# Granger-Causal Testing for Irregularly Sampled Time Series with Application to Nitrogen Signaling in Arabidopsis

Sachin Heerah[†], Roberto Molinari[‡], Stéphane Guerrier[§,¶]
and Amy Marshall-Colon[†,¶]

[†]Department of Plant Biology, University of Illinois Urbana-Champaign,
Urbana, IL 61801, USA
[‡]Department of Mathematics and Statistics, Auburn University,
Auburn, AL 36849, USA
[§] Faculty of Science & Geneva School of Economics and Management,
University of Geneva, Geneva 1205, Switzerland
[¶]Corresponding Authors

## 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 of 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 which embeds functions from the domain of engineering that allow to adapt the model's dependence structure to the specific sampling time. Using Maximum-Likelihood to estimate the parameters of 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 the *Arabidopsis thaliana* data, the proposed approach produced 3,078 significant interactions, in which 2,012 interactions have root causal genes

and 1,066 interactions have 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 thorugh 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`.
**Contact:** amymc@illinois.edu

# 1 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 for example. However, the analysis of this irregular data is challenging, among others, 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 ap-
proach to describe causal gene-gene relationships from above- (shoot) and below-
(root) ground organs of *Arabidopsis thaliana* in response to a nitrogen signal.
Through identification and bioinformatic exploration of the detected causal re-
lationships, we achieve a greater understanding of the underlying molecular and
biochemical pathways involved in the 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 un-
cover new molecules, pathways, and processes involved in the root-to-shoot-to-
root nitrogen-signal relay, providing biological insight into complex, whole-plant
nitrogen response.

## 2    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

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

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

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

In addition, let $(X_i : i = 1, \ldots, n) \in \mathbb{R}^n$ and $(Y_i : i = 1, \ldots, n) \in \mathbb{R}^n$ denote two
sequences of real-valued random variables such that we can denote a bivariate
irregularly spaced time series with $n$ time points, as $(t_i, X_i, Y_i : i = 1, \ldots, n)$,
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

3

points in time (i.e. the sequences $(t_i : i = 1, ..., n)$ 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 $F = \{F_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p\}$, with $\boldsymbol{\theta}$ being the vector containing the parameters of this model. The latter model needs to deal with irregularly spaced bivariate time series and should allow to test 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 $\rho$ 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 $\sigma^2$. This model allows to approximate many covariance structures delivering a behavior that is often reasonable for many 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 to the following:

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

where $\lambda$ therefore 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 $\rho$, 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 $\lambda$ and $\sigma^2$). For this reason, the next sections put forward a new framework for these settings.

4

## 2.1 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 centred observations as $\tilde{X}_{t_i} := X_{t_i} - \mu_i^{(x)}$ and the distance in time as $\delta_{t_i} := t_i - t_{i-1}$, with $\delta_{t_i} \in \mathbb{R}^+$ by definition. Based on this, the AR(1) model for irregularly spaced data can be represented as follows:

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

where $f(\cdot)$ is a deterministic function, possibly known up to some parameter values, that plays the same role as the constant $\rho$ but takes into account the distance between observations. For different reasons, among which estimation feasibility, the independent sequence $(W_{t_i})$ is usually considered as being Gaussian (although other distributions can be considered). Without loss of generality, we will make this assumption and therefore state that $W_{t_i} \sim \mathcal{N}(0, g(\delta_{t_i}))$ 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:

$$\boldsymbol{Z}_i = \boldsymbol{A}\left(\delta_{t_i}\right)\boldsymbol{Z}_{i-1} + \boldsymbol{V}_i, \tag{2}$$

where

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

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

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

It can be observed how the matrix $\boldsymbol{A}\left(\delta_{t_i}\right)$ 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 $\boldsymbol{A}(\delta_{t_i})$. 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 kind 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 $\boldsymbol{Z}_i$ as $l(\boldsymbol{Z}_i)$, then the likelihood function is given by

$$L(\boldsymbol{\theta}) = l(Z_1) \prod_{i=2}^{n} l(\boldsymbol{Z}_i|\boldsymbol{Z}_{i-1}), \tag{5}$$

where, using (4), we have

$$l(\boldsymbol{Z}_i|\boldsymbol{Z}_{i-1}) = \frac{1}{2\pi|\boldsymbol{\Sigma}_i|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{Z}_i - \tilde{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Z}_i - \tilde{\mu}_i)\right).$$

