the model predicted 3,078 significant granger-causal interactions. Of these, 2,012 interactions have a root causal gene while 1,066 interactions have a shoot casual gene. Of the total 1,007 different causal genes from either organ, 384 have been known or predicted to produce a mobile gene product, possibly involved in N signaling. The interactions were then globally explored using a bioinformatics pipeline that included gene ontology term analysis, network analysis, transcription factor binding, as well as exploring causal genes involved in known N-responsive signaling pathways and interactions. Further, an *A. thaliana* grafting method is put forward to validate selected bioinformatically-supported predictions. Future directions are then discussed with respect to using the time series model to integrate shoot metabolite data to root/shoot transcriptomic data to identify possible N-responsive gene-metabolite relationships.

ACKNOWLEDGMENTS

Firstly, I would like to thank my advisor, Dr. Amy Marshall-Colón, for this opportunity and I'd like to express my sincere gratitude for all her assistance and endless support throughout the completion of my master's thesis project. I'd like to especially thank Stuti Srivastava and Dr. Kavya Kannan for their support throughout this entire process from our navigating our first days of graduate school, countless days and nights setting up and performing experiments, to the hours spent earnestly supporting each other. I'd like to thank Dilkaran Singh, Jie Fu, Dr. Ghana Challa, Dr. Madara Hetti-Arachchilage, and Dr. Megan Matthews for their support and feedback, as well as all the undergraduate students of the Marshall-Colón lab for their assistance with performing lab experiments. I'd like to thank Dr. Stéphane Guerrier, Dr. Roberto Molinari and James Balamuta for their work in developing the time series model. I'd also like to thank Jennifer Quebedeaux and Jiayang Xie of the Leakey Lab as well as Chris Montes and the Ainsworth Lab for the use of their lab space, equipment and facilities. I wish to thank my committee members: Dr. Lisa Ainsworth and Dr. Donald Briskin, as well as Dr. Steven Huber for their insights and suggestions over the years. Finally, I must thank my parents, Parasram Heerah and Sharda Heerah, my brother, Pavesh Heerah, as well as my friends near and far for their love and encouragement throughout this journey. I could not have made it this far without all of you.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Nitrogen (N) is one of the necessary macronutrients for plants, and is required in the greatest amount (Marschner, 2011). Ironically, while N comprises 78% of the atmosphere, non-leguminous plants are unable to assimilate it from there as they lack a system to break the triple bond in the atmospheric N molecule (Ranjan and Yadav, 2019). As such, plants take up N from soil through reactive species such as ammonium and nitrate. During the Green Revolution, it was shown that crop yields increase with increasing amounts of N fertilizers (Sinclair and Ruffly, 2012). This crop yield response to N is in part due to improved genetics through engineering and breeding efforts (Kant et al., 2011) However, the global annual increase in yield from genetics has slowed to below 1% in wheat and rice, 1.6% in maize, while absolute yield is falling in developing countries (Fischer et al, 2009). Even considering this, global crop yield growth is projected to keep pace with an increase in N fertilizer use from 101Tg of N in 2010 to a projected 138Tg of N in 2030 (Heffer and Prud'homme 2016). This increase in N use is quite drastic given that it was previously believed before the turn of the century that N fertilizer use would not approach 130-150 million tons until 2050 (Matson, Naylor, and Ortiz-Monasterio 1998). Unfortunately, plants only take up 30-50% of available soil N (McAllister et al., 2012), with the remaining added N being lost to the environment through leaching, greenhouse gas emission, and denitrification (Good and Beatty, 2011). Given the overapplication of N fertilizers to keep pace with an increase in demand for crop yield, there is a direct economic and environmental need to increase the nitrogen use efficiency (NUE) of crops.

NUE can be broadly defined as the "grain production per unit of N available in the soil" (Moll et al., 1982). It is comprised of the two components: N uptake efficiency (NUpE) and N utilization efficiency (NUtE) (Han et al., 2015). NUE is a genetically complex phenotype and has been the target of numerous approaches to improve this trait through specific experiments that target the uptake, transport, assimilation and/or remobilization of N. However, only incremental gains have been realized (Han et al., 2016; Stahl et al., 2017; Selvaraj et al., 2017) as these improvements have only been limited in focus to investigating just one part of the NUE equation. It is now recognized that to improve NUE, a more holistic approach is needed. One such approach is to focus on genes and regulators that influence targets across N uptake, transport, assimilation and remobilization (Kant, 2018). Previous studies have attempted to do this by focusing on long-distance N-signaling between roots and shoots (Takei et al., 2004; Osugi et al., 2017), combined engineering of NUpE and NUtE (Snyman et al., 2015; DoVale et al., 2012), or simultaneous analysis of the N-responsive transcriptome, proteome and metabolome (Simons et al., 2014; Xin et al., 2019).

Communication is necessary in all biological organisms to report and coordinate systems status and response. Given that plants are sessile organisms, communication is especially necessary in order to coordinate responses between root and shoots. This communication allows plants to coordinate a response to the changing environment such as time of day, temperature, pathogens, pests and soil nutrient availability. The signals involved can travel from small distances of a few micrometers between two cells to dozens of meters from the roots to the leaves of the tallest redwood trees. Much research in plants has been performed to identify and understand these signals and their pathways. This has resulted in the identification of RNAs (Marin-Gonzalez and

Suarez-Lopez, 2012), peptides (Takahashi et al., 2019), lipids (Barbaglia and Hoffman-Benning, 2016), hormones (Blazquez et al., 2020), nutrient derived metabolites (e.g. amino acids, Dinkeloo et al., 2018), and even nutrients and ions themselves (Liu et al., 2009, Choi et al., 2017), that act as signals produced in response to abiotic and biotic perturbations. Other studies have focused on how electrical (Choi et al., 2017), hydraulic (Zwieniecki et al., 2004) and structural (Thompson, 2006) mechanisms are designed to promote and propagate these signals.

In particular, N signaling is well studied in order to shed light on how plants coordinate NUpE and NUtE. This has enabled researchers to identify targets for the improvement of NUE. Nitrate is the predominant form of N present in soils and one of the two primary forms of N taken up by plants (Kant et al., 2011). Starting from its uptake from the soil into roots, before its assimilation, nitrate can act as a long-distance signal in plants. Nitrate can be loaded into and unloaded from the xylem by nitrate transporters in the NPF gene family (Kant, 2018). Nitrate accumulation in shoots can then trigger an auxin signal that travels to the roots to repress root branching (Forde, 2002). This root-to-shoot-to-root communication that coordinates NUpE and NUtE is one of many that has been researched and observed in response to biotic and abiotic signal (Walch-Liu et al., 2005, Wang and Ruan, 2015, Chen et al., 2016). This feedback can involve a diverse array of known N signaling molecules and pathways including C-terminally encoded peptides (CEP) (Notaguchi and Okamoto, 2015) and CEP receptors (Ohkubo et al., 2017), transcription factors such as NLP7 and HY5 (Castaings et al., 2009, Chen et al., 2016), hormones including cytokinin (Ruffel et al., 2011, Kiba et al., 2011), and even RNAs (Ham and Lucas, 2017). The presence of these diverse signals involved in N signaling suggests that whole plant N signaling may be more complex than expected and most likely could involve signaling molecules and

pathways yet to be discovered (Ruffel and Gojon, 2017). These findings present new questions such as: How does the whole plant coordinate these different signaling pathways? How does the plant regulate these pathways to integrate NUpE and NUtE? And can we predict emergent properties of long-distance N-signaling through identification of causal, molecular relationships between organs?

These aforementioned studies on N signaling have been focused on NUpE and/or NUtE but do not consider integrating the data across time and space. The causal and target cells and tissues in signaling are physically separated, especially so in long distance signaling. Consequently, it takes time for a signal to propagate across this physical separation. Some studies do sample across time points in different parts of the plant and provide thorough descriptions of N signaling and response in either organ (Varala et al., 2018). While these studies do capture the transient effect of N on either NUpE or NUtE in specific organs, they fall just short of integrating the data across both time and space.

Time series data is important for elucidating the order in which biological processes and responses are activated, and is used to infer causality (Bar-Joseph et al., 2012). In biology, time series experiments are usually designed to capture the initial dynamic phase and later steady phase in a response to a perturbation. This results in unevenly spaced time points with dense early sampling and sparse late sampling (Colón et al., 2010; Krouk et al., 2010; Spellman et al., 1998; Zhu et al., 2000; Gargouri et al., 2015). Furthermore, sampling in biology is constrained to other external factors such as cost and labor. Attempts to capture transient responses to perturbations given these constraints often result in short, unevenly spaced time series data. Unfortunately, no statistical model exists to account for these unique features in biological time series data. Existing

time series models have been designed to be used for weather and financial data which are not only sampled at regular intervals, but also on the order of hundreds, if not thousands. Given these restrictions, biological studies have attempted to use other methods to analyze time series data. K-means clustering has been used to identify differentially expressed genes over developmental time in *Zea mays* (Chen et al., 2014), or those responding to drought stress in *Arabidopsis thaliana* (Bechtold et al, 2016). While clustering does identify when groups of genes are activated/repressed, it fails in predicting any causal relationships between specific gene-gene pairs. Network analysis using Bayesian networks and dynamic Bayesian networks have been used to predict relationships between genes but fall short of establishing causality. Other studies transform irregularly spaced time series data in equally spaced time series (Hamilton, 1994) such as approximating a continuous time series from the irregular data (Maller et al., 2008), or resampling to estimate missing data points at regular intervals between the observed time points (Remondini et al., 2005; Erdogan et al., 2005; Broersen & Bos, 2006; Thiebaut & Roques, 2005). Unfortunately, transforming the time series data has drawbacks that include the possibility of changing the causal relationships in multivariate time series, or data loss if time points are too close together or data dilution if time points are far apart thus biasing any estimate of statistical significance (Eckner 2014). Meanwhile, other studies have analyzed time series data but neglected to detrend the series (e.g Zhao et al., 2006; Gargouri et al., 2015). The trend of a time series is defined as an intrinsic property of the data that is driven by the same mechanisms that generate the data (Wu et al., 2007), e.g. the seasonal effect of temperature throughout the year. By detrending a time series, the remaining observations can behave independently with less correlation and little

information about the long-term trend (You et al., 2018). That is, a steady-baseline is produced around which a time series can oscillate, resulting in better predictions (Leise, 2017).

In this thesis, I present a new times series model that is suitable for time series analysis in biological systems and can establish granger-causal relationships from short, unevenly spaced time series data. This new model was applied to root and shoot transcriptome data to predict thousands of causal gene-gene relationships involved in long-distance N signaling, uncovering thousands of novel gene-gene relationships. Model predictions were then explored using a bioinformatics pipeline resulting in a number of candidate relationships for experimental validation (Chapter 2). I then report preliminary attempts to experimentally validate model-predicted causal relationships using mutant plant lines and grafting, and follow with a discussion about future directions (Chapter 3). Throughout this thesis, statistical terms are introduced that may be beyond the scope of a typical plant molecular biologist. These terms are underlined within the main text and defined in an included glossary at the end (Appendix B).

CHAPTER 2: GRANGER-CAUSAL TESTING FOR IRREGULARLY SAMPLED TIME SERIES WITH APPLICATION TO NITROGEN SIGNALING IN ARABIDOPSIS[1]

ABSTRACT

Motivation: Identification of system-wide causal relationships can contribute to our understanding of long-distance, intercellular signaling in biological organisms. Dynamic transcriptome analysis holds great potential to uncover coordinated biological processes between organs. However, many existing dynamic transcriptome studies are characterized by sparse and often unevenly spaced time points that make the identification of causal relationships across organs analytically challenging. Application of existing statistical models, designed for regular time series with abundant time points, to sparse data may fail to reveal biologically significant, causal relationships. With increasing research interest in biological time series data, there is a need for new statistical methods that are able to determine causality within and between time series data sets. Here, a statistical framework was developed to identify (Granger) causal gene-gene relationships from unevenly spaced, multivariate time series data from two different tissues of *Arabidopsis thaliana* in response to a nitrogen signal.

Results: This work delivers a statistical approach for modelling irregularly sampled bivariate signals. The approach embeds engineering functions that allow fitting of the model's dependence structure to the specific sampling time. Using Maximum-Likelihood to estimate the parameters of

[1] This chapter has been submitted for review in its entirety to the journal of 'Bioinformatics', and is available as a preprint on BioRxiv. It is referred to later in the dissertation as "Heerah, S., Molinari, R., Guerrier, S., and Marshall-Colon, A. 2020. "Granger-Causal Testing for Irregularly Sampled Time Series with Application to Nitrogen Signaling in Arabidopsis." *BioRxiv*, January, 2020.06.15. https://doi.org/10.1101/2020.06.15.152819." Some minor changes in the content has been made, and some figures and tables have been modified in comparison to the original manuscript.

this model for each bivariate time series, it is then possible to use bootstrap procedures for small samples (or asymptotics for large samples) in order to test for Granger-Causality. When applied to *Arabidopsis thaliana* data, the proposed approach produced 3,078 significant interactions, in which 2,012 interactions were from root causal genes and 1,066 interactions were from shoot causal genes. Many of the predicted causal and target genes are known players in local and long-distance nitrogen signaling, including genes encoding transcription factors, hormones, and signaling peptides. Of the 1,007 total causal genes (either organ), 384 are either known or predicted mobile transcripts, suggesting that the identified causal genes may be directly involved in long-distance nitrogen signaling through intercellular interactions. The model predictions and subsequent network analysis identified nitrogen-responsive genes that can be further tested for their specific roles in long-distance nitrogen signaling.

Availability: The method was developed with the R statistical software and is made available through the R package "irg" hosted on the GitHub repository (https://github.com/SMAC-Group/irg). A sample data set is made available as an example to apply the method and the complete *Arabidopsis thaliana* data can be found at: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE97500.

INTRODUCTION

Time series data are important for understanding the biological processes that are activated at different times and for inferring causality (Bar-Joseph et al., 2012). Many time series studies are designed to capture both dynamic and stationary phases in response to perturbations, which result in unevenly spaced time points, with dense sampling early and sparse sampling at later time points

(Spellman et al., 1998; Colón et al., 2010; Krouk et al., 2010; Gargouri et al., 2015; Zhu et al., 2000). In biology, this is a commonly used sampling scheme to efficiently capture transient transcriptional and metabolic responses. However, the analysis of this irregular data is challenging, since traditional time-lagged or cross-correlation analyses, designed for regularly spaced intervals, cannot be used. To date, it can be argued that no statistical approach has been able to comprehensively account for these unique features common to many biological time series (see e.g. Rehfeld et al., 2011).

Among the current approaches, methods designed for time-independent or regularly-spaced processes have been used to analyze unevenly-spaced time series data. For example, "static-based" clustering methods like hierarchical clustering and K-means have been used to organize and identify genes differentially expressed over developmental time in *Zea mays* (Chen et al., 2014), or in response to drought stress in *Arabidopsis thaliana* (Bechtold et al., 2016). However, clustering methods are not suitable to predict causal relationships between genes. Hence, other employed approaches include, among others, the transformation of irregularly sampled data into evenly spaced time series (Hamilton, 1994), in which the irregularity of the time interval can be approximated by forced regular intervals (Maller et al., 2008), or (resampling) strategies that estimate missing data points to fill in lags between observations (Broersen and Bos, 2006; Thiebaut and Roques, 2005; Remondini et al., 2005). Other methods directly address the irregular nature of the processes but do not consider the multivariate dependence and, consequently, the causal relation between signals (see e.g. Erdogan et al., 2005; Eyheramendy et al., 2018). These approaches have different drawbacks (Eckner, 2014) including: i) an inability to capture the variable nature of multivariate dynamic transcriptome experiments; and ii) resampling strategies

often change the (Granger) causal relationship of the multivariate time series (Bahadori and Liu, 2012). All of these approximations can lead to incorrect correlations and predictions, and are unable to determine causal relationships within or between time series. Another commonly used approach in the analysis of (biological) time series is to perform a correlation analysis which however often does not account for non-stationary features of the data (Gargouri et al., 2015; Zhao et al., 2006). Indeed, the latter form of analysis can be highly misleading if, for example, the mean and/or variance of the series change over time which can often be the case for many experimental settings.

In response to the above limitations, this work puts forward a statistical approach that provides a general framework to determine Granger-Causality (Granger, 1969) for (short) irregularly sampled bivariate signals. We use this proposed approach to describe causal gene-gene relationships from above- (shoot) and belowground (root) organs of *Arabidopsis thaliana* in response to a nitrogen signal. While the proposed approach can be used to investigate possible relationships within the same organ, this would require significant additional computational time and resources. Even if we consider the computational constraints, investigating within tissue causality would not address the goal our work: to understand the long-distance signaling and response relay to uncover how the root and the shoot coordinate their nitrogen response. Moreover, within tissue causality was recently explored by (Varala et al., 2018). Through identification and bioinformatic exploration of the detected causal relationships between organs, we achieve a greater understanding of the underlying molecular and biochemical pathways involved in the long-distance nitrogen-signal response. This increase in understanding of nitrogen-responsive biochemical pathways in different plant tissues may help to predict emergent plant properties under

nitrogen sufficiency and deficiency. Further testing of model-predicted causal relationships may uncover new molecules, pathways, and processes involved in the root-to-shoot-to-root nitrogen-signal relays, providing biological insight into complex, whole-plant nitrogen response.

GRANGER-CAUSAL ANALYSIS FOR IRREGULAR DATA

An irregularly spaced time series is a sequence of observations that are observed in time in a strictly increasing manner but where the spacing of observation times is not necessarily constant. More formally, let

$$(t_i : i = 1, \dots, n) \in T_n,$$

denote a strictly increasing time sequence of length *n* where:

$$T_n = \{(t_1 < \cdots < t_n) : t_i \in R, 1 \leq i \leq n\}$$

In addition, let $(X_i : i = 1, \dots, n) \in R^n$ and $(Y_i : i = 1, \dots, n) \in R^n$ denote two sequences of real-valued random variables (for example transcript or metabolite levels) such that we can denote a bivariate irregularly spaced time series with *n* time points, as (t_i, X_i, Y_i : i = 1,...,n), where t_i denotes the time at which X_i and Y_i are to be observed. In the context of this paper, we focus on those random sequences that are observed at the same points in time (i.e. the sequences $(t_i : i = 1, \dots, n)$ correspond for both series). However, this condition can also be relaxed as a result of the research developed in this work.

As highlighted previously, the literature on irregularly spaced time series is not abundant and methods available to practitioners for estimation and inference in these cases are lacking as well. In this section we therefore put forward a pertinent statistical model that we will denote as $F = \{F_\theta : \theta \in \Theta \subset R^p\}$, with θ being the vector containing the parameters of this model. The latter model needs to deal with irregularly spaced bivariate time series and should allow testing for Granger causal links between the series themselves. In order to achieve this goal, we firstly define $\mu_i^{(x)}$ and $\mu_i^{(y)}$ as the expected values of X_{t_i} and Y_{t_i}, respectively. These quantities represent, in the case of dynamic transcriptome and metabolome data, the natural (deterministic) variation in gene expressions due, for example, to changes in environmental conditions or natural cycles. If we were considering evenly spaced observations, it would appear reasonable to consider the class of AutoRegressive Moving Average (ARMA) models to describe the variations of (X_{t_i}) around its mean (see e.g. Box et al., 2015, for details). A commonly used model within this class, especially when dealing with small sample sizes, is the first-order autoregressive model, i.e. an AR(1), which is defined as

$$X_{t_i} - \mu_i^{(x)} = \rho\left(X_{t_{i-1}} - \mu_{i-1}^{(x)}\right) + W_{t_i},$$

where ρ represents the parameter which explains the dependence between consecutive observations and W_{t_i} is an independent sequence of random variables with a certain (finite) variance σ^2. This model approximates many covariance structures delivering a behavior that is often reasonable for biological and natural phenomena. In order to determine whether another time series (signal) has an impact on the time series under consideration, the above model can be extended as follows:

$$X_{t_i} - \mu_i^{(x)} = \rho\left(X_{t_{i-1}} - \mu_{i-1}^{(x)}\right) + \lambda\left(Y_{t_{i-1}} - \mu_{i-1}^{(y)}\right) + W_{t_i},$$

where λ represents the impact that the time series (Y_{t_i}) has on the time series (X_{t_i}). In general terms, we can say that (Y_{t_i}) *Granger-causes* (X_{t_i}) if the latter model explains the behavior of (X_{t_i}) better than the previously defined AR(1) model that only depends on the sequence (X_{t_i}). The concept of Granger Causality was introduced in Granger (1969) and the goal of the biological study considered in this work would therefore be to perform a statistical test to confirm the stronger explanatory power of the second model over the first.

