

Accepted Manuscript

Title: Review: Plant immune signaling from a network perspective

Author: Fumiaki Katagiri

PII: S0168-9452(18)30597-1

DOI: <https://doi.org/10.1016/j.plantsci.2018.07.013>

Reference: PSL 9908

To appear in: *Plant Science*

Received date: 23-5-2018

Revised date: 17-7-2018

Accepted date: 27-7-2018



Please cite this article as: Katagiri F, Review: Plant immune signaling from a network perspective, *Plant Science* (2018), <https://doi.org/10.1016/j.plantsci.2018.07.013>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Review: Plant immune signaling from a network perspective

Fumiaki Katagiri

Department of Plant and Microbial Biology, Microbial and Plant Genomics Institute, University of

Minnesota, St. Paul, MN 55108, USA

katagiri@umn.edu

Highlights

- A hybrid of mechanistic and statistical models is needed for a predictive model.
- The plant immune signaling network is necessarily resilient and tunable.
- Network resilience conceals underlying mechanisms from pathogens and from us.
- The reliability of pathogen attack information likely tunes the network output.
- Balancing network reduction and rule simplification is a key to a predictive model.

Abstract

For predictive biology of a large and complex network, important mechanistic information consists of network topology, signal-convergence rules, and signal dynamics. In practice, the key to enabling predictive modeling of a complex network is reducing it sufficiently to allow modeling without omitting important factors affecting network behavior. Here I argue that the plant immune signaling network must have high levels of resilience and tunability based on the fundamental facts that plants are evolutionarily disadvantaged relative to microbial pathogens and that unnecessary immune response is bad for plants. By reducing the middle part of the immune signaling network to a four-sector network, we previously showed that the middle part indeed has high levels of resilience and tunability and from what signaling strategies the network properties emerge. The ability to comprehensively reconstitute the reduced network enabled a reductionist approach to a resilient network, which was crucial for revealing signaling mechanisms concealed by network resilience. However, this four-sector network may be reduced too much to make a predictive model with relatively simple signal-convergence rules and probably requires deconvolution to generate a highly predictive network model with a set of simple signal-convergence rules.

Keywords

Network properties; resilience; tunability; Arabidopsis; network reconstitution; network modeling

Table of contents

1. The types of models and the types of predictions the models can make
2. Mechanistic information important for a predictive model
3. Practical applications of the network concepts to plant immune signaling
4. Network resilience

5. Network tunability
6. Types of immune response are likely selected according to the reliability of pathogen attack information.
7. Interpretations of PTI and ETI signaling and response from a network property viewpoint
8. Challenges ahead

1. The types of models and the types of predictions the models can make

One major goal of current biology research is predictive biology: we would like to make a computer model that accurately predicts the outcome of the system behavior under conditions previously unexplored. Particularly, we would like to enable such a computer modeling approach with a complex system, meaning a system with complex behavior. There are two extreme types of computer models, which I call a strictly mechanistic model and a narrow-sense statistical model. A mechanistic model explains the behavior of a system based on the principles in lower scales, such as explaining the behavior of a biological system by a mass action law acting on the molecular components of the system. In this example, the mass action law is the lower-scale principle, and the stoichiometry and the association and dissociation rates of each chemical reaction constitute the rule. Rules derived from lower-scale principles are what we call mechanisms of the system. A strict mechanistic model refers to a model in which all rules used are derived only from the lower-scale principles. A strictly mechanistic model requires all mechanistic information of the system to be implemented in the model, i.e., a strictly mechanistic model is the *in silico* copy of the actual system. In contrast, a narrow-sense statistical model does not require any mechanistic information about the system, rather the model is an arbitrary mathematical function that interpolates multiple data points that inform input-output relationships of the system. Linear regression is a simplest form of a narrow-sense statistical model. A more complex form is an implementation of machine-learning in which an arbitrary neural network structure is used.