Applying the $\log(\cdot)$ function to $L(\boldsymbol{\theta})$ and fixing $l(\boldsymbol{Z}_1)$ as constant (neglecting constant terms) we obtain the following estimating equation which defines the Maximum Likelihood Estimator (MLE):

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

where

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

6

Under a set of conditions (see Section 1 in the supplementary material), 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 $\phi \in \mathbb{R}^+$ 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 $\delta_{t_i}$ as a measure of Euclidean distance. The above model therefore gives an explicit form to the functions $f.(\cdot)$ and $g.(\cdot)$ 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 $h.(\cdot)$ that describes the behaviour of a signal based on another. Given the short time series available, we decide to impose a reasonable structure to this behaviour which allows the dependence of a signal on the other signal to grow exponentially over time (reaching its maximum) and then decay exponentially. Indeed, while we consider the impact of past values of a time series on its future values as a function only of their distance in time, we postulate 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 behaviour 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 propose to use the following function:

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

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 $\gamma \in \mathbb{R}^+$ denotes the distance in time at which the dependence of a time series on another is maximal. Finally, $\eta \in \mathbb{R}^+$ plays a similar role to $\phi$ in the previously defined function $h.(\cdot)$.

7

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 respond to the need to balance model complexity with practical feasibility, we will consider 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.

## 2.2   Testing Procedure

Once the model is defined, the goal of this work is therefore to understand which structure of the matrix $\boldsymbol{A}(\delta_{t_i})$ 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:

$$
\begin{aligned}
H_0 &: \quad \boldsymbol{A}(\delta_{t_i}) \text{ is diagonal.} \\
H_A &: \quad \boldsymbol{A}(\delta_{t_i}) \text{ is lower triangular.}
\end{aligned}
$$

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 "$\boldsymbol{A}(\delta_{t_i})$ is upper triangular" therefore reversing the direction of dependence.

The MLE defined in (6) allows to estimate 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

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

where $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\theta}}_1$ 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 $\boldsymbol{\theta}_1 \in \mathbb{R}^{p_1}$ with respect to $\boldsymbol{\theta}_0 \in \mathbb{R}^{p_0}$ (i.e. $p^\star := p_1 - p_0$). Using this distribution and the observed $LRT$ statistic one can then reject or not the null hypothesis thereby concluding whether or not a signal Granger-causes the other.

## 2.3 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 are able to 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 $h = 0$, $H \geq 100$ and a zero vector $LRT_{boot}$ of dimension $H$;

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

   1.  $h = h + 1$;

   2.  Simulate a bivariate time series $(Z_i^{(h)})$ of the same sample size as the original signals from the model $F_{\hat{\boldsymbol{\theta}}_0}$;

   3. Estimate $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ from the simulated sample $(Z_i^{(h)})$ to obtain $\hat{\boldsymbol{\theta}}_0^{(h)}$ and $\hat{\boldsymbol{\theta}}_1^{(h)}$ respectively;

   4. Compute $LRT_{boot}^{(h)} = 2\left(Q_n(\hat{\boldsymbol{\theta}}_1^{(h)}) - Q_n(\hat{\boldsymbol{\theta}}_0^{(h)})\right)$

---

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

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

If this $p$-value is smaller than a chosen level of significance $\alpha$, 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 $\alpha/(2N_X N_Y)$ if applying, for example, a Bonferroni correction.
If the sizes $N_X$ and $N_Y$ are considerable, this would require to increase 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.

## 3   Results and Discussion

The described approach was applied to the time-evolved transcriptome of Ara-
bidopsis roots and shoots (the $(X_{t_i})$ and $(Y_{t_i})$ signals respectively) whose mea-
surements were made through an experimental setup described more in detail
in the supplemental material along with the chosen pre-processing (Section 2).
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 signifi-
cantly differentially expressed root and shoot genes. Using $H = 10^3$, we apply
the procedure described in Section 2 which produces a final list of 3,078 gene
pair interactions whose details are listed in Table 1, Section 3 of supplementary
material (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 being the (Granger) causal gene, or the influencer on the ex-
pression 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 $\gamma$ 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 analyse 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".