However, these models are not well-adapted to irregularly spaced time series that are the focus of this work. For example, the parameter ρ, that measures the relation between consecutive observations, remains constant regardless of the distance in time between $X_{t_i}$and $X_{t_{i-1}}$ (as well as the parameters λ and σ^2). For this reason, the next sections put forward a new framework for these settings.

The Proposed Model

The first step required to address the problem of modelling irregularly spaced time series consists in integrating the distance in time between observations within the model specification. Assuming an appropriate technique is used to estimate $\mu_i^{(x)}$ (e.g. splines or other semi- or non-parametric approaches), we denote the centered observations as $\tilde{X}_{t_i} \coloneqq X_{t_i} - \mu_i^{(x)}$ and the distance in time as

$\delta_{t_i} \coloneqq t_i - t_{i-1}$, with $\delta_{t_i} \in R^+$by definition. Based on this, the AR(1) model for irregularly spaced data can be represented as follows:

$$(1)\ \tilde{X}_{t_i} = f(\delta_{t_i})\tilde{X}_{t_{i-1}} + W_{t_i},$$

where $f(\cdot)$ is a deterministic function, possibly known up to some parameter values, that plays the same role as the constant ρ but takes into account the distance between observations. The independent sequence (W_{t_i}) is usually considered as being Gaussian (although other distributions can be considered) for multiple reasons, one of which is estimation feasibility. A Gaussian distribution only depends on mean and covariance parameters that are independent from each other, and can therefore be defined separately. Other distributions often jointly contribute to both the mean and the variance of the process and can therefore make the definition of the likelihood more complex, and/or add more parameters to be estimated which would be theoretically and numerically impossible for short signals. Without loss of generality, we will make this assumption and therefore state that $W_{t_i} \sim \mathcal{N}\left(0, g(\delta_{t_i})\right)$ with $g(\cdot)$ being another deterministic function. Both the functions $f(\cdot)$ and $g(\cdot)$ need to respect certain properties which will be discussed further on. The model defined in (1) could be extended in several ways, for example, by considering a dependence between $\tilde{X}_{t_i}$and $\tilde{X}_{t_{i-j}}$ with $j > 1$ or between $\tilde{X}_{t_i}$and $W_{t_{i-j}}$as in general ARMA models (as well as considering non-Gaussian distributions for W_{t_i} as mentioned earlier). However, given the small sample sizes usually encountered in dynamic transcriptome and metabolome studies, it is rather unlikely that more complex models can be appropriately estimated and the model in (1) is a very reasonable approximation for more general dependence structures.

Considering the extension of AR(1) processes to irregularly spaced settings, we can consider the same extension when modelling the joint behavior of two time series. For this purpose, we define the following bivariate model, which is a natural extension of a vector AR(1) model for irregularly spaced data:

$$(2)\ \boldsymbol{Z_i} = \boldsymbol{A}(\delta_{t_i})\boldsymbol{Z_{i-1}} + \boldsymbol{V_i}$$

Where

$$\boldsymbol{Z}_i := \begin{bmatrix} X_{t_i} - \mu_i^{(x)} \\ Y_{t_i} - \mu_i^{(y)} \end{bmatrix}, \quad \boldsymbol{A}\,(\delta_{t_i}) := \begin{bmatrix} f_1(\delta_{t_i}) & h_1(\delta_{t_i}) \\ h_2(\delta_{t_i}) & f_2(\delta_{t_i}) \end{bmatrix},$$

and where $h(\cdot)$ is another deterministic function (which may depend on unknown parameters). In addition, we have $V_i \sim := \sim [W_{t_i}, U_{t_i}]^{\intercal}$ with $(V_{t_i}, i = 1, \ldots, n) \in R^{\mathbb{2}\times n}$ denoting a bivariate independent sequence with distribution $V_{t_i} \sim \mathcal{N}(0, \Sigma_i)$, with **0** being a two-dimensional zero vector and

$$(3)\quad \boldsymbol{\Sigma}_i = \begin{bmatrix} g_1(\delta_{t_i}) & 0 \\ 0 & g_2(\delta_{t_i}) \end{bmatrix}$$

It can be observed how the matrix $\boldsymbol{A}(\delta_{t_i})$ plays the main role in describing the dependence "within" and "between" the two time-series. Indeed, on one hand the functions $f_1(\delta_{t_i})$ and $f_2(\delta_{t_i})$

determine to what extent the time series depend on themselves to describe the behavior of their future observations while the functions $h_1(\delta_{t_i})$ and $h_2(\delta_{t_i})$, on the other hand, determine the degree of dependence between the two time series. Also, within this setting it is possible to recognize the idea of Granger causality where one is interested in assessing whether past values of a certain time series can significantly increase the explanation of the behavior of another time series. In general, this assessment is based on statistical tests which are typically related to characteristics of the matrix $A(\delta_{t_i})$. In fact, if this matrix is diagonal, this implies that the two time-series are independent from each other (under the Gaussian assumption) while if it is full this entails that the two time-series are also inter-dependent. Moreover, if the matrix is upper or lower triangular, this would imply that only one of the series depends on itself *and* on the other series (the latter therefore only depending on itself).

Considering the above modelling framework, there is a need to estimate the unknown parameters in the model and test whether the estimated models appear to explain the data sufficiently well to draw reliable conclusions. Firstly, to estimate these kinds of models we propose a likelihood approach based on the assumption of a jointly normal distribution of the observations which, for the bivariate series, gives the following conditional distribution:

$$\boldsymbol{Z}_i|\boldsymbol{Z}_{i-1} \sim \mathcal{N}(\tilde{\mu}_i, \boldsymbol{\Sigma}_i), \tag{4}$$

where $\boldsymbol{\Sigma}_i$ is defined in (3), and

$$\tilde{\mu}_i := \begin{bmatrix} f_1(\delta_{t_i})\tilde{X}_{t_{i-1}} + h_1(\delta_{t_i})\tilde{Y}_{t_{i-1}} \\ h_2(\delta_{t_i})\tilde{X}_{t_{i-1}} + f_2(\delta_{t_i})\tilde{Y}_{t_{i-1}} \end{bmatrix}.$$

If we denote the unconditional distribution of Z_i as $l(Z_i)$, then the likelihood function is given by

$$(5)\quad L(\boldsymbol{\theta}) = l(Z_1)\prod_{i=2}^{n} l(\boldsymbol{Z_i}|\boldsymbol{Z_{i-1}}),$$

where, using (4), we have

$$l(Z_i|Z_{i-1}) = \frac{1}{2\pi|\Sigma_i|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(Z_i - \tilde{\mu}_i)^T\Sigma_i^{-1}(Z_i - \tilde{\mu}_i)\right).$$

Applying the log(·) function to $L(\theta)$ and fixing $l(Z_1)$ as constant (neglecting constant terms) we obtain the following estimating equation which defines the Maximum Likelihood Estimator (MLE):

$$(6)\quad \hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}}\ Q_n(\boldsymbol{\theta}),$$

where

$$Q_n(\theta) = \frac{1}{n-1}\sum_{i=2}^{n}\log(\Sigma_i) + (Z_i - \tilde{\mu}_i)^T\Sigma_i^{-1}(Z_i - \tilde{\mu}_i).$$

Under a set of conditions (see Model Conditions in the Appendix), the estimator defined in (6) has appropriate statistical properties. Among these conditions there are constraints on the deterministic functions that characterize the dependence structure of the model defined in (2). For this reason, we define these functions accordingly taking from the domain of (navigation) engineering (see e.g. Titterton et al., 2004). In the latter field, a model that is often used is the discrete-time first-order Gauss-Markov model that can be defined as:

$$\tilde{X}_{t_i} = \exp\left(-\frac{\delta_{t_i}}{\phi}\right)\tilde{X}_{i-1} + W_{t_i},$$

where $\varphi \in \mathrm{R}_+$ is a parameter that determines the "range" of dependence in the data and

$$W_{t_i} \sim \mathcal{N}\left(0,\ \sigma^2\left[1 - \exp\left(-\frac{2\delta_{t_i}}{\phi}\right)\right]\right).$$

Having been mainly proposed to deal with time series measured at different frequencies, the idea behind this model is very close to the structure of an exponential model for spatial data (see e.g. Ripley, 2005). Indeed, the latter explains the dependence in space through an exponential structure and roughly corresponds to the above-mentioned Gauss-Markov process when considering δ_{ti} as a measure of Euclidean distance. The above model therefore gives an explicit form to the functions $f_{\cdot}(\cdot)$ and $g_{\cdot}(\cdot)$ mentioned earlier but of course other explicit forms can be envisaged.

While the above defined functions characterize the dependence of a time series on itself, it is still necessary to give an adequate form to the function $h_{\cdot}(\cdot)$ that describes the behavior of a

signal based on another. Given the short time series available, we decided to impose a reasonable structure to this behavior which allowed the dependence of a signal on the other signal to grow exponentially over time (reaching its maximum) and then decay exponentially. While we considered the impact of past values of a time series on its future values as a function only of their distance in time, we postulated that the impact of another time series is not constant but increases and then decreases as a function of the distance in time over the chosen experimental time-frame. This behavior can be justified from a biological point of view since genes have been shown to influence the expression of other genes in a "hit and run" manner (Doidy et al., 2016). The causal gene physically interacts with the target gene then dissociates, but the transient target gene's expression continues to be affected after the dissociation. For this reason, we proposed the following function:

$$h(\delta_{t_i}) \coloneqq \psi \exp\left[-\frac{(\delta_{t_i} - \gamma)^2}{\eta}\right],$$

where $\psi \in (-1,1)$ is a parameter that describes the "intensity" and "direction" of the dependence of a time series on the other while $\gamma \in \mathrm{R}_+$ denotes the distance in time at which the dependence of a time series on another is maximal. Finally, $\eta \in \mathrm{R}_+$ plays a similar role to φ in the previously defined function $h_{\cdot}(\cdot)$.

As stated earlier, other explicit (more complex) forms can be defined for these functions. However other forms would probably require more parameters to characterize them and would be complicated (if not impossible) to estimate in practice given the small sample sizes collected in

many experimental settings such as the one considered in this work. Hence, in order to balance model complexity with practical feasibility, we applied the above functions to understand the relationship between different root and shoot signals since they can be considered as appropriate approximations to the underlying dependence structure.

Testing Procedure

Once the model is defined, the goal of this work is therefore to understand which structure of the matrix $A(\delta_{t_i})$ in (2) best describes the observed data (e.g. diagonal, lower/upper triangular). In this perspective, we are interested in making a decision on the following set of hypotheses:

H_0 : $A(\delta_{t_i})$ is diagonal.

H_A : $A(\delta_{t_i})$ is lower triangular.

Hence, the null hypothesis H_0 states that neither signal has an impact on the other (i.e. no Granger causality in the bivariate time series) while the alternative H_A states that the first signal Granger-causes the second. This alternative can of course be changed to "$A(\delta_{t_i})$ is upper triangular" therefore reversing the direction of dependence.

The MLE defined in (6) estimates the parameters of the proposed model using the likelihood function in (5). Based on the latter, a commonly used test to determine the performance of a more "simple" model (such as the one considered in the null hypothesis stated above) with respect to a more "complex" model (such as the one in the alternative hypothesis) is the likelihood-ratio test whose statistic is given by

$$LRT \coloneqq -2\log\left(\frac{L(\tilde{\theta}_0)}{L(\tilde{\boldsymbol{\theta}}_1)}\right) = 2\left(Q_n(\tilde{\theta}_0) - Q_n(\tilde{\theta}_1)\right),$$

where θ_0 and θ_1 represent the estimated parameters of the models under the null and alternative hypothesis respectively. In order to perform this test one needs to derive the distribution of the *LRT* statistic under the null hypothesis which is asymptotically chi-squared with $p^\star$ degress of freedom, where $p^\star$represents the number of extra parameters contained in $\theta_1 \in \mathrm{R}_{p1}$ with respect to $\theta_0 \in \mathrm{R}_{p0}$ (i.e.$p^\star \coloneqq p_1 - p_0$). Using this distribution and the observed *LRT* statistic one can then test the null hypothesis thereby concluding whether or not a signal Granger-causes the other.

Implementation

As highlighted before, the sample sizes coming from target biological applications are typically small (i.e. $5 < n < 20$ time points) and it therefore seems unreasonable to make use of asymptotic properties in these cases. For this reason, Monte-Carlo-based techniques appear to be a natural alternative that consider the small sample distribution of the test statistics of interest. More specifically, we propose to use parametric bootstrap to derive the small sample distribution of the *LRT* statistic under the null hypothesis as described in Algorithm 1.

Algorithm 1: Parametric Bootstrap for LRT Statistic

Result: Estimated LRT distribution under H_0.

Initialize $\mathrm{h} = 0,\ \mathrm{H} \geq 100$ and a zero vector LRT_{boot} of dimension H;

while $h \leq H$ **do**

1. $h = h + 1$;
2. Simulate a bivariate time series $\left(Z_i^{(h)}\right)$ of the same sample size as the original signals from the model $F_{\hat{\theta}_0}$;
3. Estimate θ_0 and θ_1 from the simulated sample $\left(Z_i^{(h)}\right)$ to obtain $\widehat{\boldsymbol{\theta}}_{\boldsymbol{0}}^{(\boldsymbol{h})}$ and $\widehat{\boldsymbol{\theta}}_{\boldsymbol{1}}^{(\boldsymbol{h})}$ respectively;
4. Compute $LRT_{boot}^{(h)} = 2\left(Q_n\left(\hat{\theta}_0^{(h)}\right) - Q_n\left(\hat{\theta}_1^{(h)}\right)\right)$

The parametric bootstrap approach allows for a good approximation (for large *H*) of the *LRT* statistic distribution under the null hypothesis by using the empirical distribution of the $LRT_{boot}^{(h)}$ values. Given this distribution, it is possible to obtain an approximate *p*-value (see Davison and Hinkley, 1997) as follows

$$p\text{-value} \approx \frac{1}{H+1}\left(1 + \sum_{h=1}^{H} \mathbf{1}_{\{\boldsymbol{LRT}_{\boldsymbol{boot}}^{(\boldsymbol{h})} > \boldsymbol{LRT}\}}\right).$$

If this *p*-value is smaller than a chosen level of significance α, then we can reject the null hypothesis H_0 that there is no Granger causality in favor of the specific alternative hypothesis H_A being tested.

Given this testing framework, there are a couple of issues that need to be considered, the first of which is the computational burden of Algorithm 1. In fact, the above defined *p*-value needs to be computed for all possible bivariate signal combinations and alternative hypotheses resulting in $2 \times N_X \times N_Y$ tests, where N_X and N_Y are the number of measured expressions in the two considered

signals. Considering that the computational complexity to obtain the above *p*-value, given our assumptions, is approximately of order $\mathcal{O}(nH)$, the final algorithmic complexity of the entire procedure would be of order $\mathcal{O}(nHM)$ with $M = N_X \times N_Y$ and $N_X, N_Y \gg 10^3$. This implies that the time required to obtain the results can be considerable. Another issue consists in the multiple testing framework this procedure entails, which therefore has consequences in terms of False Discovery Rate (FDR). Indeed, each (X_{t_i}) signal is tested $2N_Y$ times (and vice-versa for the (Y_{t_i}) signals) which would require to compare the *p*-value to the level $\frac{\alpha}{(2N_X N_Y)}$ if applying, for example, a Bonferroni correction. If the sizes N_X and N_Y are considerable, this would require increasing the number of simulations *H* in a proportional manner consequently increasing the computational burden. Unless one uses the asymptotic approximation to obtain a *p*-value (which would be highly unreliable for the small sample sizes used in these settings), there is currently no way of avoiding such a computational bottleneck.

RESULTS AND DISCUSSION

The described approach was applied to the time-evolved transcriptome of Arabidopsis roots and shoots (the (X_{t_i}) and (Y_{t_i}) signals respectively) whose measurements were made through an experimental setup described more in detail in the Appendix along with the chosen pre-processing (Appendix A: Supplemental, Supplemental Methods). These signals, each of length *n* = 10 and collected at higher frequencies in the initial experimental phase, generate 1,234,264 possible gene pairs from significantly differentially expressed root and shoot genes. Using $H = 10^3$, we applied the procedure described in "Granger-Causal Analysis for Irregular Data" which produced a final list of 3,078 gene pair interactions whose details are listed in Supplemental Table 1 (only *p*-values

equal to zero were considered to reduce FDR as much as possible given computational constraints). Out of these interactions, 2,012 had a predicted root-to-shoot direction of influence meaning that the root gene was identified as the (Granger) causal gene, or the influencer on the expression of the shoot gene. The remaining 1,066 interactions had a predicted shoot-to-root direction of influence. In addition, the approach predicted 1,616 positive interactions (i.e. $\psi > 0$) and 1,462 negative interactions (i.e. $\psi < 0$). Due to the limited and irregular number of samples across time, we choose to classify the time of influence at which the maximum influence between two genes occurred (measured by the γ parameter) into three general groups: Early (0-15 min), Middle (20-45 min), and Late (60-120 min). Based on this, among the 3,078 interactions, 2,502 occur Early, 548 occur during the Middle time frame, and 28 occur Late. In the following paragraphs we analyze only some of the model-predicted interactions in terms of their known properties and/or based on how they have a coherent biological interpretation. To do so, we will use the term "causal" to indicate genes that impact another gene, the latter being referred to as "target".

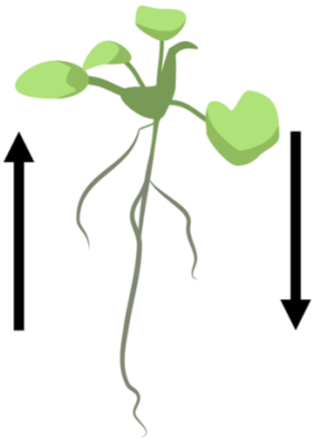
Global analysis of model-predicted interactions reveal links between biological processes and pathways

Gene Ontology (GO) term analysis was performed to understand what pathways and processes were influenced across tissues over time (see Appendix A: Supplemental, Supplemental Methods). As highlighted also in Fig. 1, at early time points (0 - 15 minutes), causal root-genes reflect the early nitrogen response, while target-shoot genes reflect post-transcriptional and translational processes (see Supplemental Tables 2 and 11). At later time points there is a shift in metabolism in which causal root-genes are involved in degradation and catabolic processes (45 - 120 minutes)

(GO enrichment p-value < 0.01), while the predicted shoot target genes are involved in peptide biosynthesis (15 - 45 minutes) (GO enrichment p-value < 0.01) and sugar/carbohydrate response and signaling (45 - 120 minutes) (GO enrichment p-value < 0.01) (see Supplemental Tables 3, 4, 12 and 13). GO analysis of the causal shoot-genes reflect the synthesis of shoot-derived signals, such as peptides and hormones, while the identified target root-genes are involved in phosphorus metabolic processes (0 - 15 minutes), lateral root development (15 - 45 minutes), and response to cytokinin (45 - 120 minutes) (see Supplemental Tables 5-10). This analysis reflects much of the current knowledge about long distance nitrogen signaling between roots and shoots (Ruffel et al., 2011; Ko and Helariutta, 2017; Poitout et al., 2018).