The two types of models are distinct regarding what type of predictions the models can accurately make. A strictly mechanistic model can make accurate predictions of the system behavior for any input values. A narrow-sense statistical model can make accurate predictions only for input values in ranges where the input-output data used for fitting or training the model were dense. For example, let's assume that we have data of the Arabidopsis hypocotyl length 7 days after germination at light intensities of 15, 30, 45, 60, and 75 $\mu\text{mole}/\text{m}^2/\text{s}$ with many replications. A narrow-sense statistical model

based on this dataset would accurately predict the hypocotyl length at 50 $\mu\text{mole}/\text{m}^2/\text{s}$, but probably not at 2 $\mu\text{mole}/\text{m}^2/\text{s}$, while a strictly mechanistic model should accurately predict the hypocotyl length at any light intensity.

Collecting sufficiently dense input-output relationship data to enable a narrow-sense statistical model may not be that difficult when the model has only one input variable and one output variable (the light intensity and the hypocotyl length, respectively, in the above example). However, the amount of input-output relationship information could explode rapidly when multiple input variables (e.g., the temperature, light quality, and day after germination in the above example) and output variables (e.g., in addition to the hypocotyl length, the cotyledon size and the number of days before true leaves emerge could also be of interest) are involved. In such a case with multiple input and output variables, a strictly mechanistic model would be much more useful. However, unless the biological system is very small, it is unrealistic to obtain all the necessary mechanistic information about the system, so it is not feasible to make a strictly mechanistic model of a large biological system. For this reason, to model a large biological system, we need to seek a hybrid between the two extreme types of models. Roughly speaking, the more mechanistic information is incorporated, the more accurate is the prediction of the model where input-output relationship information in the data is sparse. Hereafter, when predictions made by a model is discussed, I generally refer to predictions where the input-output relationship information is sparse, rather than predictions made by interpolation.

2. Mechanistic information important for a predictive model

What constitutes mechanistic information important for a predictive model when obtaining sufficiently dense data is not realistic? Here I focus on a signaling network as a model representing a biological system of interest. A signaling network can be represented by vertices (components) connected by directed edges (arrows representing signal flows). First, we need to make the network

structure as summarized as possible without strongly affecting the network behavior, so that the complexity of the summarized network structure becomes comparable to the complexity of the data that can be collected. This summarization procedure is called network reduction. An essential piece of mechanistic information in the network structure is the network topology. The network topology is the network structure from which components with only a single input edge are removed (Figs 1A and 1B). This is because we assume a monotonous input-output relationship (the output always either increases or decreases when the input increases) at a vertex and because a single-input vertex cannot cause complex network behavior. On the other hand, a vertex on which multiple input edges converge could generate a complex output profile and thus constitutes an important piece of mechanistic information. A tree-like network, which has only diverging but no converging vertices, is not a complex network however many vertices and edges it has. In contrast, a very small network could be a complex network if it has signal-converging vertices. In practice, the resulting network topology may still be too complex for a large network. In that case, edges with weakly influencing signal flows can be removed, and the topology of this further reduced network can be used. Second, because multiple converging input signals can cause complex output profiles, the rules that govern the input-output relationships at such signal-converging vertices are important. It is common in biology papers that signaling network models are represented by several gene/protein/small molecule component names, which correspond to vertices, connected by arrows for signal flows between them. The models may contain multiple signal-converging vertices without specifying the convergence rules. Although such models are typically used to explain experimental observations, the models with multiple signal-converging vertices can also explain different potential outcomes by assuming different rules at the converging vertex (Figs 1C-1F). Models without specified signal-convergence rules do not have the power to discriminate among multiple potential outcomes and, therefore, constitute weak hypotheses. Third, signal dynamics are important pieces of mechanistic information. Interesting behavior of a signaling network is typically

associated with its dynamic state, but not with steady states: we are mainly interested in how the network behaves when a triggering signal(s) is given to the network. If the timing of multiple signals reaching the signal-converging vertex is changed, the output time-course profile could be changed (Figs 1G and 1H).

In summary, the network topology, the signal-convergence rules, and the signal dynamics constitute important pieces of mechanistic information for a signaling network. The molecular identities of components in signaling pathways have been a focus in molecular biology and genetics. Knowing the molecular identities of vertices could provide practical means to perturb the network at specific vertices, such as knocking out the gene corresponding to a particular vertex to remove the vertex from the network. However, the molecular identities themselves do not constitute important mechanistic information for the purpose of predictive biology. Note that as long as we know the network topology, signal convergence rules, and signal dynamics, whether a particular vertex represents a specific protein kinase or is simply named “vertex A” does not make any difference in predicting the network behavior. This is the beauty of abstraction using network representation and thinking.