### 3.1   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 are influenced across tissues over time (see Section 2 in supple-
mental material for more details). As highlighted also in Fig. 1, at early time
points (0 - 15 minutes), causal root-genes reflect the early nitrogen response,

| 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 |

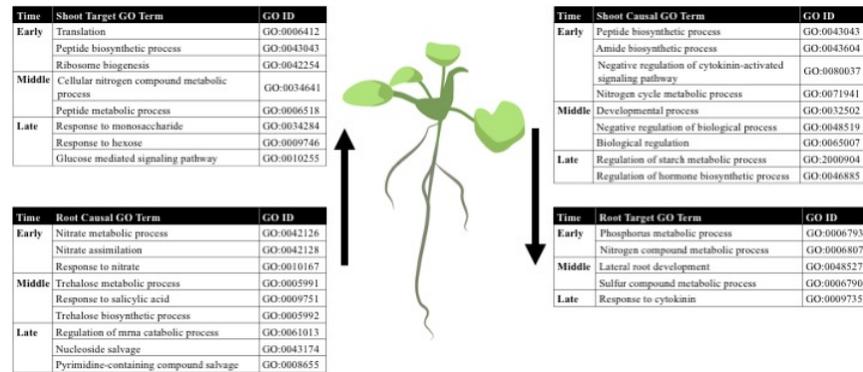| Time | Root Target GO Term | GO ID |
|------|---------------------|-------|
| Early | Phosphorus metabolic process | GO:0006793 |
| | Nitrogen compound metabolic process | GO:0006807 |
| Middle | Lateral root development | GO:0048527 |
| | Sulfur compound metabolic process | GO:0006790 |
| Late | Response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

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

while target-shoot genes reflect post-transcriptional and translational processes (see Tables 2 and 11, Section 3 of supplemental material). 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 Tables 3,4,12 and 13, Section 3 of supplemental material). 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 Tables 5-10, Section 3 of supplemental material). 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).

## 3.2 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 Section 2 in supplemental material). 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 transcrip-
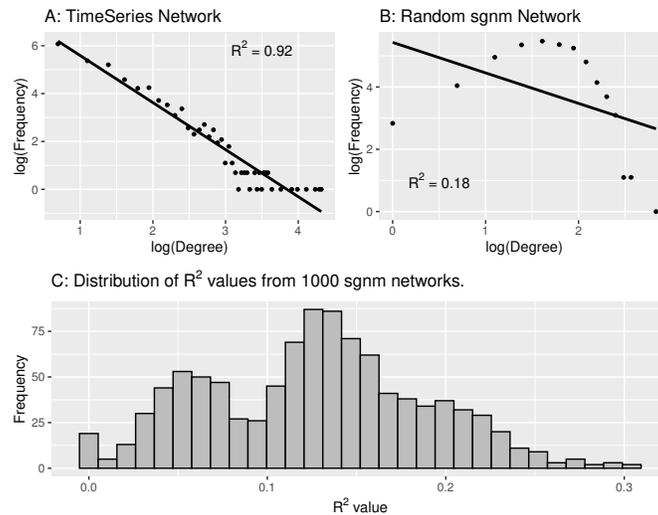
Figure 2: A: Node degree distribution for the network generated from model predictions ($R^2 = 0.92$). B: Node degree distribution for a randomly generated network ($R^2 = 0.18$). C: Histogram of the $R^2$ values for the node degree distribution of 1,000 randomly generated networks ($0 \leq R^2 < 0.35$).

tomes 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 Table 4, Section 3 of supplementary material). 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 Network File"). 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.

## 3.3 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 Tab. 1). To further investigate these predicted relationships, promoter analysis using FIMO from MEME Suite (Bailey and Machanick, 2012) was performed (as outlined in Section 2 of supplemental material). 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 heatshock 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 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.

|  | 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 with which TGA1 has been shown to physically bind to 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.