Time	Shoot Target GO Term	GO ID
Early	Translation	GO:0006412
	Peptide biosynthetic process	GO:0043043
	Ribosome biogenesis	GO:0042254
Middle	Cellular nitrogen compound metabolic process	GO:0034641
	Peptide metabolic process	GO:0006518
Late	Response to monosaccharide	GO:0034284
	Response to hexose	GO:0009746
	Glucose mediated signaling pathway	GO:0010255

Time	Shoot Causal GO Term	GO ID
Early	Peptide biosynthetic process	GO:0043043
	Amide biosynthetic process	GO:0043604
	Negative regulation of cytokinin-activated signaling pathway	GO:0080037
	Nitrogen cycle metabolic process	GO:0071941
Middle	Developmental process	GO:0032502
	Negative regulation of biological process	GO:0048519
	Biological regulation	GO:0065007
Late	Regulation of starch metabolic process	GO:2000904
	Regulation of hormone biosynthetic process	GO:0046885



Time	Root Causal GO Term	GO ID
Early	Nitrate metabolic process	GO:0042126
	Nitrate assimilation	GO:0042128
	Response to nitrate	GO:0010167
Middle	Trehalose metabolic process	GO:0005991
	Response to salicylic acid	GO:0009751
	Trehalose biosynthetic process	GO:0005992
Late	Regulation of mrna catabolic process	GO:0061013
	Nucleoside salvage	GO:0043174
	Pyrimidine-containing compound salvage	GO:0008655

Time	Root Target GO Term	GO ID
Early	Phosphorus metabolic process	GO:0006793
	Nitrogen compound metabolic process	GO:0006807
Middle	Lateral root development	GO:0048527
	Sulfur compound metabolic process	GO:0006790
Late	Response to cytokinin	GO:0009735

Figure 1: Selected enriched GO terms for root causal, shoot causal, root target and shoot target genes.

Model predictions are supported by in planta observations

A gene network was constructed where nodes (1,322 nodes) represent genes and edges (3,078 edges) constitute the model-predicted interactions described above (see Appendix A: Supplemental, Supplemental Methods). Network analysis revealed that the gene interaction network with model-defined edges closely follows a power law distribution ($R^2 = 0.92$), indicative of a scale-free biological network (Barabási, 2003; Albert, 2005). The validity of this finding was supported by a simulation of 10^3 randomly generated networks using the same number of nodes and edges whose R^2 values for the power law distribution were all between 0 and 0.35 (see Fig. 2). Network analysis for out-degree identified causal hub genes that are predicted to be highly influential in the temporal root-shoot transcriptomes in response to nitrogen treatment. Taking into consideration directionality, the top ten hubbiest genes in the network, based on out-degree, include three transcription factors previously implicated in the Arabidopsis nitrogen response: AFB3 (AT1G12820) (Vidal *et al.*, 2013b, 2014; Xu and Cai, 2019), BT1 (AT5G63160) (Vidal et al., 2013a; Araus et al., 2016; Sato et al., 2017), and WRKY38 (AT5G22570) (Scheible et al., 2004; Gaudinier et al., 2018) (see Supplemental Table 14). Other network hubs include the TF RD21A (AT1G47128) that is involved in autophagy and senescence which are key nitrogen turnover processes; and the RNA binding protein CID10 (AT3G49390), which is a poly(A) binding protein (PABP) potentially involved in mRNA stability or degradation (see Supplemental File 1). Further investigation of the interaction network revealed a number of previously identified genes and gene-gene relationships involved in local and long-distance nitrogen signaling, namely those involved in transcriptional regulation and in long-distance signaling by hormones and peptides, which are described in detail in the following sections.

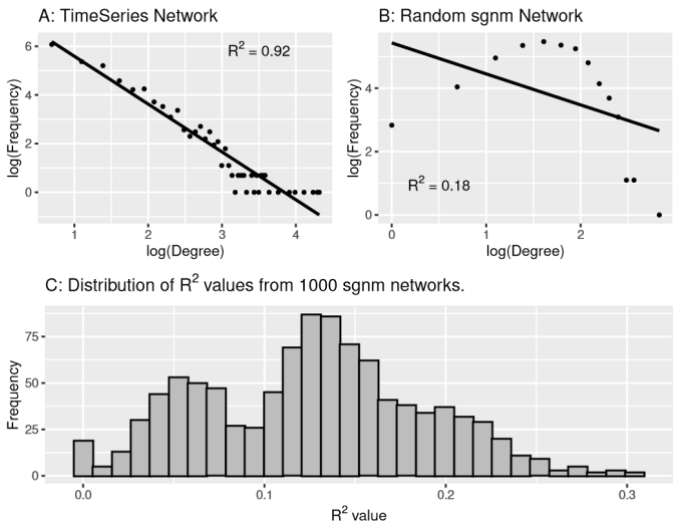


Figure 2: A: Node degree distribution for the network generated from model predictions ($R2$ = 0.92). B: Node degree distribution for a randomly generated network ($R2$ = 0.18). C: Histogram of the $R2$ values for the node degree distribution of 1,000 randomly generated networks ($0 \leq R2 <$ 0.35).

Regulators of nitrogen processes

The transcription factors TGA1 and TGA4 were shown to be involved in mediating the primary nitrate response in roots by regulating the expression of the nitrate transporters NRT1.1 and NRT2.2, and also by coordinating the root developmental response to nitrate (Alvarez et al., 2014). From our analysis, root-expressed TGA1 is predicted to influence the expression of ten shoot genes, while shoot-expressed TGA1 is predicted to influence the expression of four root genes (see Table 1). To further investigate these predicted relationships, promoter analysis using FIMO from MEME Suite (Bailey and Machanick, 2012) was performed (as outlined in Appendix A: Supplemental, Supplemental Methods). At least one TGA1 binding motif had a significant

occurrence (FIMO p-value < 0.0001) in the putative promoters of two of the targeted shoot genes: a protease-associated RING/U-Box zinc finger family protein (AT1G71980) and HSFB2A heat shock transcription factor B2A (AT5G62020). The TGA1 motif also had a significant occurrence (FIMO p-value < 0.0001) in the three root target genes: a phosphoglycerate mutase-like family protein (AT3G01310), BAP1 BON association protein 1 (AT3G61190) and a copper amine oxidase family protein (AT4G12290). DAP-seq (DNA Affinity Purification and sequencing) is an experimental technique allowing for the discovery of transcription factor binding sites on genomic DNA in vitro. A recent DAP-seq experiment showed that TGA1 actively binds to three shoot genes, AT1G71980, CIPK1 (AT3G17510) and an unknown protein (AT4G21215), as well as the three root target genes from the promoter analysis (O'Malley et al., 2016) (see Tab. 1). Furthermore, the model-predicted targets of TGA1, Phosphoglycerate mutase-like family protein (AT3G01310), and alpha/beta-Hydrolases superfamily protein (AT5G18640) were predicted to be direct targets of TGA1 in a TARGET (Transient Assay Reporting Genome-wide Effects of Transcription factors) assay experiment in root protoplasts by Brooks et al. (2019). A TARGET assay can identify candidate transcription factor targets based on TF-induced changes in gene expression (Brooks et al., 2019). These in-planta results provide support for the predicted interactions between TGA1 and its target genes within the same tissue, but additional studies will be needed to test if these interactions occur directly or indirectly between tissues.

Long-distance signaling by hormones and peptides

Cytokinin Response Factors (CRFs): Transcription factors (TF) with previously described regulatory roles in nitrogen uptake and assimilation include members of the ERF, bZIP, and NLP

	GENE ID	GENE DESCRIPTION	INFLUENCE	FIMO	DAP SEQ	TARGET
SHOOT	AT1G55890	Tetratricopeptide repeat (TPR)-like superfamily protein	Positive			
	AT1G71980	Protease-associated (PA) RING/U-box zinc finger family protein	Negative	Y	Y	
	AT1G73100	SDG19, SUVH3, SU(VAR)3-9 homolog 3	Positive			
	AT2G15230	ATLIP1, LIP1, lipase 1	Negative			
	AT3G06780	glycine-rich protein	Positive			
	AT3G17510	CIPK1, SnRK3.16, CBL-interacting protein kinase 1	Negative		Y	
	AT4G21215	unknown protein	Negative		Y	
	AT5G04840	bZIP protein	Negative			
	AT5G18640	alpha/beta-Hydrolases superfamily protein	Negative			Y
	AT5G62020	AT-HSFB2A, HSFB2A, heat shock transcription factor B2A	Negative	Y		
ROOT	AT3G01310	Phosphoglycerate mutase-like family protein	Negative	Y	Y	Y
	AT3G61190	BAP1, BON association protein 1	Negative	Y	Y	
	AT4G12290	Copper amine oxidase family protein	Positive	Y	Y	
	AT5G28770	AtbZIP63, BZO2H3, bZIP transcription factor family protein	Negative			

Table 1: TGA1 target genes in root and shoot with which genes have a TGA1 motif occurrence of $p < 0.0001$ from the FIMO promoter analysis, and genes to which TGA1 has been shown to physically bind based on DAP-seq and TARGET experiments. "Y" indicates existing evidence for a predicted interaction from a specific experiment, whereas empty cells indicate possible avenues of future investigation.

TF families (Konishi and Yanagisawa, 2013; Krapp et al., 2014; Vidal et al., 2015; Varala et al., 2018; Brooks et al., 2019). Of particular interest are the ERF TFs CRF 1-5. These CRFs were previously implicated in nitrogen signaling, targeting genes involved in nitrogen uptake and assimilation (Varala et al., 2018; Brooks et al., 2019). In our analysis, CRF5 expressed in the shoot was predicted to positively influence the expression of a heavy metal transport/detoxification

protein (AT5G03380) expressed in the root. Using Elefinder (Hudson, 2005), CRF5 has been shown to bind to the GCC-box motif (GCCGCC) (Fujimoto et al., 2000; Sakuma et al., 2002; Liang et al., 2010) which is over-represented in the 2kb promoter region of AT5G03380 (E-value = $5.85 \cdot 10^{-4}$, see Appendix A: Supplemental, Supplemental Methods), indicating potential for a physical protein-DNA binding interaction. Shoot-expressed CRF3 is a predicted target of the causal root-expressed gene AT4G34419 (an unknown protein) in which AT4G34419 positively influences the expression of CRF3. Root-expressed CRF4 is predicted to influence the expression of the shoot genes SAUR-like auxin responsive protein family (AT4G34750), and Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family (AT3G44380). CRF4 is predicted to positively influence both of these genes during the early time interval. Like CRF5, CRF4 binds to the GCC-box motif, and this motif is overrepresented in the 2kb upstream region of AT4G34750 (E-value = $1.43 \cdot 10^{-2}$, see Appendix A: Supplemental, Supplemental Methods). Root CRF4 is also predicted to negatively influence the shoot gene Homeobox Protein 6 (HB6, AT2G22430) during the middle time interval. CRF4 was shown to bind to HB6 via DAP-Seq (O'Malley et al., 2016). HB6 is a known negative regulator of the abscisic acid (ABA) signaling pathway (Himmelbach et al., 2003; Fujita et al., 2011). The ABA pathway is a phytohormone signaling pathway that was previously implicated in coordinating the long-distance nitrogen response (Kiba et al., 2011; Guan, 2017). A recent study by Varala et al. (2018) showed that CRF4 targets the TFs SNZ1 and CDF1, which in turn target HB6. The overexpression of CRF4 decreased the rate of nitrate uptake and altered root architecture in response to nitrogen treatment compared to WT plants (Varala et al., 2018). In CRF4 overexpressors, there was a decrease in primary root length and lateral root number under low nitrate conditions. Lateral root development has been

shown to be inhibited under low nitrate conditions, which trigger ABA accumulation (Signora et al., 2001; Vidal et al., 2010; Léran et al., 2015; Sun et al., 2017). Thus, the results of our analysis suggest a coherent type 4 feed-forward loop (Mangan and Alon, 2003) in which root CRF4 represses shoot HB6 which represses whole plant ABA signaling (see Fig. 3), and may have physiological consequences for the observed changes in lateral root formation (Varala et al., 2018).

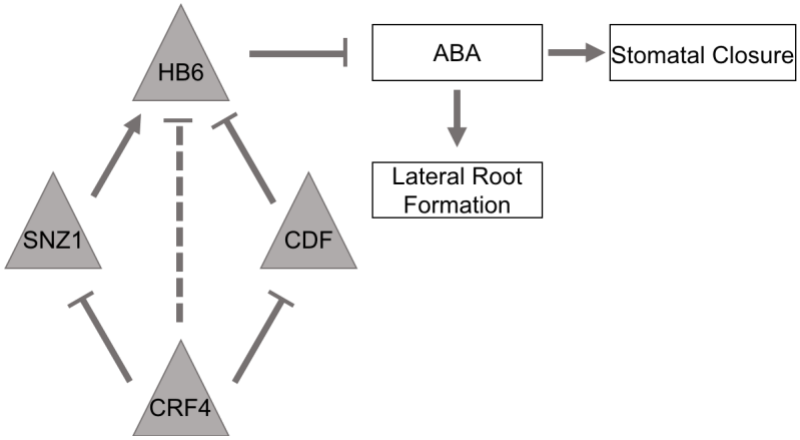


Figure 3: CRF4 interaction pathway, with flat-head arrows indicating negative interactions, pointed-head arrows indicating positive interactions, dashed arrows representing predicted interactions from the model, and solid arrows representing known interactions.

Arabidopsis Response Regulators (ARRs): The cytokinin signaling pathway is triggered by nitrogen and has been shown to be involved in the coordination of both root-to-shoot and shoot-to-root nitrogen-responses. In the shoots, cytokinins stimulate cell division and differentiation, whereas in the roots cytokinins reduce the activity of nitrogen uptake (Sakakibara et al., 2006). Cytokinins have also been shown to induce the expression of ARRs, which then regulate cytokinin

signaling through feedback (To et al., 2007, 2004). For example, ARR4 (AT1G10470) is a Type-A response regulator that negatively regulates the cytokinin response (To et al., 2007). In our study, root ARR4 is predicted to influence the expression of three shoot genes (see Supplemental Table 1), including a transmembrane amino acid transporter family protein (ATAVT1B; AT3G54830). During the middle time interval, root expressed ARR4 is predicted to negatively influence the expression of AVT1 in shoots. Yeast AVT1 homologues have been implicated in the vacuolar uptake of large neutral amino acids including glutamine, asparagine, isoleucine, and tyrosine (Russnak et al., 2001; Tone et al., 2015) where they are stored in the vacuole under high nitrogen conditions (Sekito et al., 2008). When nitrogen starvation occurs, several AVT genes are upregulated to facilitate the export of the stored amino acids from the vacuole to the cytoplasm for protein synthesis (Fujiki et al., 2017). The analysis detected a relationship between ARR4 and AVT1B suggesting a potential mechanism by which cytokinin-induced ARR4 in the root may provide a long-distance signal to regulate shoot vacuolar amino acid import under high nitrogen conditions, like those used in this study.

Peptides: Signal peptides have been implicated in the whole plant response to nitrogen (Tabata et al., 2014; Ohkubo et al., 2017; Oh et al., 2018). In the present study, seven peptides were uncovered as causal genes involved in 20 interactions (see Supplemental Table 16). ATPSK4 is a Phytosulfokine 3 precursor and was shown to influence plant growth and cellular longevity, in particular root growth (Matsubayashi et al., 2006). CLE (Clavata3/ESR-related) peptides have long been known to be involved in long distance nitrogen-signaling in legumes and have also been shown to be involved in nitrogen-signaling in Arabidopsis (Bidadi et al., 2014; Okamoto et al.,

2016). In the present study, three CLE peptides are present in the predicted long-distance signaling network; CLE3 (AT1G06225), CLE4 (AT2G31081), and CLE27 (AT3G25905). CLE3 is a predicted causal gene expressed in the shoot that influences the root-expressed gene AT5G52530 (dentin sialophosphoprotein-protein related), while CLE4 is a causal root-gene predicted to influence the expression of four shoot-expressed genes either negatively (AT5G67510 Translation protein SH3-like family protein; AT1G55890 Tetratricopeptide repeat (TPR)-like superfamily protein) or positively (AT3G61620 RRP41, 3'-5'-exoribonuclease family protein; AT5G18640 alpha/beta-Hydrolases superfamily protein). Lastly, CLE27 is a Clavata family gene that was previously shown to be repressed by auxin (Wang et al., 2016). In our study, CLE27 is a shoot expressed causal gene predicted to positively influence the expression of AT5G03380 (Heavy metal transport/detoxification superfamily protein) in the root. Devil/Rotundifolia Like (DVL) peptides are non-secretory peptides, conserved in plants, that can act as small signaling molecules and influence development in Arabidopsis (Wen et al., 2004). MTDVL1 was previously shown to be involved in symbiosis in *Medicago truncatula*, in which it has a negative regulatory role in nodulation (Combier et al., 2008). Two Devil peptides were identified in our analysis: DVL4 and DVL11. Of the four interactions involving DVL11, root DVL11 is predicted to be the causal gene influencing three shoot genes. Of these, DVL11 is predicted to positively influence the expression of ICK1, a cyclin-dependent kinase inhibitor family protein (AT2G23430). ICK1 is a known key regulator in development, and can inhibit entry into mitosis (Weinl et al., 2005). Root DVL4 is also predicted by the analysis to influence three shoot genes. Specifically, root DVL4 is predicted to positively influence shoot TCP-1/cpn60 chaperonin family protein (AT3G13470) at a middle time point. A previous study explored the transcriptional landscape of a DVL4 overexpressor line

and showed that overexpression of DVL4 resulted in the upregulation of a number of genes encoding transcription factors (Larue et al., 2010). Our re-analysis of the microarray data from this study (see Appendix A: Supplemental, Supplemental Methods) revealed that TCP1 was downregulated in DVL4 overexpressor plants compared to wild type Arabidopsis plants, providing support for a gene-gene interaction between DVL4 and TCP1 (see Fig. 4); however, this needs further exploration in the context of a nitrogen-signal.

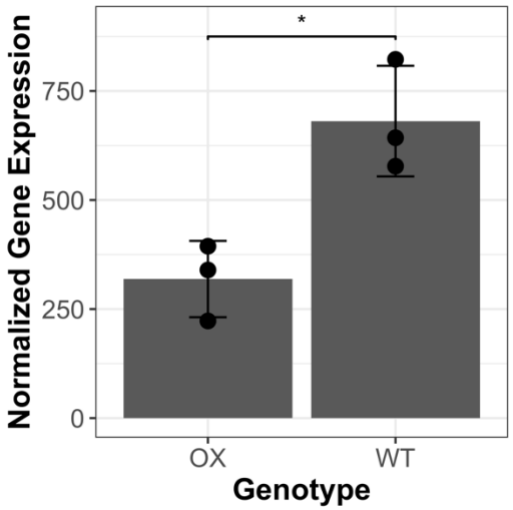


Figure 4: Bar chart of the normalized gene expression obtained from GSE8975 via GEO2R for TCP-1 in DVL4 overexpressor (OX) and wild-type (WT) Arabidopsis plants (t-test: p-value $\leq$ 0.05).