3. Practical applications of the network concepts to plant immune signaling.

Making predictions is highly challenging with the plant immune signaling network - hereafter I will mainly focus on the Arabidopsis network, for which the most mechanistic information and relevant data are available. First, the network consists of hundreds, if not thousands, of genes, proteins, and small signal molecules [1]. We need to employ network reduction. The fundamental challenge is that when the network structure at the scale of genes, proteins, and small signal molecules is not known, we do not know the network topology. Furthermore, even if the network topology were known, it could still have too many vertices to be tractable because the plant immune signaling network is highly interconnected [2].

As a practical approach we focused on four signaling sectors (subnetworks), the jasmonate (JA), ethylene (ET), PAD4, and salicylate (SA) signaling sectors. This was because each signaling sector can be removed from the network by a mutation in its hub gene and because simultaneous removal of the four signaling sectors abolished most inducible immunity in two well-defined cases [3]. We do not think that the entire immune signaling network can be reduced to the four-sector network. Rather, we interpret our results as indicating that the four-sector network forms a relative bottleneck in the immune signaling network while there are substantial parts of the network acting upstream and downstream of the four-sector network (Fig 2A) [4]. In short, we focused on a middle part of the immune signaling network and reduced the middle part to a network consisting of four vertices. Technical advantages of the four-sector network are that the four vertices can be experimentally perturbed individually or in combinations and the output of each sector can be measured by the associated hormone levels or marker gene mRNA levels. These advantages enabled us to collect data that supported inference of network topology, signal-convergence rules and signal dynamics in the four-sector network [5]. Although the precision of the inference was limited mainly due to limited time resolution of the data, we were able to identify the mechanisms underlying two important properties of the network, resilience (formerly we used the term robustness for this property) and tunability.

4. Network resilience

Network resilience is defined by the stability of the network output when part of the network is disabled. In the immune signaling network, pathogens not only trigger the network response but also produce effectors that attack the network. This pathogen attack differentiates the immune signaling network from other signaling networks (Fig. 3). Many pathogens are microbes and can evolve much faster than plants. Therefore, the immune signaling network cannot use adaptive evolution as a main mechanism to fend off pathogen assault: straightforward evolutionary arms-races do not favor the

plant. The network needs to have some mechanism to buy time to make co-evolution with pathogens feasible. As discussed below, network resilience can provide a time-buying mechanism, and thus, the plant immune signaling network needs to be highly resilient.

Network resilience can effectively conceal the underlying signaling mechanism from pathogen evolution by presenting a deceptive selection landscape to pathogens (Fig 4). We showed that both JA-mediated signaling and SA-mediated signaling (the latter acting with PAD4-mediated signaling) positively contribute to two types of inducible immunity against the bacterial pathogen *Pseudomonas syringae* [3]. However, the positive contribution of the JA-mediated signaling was weaker. In addition, JA- and SA-mediated signaling negatively regulate each other. This compensating relationship between JA- and SA-mediated signaling results in the selection landscape of the pathogen, which is a summation of JA- and SA-mediated signaling contributions (red curve in Fig 4). Note that in this figure, lower parts of the landscape curve represent better fit of the pathogen as the landscape curve represents the overall plant immunity level. Through rapid adaptive evolution, often including horizontal transfer of useful effector genes, the pathogen manipulates the plant's JA-SA balance toward the minimum of the selection landscape. Although the plant is more susceptible to the highly adapted pathogen strain, which corresponds to the minimum of the selection landscape, than the original strain, a substantial level of immunity is yet to be compromised by the pathogen ("concealed immunity" in Fig 4). It would be difficult for the pathogen to break the concealed immunity because the pathogen would need to "decipher" the underlying compensatory relationships between JA-mediated and SA-mediated immunity and manipulate the JA and SA sector activities independently, instead of just manipulating one of the JA and SA sectors to shift the JA-SA balance. This difficulty in breaking the concealed immunity could buy time for the plant to co-evolve a specific counter measure to this pathogen. Not just the JA and SA sectors but multiple immune signaling mechanisms negatively regulate others [2]. Thus, it is conceivable that the actual selection landscape presented to a pathogen is much more complex than that in Fig 4,

and the underlying immune signaling mechanisms are likely much better concealed than the situation illustrated in Fig 4.