## 3.4   Long-distance signaling by hormones and peptides

**Cytoknin Response Factors (CRFs):**   Transcription factors (TF) with previously described regulatory roles in nitrogen uptake and assimilation include members of the ERF, bZIP, and NLP 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 Section 2 of supplemental material), 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 Section 2 of supplemental material). 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 signalling 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).
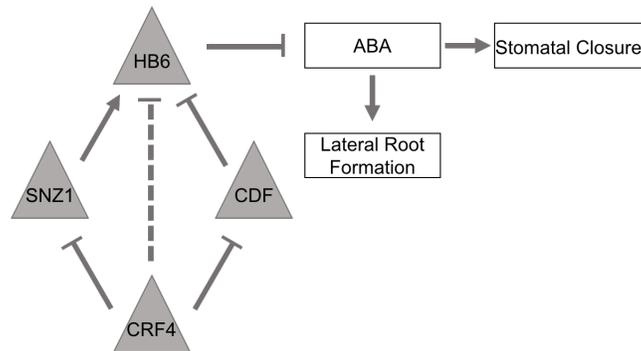
Figure 3: CRF4 interaction pathway, with flat-head arrows indicating negative interactions, pointed-head arrows indicating positive interactions, dashed arrows representing predicted interactions from the model, and solid arrows representing known interactions.

**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 Table 1, Section 3 of supplementary material), 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 Table 16, Section 3 of supplementary material). ATPSK4

15

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-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 the 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 over-expression of DVL4 resulted in the up-regualtion of a number of genes encoding transcription factors (Larue *et al.*, 2010). Our re-analysis of the microarray data from this study (see Section 2 of supplemental material) 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.

## 3.5 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.
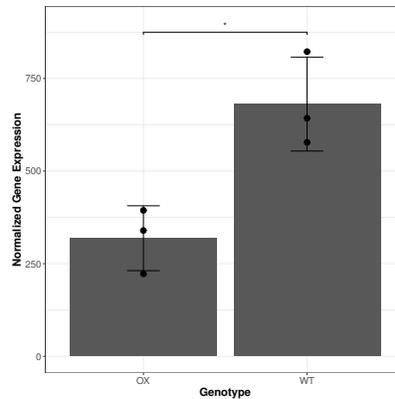
16

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 < 0.05).

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 Section 2 in supplemental material) 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 Table 17, Section 3 of supplemental material). 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 have hypothesised 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).

## 4    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 to enhance 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.

19

# References

Albert, R. (2005). Scale-free networks in cell biology. *Journal of cell science*, **118**(21), 4947–4957.

Alvarez, J. M., Riveras, E., Vidal, E. A., Gras, D. E., Contreras-López, O., Tamayo, K. P., Aceituno, F., Gómez, I., Ruffel, S., Lejay, L., *et al.* (2014). Systems approach identifies tga 1 and tga 4 transcription factors as important regulatory components of the nitrate response of a rabidopsis thaliana roots. *The Plant Journal*, **80**(1), 1–13.

Araus, V., Vidal, E. A., Puelma, T., Alamos, S., Mieulet, D., Guiderdoni, E., and Gutiérrez, R. A. (2016). Members of btb gene family regulate negatively nitrate uptake and nitrogen use efficiency in arabidopsis thaliana and oryza sativa. *Plant physiology*.

Bahadori, M. T. and Liu, Y. (2012). Granger causality analysis in irregular time series. In *Proceedings of the 2012 SIAM International Conference on Data Mining*, pages 660–671. SIAM.

Bailey, T. L. and Machanick, P. (2012). Inferring direct dna binding from chip-seq. *Nucleic acids research*, **40**(17), e128–e128.

Banerjee, S., Neveu, P., and Kosik, K. S. (2009). A coordinated local translational control point at the synapse involving relief from silencing and mov10 degradation. *Neuron*, **64**(6), 871–884.

Bar-Joseph, Z., Gitter, A., and Simon, I. (2012). Studying and modelling dynamic biological processes using time-series gene expression data. *Nature Reviews Genetics*, **13**(8), 552–564.

Barabási, A.-L. (2003). Linked: The new science of networks.

Bechtold, U., Penfold, C. A., Jenkins, D. J., Legaie, R., Moore, J. D., Lawson, T., Matthews, J. S., Vialet-Chabrand, S. R., Baxter, L., Subramaniam, S., *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), 345–366.

