Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Plant Biology

Modelling signaling networks underlying plant defence

Oliver Windram[1] and Katherine J Denby[2]

Transcriptional reprogramming plays a significant role in governing plant responses to pathogens. The underlying regulatory networks are complex and dynamic, responding to numerous input signals. Most network modelling studies to date have used large-scale expression data sets from public repositories but defence network models with predictive ability have also been inferred from single time series data sets, and sophisticated biological insights generated from focused experiments containing multiple network perturbations. Using multiple network inference methods, or combining network inference with additional data, such as promoter motifs, can enhance the ability of the model to predict gene function or regulatory relationships. Network topology can highlight key signaling components and provides a systems level understanding of plant defence.

Addresses
[1] Department of Life Sciences, Imperial College London, Silwood Park SL5 7PY, UK
[2] School of Life Sciences and Warwick Systems Biology Centre, University of Warwick, CV4 7AL, UK

Corresponding author: Denby, Katherine J (k.j.denby@warwick.ac.uk)

Current Opinion in Plant Biology 2015, **27**:165–171

This review comes from a themed issue on **Cell signalling and gene regulation**

Edited by **Xiaofeng Cao** and **Blake C Meyers**

For a complete overview see the Issue and the Editorial

Available online 21st August 2015

http://dx.doi.org/10.1016/j.pbi.2015.07.007

Introduction

Computational and mathematical modelling of biological data is not a new approach in plant science. For example, many multi-component modelling strategies are used in agriculture to support decision-making and predict crop yields, and evolution and ecology are two fields where modelling has had a significant impact for many decades. However, it is only relatively recently that modelling approaches have been used to understand molecular signaling pathways underlying plant defence responses and in many cases the significance, or accuracy, of insights from the models have not been tested. In this review we highlight recent advances in modelling of plant defence signaling and the types of questions that can be addressed (Table 1). Where appropriate we have outlined techniques that have been applied to other areas of plant science, typically abiotic stress responses. For greater detail on specific modelling methods please see alternative reviews [1,2].

Predicting defence gene function

Modelling is often used to predict new functions for genes or gene products. Networks can be constructed linking genes on the basis of co-expression across a set of transcriptome data [3] and a guilt-by-association strategy used to predict gene function. This assumes that genes closely associated with a gene of known function may share that function, and can be complemented by prioritising unknown hub genes for experimental validation [4] (see Box 1 for explanation of network terms). The data sets used for such networks can be condition-independent (i.e. no selection for data relevant to the biological process being investigated) or condition-dependent. A recent genome-wide co-expression network in Arabidopsis was generated using nearly 900 microarray data sets and includes over 18 000 genes [5]. Several network modules were induced in response to biotic stress and hormone treatment suggesting a role for that hormone within the defence regulation of that module, and two modules were specifically repressed in the presence of *Pseudomonas syringae* effector proteins. Genes of unknown function within these modules are potentially novel players in the plant defence response. Functional association networks extend the co-expression concept and incorporate multiple large-scale data sets to enhance their predictive ability. An early plant functional association network, AraNet [6], used transcriptome data, experimentally determined protein–protein interactions, and protein sequence information as well as a variety of gene–gene association data inferred from other organisms including mouse, yeast and human. This network successfully predicted seed pigmentation, drought tolerance and lateral root formation roles for novel genes.

Co-expression analysis is a popular method for gene discovery given its ease of implementation and the ability to utilize gene expression data from a large range of studies such as those in publically available microarray compendia. Two such recent studies have attempted to predict genes with a role in the plant defence response [7,8]. Both used large collections of expression data from pathogen infections of Arabidopsis to infer co-expression networks, with Tully *et al.* [8] combining co-expression network inference with a motif discovery tool, tailor-made to handle large groups of genes, to predict causality within the network. Both studies suggested that hub genes and nodes with high betweenness centrality (Box 1) play important roles in the plant immune

Table 1

Types of network modelling strategies used to study the plant defence response. We indicate the biological question(s) that can be addressed by each approach, potential advantages and disadvantages of the different methodologies and an example

Method	Advantages	Disadvantages	Example
Predicting gene function			
Co-expression networks	Can use existing expression data, may be genome-wide, can predict novel gene function	Generally need large data sets hence more relevant for model organisms, predict gene function not regulation	Ransbotyn *et al*. (2014) and Tully *et al*. (2014)
Functional association networks	Genome-wide, cross species predictive ability	Extensive data sets still needed so more useful for model organisms, predict gene function not regulation	Lee *et al*. (2015)
Inferring regulatory relationships			
Static regulatory network model — using extensive genome-wide data sets	Genome-wide, predicts regulatory relationships, network topology can predict gene function	Requires extensive data sets so only appropriate for model organisms, requires known binding motifs for causal relationships	Vermeirssen *et al*. (2015)
Static regulatory network model — using mutant data	Can predict regulatory interactions, can generate a useful network from a limited data set, network nodes do not need to be transcriptionally regulated	genotypes available limit network size, interactions are undirected, network does not predict causal regulatory relationships	Sato *et al*. (2010)
Dynamic regulatory network model — inferred from time series data	Predicts causal regulatory interactions, can be inferred from a single data set, network topology can predict gene function	Genome-wide networks are computationally expensive, requires high-resolution time series expression data	Windram *et al*. (2012)
Investigating regulation and organising principles of networks			
Multiple regression model	Used elegant small-scale data set, data from multiple perturbations, feasible for non-model organisms	Requires some knowledge of the biological process to design suitable experiment, small scale network	Kim *et al*. (2014)
Machine-learning approach	Allows capturing of multiple network states of system, relatively computationally inexpensive, genome-wide	Quality and suitability of gold standard data for classification can impact model outcome	Dong *et al*. (2015)

response. Betweenness centrality is a measure of how important a node is in linking poorly connected parts of a network, and these information bridges are crucial to information flow within the network [9,10].

A remarkable gene discovery rate (for key regulators of abiotic stress responses) was obtained combining co-expression analysis with a gene expression diversity measure [11••]. Using a compendium of expression data following multiple abiotic stress treatments, genes were scored on the basis of how varied their differential expression was across stresses, and on how reproducible expression was within independent experiments of the same stress. A set of high-scoring regulatory proteins (for example, transcription factors (TFs), kinases and phosphatases) was selected and incorporated into a co-expression network. Within this network, modules (Box 1) were ranked based on their expression diversity and the presence of known stress regulators with individual genes within these modules prioritised on the prevalence of homozygous T-DNA knockout lines. Impressively this ranking-modelling method correctly predicted phenotypes for 62% of the 42 regulators. Phenotypic predictions based on the gene's score alone had a success rate of 36% revealing the power of the guilt-by-association network approach.

Condition-dependent approaches rely on sufficient data sets being available but may be more feasible for non-model organisms. Four data sets analysing the citrus transcriptome after *Candidatus* Liberibacter asiaticus (Las) infection were used by Zheng and Zhao [12] to construct a co-expression network. Although predictions from this analysis were not experimentally tested, many of the hub genes were orthologues of known defence regulators in Arabidopsis providing some validation to the network. The majority of condition-dependent co-expression studies focus on experiments from a single type of treatment, whereas in natural environments plants are regularly exposed to multiple stresses simultaneously. A novel combinatorial biotic and abiotic stress study revealed that up to 60% of differential gene expression in dual stress treatments was not observed from expression in single stress treatments [13]. This work highlights how we need to broaden our experimental horizons, and/or predictive modelling abilities, to ensure we are capturing and inferring biological meaning with relevance in the real world.