Network resilience conceals the underlying signaling mechanisms not only from pathogen evolution but also from us, researchers. Imagine that a researcher tries to study the network by completely removing the SA contribution in Fig 4. The effect on immunity is not just loss of the SA contribution but also compensation by the JA contribution as JA signaling is released from inhibition by SA signaling. Without knowledge of a positive JA contribution and JA signaling inhibition by SA signaling, the researcher cannot make a correct mechanistic interpretation of the result. This demonstrates that the mechanistic interpretation of the effect of removing one vertex from a complex network is not simple: it is not like studying a pathway with no converging signals, where the function of a vertex is the opposite of the effect observed by removing the vertex. This is the reason we took the opposite approach in studying the four-sector network. Instead of comparing the effect of removing sectors to the intact (wild-type) network, we compared the effect of restoring sectors to the completely broken (quadruple mutant) network. We call this approach signaling allocation analysis [3] or, more recently, network reconstitution [6,7]. By network reconstitution, the independent effects of single sectors and the interactions among the sectors are quantified, which allows much simpler mechanistic interpretations of the mechanisms underlying the network behavior.

We have shown that transcriptome response to an immune elicitor is highly resilient against perturbations to the signaling sectors [6]. Using the same set of exhaustively combinatorial sector perturbation lines of Arabidopsis, the level of resilience in other molecular processes involved in inducible immunity can be determined as well.

5. Network tunability

A simple way to build a resilient network is to make a network with switch-like behavior (i.e., a network with only two output states, ON or OFF). For example, various networks that mediate developmental programs often use networks with switch-like behavior. However, the immune signaling network needs more quantitative control of its outputs. This is because immune response is expensive so that misfiring of immune response would result in a strong negative impact on plant fitness. At the same time, it is difficult to determine when and what level of immune response is needed, as the reliability of pathogen attack information is not perfect. The reliability of the information is particularly limited at the early stages of interactions with microbes. If strong immune response were induced every time microbes, whether pathogens or non-pathogens, are detected by a plant, the negative impact of misfiring it against non-pathogens would surpass the benefit of completely stopping occasional pathogen attacks. Even more subtle control of immune response should be beneficial: plants that induce just enough immune response according to the potency of pathogens would have better fitness.

The level of the outputs and the spectrum of the outputs need to be tuned. Roughly speaking, a single large immune signaling network deals with a wide variety of pathogens. This large signaling network controls hundreds of molecular defenses. A spectrum of defenses is defined by information about the induced level of each defense. While there is no clear-cut specialization of the immune signaling mechanisms to particular pathogens, the spectrum of defenses could be shifted in one way or another. Figs 2B and 2C illustrate shifts in a spectrum of defenses.

We showed that the four-sector network has multiple input ports (shaded circles on the top of the 4-sector network boxes in Fig 2) and that the level and timing pattern of inputs to the multiple input ports can tune the output level of this highly resilient network (Figs 2B and 2C) [5]. It is likely that modulation of the signal input pattern can tune the spectrum of defenses as well since the major defense genes regulated by different signaling sectors are different [8].

An important difference between situations in which immune response is triggered by well-defined chemical signals or infection by a pathogen is that with a pathogen, multiple pathogen attack signals of varying level and timing can be recognized by plant cells. This pathogen signal pattern can be used as a signature to classify the type of the pathogen. Then, the upstream network can generate the input pattern to the multiple input ports of the four-sector network according to the pathogen signal signature (Fig 2D). This computation of the pathogen signal pattern to the four-sector network signal input pattern is likely selected probabilistically: the level of fitness loss caused by the pathogen if effective plant immunity is or is not induced and the level of fitness loss associated with misfiring of immune response are weighted by various probabilities, including the probability that this type of pathogen attack occurs in the natural environment and the probability that this classification of the pathogen type is correct. The network tunability based on patterns among multiple input signals can enable such probabilistic selection of the output tuning.