Model predictions contain an over-representation of mobile causal gene products

The proposed approach, as stated previously, aims at understanding if the expression of one gene influences the expression of its target gene through the notion of Granger-Causality. Biologically, this influence may be direct or indirect. It has previously been shown that mobile mRNAs that

originate from one cell type or organ can translocate to another cell-type or organ and have functional activity there (Lough and Lucas, 2006; Banerjee et al., 2009; Luo et al., 2018). To identify potential direct, long-distance interactions, we took advantage of two recent publications (Thieme et al., 2015; Guan et al., 2016) with extensive lists of experimentally determined mobile mRNAs that travel from root-to-shoot and from shoot-to-root. The lists of directional, causal genes from our model were intersected with the mobile transcripts identified by these studies. This analysis provided support for 204 causal genes involved in 340 predicted root-to-shoot, and 241 predicted shoot-to-root relationships; meaning that the direction of influence of the causal gene was the same in our analysis as that experimentally determined by these studies. An over-representation analysis (see Appendix A: Supplemental, Supplemental Methods) was performed with the following hypotheses: "H_0: the proposed approach (model) is equivalent to detecting known mobile transcripts randomly" and (alternative) "H_A: the proposed approach (model) detects more known mobile transcripts than random selection". In this case the *p*-value is 0 allowing us to reject the null hypothesis and hence the model is able to detect mobile transcripts which are potentially able to interact directly with their target genes. At least 36 of the total causal genes are known RNA-binding proteins (Marondedze et al., 2016), and 21 of these are mobile (see Supplemental Table 17). In general, RNA-binding proteins can form ribonucleoprotein complexes (RNPs) that facilitate phloem transport and long-distance trafficking of RNA molecules (Ham et al., 2009; Kehr and Kragler, 2018). An additional 79 causal genes involved in 203 relationships (121 root-to-shoot and 82 shoot-to-root) have not been experimentally shown to be mobile but are predicted to produce an mRNA molecule that possesses a t-RNA like motif. Guan et al. (2016) also hypothesized that some mRNA have a tRNA-like structure in their sequence. This allows the

mRNA to fold into a tRNA-like shape that confers some stability to the mRNA strand. This stability allows the mRNA to move long distances in the plant. These results suggest that a large proportion of the model-predicted causal genes have the potential to influence the expression of its target gene (directly or indirectly) via long-distance vascular trafficking. One example of a model-predicted gene interaction that may function through interaction of a mobile causal gene with its target is the relationship between root derived aconitase 2 (ACO2), predicted to have a negative influence on the expression of malate dehydrogenase (MDH2) in the shoot. ACO2 is the only isoform of aconitase that is specifically induced by nitrogen treatment. Root ACO2 is involved in the TCA cycle, while shoot MDH2 is localized in the mitochondria and involved in gluconeogenesis. One possibility is that a direct or downstream gene product of root ACO2 represses shoot MDH2, resulting in possible down-regulation of shoot gluconeogenesis in response to a large, transient nitrogen signal. Although the specific mechanism of this relationship needs experimental exploration, it is partially supported by existing data describing the tight relationship between carbon and nitrogen metabolism to maintain whole plant C:N balance (Palenchar et al., 2004; Zheng, 2009; Goel et al., 2016). Alternatively, aconitase, an iron-sulphur protein, has been shown to be a bifunctional enzyme/RNA-binding protein that binds to iron-responsive elements in target RNA to stabilize the transcript and function in iron homeostasis (Hentze and Argos, 1991). Our analysis predicted a positive relationship between ACO2 (causal root) and Ironman 1 (target shoot), an Fe-uptake inducing peptide 3 that is involved in the regulation of iron deficiency response genes (Grillet et al., 2018). It was previously shown that nitrogen treatment induces the expression of genes involved in iron uptake, transport, and homeostasis in plants (Wang et al., 2000, 2003), and that the form of nitrogen taken up by roots

influences the amount of iron accumulation in leaves (Zou et al., 2001). There is also a well-established relationship between nitrogen and Fe pathways since Fe is a component of many enzymes involved in nitrate assimilation (Wang et al., 2003).

CONCLUSIONS

This work puts forward an approach to perform Granger-Causal analysis for (small-sample) irregularly spaced bivariate signals which overcomes existing limitations in the analysis of biological time series data following this common sampling scheme. Based on this new framework, (Granger) causal relationships were detected and whole-organism molecular response to a nitrogen signal were predicted. The survey of genes with predicted temporal cause-and effect relationships enabled discovery of coordinated biological processes and chemical pathways that communicate the nitrogen-signal between roots and shoots of plants. These coordinated processes can now be further investigated to identify potential regulatory bottlenecks that influence whole plant nitrogen uptake/utilization efficiency. The abundance of genes involved in the known transcriptional nitrogen-response (nitrogen-transport and assimilation) as both causal and target genes indicate that the proposed approach was able to capture whole-plant response to a transient nitrogen-treatment across tissues. The predicted cross-organ dependencies provide insights and hypotheses about potential signaling cascades that are triggered sequentially as the nitrogen-signal propagates from roots-to-shoots-to-roots. Importantly, regulatory factors that have not previously been implicated in whole plant nitrogen-response were highlighted by the proposed approach. These novel factors can be targets for engineering improvements in plant nitrogen uptake/utilization efficiency. The findings from this research will have implications for predicting

causal molecular relationships that influence intercellular, long-distance nitrogen-signaling, and the methodological framework proposed in this work is applicable to researchers struggling with meaningful integration of dynamic, system-wide transcriptome data.

CHAPTER 3: EXPERIMENTAL VALIDATION OF SELECTED INTERACTIONS AND FUTURE DIRECTIONS

ABSTRACT

The time series model predicted 3,078 N-responsive gene-gene interactions. These interactions were globally analyzed using a bioinformatics pipeline to identify causal N-responsive genes for further analysis. Grafting and physiological measurements were used to validate selected gene-gene interactions. CRF4 knockout (KO) seedlings were grafted to Col-0 seedlings. Grafting had low success rate, with surviving grafts producing adventitious roots from the scion, above the graft junction. Physiological observations of 5-week-old CRF4 KO, CRF4 overexpressor (OX) and Col-0 plants showed that CRF4 OX plants had the largest rosette, while the leaves on CRF4 KO plants had curled edges indicating water loss. Gas exchange measurements at 8 weeks showed that while there was a 1.4 and 1.9-fold increase of CRF4 KO and CRF4 OX stomatal conductance to Col-0 stomatal conductance, an ANOVA analysis revealed that this change was not significant. Future directions are discussed with respect to using the time series model to integrate shoot metabolite data with root and shoot transcriptomic data to identify possible gene-metabolite interactions involved in the N response.

INTRODUCTION

The predicted interactions from the time-series model were attributed as biologically meaningful by applying statistical constraints, resulting in 3,078 predicted gene-gene interactions. However, further *in-silico* and *in-planta* validations can not only increase the confidence in the model

predictions but also provide the experimental data to improve the model. One of the major challenges, however, is to individually test these thousands of predictions. In Chapter 2, a bioinformatic approach was used to validate predictions supported by existing literature, for genes that produce known mobile gene products (Guan et al., 2017), transcription factor-target interactions (Brooks et al., 2019), and genes involved in known signaling pathways (Varala et al., 2018, Brooks et al., 2019). This analysis identified promising gene-gene predictions that are ideal candidates for experimental validation using different methods including, grafting and physiological measurements of mutant plants.

Cytokinin response factor 4 (CRF4), as described in Chapter 1, has been implicated as a key transcription factor hub by various studies to be involved in N uptake and assimilation (Varala et al., 2018, Brooks et al., 2019). Varala et al. (2018) have also shown that CRF4 regulates genes that alter biomass, root development and nitrate uptake under low N conditions. From the model predictions, root CRF4 was predicted to negatively influence shoot Homeobox Protein 6 (HB6) at a middle time point. Furthermore, DAP-seq experiments have shown that CRF4 binds to HB6 (O'Malley et al., 2016). HB6 is a known regulator of the abscisic acid (ABA) hormone signaling pathway where it represses ABA production (Himmelbach et al, 2003). The repression of ABA by HB6 has been shown to result in repression of stomatal closure in the leaves (Lechner et al., 2011). Therefore, CRF4 should positively regulate stomatal closure. Repression or absence of CRF4 gene expression should result in reduced stomatal closure through its relationship with HB6 as illustrated in Figure 3, Chapter 2.

In this chapter, I will report my preliminary attempts at experimentally validating model-predicted relationships using mutant plant lines for CRF4, and I will discuss future directions that

integrate metabolomic and transcriptomic data into the model to predict new gene-metabolite interactions in the long-distance N signaling pathways.

METHODS

Plant Material and Preparation

Arabidopsis thaliana seed stocks for a CRF4 overexpressor (OX) and CRF4 knockout (KO) line, both in Col-0 background, were donated by Aaron Rashotte from Auburn University. Seeds were surface sterilized using a sterilization mix of 4mL of 100% ethanol, 1mL of bleach and 3mL of water. To each 1.5mL Eppendorf tube containing seeds for the experiments, 1mL of this solution was added and the tubes were mixed by inversion before incubation for 8 minutes. The solution was discarded, and the seeds were washed three times with 70% ethanol. The seeds were air-dried in a laminar flow hood for 30 minutes. The seeds were stratified in 1mL of sterile water, tube wrapped in foil and stored at 4°C for two days before sowing on plates.

Grafting

Sterilized *A. thaliana* seeds were grown on 50mL of a half-strength Murashige and Skoog (1/2 MS) modified basal-salt mixture (Phytotech Labs M524) containing 0.5% w/v sucrose and 2% w/v agar. The seedlings were grown for 6 days under short-day conditions (8hrs light, 16hrs dark, 22°C) and the plates were oriented at an angle of about 5-10 degrees. The seedlings were then grafted following the methods described previously by Marsch-Martinez (2013). Briefly, cutting plates containing approximately 5mL of the MS media, described above, were used with 1% agar added. A thin strip was cut out to provide a solid surface for cutting, while the agar was used to

support the roots of the seedlings. The cotyledons were first cut from the seedlings, followed by a horizontal cut across the hypocotyl. The stock (root) was then moved to a new plate with MS agar, followed by the corresponding scion (shoot) to be attached. The scion and the stock were joined using tweezers and the end of a plastic dropper. The grafted seedlings were then grown for 10 days under short day conditions with the plates angled at 5-10 degrees. On the tenth day, any adventitious roots that formed were cut.

CRF4 Stomatal Conductance

CRF4 OX, CRF4 KO and Col-0 seeds were sown onto soil. Plants were grown under short-day conditions (8hrs light, 16hrs dark, 22°C). Plants were watered as needed, with every other watering with ½ MS solution containing 1% w/v sucrose. Due to a temperature issue with the original growth room, the plants were then moved to a new growth chamber after 2 weeks. The plants were grown for an additional 6 weeks. Stomatal conductance measurements were taken over two days using a portable infrared gas analyzer (LI-6400; Licor Biosciences, Inc., Lincoln, NE, http://www.licor.com) by setting the chamber to reflect ambient temperature, relative humidity set to 65% and light intensity at 400umol$m^{-2}s^{-1}$. The tenth leaf on each plant, with three plants from each genotype, was measured. After the measurements on the first day, the plants were watered, and measurements were repeated on the eleventh leaf the next day. The measurements were statistically analyzed in R using ANOVA.

Metabolite Extraction

Shoot samples were obtained from the time series experiment performed by Varala et al. (2019) as described in Chapter 2. Metabolites were extracted based on the method previously outlined by Lisec et al. (2006). Approximately 100mg of frozen tissue was added to screw cap tubes containing stainless steel beads. The tissue was homogenized at $4ms^{-1}$ in 30 second intervals until the tissue was a fine powder. Between intervals, samples were returned to liquid nitrogen to prevent thawing. 1400µL of 100% methanol pre-cooled at -20°C was added to each sample followed 10µL of 0.2mg mL^{-1} Ribitol was added as the internal standard. The samples were vortexed for 10seconds and then shaken for 10min at 70°C at 950rpm. Afterwards, the samples were centrifuged for 10min at 11,000g and the supernatant was transferred to a Schott glass vial. 750µL of -20°C chloroform and 1500µL of 4°C dH2O was added to the samples and tubes vortexed for 10 seconds. The samples were centrifuged for 15min at 2,200g. The entire upper phase was transferred into fresh 1.5mL tubes and dried in a vacuum concentrator before being stored at -80°C.

RESULTS

Grafting

Grafting CRF4 mutant seedlings to Col-0 seedlings provided limited success. In one instance of grafting 20 CRF4 KO seedlings to 20 Col-0 seedlings, only 60% (12 samples) had joined at the graft junction with all grafted seedlings producing adventitious roots (Figure 5). Attempting to remove these adventitious roots proved to be a difficult challenge as the adventitious roots sometimes wrapped over and around the hypocotyl. Often adjusting the grafted plant to get a better angle to cut the adventitious root resulted in breaking the graft junction. If the junction survived, the junction broke from trying to cut off the adventitious roots. Furthermore, adventitious roots

often reappeared at a later date in greater numbers (Figure 6). In other events, as the first leaves emerged the scion separated from the stock.

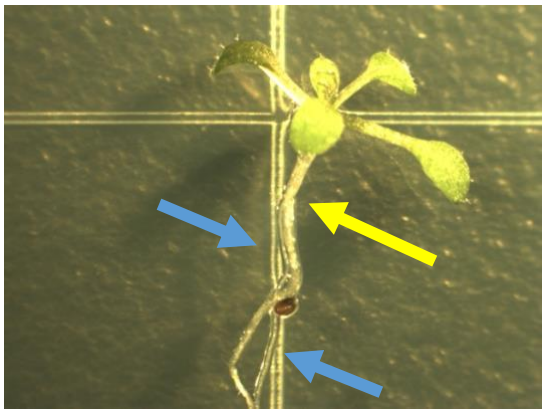


Figure 5: Image of a Col-0 scion:CRF4 KO stock 10 days after grafting. Yellow arrow points to the graft junction, while the blue arrows point to the adventitious root that formed from the scion after grafting.

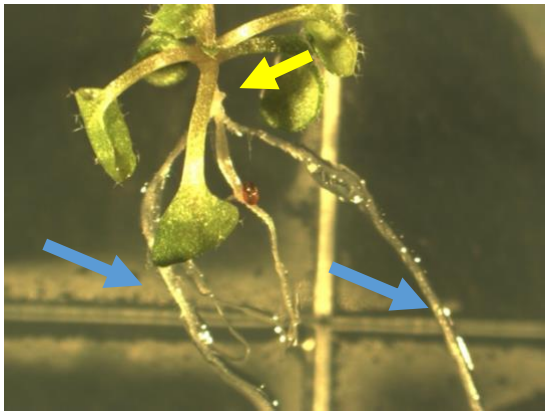


Figure 6: Image of a Col-0 scion:CRF4 KO stock 17 days after grafting (7 days after cutting first set of adventitious roots). Yellow arrow points to the graft junction, while blue arrows point to the adventitious roots.

CRF4 Physiological Observations and Stomatal Conductance

CRF4 overexpressor showed the largest rosette after 5 weeks, while Col-0 plants had long petioles. CRF4 knockout plants had smaller rosettes and the leaves were curled (Figure 7). Gas exchange measurements were conducted on 8-week-old plants over two days. On the first day before watering, CRF4 KO and CRF4 OX stomatal conductance had a 1.4 and 1.9-fold change compared to Col-0 wild-type control. However, an ANOVA analysis showed that there was no significant difference between the day of measurements (p-value = 0.460), genotypes (p-value = 0.140) or from an interaction of the two (p-value= 0.631) (Figure 8).

Figure 7: Image showing CRF4 OX (left), CRF4 KO (middle) and Col-0 (right) plants at 5 weeks old. The CRF4 OX plant has the largest rosette while Col-0 has the smallest leaves but longest petioles. The CRF4 KO leaves are curling at the edges indicating water loss.

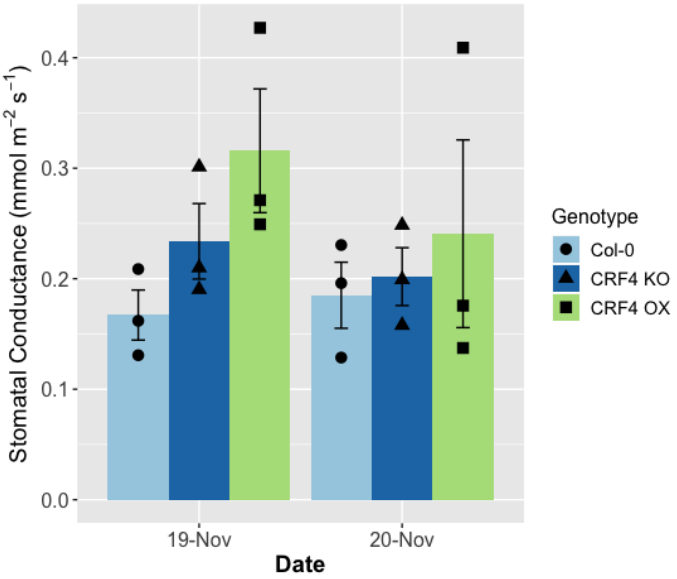


Figure 8: Graph showing average stomatal conductance (mmol m^{-2} s^{-1}) of CRF4 KO (KO), CRF4 OX (OX) and Col-0 plants before (19-Nov) and after (20-Nov) watering. While there was a 1.4- and 1.9-fold change in stomatal conductance in CRF4 KO and CRF4OX plants compared to Col-0, there was no significant change between values from an ANOVA analysis.

Discussion and Future Directions

Grafting is considered an excellent tool to investigate long-distance signaling in plants (Marsch-Martinez et al., 2013). However, grafting presents its own sets of challenges. With the current success rate, the best approach is to graft as many seedlings as possible, with the limitation of both the difficulty and time needed to graft seedlings together. Furthermore, with the challenges of the growth of adventitious roots post-graft, separation of scion and stock from the growth of the leaves, and simply graft junctions not joining, the chance of obtaining successful grafts was low. The methods used for grafting (seedling age, sugar concentration) represented the scenarios described by Marsch-Martinez (2013) that resulted in the highest success rates. Considering that genotype

may play a role in the ideal conditions for a successful graft, further experiments with grafting should consider using seedlings at a different developmental age and/or using different sucrose concentrations.

It was predicted that CRF4 OX plants would have higher stomatal closure, while CRF4 KO plants would have less stomatal closure compared to wild type plants. This was represented in the observed phenotype for the CRF4 KO plants. The leaves in the CRF4 KO plants were curled, compared to the leaves of the CRF4 OX and the Col-0 plants. This phenotype is consistent with another study that used ABA1 (AT5G67030, ABA Deficient 1) knock-out mutants to show that loss of ABA results in altered leaf morphology such as leaf curling (Barrero et al., 2005). Given that the degree of stomatal closure is tied to amount water loss in plants, the LICOR experiment was set up over a period of two days to obtain results before and after watering. However, there was no significant difference in stomatal conductance among the genotypes. There also was no significant change in stomatal conductance across the two days. Furthermore, there was a temperature issue in the growth chamber in the first 3 weeks. Temperatures approached 30°C which is detrimental for the seedlings' early growth and development. The seedlings were moved to new chambers with temperatures set to 22°C. Unfortunately, two thirds of the seedlings did not survive while the remaining plants were used for the initial analysis. Regardless, these results from the preliminary analysis open up a number of variables which can be looked at for the future experiments. The stomatal behavior is altered in response to temperature and the early developmental stages of these plants underwent heat stress, which could impact the stomatal density which in turn would affect the conductance (Samakovli at al., 2020).

Several recent attempts have been made to correlate metabolite levels with gene expression. Although there has been some success in yeast and bacterial systems (Krömer et al., 2004; Nacher et al., 2006; Seggewiß et al., 2006), it has been more difficult to prove a meaningful connection between transcript and metabolites in higher organisms (Baxter et al., 2007; Hirai et al., 2004; Howell et al., 2009; Lehmann et al., 2009; Osuna et al., 2007; Sweetlove and Fernie, 2005). The lack of strong, universal correlation between metabolite and transcript levels may be due to a lag between the inception of transcription and the accumulation of the metabolic product (Scheible et al., 2004) that has not been accurately captured by the sampling techniques and time courses utilized in these studies.