Bidadi, H., Matsuoka, K., Sage-Ono, K., Fukushima, J., Pitaksaringkarn, W., Asahina, M., Yamaguchi, S., Sawa, S., Fukuda, H., Matsubayashi, Y., *et al.* (2014). Cle6 expression recovers gibberellin deficiency to promote shoot growth in arabidopsis. *The Plant Journal*, **78**(2), 241–252.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Broersen, P. M. and Bos, R. (2006). Estimating time-series models from irregularly spaced data. *IEEE transactions on instrumentation and measurement*, **55**(4), 1124–1131.

Brooks, M. D., Cirrone, J., Pasquino, A. V., Alvarez, J. M., Swift, J., Mittal, S., Juang, C.-L., Varala, K., Gutiérrez, R. A., Krouk, G., *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.

Chen, J., Zeng, B., Zhang, M., Xie, S., Wang, G., Hauck, A., and Lai, J. (2014). Dynamic transcriptome landscape of maize embryo and endosperm development. *Plant physiology*, **166**(1), 252–264.

Colón, A. M., Sengupta, N., Rhodes, D., Dudareva, N., and Morgan, J. (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, J.-P., Küster, H., Journet, E.-P., Hohnjec, N., Gamas, P., and Niebel, A. (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), 1118–1127.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application*, volume 1. Cambridge university press.

Doidy, J., Li, Y., Neymotin, B., Edwards, M. B., Varala, K., Gresham, D., and Coruzzi, G. M. (2016). "hit-and-run" transcription: de novo transcription initiated by a transient bzip1 "hit" persists after the "run". *BMC genomics*, **17**(1), 92.

Eckner, A. (2014). A framework for the analysis of unevenly spaced time series data. *Preprint*.

Erdogan, E., Ma, S., Beygelzimer, A., and Rish, I. (2005). Statistical models for unequally spaced time series. In *proceedings of the 2005 SIAM International conference on data mining*, pages 626–630. SIAM.

Eyheramendy, S., Elorrieta, F., and Palma, W. (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), 4311–4322.

Fujiki, Y., Teshima, H., Kashiwao, S., Kawano-Kawada, M., Ohsumi, Y., Kakinuma, Y., and Sekito, T. (2017). Functional identification of atavt3, a family of vacuolar amino acid transporters, in arabidopsis. *FEBS letters*, **591**(1), 5–15.

Fujimoto, S. Y., Ohta, M., Usui, A., Shinshi, H., and Ohme-Takagi, M. (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, Y., Fujita, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2011). Aba-mediated transcriptional regulation in response to osmotic stress in plants. *Journal of plant research*, **124**(4), 509–525.

Gargouri, M., Park, J.-J., Holguin, F. O., Kim, M.-J., Wang, H., Deshpande, R. R., Shachar-Hill, Y., Hicks, L. M., and Gang, D. R. (2015). Identification of regulatory network hubs that control lipid metabolism in chlamydomonas reinhardtii. *Journal of experimental botany*, **66**(15), 4551–4566.

Gaudinier, A., Rodriguez-Medina, J., Zhang, L., Olson, A., Liseron-Monfils, C., Bågman, A.-M., Foret, J., Abbitt, S., Tang, M., Li, B., *et al.* (2018). Transcriptional regulation of nitrogen-associated metabolism and growth. *Nature*, **563**(7730), 259–264.

Goel, P., Bhuria, M., Kaushal, M., and Singh, A. K. (2016). Carbon: Nitrogen interaction regulates expression of genes involved in n-uptake and assimilation in brassica juncea l. *PloS one*, **11**(9).

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438.

Grillet, L., Lan, P., Li, W., Mokkapati, G., and Schmidt, W. (2018). Iron man is a ubiquitous family of peptides that control iron transport in plants. *Nature plants*, **4**(11), 953–963.

Guan, D., Yan, B., Thieme, C., Hua, J., Zhu, H., Boheler, K. R., Zhao, Z., Kragler, F., Xia, Y., and Zhang, S. (2016). Plamom: a comprehensive database compiles plant mobile macromolecules. *Nucleic acids research*, page gkw988.

Guan, P. (2017). Dancing with hormones: A current perspective of nitrate signaling and regulation in arabidopsis. *Frontiers in plant science*, **8**, 1697.

Ham, B.-K., Brandom, J. L., Xoconostle-Cázares, B., Ringgold, V., Lough, T. J., and Lucas, W. J. (2009). A polypyrimidine tract binding protein, pumpkin rbp50, forms the basis of a phloem-mobile ribonucleoprotein complex. *The Plant Cell*, **21**(1), 197–215.