6. Types of immune response are likely selected according to the reliability of pathogen attack information.

Pattern-Triggered Immunity (PTI) and Effector-Triggered Immunity (ETI) are two well-characterized modes of plant immunity, which are generally effective against biotrophic pathogens and hemibiotrophic pathogens in the early biotrophic phase [9]. PTI and ETI are defined by the types of molecules that trigger them [10]. PTI is triggered when molecular patterns directly or indirectly derived from microbes, such as structural molecules of microbes, products of microbe-originated enzymes, and plant molecules produced in response to damage, are recognized by plant membrane receptors called pattern recognition receptors. ETI is triggered when pathogen effectors, which are delivered by pathogens and can inhibit plant immune signaling or promote pathogen virulence in other ways, are directly or indirectly recognized by resistance (R) proteins inside the plant cell. Since the receptors for

PTI and ETI are of different types and have different cellular localization, the upstream immune signaling networks are likely different between PTI and ETI. Although the middle part of the network, the four-sector network, is highly shared between PTI and ETI signaling, the four-sector network is used differently in PTI than in ETI signaling [3]: the main difference between PTI and ETI signaling from the middle to the downstream network is not the machinery used but how the common machinery is used.

Responses associated with PTI typically include very fast responses (on the order of minutes), but PTI typically takes long time to develop strong immunity (on the order of 1 day). Responses associated with ETI typically take a little longer to initiate (on the order of a few hours) but reach strongly induced levels very rapidly (on the order of another few hours). Although for the sake of simplicity, these typical responses or signaling processes are called PTI and ETI responses or signaling in the subsequent paragraphs, it should be noted that there are PTI or ETI cases in which the immunity mode dissociates from its typical responses. This dissociation is nothing surprising because PTI and ETI are not defined by the types of immune responses. For example, the rice XA21 receptor kinase recognizes a sulfated peptide, RaxX21-sY, from the bacterial pathogen *Xanthomonas oryzae*, and the recognition results in strong immunity to *X. oryzae* [11]. Since XA21 recognizes RaxX21 in the apoplast, this recognition mode is akin to PTI. However, the associated responses are more like typical ETI responses. It clearly shows that the particular types of receptors, either for PTI or ETI, are not hardwired to the particular types of responses. Rather, the response types are likely selected according to the reliability of pathogen attack information. Many molecular patterns that trigger PTI are associated with certain taxa of microbes and do not particularly distinguish pathogens from non-pathogens. For example, a 22 amino acid peptide region in flagellin (flg22) is a well-characterized PTI-triggering molecular pattern in Arabidopsis and well-conserved among Gram-negative bacteria, whether pathogenic or non-pathogenic to Arabidopsis [12]. In contrast, detection of pathogen effectors inside the plant cell that lead to ETI induction indicates that it is very likely a potent pathogen is present. Thus, lower reliability of pathogen attack information is

associated with typical PTI response, and higher reliability is associated with typical ETI response. In the case of PTI triggered by RaxX21-sY, the distribution of RaxX21-sY among bacteria is highly biased toward potential plant pathogens [11]. Thus, the reliability of RaxX21-sY as pathogen attack information is much higher than that of a molecular pattern that is nondiscriminatory between pathogens and non-pathogens, such as flg22, and consequently more ETI-like response and signaling were likely selected in PTI triggered by RaxX21-sY.

I often use a guard-dog analogy to explain typical PTI and ETI responses. PTI and ETI responses are analogous to “Fifi” the Chihuahua and “Max” the German Shepherd, respectively. Both Fifi and Max are dogs (analogous to PTI and ETI using the highly overlapping machinery; in the following, analogous concepts in immunity are shown in the parentheses following the guard-dog analogies). However, since they are extreme representatives among dog breeds, there are many qualitative differences between them (typical PTI and ETI responses are pretty different). Note that there are dog breeds that can be placed between Fifi and Max as well (potentially selectable output levels and spectra of the immune signaling network). Whether the different output levels and spectra of the immune signaling network are fundamentally continuous or whether there are some boundaries among different dynamical domains in the same network is completely unknown at this point.