Amino acids are known to be major components of xylem and phloem sap. In 1989, Cooper and Clarkson said that the cycling of amino acids between the root and shoot via the vascular system integrate the N status of the whole plant (Cooper et al., 1989). Other studies have shown that under exogenous application of amino acids to the roots of soybean and Arabidopsis, nitrate uptake and NRT2.1 expression is repressed (Zhuo et al., 1999, Muller et al., 1992, Nazoa et al., 2003).

N metabolites have been shown to regulate the expression of genes involved in N uptake, acting as a feedback mechanism. It takes time for N metabolites to be produced in response to N and to accumulate. However, existing pools of metabolite in the cell also define the rate and amount of alteration required for the metabolite pools in response to a stimulus or perturbation (Stitt et al., 2010). When amino acids accumulate, they can act as long-distance signals communicating the N status of the plant (Liu et al., 2009, Nunes-Nesi et al., 2010). In the case of root HY5, sucrose can also act as a long-distance signal in tandem with the shoot HY5 protein

(Chen et al., 2016). However, despite this evidence for the regulation of N uptake by downstream metabolites, the underlying mechanisms of how metabolites regulate N uptake and N transporter gene expression is poorly understood (Nacry et al., 2013). Furthermore, there is conflicting evidence over whether or not phloem metabolites, such as amino acids, even regulate N uptake as it was shown previously that between N treatments, phloem amino acid levels do not change, and are not related to the regulation of N uptake (Tillard et al., 1998).

One of the major advantages of using the existing time-series transcriptome data is that it will establish a temporal dependency between metabolite levels and gene expression levels. This functional dependency can be very easily masked when comparing static metabolite levels with static transcriptome data, as it fails to capture this time delay response of the adjustment in the metabolite levels (Stitt et al., 2010). To overcome this challenge, the new model would then be used to integrate the time series metabolite data with the existing transcriptome data in Chapter 2.

The shoot time series metabolite data would be obtained from the extracted metabolite samples using GC-MS and/or LC-MS analysis. Just like the transcriptome data in Chapter 2, the shoot time series metabolite data would then be detrended before being integrated into the model. The predicted causal relationships for shoot gene-shoot metabolite and root gene-shoot metabolite interactions would then be globally analyzed using a bioinformatic pipeline similar to that in Chapter 2, and eventually experimentally validated using grafting and/or other methods.

CONCLUSION

While the results from the preliminary grafting and stomatal conductance experiments fall short of validating the CRF4 predicted interactions, they do provide a path forward for validating

predicted gene-gene interactions from Chapter 2. With the integration of metabolomic and transcriptomic data in the time series model, new gene-metabolite interactions can be identified. Together, the resulting predictions and validations can provide avenues forward for further research in N signaling and NUE, hopefully allowing us to break past the current limitations in genetic engineering so as to improve crop yields.

REFERENCES

Albert, Reka. 2005. “Scale-Free Networks in Cell Biology.” *Journal of Cell Science* 118 (21): 4947–57.

Alvarez, José M, Eleodoro Riveras, Elena A Vidal, Diana E Gras, Orlando Contreras-López, Karem P Tamayo, Felipe Aceituno, et al. 2014. “Systems Approach Identifies TGA 1 and TGA 4 Transcription Factors as Important Regulatory Components of the Nitrate Response of Arabidopsis Thaliana Roots.” *The Plant Journal* 80 (1): 1–13.

Araus, Viviana, Elena A Vidal, Tomas Puelma, Simon Alamos, Delphine Mieulet, Emmanuel Guiderdoni, and Rodrigo A Gutiérrez. 2016. “Members of BTB Gene Family Regulate Negatively Nitrate Uptake and Nitrogen Use Efficiency in Arabidopsis Thaliana and Oryza Sativa.” *Plant Physiology*, April, pp.01731.2015. https://doi.org/10.1104/pp.15.01731.

Araya, Takao, Nicolaus von Wirén, and Hideki Takahashi. 2014. “CLE Peptides Regulate Lateral Root Development in Response to Nitrogen Nutritional Status of Plants.” *Plant Signaling & Behavior* 9 (7): 2029–34.

Ashburner, Michael, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, et al. 2000. “Gene Ontology: Tool for the Unification of Biology.” *Nature Genetics* 25 (1): 25–29. https://doi.org/10.1038/75556.

Ayre Brian G. 2011. “Membrane-Transport Systems for Sucrose in Relation to Whole-Plant Carbon Partitioning.” *Molecular Plant* 4 (3): 377–94. https://doi.org/10.1093/MP.

Bahadori, Mohammad Taha, and Yan Liu. 2012. “Granger Causality Analysis in Irregular Time Series.” In *Proceedings of the 2012 SIAM International Conference on Data Mining*, 660–71.

Bailey, Timothy L., and Philip MacHanick. 2012. “Inferring Direct DNA Binding from ChIP-Seq.” *Nucleic Acids Research* 40 (17). https://doi.org/10.1093/nar/gks433.

Banerjee, Aditya, and Aryadeep Roychoudhury. 2017. “Abscisic-Acid-Dependent Basic Leucine Zipper (BZIP) Transcription Factors in Plant Abiotic Stress.” *Protoplasma* 254 (1): 3–16.

Banerjee, Sourav, Pierre Neveu, and Kenneth S Kosik. 2009. “A Coordinated Local Translational Control Point at the Synapse Involving Relief from Silencing and MOV10 Degradation.” *Neuron* 64 (6): 871–84.

Bar-Joseph, Ziv, Anthony Gitter, and Itamar Simon. 2012. “Studying and Modelling Dynamic Biological Processes Using Time-Series Gene Expression Data.” *Nature Reviews Genetics* 13 (8): 552–64.

Barabási, Albert-László. 2003. “Linked: The New Science of Networks.” American Association of Physics Teachers.

Barbaglia, Allison M., and Susanne Hoffmann-Benning. 2016. “Long-Distance Lipid Signaling and Its Role in Plant Development and Stress Response.” *Subcellular Biochemistry* 86 (March): 339–61. https://doi.org/10.1007/978-3-319-25979-6_14.

Barrero, José María, Pedro Piqueras, Miguel González-Guzmán, Ramón Serrano, Pedro L. Rodríguez, María Rosa Ponce, and José Luis Micol. 2005. "A Mutational Analysis of the ABA1 Gene of Arabidopsis Thaliana Highlights the Involvement of ABA in Vegetative Development." *Journal of Experimental Botany* 56 (418): 2071–83. https://doi.org/10.1093/jxb/eri206.

Bauer, Markus, Gunnar W Klau, and Knut Reinert. 2007. "Accurate Multiple Sequence-Structure Alignment of RNA Sequences Using Combinatorial Optimization." *BMC Bioinformatics* 8: 271. https://doi.org/10.1186/1471-2105-8-271.

Baxter, Charles J., Henning Redestig, Nicolas Schauer, Dirk Repsilber, Kiran R. Patil, Jens Nielsen, Joachim Selbig, Junli Liu, Alisdair R. Fernie, and Lee J. Sweetlove. 2007. "The Metabolic Response of Heterotrophic Arabidopsis Cells to Oxidative Stress." *Plant Physiology* 143 (1): 312–25. https://doi.org/10.1104/pp.106.090431.

Bechtold, Ulrike, Christopher A Penfold, Dafyd J Jenkins, Roxane Legaie, Jonathan D Moore, Tracy Lawson, Jack S A Matthews, et al. 2016. "Time-Series Transcriptomics Reveals That AGAMOUS-LIKE22 Affects Primary Metabolism and Developmental Processes in Drought-Stressed Arabidopsis." *The Plant Cell* 28 (2): 345–66.

Bidadi, Haniyeh, Keita Matsuoka, Kimiyo Sage-Ono, Jun Fukushima, Weerasak Pitaksaringkarn, Masashi Asahina, Shinjiro Yamaguchi, et al. 2014. "CLE6 Expression Recovers Gibberellin Deficiency to Promote Shoot Growth in Arabidopsis." *The Plant Journal* 78 (2): 241–52.

Blázquez, Miguel A., David C. Nelson, and Dolf Weijers. 2020. "Evolution of Plant Hormone

Response Pathways." *Annual Review of Plant Biology* 71 (1): 327–53. https://doi.org/10.1146/annurev-arplant-050718-100309.

Box, George E P, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. 2015. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.

Brenner, Eric D, Rainer Stahlberg, Stefano Mancuso, Jorge Vivanco, František Baluška, and Elizabeth Van Volkenburgh. 2006. "Plant Neurobiology: An Integrated View of Plant Signaling." *Trends in Plant Science* 11 (8): 413–19.

Broersen, Piet M T, and Robert Bos. 2006. "Estimating Time-Series Models from Irregularly Spaced Data." *IEEE Transactions on Instrumentation and Measurement* 55 (4): 1124–31.

Brooks, Matthew D, Jacopo Cirrone, Angelo V Pasquino, Jose M Alvarez, Joseph Swift, Shipra Mittal, Che-Lun Juang, et al. 2019. "Network Walking Charts Transcriptional Dynamics of Nitrogen Signaling by Integrating Validated and Predicted Genome-Wide Interactions." *Nature Communications* 10 (1): 1–13.

Castaings, Loren, Antonio Camargo, Delphine Pocholle, Virginie Gaudon, Yves Texier, Stéphanie Boutet-Mercey, Ludivine Taconnat, et al. 2009. "The Nodule Inception-like Protein 7 Modulates Nitrate Sensing and Metabolism in Arabidopsis." *Plant Journal* 57 (3): 426–35. https://doi.org/10.1111/j.1365-313X.2008.03695.x.

Chen, Zhen, Ping-Xia Zhao, Zi-Qing Miao, Guo-Feng Qi, Zhen Wang, Yang Yuan, Nisar Ahmad, et al. 2019. "SULTR3s Function in Chloroplast Sulfate Uptake and Affect ABA Biosynthesis

and the Stress Response." *Plant Physiology* 180 (1): 593–604.

Chen, Xiangbin, Qinfang Yao, Xiuhua Gao, Caifu Jiang, Nicholas P Harberd, and Xiangdong Fu. 2016. "Shoot-to-Root Mobile Transcription Factor HY5 Coordinates Plant Carbon and Nitrogen Acquisition." *Current Biology* 26 (5): 640–46.

Chen, Jian, Biao Zeng, Mei Zhang, Shaojun Xie, Gaokui Wang, Andrew Hauck, and Jinsheng Lai. 2014. "Dynamic Transcriptome Landscape of Maize Embryo and Endosperm Development." *Plant Physiology* 166 (1): 252–64.

Choi, Won-Gyu, Gad Miller, Ian Wallace, Jeffrey Harper, Ron Mittler, and Simon Gilroy. 2017. "Orchestrating Rapid Long-Distance Signaling in Plants with Ca $^{2+}$, ROS and Electrical Signals." *The Plant Journal* 90 (4): 698–707. https://doi.org/10.1111/tpj.13492.

Colón, Amy Marshall, Neelanjan Sengupta, David Rhodes, Natalia Dudareva, and John Morgan. 2010. "A Kinetic Model Describes Metabolic Response to Perturbations and Distribution of Flux Control in the Benzenoid Network of Petunia Hybrida." *The Plant Journal* 62 (1): 64–76.

Combier, Jean-Philippe, Helge Küster, Etienne-Pascal Journet, Natalija Hohnjec, Pascal Gamas, and Andreas Niebel. 2008. "Evidence for the Involvement in Nodulation of the Two Small Putative Regulatory Peptide-Encoding Genes MtRALFL1 and MtDVL1." *Molecular Plant-Microbe Interactions* 21 (8): 1118–27.

Consortium, Gene Ontology. 2019. "The Gene Ontology Resource: 20 Years and Still GOing

Strong.” *Nucleic Acids Research* 47 (D1): D330--D338.

Cooper, H D, and D T Clarkson. 1989. “Cycling of Amino-Nitrogen and Other Nutrients between Shoots and Roots in Cereals - a Possible Mechanism Integrating Shoot and Root in the Regulation of Nutrient Uptake.” *Journal of Experimental Botany* 40 (216): 753–62. https://doi.org/10.1093/jxb/40.7.753.

Davison, Anthony Christopher, and David Victor Hinkley. 1997. *Bootstrap Methods and Their Application*. Vol. 1. Cambridge university press.

Dinkeloo, Kasia, Shelton Boyd, and Guillaume Pilot. 2018. “Update on Amino Acid Transporter Functions and on Possible Amino Acid Sensing Mechanisms in Plants.” *Seminars in Cell and Developmental Biology*. Elsevier Ltd. https://doi.org/10.1016/j.semcdb.2017.07.010.

Doidy, Joan, Ying Li, Benjamin Neymotin, Molly B Edwards, Kranthi Varala, David Gresham, and Gloria M Coruzzi. 2016. “``Hit-and-Run'' Transcription: De Novo Transcription Initiated by a Transient BZIP1 ``hit'' Persists after the ``run''.” *BMC Genomics* 17 (1): 92.

Dovale, Júlio César, Rodrigo Oliveira Delima, and Roberto Fritsche-Neto. 2012. “Breeding for Nitrogen Use Efficiency.” In *Plant Breeding for Abiotic Stress Tolerance*, 9783642305:53–65. Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30553-5_4.

Droujinine, Ilia A, Dong Yan, and Norbert Perrimon. 2014. “A Sharp End to Sugary Wingless Travels.” *Journal of Cell Biology* 206 (7): 819–21.

Eckner, Andreas. 2014. “A Framework for the Analysis of Unevenly Spaced Time Series Data.” *Preprint*.

Erdogan, Emre, Sheng Ma, Alina Beygelzimer, and Irina Rish. 2005. “Statistical Models for Unequally Spaced Time Series.” In *Proceedings of the 2005 SIAM International Conference on Data Mining*, 626–30.

Eyheramendy, Susana, Felipe Elorrieta, and Wilfredo Palma. 2018. “An Irregular Discrete Time Series Model to Identify Residuals with Autocorrelation in Astronomical Light Curves.” *Monthly Notices of the Royal Astronomical Society* 481 (4): 4311–22.

Fischer, R.A., Derek R. Byerlee, and Gregory O. Edmeades. 2009. “Can Technology Deliver on the Yield Challenge to 2050?” https://doi.org/10.22004/AG.ECON.55481.

Forde, Brian G. 2002. “The Role of Long-distance Signalling in Plant Responses to Nitrate and Other Nutrients.” *Journal of Experimental Botany* 53 (366): 39–43. https://doi.org/10.1093/JEXBOT/53.366.39.

Friedrich, Joachim, Thomas Dandekar, Matthias Wolf, and Tobias Müller. 2005. “ProfDist: A Tool for the Construction of Large Phylogenetic Trees Based on Profile Distances.” *Bionformatics* 21 (9): 2108–9. https://doi.org/10.1093/bioinformatics/bti289.

Fujiki, Yuki, Hiromitsu Teshima, Shinji Kashiwao, Miyuki Kawano-Kawada, Yoshinori Ohsumi, Yoshimi Kakinuma, and Takayuki Sekito. 2017. “Functional Identification of AtAVT3, a Family of Vacuolar Amino Acid Transporters, in Arabidopsis.” *FEBS Letters* 591 (1): 5–15.

Fujimoto, Susan Y, Masaru Ohta, Akemi Usui, Hideaki Shinshi, and Masaru Ohme-Takagi. 2000. "Arabidopsis Ethylene-Responsive Element Binding Factors Act as Transcriptional Activators or Repressors of GCC Box--Mediated Gene Expression." *The Plant Cell* 12 (3): 393–404.

Fujita, Yasunari, Miki Fujita, Kazuo Shinozaki, and Kazuko Yamaguchi-Shinozaki. 2011. "ABA-Mediated Transcriptional Regulation in Response to Osmotic Stress in Plants." *Journal of Plant Research* 124 (4): 509–25.

Gargouri, Mahmoud, Jeong-Jin Park, F Omar Holguin, Min-Jeong Kim, Hongxia Wang, Rahul R Deshpande, Yair Shachar-Hill, Leslie M Hicks, and David R Gang. 2015. "Identification of Regulatory Network Hubs That Control Lipid Metabolism in Chlamydomonas Reinhardtii." *Journal of Experimental Botany* 66 (15): 4551–66.

Gascuel, O. 1997. "BIONJ: An Improved Version of the NJ Algorithm Based on a Simple Model of Sequence Data." *Mol Biol Evol* 14 (7): 685–95.

Gaudinier, Allison, Joel Rodriguez-Medina, Lifang Zhang, Andrew Olson, Christophe Liseron-Monfils, Anne-Maarit Bågman, Jessica Foret, et al. 2018. "Transcriptional Regulation of Nitrogen-Associated Metabolism and Growth." *Nature* 563 (7730): 259–64.

George, Eckhard, Horst Marschner, and Iver Jakobsen. 1995. "Role of Arbuscular Mycorrhizal Fungi in Uptake of Phosphorus and Nitrogen from Soil." *Critical Reviews in Biotechnology* 15 (3–4): 257–70.

Gerlach, Daniel, Matthias Wolf, Thomas Dandekar, Tobias Müller, Andreas Pokorny, and Sven Rahmann. 2007. “Deep Metazoan Phylogeny.” *In Silico Biol* 7 (2): 151–54.

Glass, Anthony D. M. 2003. “Nitrogen Use Efficiency of Crop Plants: Physiological Constraints upon Nitrogen Absorption.” *Critical Reviews in Plant Sciences* 22 (5): 453–70. https://doi.org/10.1080/07352680390243512.

Goel, Parul, Monika Bhuria, Mamta Kaushal, and Anil Kumar Singh. 2016. “Carbon: Nitrogen Interaction Regulates Expression of Genes Involved in N-Uptake and Assimilation in Brassica Juncea L.” *PloS One* 11 (9).

Good, Allen G., and Perrin H. Beatty. 2011. “Fertilizing Nature: A Tragedy of Excess in the Commons.” *PLoS Biology* 9 (8). https://doi.org/10.1371/journal.pbio.1001124.

Grajales, Alejandro, Catalina Aguilar, and Juan A Sánchez. 2007. “Phylogenetic Reconstruction Using Secondary Structures of Internal Transcribed Spacer 2 (ITS2, RDNA): Finding the Molecular and Morphological Gap in Caribbean Gorgonian Corals.” *BMC Evol Biol* 7: 90. https://doi.org/10.1186/1471-2148-7-90.

Granger, Clive W J. 1969. “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods.” *Econometrica: Journal of the Econometric Society*, 424–38.

Grillet, Louis, Ping Lan, Wenfeng Li, Girish Mokkapati, and Wolfgang Schmidt. 2018. “IRON MAN Is a Ubiquitous Family of Peptides That Control Iron Transport in Plants.” *Nature Plants* 4 (11): 953–63.

Guan, Daogang, Bin Yan, Christoph Thieme, Jingmin Hua, Hailong Zhu, Kenneth R Boheler, Zhongying Zhao, Friedrich Kragler, Yiji Xia, and Shoudong Zhang. 2016. “PlaMoM: A Comprehensive Database Compiles Plant Mobile Macromolecules.” *Nucleic Acids Research*, gkw988.

Guan, Peizhu. 2017. “Dancing with Hormones: A Current Perspective of Nitrate Signaling and Regulation in Arabidopsis.” *Frontiers in Plant Science* 8: 1697.

Ham, Byung-Kook, Jeri L Brandom, Beatriz Xoconostle-Cázares, Vanessa Ringgold, Tony J Lough, and William J Lucas. 2009. “A Polypyrimidine Tract Binding Protein, Pumpkin RBP50, Forms the Basis of a Phloem-Mobile Ribonucleoprotein Complex.” *The Plant Cell* 21 (1): 197–215.