Hamilton, J. D. (1994). *Time series analysis*, volume 2. Princeton New Jersey.

Hentze, M. W. and Argos, P. (1991). Homology between ire-bp, a regulatory rna-binding protein, aconitase, and isopropylmalate isomerase. *Nucleic acids research*, **19**(8), 1739–1740.

Himmelbach, A., Yang, Y., and Grill, E. (2003). Relay and control of abscisic acid signaling. *Current opinion in plant biology*, **6**(5), 470–479.

Hudson, M. (2005). Elefinder.

Kehr, J. and Kragler, F. (2018). Long distance rna movement. *New Phytologist*, **218**(1), 29–40.

Kiba, T., Kudo, T., Kojima, M., and Sakakibara, H. (2011). Hormonal control of nitrogen acquisition: roles of auxin, abscisic acid, and cytokinin. *Journal of experimental botany*, **62**(4), 1399–1409.

Ko, D. and Helariutta, Y. (2017). Shoot–root communication in flowering plants. *Current Biology*, **27**(17), R973–R978.

Konishi, M. and Yanagisawa, S. (2013). Arabidopsis nin-like transcription factors have a central role in nitrate signalling. *Nature communications*, **4**(1), 1–9.

Krapp, A., David, L. C., Chardin, C., Girin, T., Marmagne, A., Leprince, A.-S., Chaillou, S., Ferrario-Méry, S., Meyer, C., and Daniel-Vedele, F. (2014). Nitrate transport and signalling in arabidopsis. *Journal of experimental botany*, **65**(3), 789–798.

Krouk, G., Mirowski, P., LeCun, Y., Shasha, D. E., and Coruzzi, G. M. (2010). Predictive network modeling of the high-resolution dynamic plant transcriptome in response to nitrate. *Genome biology*, **11**(12), R123.

Larue, C. T., Wen, J., and Walker, J. C. (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.

Léran, S., Edel, K. H., Pervent, M., Hashimoto, K., Corratgé-Faillie, C., Offenborn, J. N., Tillard, P., Gojon, A., Kudla, J., and Lacombe, B. (2015). Nitrate sensing and uptake in arabidopsis are enhanced by abi2, a phosphatase inactivated by the stress hormone abscisic acid. *Sci. Signal.*, **8**(375), ra43–ra43.

Liang, Y. S., Ermawati, N., Cha, J.-Y., Jung, M. H., Suudi, M., Kim, M. G., Ha, S.-H., Park, C.-G., and Son, D. (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), 142–148.

Lough, T. J. and Lucas, W. J. (2006). Integrative plant biology: role of phloem long-distance macromolecular trafficking. *Annu. Rev. Plant Biol.*, **57**, 203–232.

Luo, K.-R., Huang, N.-C., and Yu, T.-S. (2018). Selective targeting of mobile mrnas to plasmodesmata for cell-to-cell movement. *Plant physiology*, **177**(2), 604–614.

Maller, R. A., Müller, G., Szimayer, A., *et al.* (2008). Garch modelling in continuous time for irregularly spaced time series data. *Bernoulli*, **14**(2), 519–542.

Mangan, S. and Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences*, **100**(21), 11980–11985.

Marondedze, C., Thomas, L., Serrano, N. L., Lilley, K. S., and Gehring, C. (2016). The rna-binding protein repertoire of arabidopsis thaliana. *Scientific reports*, **6**(1), 1–13.

Matsubayashi, Y., Ogawa, M., Kihara, H., Niwa, M., and Sakagami, Y. (2006). Disruption and overexpression of arabidopsis phytosulfokine receptor gene affects cellular longevity and potential for growth. *Plant Physiology*, **142**(1), 45–53.

Oh, E., Seo, P. J., and Kim, J. (2018). Signaling peptides and receptors coordinating plant root development. *Trends in plant science*, **23**(4), 337–351.

Ohkubo, Y., Tanaka, M., Tabata, R., Ogawa-Ohnishi, M., and Matsubayashi, Y. (2017). Shoot-to-root mobile polypeptides involved in systemic regulation of nitrogen acquisition. *Nature Plants*, **3**(4), 1–6.