When you see a suspicious person, who you are confident is criminal and dangerous, you want to send Max: “Go get him, Max!” However, if you made a mistake and actually the person was innocent, you get sued big time (a high cost of misfiring immunity). You want to send Max only when you are highly confident about the person (highly reliable information about pathogen attack). When you are not confident, you should send Fifi and have Fifi yap around the person instead (typical PTI response only). This could be sufficient to chase away a petty criminal (a non-host pathogen). Even when it was a mistake and the person was innocent, the person would just say, “Oh, silly doggie.” (a low cost of misfiring immunity). Usually information about the person is limited, particularly in the early stage, so

sending Fifi first is a smart thing to do. Send Max later if it is necessary. That is what plants usually do: induce PTI first and induce ETI later.

7. Interpretations of PTI and ETI signaling and response from a network property viewpoint.

Since the signaling machinery highly overlaps, it is very likely that PTI and ETI signaling interact. In nature, ETI is usually accompanied by PTI because a microbe that delivers an effector that triggers ETI presents molecular patterns that trigger PTI as well. Of course, a typical natural situation is much more complex: other effectors delivered could interfere with immune signaling for PTI and ETI, which is called Effector-Triggered Susceptibility (ETS) [10]; other microbes and environmental factors could affect immunity; etc. To allow mechanistic interpretations of observations, we should take a strategy commonly taken to simplify a complex problem, i.e., a reductionist approach. That is what we did in [13]. To focus on PTI and ETI but to exclude influence of ETS and other factors, we used a single molecular pattern at a time to trigger PTI and a single pathogen effector at a time to trigger ETI. To simplify the middle part of the immune signaling network, the four sectors were removed leaving a remaining functional sector, which we called the ETI-mediating, PTI-inhibited Sector (EMPIS; Fig 2A).

The things we discovered in [13] are: (1) EMPIS can mediate ETI signaling for hypersensitive cell death response (HR) and ETI-associated gene expression change even when the four sectors are disabled; (2) this ETI signal-mediating function of EMPIS is inhibited by PTI signaling; (3) Whether the ETI signal reaches EMPIS before or after inhibition of EMPIS by PTI signaling depends on the ETI triggering effectors and the consequential ETI signaling rate. Points (1) and (2) explain why EMPIS was not detected in our earlier study in which ETI was triggered by AvrRpt2 delivered from a bacterium [3]: molecular patterns unavoidably presented by the bacterium inhibited EMPIS. Point (3) explained why the HR phenotype caused by different ETI-triggering effectors delivered from a bacterial strain has

different sensitivities to removal of the four sectors, complete loss of ETI responses in the case of AvrRpt2 to no effect on ETI responses in the case of AvrRpm1.

These discoveries can be interpreted in a network resilience and tunability context using the guard-dog analogy. When Fifi is sent first, can we get Fifi's feedback in deciding whether or not to send Max, i.e., use Fifi as a scout? If Fifi is happily running back, saying, "I did it! I chased the person away!", you can send Max to bed and save the cost of a big meal Max would demand after work (If PTI is effective, i.e., not compromised by PTS, it is better to stop inducing ETI to reduce the negative impact of unnecessary immunity). In some cases you may recognize the person as a particularly dangerous criminal based on previous experience (through natural selection). In this case, Fifi is not a match to the person and you should send Max even before Fifi reports back (for some ETI triggering effectors, the signal to induce ETI responses goes through EMPIS before EMPIS has been inhibited by PTI signaling). This interpretation of Point (3) suggests that pathogen effectors that trigger faster ETI signaling are more potent than those that trigger slower ETI signaling.

The idea of using Fifi as a scout also has a risk. Can we trust Fifi's report? Is it smart to make the potentially dangerous decision of sending Max to bed based only on Fifi's report? Fifi may be happily running back because the person gave Fifi a doggie treat but not because Fifi successfully chased the person away (a pathogen effector could fake an effective PTI signal although PTI is actually compromised). A potentially dangerous decision should be made conservatively. The four sectors needed to be disabled to observe the EMPIS function because the four sectors and EMPIS compensate each other for ETI signaling, i.e., four sectors and EMPIS together form a highly resilient network. Thus, in the intact (wild-type) immune signaling network, every signaling sector of the five needs to be disabled simultaneously to abolish the ETI response. It is conceivable that different measures of immune success, such as cell integrity and decrease of the ETI-inducing effector amount, can inhibit different sectors and that ETI response is abolished only when all the different measures report immune success,

which is highly conservative decision making. A highly resilient signaling network can be used for conservative decision making about aborting the function of the network. To test this hypothesis, each signaling sector needs to be isolated first and then measures of immune success that inhibit the sector needs to be identified.