Ham, Byung-Kook, and William J. Lucas. 2017. “Phloem-Mobile RNAs as Systemic Signaling Agents.” *Annual Review of Plant Biology* 68 (1): 173–95. https://doi.org/10.1146/annurev-arplant-042916-041139.

Hamilton, James D. 1994. *Time Series Analysis*. Vol. 2. Princeton New Jersey.

Han, Mei, Mamoru Okamoto, Perrin H. Beatty, Steven J. Rothstein, and Allen G. Good. 2015. “The Genetics of Nitrogen Use Efficiency in Crop Plants.” *Annual Review of Genetics* 49 (1): 269–89. https://doi.org/10.1146/annurev-genet-112414-055037.

Han, Yong-Liang, Hai-Xing Song, Qiong Liao, Yin Yu, Shao-Fen Jian, Joe Eugene Lepo, Qiang Liu, et al. 2016. “Nitrogen Use Efficiency Is Mediated by Vacuolar Nitrate Sequestration Capacity in Roots of Brassica Napus.” *Plant Physiology* 170 (3): 1684–98.

Heerah, Sachin, Roberto Molinari, Stéphane Guerrier, and Amy Marshall-Colon. 2020. “Granger-Causal Testing for Irregularly Sampled Time Series with Application to Nitrogen Signaling in Arabidopsis.” *BioRxiv*, January, 2020.06.15.152819. https://doi.org/10.1101/2020.06.15.152819.

Heffer, Patrick, and Michel Prud’homme. 2016. “Global Nitrogen Fertiliser Demand and Supply: Trend, Current Level and Outlook Figure 1: Evolution of Global Fertilizer Consumption by Nutrient (Tg N+P2O5+K2O) (IFA, 2016a).” www.ini2016.com.

Hentze, Matthias W, and Patrick Argos. 1991. “Homology between IRE-BP, a Regulatory RNA-Binding Protein, Aconitase, and Isopropylmalate Isomerase.” *Nucleic Acids Research* 19 (8): 1739–40.

Himmelbach, Axel, Yi Yang, and Erwin Grill. 2003. “Relay and Control of Abscisic Acid Signaling.” *Current Opinion in Plant Biology* 6 (5): 470–79.

Hirai, M. Y., M. Yano, D. B. Goodenowe, S. Kanaya, T. Kimura, M. Awazuhara, M. Arita, T. Fujiwara, and K. Saito. 2004. “From The Cover: Integration of Transcriptomics and Metabolomics for Understanding of Global Responses to Nutritional Stresses in Arabidopsis Thaliana.” *Proceedings of the National Academy of Sciences* 101 (27): 10205–10. https://doi.org/10.1073/pnas.0403218101.

Howell, Katharine A., Reena Narsai, Adam Carroll, Aneta Ivanova, Marc Lohse, Björn Usadel, A. Harvey Millar, and James Whelan. 2009. “Mapping Metabolic and Transcript Temporal Switches during Germination in Rice Highlights Specific Transcription Factors and the Role of RNA Instability in the Germination Process.” *Plant Physiology* 149 (2): 961–80. https://doi.org/10.1104/pp.108.129874.

Hudelot, Cendrine, Vivek Gowri-Shankar, Howsun Jow, Magnus Rattray, and Paul G Higgs. 2003. “RNA-Based Phylogenetic Methods: Application to Mammalian Mitochondrial RNA Sequences.” *Molecular Phylogenetics and Evolution* 28 (2): 241–52. https://doi.org/10.1016/S1055-7903(03)00061-7.

Hudson, Matt. 2005. “Elefinder.”

Jow, H, C Hudelot, M Rattray, and P G Higgs. 2002. “Bayesian Phylogenetics Using an RNA Substitution Model Applied to Early Mammalian Evolution.” *Mol Biol Evol* 19 (9): 1591–1601.

Jukes, T, and C R Cantor. 1969. “Evolution of Protein Molecules.” In *Mammalian Protein Metabolism*, edited by H Munro, 21–132. New York, USA: Academic Press.

Kant, Surya. 2018. “Understanding Nitrate Uptake, Signaling and Remobilisation for Improving Plant Nitrogen Use Efficiency.” *Seminars in Cell and Developmental Biology*. Elsevier Ltd. https://doi.org/10.1016/j.semcdb.2017.08.034.

Kant, Surya, Yong-Mei Bi, and Steven J. Rothstein. 2011. “Understanding Plant Response to Nitrogen Limitation for the Improvement of Crop Nitrogen Use Efficiency.” *Journal of Experimental Botany* 62 (4): 1499–1509. https://doi.org/10.1093/jxb/erq297.

Kataoka, Tatsuhiko, Naomi Hayashi, Tomoyuki Yamaya, and Hideki Takahashi. 2004. “Root-to-Shoot Transport of Sulfate in Arabidopsis. Evidence for the Role of SULTR3; 5 as a Component of Low-Affinity Sulfate Transport System in the Root Vasculature.” *Plant Physiology* 136 (4): 4198–4204.

Kehr, Julia, and Friedrich Kragler. 2018. “Long Distance RNA Movement.” *New Phytologist* 218 (1): 29–40.

Kiba, Takatoshi, Toru Kudo, Mikiko Kojima, and Hitoshi Sakakibara. 2011. “Hormonal Control of Nitrogen Acquisition: Roles of Auxin, Abscisic Acid, and Cytokinin.” *Journal of Experimental Botany* 62 (4): 1399–1409. https://doi.org/10.1093/jxb/erq410.

Ko, Donghwi, and Ykä Helariutta. 2017. “Shoot–Root Communication in Flowering Plants.” *Current Biology*. Cell Press. https://doi.org/10.1016/j.cub.2017.06.054.

Konishi, Mineko, and Shuichi Yanagisawa. 2013. “Arabidopsis NIN-like Transcription Factors Have a Central Role in Nitrate Signalling.” *Nature Communications* 4 (1): 1–9.

Krapp, Anne, Laure C David, Camille Chardin, Thomas Girin, Anne Marmagne, Anne-Sophie Leprince, Sylvain Chaillou, Sylvie Ferrario-Méry, Christian Meyer, and Françoise Daniel-

Vedele. 2014. “Nitrate Transport and Signalling in Arabidopsis.” *Journal of Experimental Botany* 65 (3): 789–98.

Krömer, Jens Olaf, Oliver Sorgenfrei, Kai Klopprogge, Elmar Heinzle, and Christoph Wittmann. 2004. “In-Depth Profiling of Lysine-Producing Corynebacterium Glutamicum by Combined Analysis of the Transcriptome, Metabolome, and Fluxome.” *Journal of Bacteriology* 186 (6): 1769–84. https://doi.org/10.1128/JB.186.6.1769-1784.2004.

Krouk, Gabriel, Piotr Mirowski, Yann LeCun, Dennis E Shasha, and Gloria M Coruzzi. 2010. “Predictive Network Modeling of the High-Resolution Dynamic Plant Transcriptome in Response to Nitrate.” *Genome Biology* 11 (12): R123.

Lappartient, Anne G, J John Vidmar, Thomas Leustek, Anthony D M Glass, and Bruno Touraine. 1999. “Inter-Organ Signaling in Plants: Regulation of ATP Sulfurylase and Sulfate Transporter Genes Expression in Roots Mediated by Phloem-Translocated Compound.” *The Plant Journal* 18 (1): 89–95.

Larue, Clayton T, Jiangqi Wen, and John C Walker. 2010. “Interactions between a NAC-Domain Transcription Factor and the Putative Small Protein Encoding DVL/ROT Gene Family.” *Plant Molecular Biology Reporter* 28 (1): 162.

Lechner, E., N. Leonhardt, H. Eisler, Y. Parmentier, M. Alioua, H. Jacquet, J. Leung, and P. Genschik. 2011. “MATH/BTB CRL3 Receptors Target the Homeodomain-Leucine Zipper ATHB6 to Modulate Abscisic Acid Signaling.” *Developmental Cell* 21 (6): 1116–28. https://doi.org/10.1016/j.devcel.2011.10.018.

Lehmann, Martin, Markus Schwarzländer, Toshihiro Obata, Supaart Sirikantaramas, Meike Burow, Carl Erik Olsen, Takayuki Tohge, et al. 2009. “The Metabolic Response of Arabidopsis Roots to Oxidative Stress Is Distinct from That of Heterotrophic Cells in Culture and Highlights a Complex Relationship between the Levels of Transcripts, Metabolites, and Flux.” *Molecular Plant* 2 (3): 390–406. https://doi.org/10.1093/mp/ssn080.

Leise, Tanya L. 2017. “Analysis of Nonstationary Time Series for Biological Rhythms Research.” *Journal of Biological Rhythms* 32 (3): 187–94. https://doi.org/10.1177/0748730417709105.

Léran, Sophie, Kai H. Edel, Marjorie Pervent, Kenji Hashimoto, Claire Corratgé-Faillie, Jan Niklas Offenborn, Pascal Tillard, Alain Gojon, Jörg Kudla, and Benoît Lacombe. 2015. “Nitrate Sensing and Uptake in Arabidopsis Are Enhanced by ABI2, a Phosphatase Inactivated by the Stress Hormone Abscisic Acid.” *Science Signaling* 8 (375): ra43–ra43. https://doi.org/10.1126/scisignal.aaa4829.

Leustek, Thomas, and Kazuki Saito. 1999. “Sulfate Transport and Assimilation in Plants.” *Plant Physiology* 120 (3): 637–44.

Li, Chun-jian, Xiao-ping Zhu, and Fu-suo Zhang. 2000. “Role of Shoot in Regulation of Iron Deficiency Responses in Cucumber and Bean Plants.” *Journal of Plant Nutrition* 23 (11–12): 1809–18.

Liang, Ying Shi, Netty Ermawati, Joon-Yung Cha, Min Hee Jung, Mukhamad Su’udi, Min Gab Kim, Sun-Hwa Ha, Chung-Gyoo Park, and Daeyoung Son. 2010. “Overexpression of an

AP2/ERF-Type Transcription Factor CRF5 Confers Pathogen Resistance to Arabidopsis Plants." *Journal of the Korean Society for Applied Biological Chemistry* 53 (2): 142–48.

Lisec, Jan, Nicolas Schauer, Joachim Kopka, Lothar Willmitzer, and Alisdair R. Fernie. 2006. "Gas Chromatography Mass Spectrometry-Based Metabolite Profiling in Plants." *Nature Protocols* 1 (1): 387–96. https://doi.org/10.1038/nprot.2006.59.

Liu, Qian, Xiangbin Chen, Kun Wu, and Xiangdong Fu. 2015. "Nitrogen Signaling and Use Efficiency in Plants: What's New?" *Current Opinion in Plant Biology*. Elsevier Ltd. https://doi.org/10.1016/j.pbi.2015.08.002.

Liu, Tzu-Yin, Chiung-Yun Chang, and Tzyy-Jen Chiou. 2009. "The Long-Distance Signaling of Mineral Macronutrients." *Current Opinion in Plant Biology* 12 (3): 312–19. https://doi.org/10.1016/j.pbi.2009.04.004.

Lock, Yee Ying. 2011. "Engineering Nitrogen Use Efficiency in Oryza Sativa by the Developmental Over-Expression of Barley Alanine Aminotransferase Using a Novel Rice Promoter." https://doi.org/10.7939/R3PK5T.

Lough, Tony J, and William J Lucas. 2006. "INTEGRATIVE PLANT BIOLOGY: Role of Phloem Long-Distance Macromolecular Trafficking." *Annual Review of Plant Biology* 57 (1): 203–32. https://doi.org/10.1146/annurev.arplant.56.032604.144145.

Luna, Estrella, Toby J.A. Bruce, Michael R Roberts, Victor Flors, and Jurriaan Ton. 2012.

"Next-Generation Systemic Acquired Resistance." *Plant Physiology* 158 (2): 844–53. https://doi.org/10.1104/pp.111.187468.

Luo, Kai-Ren, Nien-Chen Huang, and Tien-Shin Yu. 2018. "Selective Targeting of Mobile MRNAs to Plasmodesmata for Cell-to-Cell Movement." *Plant Physiology* 177 (2): 604–14. https://doi.org/10.1104/pp.18.00107.

Maierhofer, Tobias, Christof Lind, Stefanie Hüttl, Sönke Scherzer, Melanie Papenfuß, Judy Simon, Khaled A S Al-Rasheid, et al. 2014. "A Single-Pore Residue Renders the Arabidopsis Root Anion Channel SLAH2 Highly Nitrate Selective." *The Plant Cell* 26 (6): 2554–67.

Maller, Ross A, Gernot Müller, Alex Szimayer, and others. 2008. "GARCH Modelling in Continuous Time for Irregularly Spaced Time Series Data." *Bernoulli* 14 (2): 519–42.

Mangan, Shmoolik, and Uri Alon. 2003. "Structure and Function of the Feed-Forward Loop Network Motif." *Proceedings of the National Academy of Sciences* 100 (21): 11980–85.

Marín-González, Esther, and Paula Suárez-López. 2012. "'And yet It Moves': Cell-to-Cell and Long-Distance Signaling by Plant MicroRNAs." *Plant Science*. Elsevier. https://doi.org/10.1016/j.plantsci.2012.07.009.

Marondedze, Claudius, Ludivine Thomas, Natalia L Serrano, Kathryn S Lilley, and Chris Gehring. 2016. "The RNA-Binding Protein Repertoire of *Arabidopsis thaliana*." *Scientific Reports* 6 (1): 1–13.

Marsch-Martínez, Nayelli, John Franken, Karla L. Gonzalez-Aguilera, Stefan de Folter, Gerco Angenent, and Elena R. Alvarez-Buylla. 2013. "An Efficient Flat-Surface Collar-Free Grafting Method for *Arabidopsis thaliana* Seedlings." *Plant Methods* 9 (1): 14. https://doi.org/10.1186/1746-4811-9-14.

Marschner, Horst. 2011. *Marschner's Mineral Nutrition of Higher Plants*. Academic press.

Matimati, Ignatious, G Anthony Verboom, and Michael D Cramer. 2014. "Nitrogen Regulation of Transpiration Controls Mass-Flow Acquisition of Nutrients." *Journal of Experimental Botany* 65 (1): 159–68.

Matson, Pamela A., Rosamond Naylor, and Ivan Ortiz-Monasterio. 1998. "Integration of Environmental, Agronomic, and Economic Aspects of Fertilizer Management." *Science* 280 (5360): 112–15. https://doi.org/10.1126/science.280.5360.112.

Matsubayashi, Yoshikatsu, Mari Ogawa, Hitomi Kihara, Masaaki Niwa, and Youji Sakagami. 2006. "Disruption and Overexpression of Arabidopsis Phytosulfokine Receptor Gene Affects Cellular Longevity and Potential for Growth." *Plant Physiology* 142 (1): 45–53.

McAllister, Chandra H, Perrin H Beatty, and Allen G Good. 2012. "Engineering Nitrogen Use Efficient Crop Plants: The Current Status." *Plant Biotechnology Journal* 10 (9): 1011–25.

Melaragno, Jerry E, Bharati Mehrotra, and Annette W Coleman. 1993. "Relationship between Endopolyploidy and Cell Size in Epidermal Tissue of Arabidopsis." *The Plant Cell* 5 (11): 1661–68.

Mi, Huaiyu, Anushya Muruganujan, Dustin Ebert, Xiaosong Huang, and Paul D Thomas. 2019. "PANTHER Version 14: More Genomes, a New PANTHER GO-Slim and Improvements in Enrichment Analysis Tools." *Nucleic Acids Research* 47 (D1): D419--D426.

Moll, R. H., E. J. Kamprath, and W. A. Jackson. 1982. "Analysis and Interpretation of Factors Which Contribute to Efficiency of Nitrogen Utilization 1." *Agronomy Journal* 74 (3): 562–64. https://doi.org/10.2134/agronj1982.00021962007400030037x.

Muller, Bertrand, and Bruno Touraine. 1992. "Inhibition of NO3 Uptake by Various Phloem-Translocated Amino Acids in Soybean Seedlings." *Journal of Experimental Botany* 43 (5): 617–23. https://doi.org/10.1093/jxb/43.5.617.

Müller, Tobias, Sven Rahmann, Thomas Dandekar, and Matthias Wolf. 2004. "Accurate and Robust Phylogeny Estimation Based on Profile Distances: A Study of the Chlorophyceae (Chlorophyta)." *BMC Evolutionary Biology* 4 (June): 20. https://doi.org/10.1186/1471-2148-4-20.

Müller, Tobias, Rainer Spang, and Martin Vingron. 2002. "Estimating Amino Acid Substitution Models: A Comparison of Dayhoff's Estimator, the Resolvent Approach and a Maximum Likelihood Method." *Molecular Biology and Evolution* 19 (1): 8–13.

Müller, Tobias, and Vingron, Martin. 2000. "Modeling Amino Acid Replacement." *Journal of Computational Biology* 7 (6): 761–76. https://doi.org/10.1089/10665270050514918.

Nacher, J. C., J.-M. Schwartz, M. Kanehisa, and T. Akutsu. 2006. “Identification of Metabolic Units Induced by Environmental Signals.” *Bioinformatics* 22 (14): e375–83. https://doi.org/10.1093/bioinformatics/btl202.

Nacry, Philippe, Eléonore Bouguyon, and Alain Gojon. 2013. “Nitrogen Acquisition by Roots: Physiological and Developmental Mechanisms Ensuring Plant Adaptation to a Fluctuating Resource.” *Plant and Soil* 370 (1–2): 1–29. https://doi.org/10.1007/s11104-013-1645-9.

Nazoa, Patricia, J. John Vidmar, Timothy J. Tranbarger, Karine Mouline, Isabelle Damiani, Pascal Tillard, Degen Zhuo, Anthony D.M. Glass, and Bruno Touraine. 2003. “Regulation of the Nitrate Transporter Gene AtNRT2.1 in *Arabidopsis thaliana*: Responses to Nitrate, Amino Acids and Developmental Stage.” *Plant Molecular Biology* 52 (3): 689–703. https://doi.org/10.1023/A:1024899808018.

Notaguchi, Michitaka, and Satoru Okamoto. 2015. “Dynamics of Long-Distance Signaling via Plant Vascular Tissues.” *Frontiers in Plant Science* 6 (MAR): 161. https://doi.org/10.3389/fpls.2015.00161.

Nunes-Nesi, Adriano, Alisdair R. Fernie, and Mark Stitt. 2010. “Metabolic and Signaling Aspects Underpinning the Regulation of Plant Carbon Nitrogen Interactions.” *Molecular Plant* 3 (6): 973–96. https://doi.org/10.1093/mp/ssq049.

O’Malley, Ronan C., Shao Shan Carol Huang, Liang Song, Mathew G. Lewsey, Anna Bartlett, Joseph R. Nery, Mary Galli, Andrea Gallavotti, and Joseph R. Ecker. 2016. “Cistrome and

Epicistrome Features Shape the Regulatory DNA Landscape." *Cell* 165 (5): 1280–92. https://doi.org/10.1016/j.cell.2016.04.038.

Oh, Eunkyoo, Pil Joon Seo, and Jungmook Kim. 2018. "Signaling Peptides and Receptors Coordinating Plant Root Development." *Trends in Plant Science* 23 (4): 337–51.

Ohkubo, Yuri, Mina Tanaka, Ryo Tabata, Mari Ogawa-Ohnishi, and Yoshikatsu Matsubayashi. 2017. "Shoot-to-Root Mobile Polypeptides Involved in Systemic Regulation of Nitrogen Acquisition." *Nature Plants* 3 (4): 1–6. https://doi.org/10.1038/nplants.2017.29.

Okamoto, Satoru, Ryo Tabata, and Yoshikatsu Matsubayashi. 2016. "Long-Distance Peptide Signaling Essential for Nutrient Homeostasis in Plants." *Current Opinion in Plant Biology* 34: 35–40.