Okamoto, S., Tabata, R., and Matsubayashi, Y. (2016). Long-distance peptide signaling essential for nutrient homeostasis in plants. *Current opinion in plant biology*, **34**, 35–40.

O'Malley, R. C., Huang, S.-s. C., Song, L., Lewsey, M. G., Bartlett, A., Nery, J. R., Galli, M., Gallavotti, A., and Ecker, J. R. (2016). Cistrome and epicistrome features shape the regulatory dna landscape. *Cell*, **165**(5), 1280–1292.

Palenchar, P. M., Kouranov, A., Lejay, L. V., and Coruzzi, G. M. (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.

Poitout, A., Crabos, A., Petřík, I., Novák, O., Krouk, G., Lacombe, B., and Ruffel, S. (2018). Responses to systemic nitrogen signaling in arabidopsis roots involve trans-zeatin in shoots. *The Plant Cell*, **30**(6), 1243–1257.

Rehfeld, K., Marwan, N., Heitzig, J., and Kurths, J. (2011). Comparison of correlation analysis techniques for irregularly sampled time series. *Nonlinear Processes in Geophysics*, **18**(3), 389–404.

Remondini, D., O'connell, B., Intrator, N., Sedivy, J., Neretti, N., Castellani, G., and Cooper, L. (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), 6902–6906.

Ripley, B. D. (2005). *Spatial statistics*, volume 575. John Wiley & Sons.

Ruffel, S., Krouk, G., Ristova, D., Shasha, D., Birnbaum, K. D., and Coruzzi, G. M. (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*, **108**(45), 18524–18529.

Russnak, R., Konczal, D., and McIntire, S. L. (2001). A family of yeast proteins mediating bidirectional vacuolar amino acid transport. *Journal of Biological Chemistry*, **276**(26), 23849–23857.

Sakakibara, H., Takei, K., and Hirose, N. (2006). Interactions between nitrogen and cytokinin in the regulation of metabolism and development. *Trends in plant science*, **11**(9), 440–448.

Sakuma, Y., Liu, Q., Dubouzet, J. G., Abe, H., Shinozaki, K., and Yamaguchi-Shinozaki, K. (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.

Sato, T., Maekawa, S., Konishi, M., Yoshioka, N., Sasaki, Y., Maeda, H., Ishida, T., Kato, Y., Yamaguchi, J., and Yanagisawa, S. (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), 380–386.

Scheible, W.-R., Morcuende, R., Czechowski, T., Fritz, C., Osuna, D., Palacios-Rojas, N., Schindelasch, D., Thimm, O., Udvardi, M. K., and Stitt, M. (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), 2483–2499.

Sekito, T., Fujiki, Y., Ohsumi, Y., and Kakinuma, Y. (2008). Novel families of vacuolar amino acid transporters. *IUBMB life*, **60**(8), 519–525.

Signora, L., De Smet, I., Foyer, C. H., and Zhang, H. (2001). Aba plays a central role in mediating the regulatory effects of nitrate on root branching in arabidopsis. *The Plant Journal*, **28**(6), 655–662.

Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D., and Futcher, B. (1998). Comprehensive identification of cell cycle–regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization. *Molecular biology of the cell*, **9**(12), 3273–3297.

Sun, C.-H., Yu, J.-Q., and Hu, D.-G. (2017). Nitrate: a crucial signal during lateral roots development. *Frontiers in plant science*, **8**, 485.

Tabata, R., Sumida, K., Yoshii, T., Ohyama, K., Shinohara, H., and Matsubayashi, Y. (2014). Perception of root-derived peptides by shoot lrr-rks mediates systemic n-demand signaling. *Science*, **346**(6207), 343–346.

Thiebaut, C. and Roques, S. (2005). Time-scale and time-frequency analyses of irregularly sampled astronomical time series. *EURASIP Journal on Advances in Signal Processing*, **2005**(15), 852587.

Thieme, C. J., Rojas-Triana, M., Stecyk, E., Schudoma, C., Zhang, W., Yang, L., Miñambres, M., Walther, D., Schulze, W. X., Paz-Ares, J., *et al.* (2015). Endogenous arabidopsis messenger rnas transported to distant tissues. *Nature Plants*, **1**(4), 15025.