8. Challenges ahead

Our work based on the four-sector network has been successful in discovery of a resilience-concealed sector, quantification of network properties, identification of signaling strategies underlying the network properties, and quantitative interpretations of the network behavior. However, I must admit that it has not generated many new quantitative predictions. I will discuss what is needed to enable a quantitatively predictive model based on the modeling fundamentals discussed at the beginning of this paper.

The four-sector network is not a network reduced from the true network structure. Instead it was defined by experimental data indicating that the four sectors form a relative bottleneck in immune signaling. Thus, the network topology needs to be determined based on the data. Although the network reconstitution approach allowed network topology and signal-convergence rule inferences [5], they were time-dependent, i.e., they were not determined separately from signal dynamics. This is because three pieces of mechanistic information are not independent of each other when inferred from the data. For example, when no significant signal flow from vertices A to B was observed, is it because the vertices are not connected, which is a topology issue, or because no signal was transmitted at that particular time point with the particular trigger, which is a dynamics issue? To untangle three pieces of information, we need to know the exact input and output values of each sector during the entire time range of interest. Defining these is not a simple task.

First, definitions of exact inputs and outputs of a sector are unclear because a sector is an experimentally-defined, highly summarized subnetwork. Since it was not structurally defined, the boundary of a sector within the network structure is not clear. If the boundary were clear, the edges coming into and going out of the boundary could be unambiguously defined as inputs and outputs, respectively, of the sector. Second, since a sector was not topologically defined, there could be feedback and/or feedforward loop motifs within a sector, which can make behavior within a sector complex. Even without such motifs, a signal time-course could be delayed by a single-input vertex, i.e., even a vertex that would be removed from the network topology could affect signal dynamics. A signal-convergence rule at the sector vertex could be very complex since all potential signal modulations made within the sector should be included in the convergence rule.

To overcome these complexities, I think that the four-sector network model needs to be deconvoluted to a model with some more vertices, resulting in a network of vertices with relatively simple input-output rules. We need two things to enable the deconvolution procedure. First, we need to define the lower-scale principle to incorporate the mechanistic modeling aspect of the hybrid model. Since this model has an arbitrary scale, there is no known fundamental lower-scale principle – this is different from modeling at the chemical reaction scale. Since the model is not expected to be strictly mechanistic, the lower-scale principle does not have to be exact, but it just needs to be a good approximation. The principle should be expressed in a particular mathematical function form that is mechanistically justifiable and has as few parameters as possible. Once the mathematical function form is defined, the deconvolution procedure is to make a model with the minimum number of additional vertices (i.e., the most parsimonious network), while each vertex operates based on the mathematical function with a different set of parameter values. The parameter values and the network topology at each vertex constitute the rules. Second, we need data to support the deconvolution procedure, which are input-output relationship data of most, if not all, vertices. How can we collect such data when we do

not know all the vertices in advance? This is where deep profiling technologies, such as mRNA profiling (RNA-Seq), help. In this case, we do not use deep profiling data as causal information (e.g., using the mRNA level of a gene as a proxy of the activity level of the function encoded by the gene) instead we use the data as detailed descriptions of the network state. For the latter purpose, probably mRNA profiling is the best deep profiling technology of choice considering data accuracy, information content, and cost. We anticipate that the mRNA levels of particular genes report the activity levels of specific parts of the network, possibly with some delay. It is conceivable that mRNA profiling data contain input-output relationship data of most vertices. Selecting the genes whose mRNA levels are associated with a particular sector (the marker genes of parts within the sector) is simple. For example, the marker genes within the JA sector are those whose mRNA levels respond to the trigger in the triple mutant in which only the JA sector is functional (triple mutations in the ET, PAD4, and SA sectors) but do not respond in the other triple mutants. Then limited time-series mRNA level patterns among the selected genes are considered to correspond to the activity time-series patterns somewhere within the JA sector.