Osugi, Asami, Mikiko Kojima, Yumiko Takebayashi, Nanae Ueda, Takatoshi Kiba, and Hitoshi Sakakibara. 2017. "Systemic Transport of Trans-Zeatin and Its Precursor Have Differing Roles in Arabidopsis Shoots." *Nature Plants* 3 (8): 1–6. https://doi.org/10.1038/nplants.2017.112.

Osuna, Daniel, Bjorn Usadel, Rosa Morcuende, Yves Gibon, Oliver E. Bläsing, Melanie Höhne, Manuela Günter, et al. 2007. "Temporal Responses of Transcripts, Enzyme Activities and Metabolites after Adding Sucrose to Carbon-Deprived Arabidopsis Seedlings." *The Plant Journal* 49 (3): 463–91. https://doi.org/10.1111/j.1365-313X.2006.02979.x.

Palenchar, Peter M, Andrei Kouranov, Laurence V Lejay, and Gloria M Coruzzi. 2004. "Genome-Wide Patterns of Carbon and Nitrogen Regulation of Gene Expression Validate the Combined Carbon and Nitrogen (CN)-Signaling Hypothesis in Plants." *Genome Biology* 5 (11): R91.

Perchlik, Molly, and Mechthild Tegeder. 2017. "Improving Plant Nitrogen Use Efficiency through Alteration of Amino Acid Transport Processes." *Plant Physiology* 175 (1): 235–47. https://doi.org/10.1104/pp.17.00608.

Poitout, Arthur, Amandine Crabos, Ivan Pet\vr\'\ik, Ondrej Novák, Gabriel Krouk, Benoît Lacombe, and Sandrine Ruffel. 2018. "Responses to Systemic Nitrogen Signaling in Arabidopsis Roots Involve Trans-Zeatin in Shoots." *The Plant Cell* 30 (6): 1243–57.

Rahmann, Sven, Tobias Muller, Thomas Dandekar, and Matthias Wolf. 2006. "Efficient and Robust Analysis of Large Phylogenetic Datasets." In *Advanced Data Mining Technologies in Bioinformatics*, edited by Hui-Huang Hsu, 104–17. Hershey, PA, USA: Idea Group, Inc.

Ranjan, Rumesh, and Rajbir Yadav. 2019. "Targeting Nitrogen Use Efficiency for Sustained Production of Cereal Crops." *Journal of Plant Nutrition*. Taylor and Francis Inc. https://doi.org/10.1080/01904167.2019.1589497.

Rehfeld, Kira, Norbert Marwan, Jobst Heitzig, and Jürgen Kurths. 2011. "Comparison of Correlation Analysis Techniques for Irregularly Sampled Time Series." *Nonlinear Processes in Geophysics* 18 (3): 389–404.

Remondini, D, B O'connell, N Intrator, J M Sedivy, N Neretti, G C Castellani, and L N Cooper. 2005. "Targeting C-Myc-Activated Genes with a Correlation Method: Detection of Global Changes in Large Gene Expression Network Dynamics." *Proceedings of the National Academy of Sciences* 102 (19): 6902–6.

Ripley, Brian D. 2005. *Spatial Statistics*. Vol. 575. John Wiley & Sons.

Robertson, G. Philip, and Peter M. Vitousek. 2009. "Nitrogen in Agriculture: Balancing the Cost of an Essential Resource." *Annual Review of Environment and Resources* 34 (1): 97–125. https://doi.org/10.1146/annurev.environ.032108.105046.

Ruffel, Sandrine, and Alain Gojon. 2017. "Systemic Nutrient Signalling: On the Road for Nitrate." *Nature Plants*. Palgrave Macmillan Ltd. https://doi.org/10.1038/nplants.2017.40.

Ruffel, Sandrine, Gabriel Krouk, Daniela Ristova, Dennis Shasha, Kenneth D. Birnbaum, and Gloria M. Coruzzi. 2011. "Nitrogen Economics of Root Foraging: Transitive Closure of the Nitrate-Cytokinin Relay and Distinct Systemic Signaling for N Supply vs. Demand." *Proceedings of the National Academy of Sciences of the United States of America* 108 (45): 18524–29. https://doi.org/10.1073/pnas.1108684108.

Russnak, Roland, David Konczal, and Steven L. McIntire. 2001. "A Family of Yeast Proteins Mediating Bidirectional Vacuolar Amino Acid Transport." *Journal of Biological Chemistry* 276 (26): 23849–57. https://doi.org/10.1074/jbc.M008028200.

Saeed, AIa, Vasily Sharov, Joe White, Jerry Li, Wei Liang, Nirmal Bhagabati, J Braisted, et al. 2003. “TM4: A Free, Open-Source System for Microarray Data Management and Analysis.” *Biotechniques* 34 (2): 374–78.

Saito, Shunya, and Nobuyuki Uozumi. 2019. “Guard Cell Membrane Anion Transport Systems and Their Regulatory Components: An Elaborate Mechanism Controlling Stress-Induced Stomatal Closure.” *Plants* 8 (1): 9.

Saitou, N, and M Nei. 1987. “The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees.” *Molecular Biology and Evolution* 4 (4): 406–25. https://doi.org/10.1093/oxfordjournals.molbev.a040454.

Sakakibara, Hitoshi, Kentaro Takei, and Naoya Hirose. 2006. “Interactions between Nitrogen and Cytokinin in the Regulation of Metabolism and Development.” *Trends in Plant Science*. https://doi.org/10.1016/j.tplants.2006.07.004.

Sakuma, Yoh, Qiang Liu, Joseph G Dubouzet, Hiroshi Abe, Kazuo Shinozaki, and Kazuko Yamaguchi-Shinozaki. 2002. “DNA-Binding Specificity of the ERF/AP2 Domain of Arabidopsis DREBs, Transcription Factors Involved in Dehydration-and Cold-Inducible Gene Expression.” *Biochemical and Biophysical Research Communications* 290 (3): 998–1009.

Samakovli, Despina, Tereza Tichá, Tereza Vavrdová, Miroslav Ovečka, Ivan Luptovčiak, Veronika Zapletalová, Anna Kuchařová, et al. 2020. “YODA-HSP90 Module Regulates Phosphorylation-Dependent Inactivation of SPEECHLESS to Control Stomatal Development

under Acute Heat Stress in Arabidopsis." *Molecular Plant* 13 (4): 612–33. https://doi.org/10.1016/j.molp.2020.01.001.

Sato, Takeo, Shugo Maekawa, Mineko Konishi, Nozomi Yoshioka, Yuki Sasaki, Haruna Maeda, Tetsuya Ishida, Yuki Kato, Junji Yamaguchi, and Shuichi Yanagisawa. 2017. "Direct Transcriptional Activation of BT Genes by NLP Transcription Factors Is a Key Component of the Nitrate Response in Arabidopsis." *Biochemical and Biophysical Research Communications* 483 (1): 380–86.

Sauter, Angela, William J Davies, and Wolfram Hartung. 2001. "The Long-Distance Abscisic Acid Signal in the Droughted Plant: The Fate of the Hormone on Its Way from Root to Shoot." *Journal of Experimental Botany* 52 (363): 1991–97.

Scheible, W.-R. 2004. "Genome-Wide Reprogramming of Primary and Secondary Metabolism, Protein Synthesis, Cellular Growth Processes, and the Regulatory Infrastructure of Arabidopsis in Response to Nitrogen." *PLANT PHYSIOLOGY* 136 (1): 2483–99. https://doi.org/10.1104/pp.104.047019.

Schöniger, Michael, and Arndt Von Haeseler. 1994. "A Stochastic Model for the Evolution of Autocorrelated DNA Sequences." *Molecular Phylogenetics and Evolution* 3 (3): 240–47. https://doi.org/10.1006/mpev.1994.1026.

Schultz, Jörg, Stefanie Maisel, Daniel Gerlach, Tobias Müller, and Matthias Wolf. 2005. "A Common Core of Secondary Structure of the Internal Transcribed Spacer 2 (ITS2) throughout the Eukaryota." *RNA* 11 (4): 361–64. https://doi.org/10.1261/rna.7204505.

Seggewiß, Jochen, Karsten Becker, Oliver Kotte, Martin Eisenacher, Mohammad Reza Khoschkhoi Yazdi, Andreas Fischer, Peter McNamara, et al. 2006. “Reporter Metabolite Analysis of Transcriptional Profiles of a Staphylococcus Aureus Strain with Normal Phenotype and Its Isogenic HemB Mutant Displaying the Small-Colony-Variant Phenotype.” *Journal of Bacteriology* 188 (22): 7765–77. https://doi.org/10.1128/JB.00774-06.

Seibel, Philipp N, Tobias Müller, Thomas Dandekar, Jörg Schultz, and Matthias Wolf. 2006. “4SALE--a Tool for Synchronous RNA Sequence and Secondary Structure Alignment and Editing.” *BMC Bioinformatics* 7: 498. https://doi.org/10.1186/1471-2105-7-498.

Sekito, Takayuki, Yuki Fujiki, Yoshinori Ohsumi, and Yoshimi Kakinuma. 2008. “Novel Families of Vacuolar Amino Acid Transporters.” *IUBMB Life* 60 (8): 519–25.

Selig, Christian, Matthias Wolf, T. Muller, Thomas Dandekar, and Jörg Schultz. 2007. “The ITS2 Database II: Homology Modelling RNA Structure for Molecular Systematics.” *Nucleic Acids Research* 36 (Issue suppl_1): D377–80. https://doi.org/10.1093/nar/gkm827.

Selvaraj, Michael Gomez, Milton Orlando Valencia, Satoshi Ogawa, Yingzhi Lu, Liying Wu, Christopher Downs, Wayne Skinner, et al. 2017. “Development and Field Performance of Nitrogen Use Efficient Rice Lines for Africa.” *Plant Biotechnology Journal* 15 (6): 775–87. https://doi.org/10.1111/pbi.12675.

Shimshi, Daniel. 1970. “The Effect of Nitrogen Supply on Transpiration and Stomatal Behaviour of Beans (Phaseolus Vulgaris L.).” *New Phytologist* 69 (2): 405–12.

Siebert, Sven, and Rolf Backofen. 2005. “MARNA: Multiple Alignment and Consensus Structure Prediction of RNAs Based on Sequence Structure Comparisons.” *Bioinformatics* 21 (16): 3352–59. https://doi.org/10.1093/bioinformatics/bti550.

Signora, Laurent, Ive De Smet, Christine H. Foyer, and Hanma Zhang. 2001. “ABA Plays a Central Role in Mediating the Regulatory Effects of Nitrate on Root Branching in Arabidopsis.” *Plant Journal* 28 (6): 655–62. https://doi.org/10.1046/j.1365-313x.2001.01185.x.

Simons, Margaret, Rajib Saha, Lenaig Guillard, Gilles Clement, Patrick Armengaud, Rafael Cañas, Costas D Maranas, Peter J Lea, and Bertrand Hirel. 2014. “Nitrogen-Use Efficiency in Maize (Zea Mays L.): From ‘Omics’ Studies to Metabolic Modelling.” *Journal of Experimental Botany* 65 (19): 5657–71. https://doi.org/10.1093/jxb/eru227.

Sinclair, Thomas R., and Thomas W. Rufty. 2012. “Nitrogen and Water Resources Commonly Limit Crop Yield Increases, Not Necessarily Plant Genetics.” *Global Food Security*. Elsevier. https://doi.org/10.1016/j.gfs.2012.07.001.

Smith, Andrew D, Thomas W H Lui, and Elisabeth R M Tillier. 2004. “Empirical Models for Substitution in Ribosomal RNA.” *Molecular Biology and Evolution* 21 (3): 419–27. https://doi.org/10.1093/molbev/msh029.

Snyman, S. J., E. Hajari, M. P. Watt, Y. Lu, and J. C. Kridl. 2015. “Improved Nitrogen Use Efficiency in Transgenic Sugarcane: Phenotypic Assessment in a Pot Trial under Low Nitrogen Conditions.” *Plant Cell Reports*. Springer Verlag. https://doi.org/10.1007/s00299-015-1768-y.

Sobeih, Wagdy Y, Ian C Dodd, Mark A Bacon, Donald Grierson, and William J Davies. 2004. "Long-Distance Signals Regulating Stomatal Conductance and Leaf Growth in Tomato (Lycopersicon Esculentum) Plants Subjected to Partial Root-Zone Drying." *Journal of Experimental Botany* 55 (407): 2353–63.

Spellman, Paul T, Gavin Sherlock, Michael Q Zhang, Vishwanath R Iyer, Kirk Anders, Michael B Eisen, Patrick O Brown, David Botstein, and Bruce Futcher. 1998. "Comprehensive Identification of Cell Cycle--Regulated Genes of the Yeast Saccharomyces Cerevisiae by Microarray Hybridization." *Molecular Biology of the Cell* 9 (12): 3273–97.

Stahl, Andreas, Mara Pfeifer, Matthias Frisch, Benjamin Wittkop, and Rod J. Snowdon. 2017. "Recent Genetic Gains in Nitrogen Use Efficiency in Oilseed Rape." *Frontiers in Plant Science* 8 (June): 963. https://doi.org/10.3389/fpls.2017.00963.

Stitt, Mark, Ronan Sulpice, and Joost Keurentjes. 2010. "Metabolic Networks: How to Identify Key Components in the Regulation of Metabolism and Growth." *Plant Physiology* 152 (2): 428–44. https://doi.org/10.1104/pp.109.150821.

Sun, Cui-Hui, Jian-Qiang Yu, and Da-Gang Hu. 2017. "Nitrate: A Crucial Signal during Lateral Roots Development." *Frontiers in Plant Science* 8: 485.

Sweetlove, Lee J., and Alisdair R. Fernie. 2005. "Regulation of Metabolic Networks: Understanding Metabolic Complexity in the Systems Biology Era." *New Phytologist* 168 (1): 9–24. https://doi.org/10.1111/j.1469-8137.2005.01513.x.

Tabata, Ryo, Kumiko Sumida, Tomoaki Yoshii, Kentaro Ohyama, Hidefumi Shinohara, and Yoshikatsu Matsubayashi. 2014. “Perception of Root-Derived Peptides by Shoot LRR-RKs Mediates Systemic N-Demand Signaling.” *Science* 346 (6207): 343–46.

Takahashi, Fuminori, Kousuke Hanada, Takayuki Kondo, and Kazuo Shinozaki. 2019. “Hormone-like Peptides and Small Coding Genes in Plant Stress Signaling and Development.” *Current Opinion in Plant Biology*. Elsevier Ltd. https://doi.org/10.1016/j.pbi.2019.05.011.

Takei, Kentaro, Nanae Ueda, Koh Aoki, Takashi Kuromori, Takashi Hirayama, Kazuo Shinozaki, Tomoyuki Yamaya, and Hitoshi Sakakibara. 2004. “AtIPT3 Is a Key Determinant of Nitrate-Dependent Cytokinin Biosynthesis in Arabidopsis.” *Plant and Cell Physiology* 45 (8): 1053–62. https://doi.org/10.1093/PCP/PCH119.

Thiebaut, Carole, and Sylvie Roques. 2005. “Time-Scale and Time-Frequency Analyses of Irregularly Sampled Astronomical Time Series.” *EURASIP Journal on Advances in Signal Processing* 2005 (15): 852587.

Thieme, Christoph J., Monica Rojas-Triana, Ewelina Stecyk, Christian Schudoma, Wenna Zhang, Lei Yang, Miguel Miñambres, et al. 2015. “Endogenous Arabidopsis Messenger RNAs Transported to Distant Tissues.” *Nature Plants* 1 (4): 15025. https://doi.org/10.1038/nplants.2015.25.

Thompson, Matthew V. 2006. “Phloem: The Long and the Short of It.” *Trends in Plant Science* 11 (1): 26–32. https://doi.org/10.1016/j.tplants.2005.11.009.

Tillard, Pascal, Lucien Passama, and Alain Gojon. 1998. “Are Phloem Amino Acids Involved in the Shoot to Root Control of NO3/-Uptake in Ricinus Communis Plants?” *Journal of Experimental Botany* 49 (325): 1371–79. https://doi.org/10.1093/jxb/49.325.1371.

Titterton, David, John L Weston, and John Weston. 2004. *Strapdown Inertial Navigation Technology*. Vol. 17. IET.

To, Jennifer P C, Jean Deruère, Bridey B Maxwell, Veronica F Morris, Claire E Hutchison, Fernando J Ferreira, G Eric Schaller, and Joseph J Kieber. 2007. “Cytokinin Regulates Type-A Arabidopsis Response Regulator Activity and Protein Stability via Two-Component Phosphorelay.” *The Plant Cell* 19 (12): 3901–14.

To, Jennifer P C, Georg Haberer, Fernando J Ferreira, Jean Deruere, Michael G Mason, G Eric Schaller, Jose M Alonso, Joseph R Ecker, and Joseph J Kieber. 2004. “Type-A Arabidopsis Response Regulators Are Partially Redundant Negative Regulators of Cytokinin Signaling.” *The Plant Cell* 16 (3): 658–71.

Tone, Junichi, Ayumi Yoshimura, Kunio Manabe, Nami Murao, Takayuki Sekito, Miyuki Kawano-Kawada, and Yoshimi Kakinuma. 2015. “Characterization of Avt1p as a Vacuolar Proton/Amino Acid Antiporter in Saccharomyces Cerevisiae.” *Bioscience, Biotechnology and Biochemistry* 79 (5): 782–89. https://doi.org/10.1080/09168451.2014.998621.

Trolltech. 2008. “Http://Trolltech.Com/Products/Qt/.” http://trolltech.com/products/qt/.

Brûle, Sybille van den, and Cheryl C Smart. 2002. “The Plant PDR Family of ABC Transporters.” *Planta* 216 (1): 95–106.

Vanormelingen, P, E Hegewald, A Braband, M Kitschke, T Friedl, K Sabbe, and W Vyverman. 2007. “The Systematics of a Small Spineless Desmodesmus Species, D-Costato-Granulatus (Sphaeropleales, Chlorophyceae), Based on ITS2 RDNA Sequence Analyses and Cell Wall Morphology.” *Journal of Phyclology* 43 (2): 378–96.

Varala, Kranthi, Amy Marshall-Colón, Jacopo Cirrone, Matthew D Brooks, Angelo V Pasquino, Sophie Léran, Shipra Mittal, et al. 2018. “Temporal Transcriptional Logic of Dynamic Regulatory Networks Underlying Nitrogen Signaling and Use in Plants.” *Proceedings of the National Academy of Sciences* 115 (25): 6494–99.

Verrier, Paul J., David Bird, Bo Burla, Elie Dassa, Cyrille Forestier, Markus Geisler, Markus Klein, et al. 2008. “Plant ABC Proteins - a Unified Nomenclature and Updated Inventory.” *Trends in Plant Science* 13 (4): 151–59. https://doi.org/10.1016/j.tplants.2008.02.001.

Vidal, Elena A., José M. Álvarez, Tomás C. Moyano, and Rodrigo A. Gutiérrez. 2015. “Transcriptional Networks in the Nitrate Response of Arabidopsis Thaliana.” *Current Opinion in Plant Biology*. Elsevier Ltd. https://doi.org/10.1016/j.pbi.2015.06.010.

Vidal, Elena A, José M Álvarez, and Rodrigo A Gutiérrez. 2014. “Nitrate Regulation of AFB3 and NAC4 Gene Expression in Arabidopsis Roots Depends on NRT1. 1 Nitrate Transport Function.” *Plant Signaling & Behavior* 9 (6): e28501.

Vidal, Elena A, Tomás C Moyano, Gabriel Krouk, Manpreet S Katari, Milos Tanurdzic, W Richard McCombie, Gloria M Coruzzi, and Rodrigo A Gutiérrez. 2013a. “Integrated RNA-Seq and SRNA-Seq Analysis Identifies Novel Nitrate-Responsive Genes in Arabidopsis Thaliana Roots.” *BMC Genomics* 14 (1): 701.