Titterton, D., Weston, J. L., and Weston, J. (2004). *Strapdown inertial navigation technology*, volume 17. IET.

To, J. P., Haberer, G., Ferreira, F. J., Deruere, J., Mason, M. G., Schaller, G. E., Alonso, J. M., Ecker, J. R., and Kieber, J. J. (2004). Type-a arabidopsis response regulators are partially redundant negative regulators of cytokinin signaling. *The Plant Cell*, **16**(3), 658–671.

To, J. P., Deruère, J., Maxwell, B. B., Morris, V. F., Hutchison, C. E., Ferreira, F. J., Schaller, G. E., and Kieber, J. J. (2007). Cytokinin regulates type-a arabidopsis response regulator activity and protein stability via two-component phosphorelay. *The Plant Cell*, **19**(12), 3901–3914.

Tone, J., Yoshimura, A., Manabe, K., Murao, N., Sekito, T., Kawano-Kawada, M., and Kakinuma, Y. (2015). Characterization of avt1p as a vacuolar proton/amino acid antiporter in saccharomyces cerevisiae. *Bioscience, biotechnology, and biochemistry*, **79**(5), 782–789.

Varala, K., Marshall-Colón, A., Cirrone, J., Brooks, M. D., Pasquino, A. V., Léran, S., Mittal, S., Rock, T. M., Edwards, M. B., Kim, G. J., *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), 6494–6499.

Vidal, E. A., Araus, V., Lu, C., Parry, G., Green, P. J., Coruzzi, G. M., and Gutiérrez, R. A. (2010). Nitrate-responsive mir393/afb3 regulatory module controls root system architecture in arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, **107**(9), 4477–4482.

Vidal, E. A., Moyano, T. C., Krouk, G., Katari, M. S., Tanurdzic, M., McCombie, W. R., Coruzzi, G. M., and Gutiérrez, R. A. (2013a). Integrated rna-seq and srna-seq analysis identifies novel nitrate-responsive genes in arabidopsis thaliana roots. *BMC genomics*, **14**(1), 701.

Vidal, E. A., Moyano, T. C., Riveras, E., Contreras-López, O., and Gutiérrez, R. A. (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), 12840–12845.

Vidal, E. A., Álvarez, J. M., and Gutiérrez, R. A. (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, E. A., Álvarez, J. M., Moyano, T. C., and Gutiérrez, R. A. (2015). Transcriptional networks in the nitrate response of arabidopsis thaliana. *Current opinion in plant biology*, **27**, 125–132.

Wang, G., Zhang, G., and Wu, M. (2016). Cle peptide signaling and crosstalk with phytohormones and environmental stimuli. *Frontiers in plant science*, **6**, 1211.

Wang, R., Guegler, K., LaBrie, S. T., and Crawford, N. M. (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.

Wang, R., Okamoto, M., Xing, X., and Crawford, N. M. (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), 556–567.

Weinl, C., Marquardt, S., Kuijt, S. J., Nowack, M. K., Jakoby, M. J., Hülskamp, M., and Schnittger, A. (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), 1704–1722.

Wen, J., Lease, K. A., and Walker, J. C. (2004). Dvl, a novel class of small polypeptides: overexpression alters arabidopsis development. *The Plant Journal*, **37**(5), 668–677.

Xu, P. and Cai, W. (2019). Nitrate-responsive obp4-xth9 regulatory module controls lateral root development in arabidopsis thaliana. *PLoS genetics*, **15**(10).

Zhao, W., Serpedin, E., and Dougherty, E. R. (2006). Inferring gene regulatory networks from time series data using the minimum description length principle. *Bioinformatics*, **22**(17), 2129–2135.

Zheng, Z.-L. (2009). Carbon and nitrogen nutrient balance signaling in plants. *Plant Signaling & Behavior*, **4**(7), 584–591.

Zhu, G., Spellman, P. T., Volpe, T., Brown, P. O., Botstein, D., Davis, T. N., and Futcher, B. (2000). Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth. *Nature*, **406**(6791), 90–94.

Zou, C., Shen, J., Zhang, F., Guo, S., Rengel, Z., and Tang, C. (2001). Impact of nitrogen form on iron uptake and distribution in maize seedlings in solution culture. *Plant and soil*, **235**(2), 143–149.