Vidal, Elena A., Tomás C. Moyano, Eleodoro Riveras, Orlando Contreras-López, and Rodrigo A. Gutiérrez. 2013b. “Systems Approaches Map Regulatory Networks Downstream of the Auxin Receptor AFB3 in the Nitrate Response of Arabidopsis Thaliana Roots.” *Proceedings of the National Academy of Sciences* 110 (31): 12840–45. https://doi.org/10.1073/pnas.1310937110.

Vidal, Elena A, Viviana Araus, Cheng Lu, Geraint Parry, Pamela J Green, Gloria M Coruzzi, and Rodrigo A Gutiérrez. 2010. “Nitrate-Responsive MiR393/AFB3 Regulatory Module Controls Root System Architecture in Arabidopsis Thaliana.” *Proceedings of the National Academy of Sciences* 107 (9): 4477–82.

Vidal, Elena A, and Rodrigo A Gutiérrez. 2008. “A Systems View of Nitrogen Nutrient and Metabolite Responses in Arabidopsis.” *Current Opinion in Plant Biology* 11 (5): 521–29. https://doi.org/10.1016/J.PBI.2008.07.003.

Walch-Liu, P, S Filleur, Y Gan, and B G Forde. 2005. “Signaling Mechanisms Integrating Root and Shoot Responses to Changes in the Nitrogen Supply.” *Photosynthesis Research* 83 (2): 239–50. https://doi.org/10.1007/s11120-004-2080-9.

Wang, Guodong, Guohua Zhang, and Mengyao Wu. 2016. “CLE Peptide Signaling and Crosstalk with Phytohormones and Environmental Stimuli.” *Frontiers in Plant Science* 6: 1211.

Wang, R., Mamoru Okamoto, Xiujuan Xing, Nigel M Crawford, Daniel Osuna, Natalia Palacios-Rojas, Dana Schindelasch, Oliver Thimm, Michael K. Udvardi, and Mark Stitt. 2003. “Microarray Analysis of the Nitrate Response in Arabidopsis Roots and Shoots Reveals over 1,000 Rapidly Responding Genes and New Linkages to Glucose, Trehalose-6-Phosphate, Iron, and Sulfate Metabolism.” *PLANT PHYSIOLOGY* 132 (2): 556–67. https://doi.org/10.1104/pp.103.021253.

Wang, Rongchen, Karl Guegler, Samuel T LaBrie, and Nigel M Crawford. 2000. “Genomic Analysis of a Nutrient Response in Arabidopsis Reveals Diverse Expression Patterns and Novel Metabolic and Potential Regulatory Genes Induced by Nitrate.” *The Plant Cell* 12 (8): 1491–1509.

Weinl, Christina, Sebastian Marquardt, Suzanne J H Kuijt, Moritz K Nowack, Marc J Jakoby, Martin Hülskamp, and Arp Schnittger. 2005. “Novel Functions of Plant Cyclin-Dependent Kinase Inhibitors, ICK1/KRP1, Can Act Non-Cell-Autonomously and Inhibit Entry into Mitosis.” *The Plant Cell* 17 (6): 1704–22.

Wen, Jiangqi, Kevin A. Lease, and John C. Walker. 2004. “DVL, a Novel Class of Small Polypeptides: Overexpression Alters Arabidopsis Development.” *Plant Journal* 37 (5): 668–77. https://doi.org/10.1111/j.1365-313X.2003.01994.x.

Wheeler, D L, C Chappey, A E Lash, D D Leipe, T L Madden, G D Schuler, T A Tatusova, and B A Rapp. 2000. “Database Resources of the National Center for Biotechnology Information.” *Nucleic Acids Research* 28 (1): 10–14. https://doi.org/10.1093/nar/28.1.10.

Wolf, Matthias, Marco Achtziger, Jörg Schultz, Thomas Dandekar, and Tobias Müller. 2005. “Homology Modeling Revealed More than 20,000 RRNA Internal Transcribed Spacer 2 (ITS2) Secondary Structures.” *RNA* 11 (11): 1616–23. https://doi.org/10.1261/rna.2144205.

Wu, Zhaohua, Norden E. Huang, Steven R. Long, and Chung Kang Peng. 2007. “On the Trend, Detrending, and Variability of Nonlinear and Nonstationary Time Series.” *Proceedings of the National Academy of Sciences of the United States of America* 104 (38): 14889–94. https://doi.org/10.1073/pnas.0701020104.

Xin, Wei, Lina Zhang, Wenzhong Zhang, Jiping Gao, Jun Yi, Xiaoxi Zhen, Ziang Li, Ying Zhao, Chengcheng Peng, and Chen Zhao. 2019. “An Integrated Analysis of the Rice Transcriptome and Metabolome Reveals Differential Regulation of Carbon and Nitrogen Metabolism in Response to Nitrogen Availability.” *International Journal of Molecular Sciences* 20 (9): 2349. https://doi.org/10.3390/ijms20092349.

Xu, Peipei, and Weiming Cai. 2019. “Nitrate-Responsive OBP4-XTH9 Regulatory Module Controls Lateral Root Development in Arabidopsis Thaliana.” *PLoS Genetics* 15 (10).

You, Cheng, Dennis K. J. Lin, and S. Stanley Young. 2018. “Time Series Smoother for Effect Detection.” Edited by Xiaofeng Wang. *PLOS ONE* 13 (4): e0195360. https://doi.org/10.1371/journal.pone.0195360.

Zhang, Hanma, Andrea Jennings, Peter W Barlow, and Brian G Forde. 1999. “Dual Pathways for Regulation of Root Branching by Nitrate.” *Proceedings of the National Academy of Sciences* 96 (11): 6529–34.

Zhao, Wentao, Erchin Serpedin, and Edward R Dougherty. 2006. “Inferring Gene Regulatory Networks from Time Series Data Using the Minimum Description Length Principle.” *Bioinformatics* 22 (17): 2129–35.

Zheng, Zhi-Liang. 2009. “Carbon and Nitrogen Nutrient Balance Signaling in Plants.” *Plant Signaling & Behavior* 4 (7): 584–91.

Zhong, Silin, Je-Gun Joung, Yi Zheng, Yun-ru Chen, Bao Liu, Ying Shao, Jenny Z Xiang, Zhangjun Fei, and James J Giovannoni. 2011. “High-Throughput Illumina Strand-Specific RNA Sequencing Library Preparation.” *Cold Spring Harbor Protocols* 2011 (8): pdb--prot5652.

Zhuo, Degen, Mamoru Okamoto, J. John Vidmar, and Anthony D. M. Glass. 1999. “Regulation of a Putative High-Affinity Nitrate Transporter (Nrt2;1At) in Roots Of Arabidopsis Thaliana.” *The Plant Journal* 17 (5): 563–68. https://doi.org/10.1046/j.1365-313X.1999.00396.x.

Zou, C, J Shen, F Zhang, S Guo, Zed Rengel, and Caixian Tang. 2001. “Impact of Nitrogen Form on Iron Uptake and Distribution in Maize Seedlings in Solution Culture.” *Plant and Soil* 235 (2): 143–49.

Zwieniecki, Maciej A., Peter J. Melcher, Taylor S. Feild, and N. Michele Holbrook. 2004. “A Potential Role for Xylem–Phloem Interactions in the Hydraulic Architecture of Trees: Effects of Phloem Girdling on Xylem Hydraulic Conductance.” *Tree Physiology* 24 (8): 911–17. https://doi.org/10.1093/TREEPHYS/24.8.911.

APPENDIX A: SUPPLEMENTAL

MODEL CONDITIONS

Knowing that the parameters of interest θ are included in the terms $\tilde{\mu}_i$ and Σ_i, a series of conditions and assumptions need to be considered to estimate these parameters in a statistically sound manner. Aside from the assumption of normality mentioned above, the other assumptions are as follows:

(**A1)** The functions $f_j(\cdot), h_j(\cdot)$ and $g_j(\cdot)$ have known structures for j = 1,2.

(**A2)** The mapping $\theta \rightarrow F_\theta$ is injective.

(**A3)** The process (t_i, Z_i) is such that:

(i) $E[\boldsymbol{Z_{i_j}}] = 0, \forall i \in N_+$.

(ii) $cov(\boldsymbol{Z_{i_j}}, \boldsymbol{Z_{i_k}}) = \boldsymbol{\Sigma}(|t_i - t_k|)$, where $\boldsymbol{\Sigma}(|t_i - t_k|)$ is a nonsingular covariance matrix for all $i,j \in N_+$.

(**A4)** The functions $f_j(\cdot)$ and $h_j(\cdot)$, j = 1,2, jointly denoted as $m_j(\cdot)$, are twice continuously differentiable and such that:

(i) $m_j(0) = 1$ and $\lim_{\delta \to \infty} m_j(\delta) = 0$

(ii) $0 \leq |m_j(\delta)| \leq 1, \forall\ \delta \in R_+$.

(**A5)** The functions $g_j(\cdot)$, j = 1,2, are twice continuously differentiable and such that:

(i) $g_j(0) = 0$ and $\lim_{\delta \to \infty} g_j(\delta) = \sigma_j^2$, where $0 < \sigma_{j2} < \infty$.

(ii) $g_j(\delta) < g_j(\delta + h), \forall\ \delta, h \in R^+$

The above assumptions, along with other standard regularity conditions, are necessary in order to correctly estimate the model parameters and to test whether these parameters are significant.

SUPPLEMENTAL METHODS

Plant growth conditions and sampling

The time-evolved transcriptome of Arabidopsis roots and shoots was obtained as described in detail by Varala et al. (2018). Briefly, *Arabidopsis thaliana* (Col-0) seeds were grown hydroponically on 1 mM KNO_3^- or two weeks and then transiently treated with nitrogen (N) (20 mM KNO_3^- plus 20 mM NH4NO3) or control (20 mM KCl) for two hours. Samples were harvested at times 0, 5, 10, 15, 20, 30, 45, 60, 90, and 120 minutes, in which three replicates of roots and shoots were separated at harvest and immediately frozen in liquid nitrogen (see Fig.9).

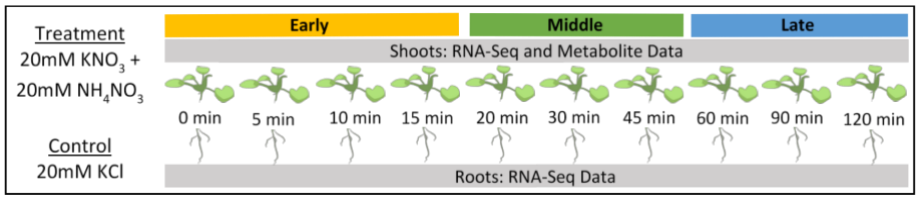


Figure 9: Tissue sampling scheme.

Transcriptome analysis

As described in Varala et al. (2018), total RNA was extracted from approximately 100 mg of tissue using the Qiagen RNeasy Kit. RNA was then processed for paired-end Illumina Sequencing using standard protocols (Zhong et al., 2011). The upper quartile normalization method from the EDASeq package in R was used to normalize the gene expression counts for both shoots and

roots. The set of Differentially Expressed Genes (DEGs) (in response to N-treatment vs control over time) was derived using the spline fitting model in the limma package in R. The shoot gene set was determined at an FDR adjusted *p*-value < 10^{-5}, resulting in 2173 shoot DEGs, while the root gene set was determined at an FDR adjusted *p*-value < 10^{-4}, resulting in 568 root DEGs (see Varala et al., 2018).

Data pre-processing

The replicates were first combined by taking the average gene expression for each gene at each time point. This gave one single time series expression for each gene. Root genes and shoot genes were clustered separately using MultiExperiment Viewer (MeV) (Saeed et al., 2003). The gene time series were imported and normalized using the "Normalize Genes/Rows" function in MeV. This transforms the gene expression values using the mean and standard deviation of each time series. The genes were then clustered using the QT clustering algorithm, setting the maximum threshold to 0.25, and the minimum cluster population to 5. The list of genes in each cluster was then exported. In each cluster, the average gene expression at each time point was subtracted from the gene expression value at that time point for each time series. This method resulted in a de-trending of the root and shoot time series (thereby removing the mean $\mu_i^{(x)}$).

Bioinformatic validation of the proposed method

In this section we provide an overview of the validation procedures followed in support of the results discussed in Results and Discussion in Chapter 2. Indeed, the proposed method merely suggests that a transcript level in one tissue is a result of a transcript in another tissue, but the

predicted relationships are not necessarily direct. Functional validation of predicted relationships focused on those that are likely caused by a direct, or potentially physical, interaction, such as protein:protein; protein:DNA; protein:metabolite, etc. A bioinformatics pipeline was developed to provide support for predicted relationships using network analysis, gene ontology, and text-mining to narrow down a manageable list of candidate genes for experimental testing.

Gene Ontology (GO) Term Analysis

This analysis was performed using the GO enrichment analysis tool from the Gene Ontology Consortium (Ashburner et al., 2000; Consortium, 2019; Mi et al., 2019). This tool returns a *p*-value from a Fisher's exact test in which the null hypothesis is that the biological functions of the genes are distributed evenly throughout the subset of genes as compared to the whole genome. A significant *p*-value indicates that the corresponding GO term appears more frequently than expected in the gene list compared to the overall genome. In a GO analysis there are many class-subclass relationships, i.e. the GO term "nitrogen fixation" is a subclass of the "nitrogen cycle metabolic process". A Benjamini-Hochberg False Discovery Rate (FDR) correction is used to correct for the multiple testing and a cut-off of 0.05 is suggested by default for significant results (Mi et al., 2019). GO terms were filtered using an FDR cut-off of 0.05 except in cases where there were too few GO terms (cutoff = 0.1) or no GO terms at the cutoff (no cutoff value).

Network Analysis

Directed networks were generated where genes are represented as nodes, and the directional dependence, as determined by the model, is represented as edges between nodes. Biological

networks have been shown to exhibit scale-free behavior such as the distribution in the network following a power-law (Albert, 2005). To provide some support towards the hypothesis that the proposed model-based network respects this feature, the predicted network was compared to random networks to determine how well it followed the power law for scale-free biological networks. 10^3 random networks with the same number of nodes and edges were generated in R using the sample_gnm function as part of the iGraph package. For each generated random network, the R^2 was calculated for the degree, in-degree and out-degree.

Promoter Analysis

The 2KB upstream region was obtained using Elefinder (Hudson, 2005). These regions were then used in Elefinder to determine over-represented transcription factor binding motifs. The results returned an E-value which is the likelihood of the result being returned by chance based on a binomial distribution. To search the 2KB upstream region for the significant occurrence, the FIMO tool from the MEME Suite (Bailey and Machanick, 2012) was used. Transcription factor binding motifs were first retrieved from the Plant Cistrome and EpiCistrome database (O'Malley et al., 2016). Using FIMO, promoter regions obtained from Elefinder were then searched for the specific motif using the default settings.

Nitrogen Response

For purposes of validation, particular attention was given to those genes previously implicated in the nitrogen response such as peptides (Araya et al., 2014) and those involved in cytokinin

biosynthesis (Takei et al. 2004), carbon/nitrogen balance (Palenchar et al., 2004), primary nitrogen metabolism and nitrogen transport (Perchlik and Tegeder, 2017; Krapp et al., 2014).

Microarray Data Analysis

GEO was searched for datasets with mutants of candidate causal genes. GSE8975, a DVL4 overexpression experiment, was analyzed using GEO2R using the default settings. The results were scanned to see if any target genes were differentially expressed between the wild type and mutant plants (p-value < 0.05).

Mobile Causal Gene Testing

In order to understand how well the proposed method detects known mobile causal genes, we performed a bootstrap procedure in which we considered all possible expression pairs (among all tested root and shoot expressions) and, from these pairs, we randomly selected the same amount of Granger-causal pairs detected by our method. Among these we then randomly selected the causal gene in each pair and, once the list of causal genes was completed in this manner, this was compared to the list of known mobile genes. The latter list was obtained from the PlaMoM (Plant Mobile Macromolecules) database (Guan et al., 2016) and is made up of genes that produce a mobile product that has been previously experimentally shown to move from either root to shoot, shoot to root, or in both directions. Following this approach, we then counted the number of causal genes in the randomly selected list that also appeared in the list of known mobile genes. This procedure was repeated 10^3 times and this distribution of counts was compared to the number of causal genes detected by our approach. This showed that the number of known mobile causal

genes detected by the proposed method is always significantly higher than a method that simply randomly samples the same number of causal genes thereby supporting the validity of the proposed analysis.

SUPPLEMENTAL NETWORK

A supplemental network file containing the network generated from the network analysis described above is made available as a standalone file called “Supplemental File 1.cys”

SUPPLEMENTAL TABLES

The supplemental tables are made available in a supplemental Excel file called “Supplemental Tables.xlsx”. On the next page is a table that collects the name of each sheet in the file and what it contains.

Sheet Name	**Description**
Supplemental Table 1	Table showing predicted 3078 root-shoot interactions
Supplemental Table 2	GO Terms for causal root genes at early time points.
Supplemental Table 3	GO Terms for causal root genes at middle time points.
Supplemental Table 4	GO Terms for causal root genes at late time points.
Supplemental Table 5	GO Terms for causal shoot genes at early time points.
Supplemental Table 6	GO Terms for causal shoot genes at middle time points.
Supplemental Table 7	GO Terms for causal shoot genes at late time points.
Supplemental Table 8	GO Terms for target root genes at early time points.
Supplemental Table 9	GO Terms for target root genes at middle time points.
Supplemental Table 10	GO Terms for target root genes at late time points.
Supplemental Table 11	GO Terms for target shoot genes at early time points.
Supplemental Table 12	GO Terms for target shoot genes at middle time points.
Supplemental Table 13	GO Terms for target shoot genes at late time points.
Supplemental Table 14	Table showing top 10 hubbiest genes by out-degree, and their network node properties.
Supplemental Table 15	Nitrogen signal responsive gene families and their members appearing in the predicted model interactions.
Supplemental Table 16	List of interactions in which the causal gene is a known signaling peptide.
Supplemental Table 17	List of RNA-binding proteins and their known mobility according to the PLAMOM database.

Table 2: Summary of tables contained in supplemental Excel file.

APPENDIX B: GLOSSARY

AutoRegressive (AR) Model: This is a regression model used to predict future values in a time series. The order of the regression model defines the number of preceding values used to predict the current value. For example, a first order autoregressive model, AR(1), predicts the current value using the immediately preceding value in the time series.

AutoRegressive Moving Average (ARMA) Model: A linear combination of the autoregressive model and the moving average (MA) model used in time series analysis. The MA model uses the previous white noise terms unlike the AR model which uses previous time series values.

Dependence Structure: This refers to the association of observations with variables at previous time points.

Deterministic Function: A function that always returns the same results when the same specific inputs are used.

Granger Causality: Consider the two time-series X_t and Y_t. If we are better able to predict X_t using Yt, as opposed to not using any information from Y_t, then we can say that $Y_t$is causing X_t.

Gauss-Markov Model/Process: A stochastic model with a Gaussian distribution and is considered to have a Markov property. A time series can be described as having a Markov property if $X_t$has all the information necessary for predicting X_{t+1}.

Likelihood Function: A probability density of the data viewed as a function of the parameters.

Maximum Likelihood Estimator: A framework that involves maximizing the likelihood function in order to find the parameters that best define the observed data.

Monte-Carlo-based Techniques: A collection of computational techniques that use simulated random numbers for the estimation of functions in a probability distribution.

Multivariate Time Series: This is a time series that is dependent on more than one variable. For example, for a time series X_t, the value of X at time t+1 is dependent not only on X_t, but also on the information in other time series.