Toward predictive biology of the four-sector part of the immune signaling network, we need: (1) the simple mathematical function form as the lower-scale principle and (2) dense time-series mRNA profile data from exhaustively combinatorial network perturbations. For (2), it will be helpful if the data are generated with multiple different triggers, which generate different patterns of input signals to the multiple input ports of the four-sector network (Figs 2B and 2C). Then we need algorithms: (3) to find the most parsimonious network when mathematical function form (1) is applied to every vertex; (4) to calculate input-output level values of most vertices in the networks found in (3) based on data (2). Although the four-sector network is only a middle part of the entire immune signaling network, it is already a large and complex network. With the road map discussed above, I envision that a highly predictive model for this large complex signaling network is within reach. Furthermore, success in

predictive biology of this signaling network will lay a framework that can be broadly applicable to predictive biology of other complex signaling networks at any scales.

Acknowledgement

I thank Jane Glazebrook for critical reading of the manuscript.

Funding

This work was supported by the National Science Foundation [grant numbers MCB-1518058, IOS-1645460].

References

- [1] M.S. Mukhtar, A.R. Carvunis, M. Dreze, P. Eppele, J. Steinbrenner, J. Moore, M. Tasan, M. Galli, T. Hao, M.T. Nishimura, S.J. Pevzner, S.E. Donovan, L. Ghamsari, B. Santhanam, V. Romero, M.M. Poulin, F. Gebreab, B.J. Gutierrez, S. Tam, D. Monachello, M. Boxem, C.J. Harbort, N. McDonald, L. Gai, H. Chen, Y. He, J. Vandenhoute, F.P. Roth, D.E. Hill, J.R. Ecker, M. Vidal, J. Beynon, P. Braun, J.L. Dangl, A. Cabral, G. Van Den Ackerveken, J. Bator, R. Yatusovich, S. Katou, J. Parker, G. Fabro, J. Jones, M. Coates, T. Payne, Independently evolved virulence effectors converge onto hubs in a plant immune system network, *Science* (80-.). 333 (2011) 596–601.
doi:10.1126/science.1203659.
- [2] M. Sato, K. Tsuda, L. Wang, J. Collier, Y. Watanabe, J. Glazebrook, F. Katagiri, Network modeling reveals prevalent negative regulatory relationships between signaling sectors in arabidopsis immune signaling, *PLoS Pathog.* 6 (2010). doi:10.1371/journal.ppat.1001011.
- [3] K. Tsuda, M. Sato, T. Stoddard, J. Glazebrook, F. Katagiri, Network Properties of Robust Immunity in Plants, *PLoS Genet.* 5 (2009) e1000772. doi:10.1371/journal.pgen.1000772.

- [4] R.A. Hillmer, F. Katagiri, Toward predictive modeling of large and complex biological signaling networks, *Physiol. Mol. Plant Pathol.* 95 (2016) 77–83. doi:10.1016/j.pmpp.2016.01.009.
- [5] Y. Kim, K. Tsuda, D. Igarashi, R.A. Hillmer, H. Sakakibara, C.L. Myers, F. Katagiri, Mechanisms underlying robustness and tunability in a plant immune signaling network., *Cell Host & Microbe*. 15 (2014) 84–94. doi:10.1016/j.chom.2013.12.002.
- [6] R.A. Hillmer, K. Tsuda, G. Rallapalli, S. Asai, W. Truman, M.D. Papke, H. Sakakibara, J.D.G. Jones, C.L. Myers, F. Katagiri, The highly buffered Arabidopsis immune signaling network conceals the functions of its components, *PLoS Genet.* 13 (2017). doi:10.1371/journal.pgen.1006639.
- [7] F. Katagiri, Network reconstitution for quantitative subnetwork interaction analysis, in: *Methods Mol. Biol.*, 2017: pp. 223–231. doi:10.1007/978-1-4939-6859-6_18.
- [8] C.M.J. Pieterse, D. Van der Does, C. Zamioudis, A. Leon-Reyes, S.C.M. Van Wees, Hormonal modulation of plant immunity., *Annu. Rev. Cell Dev. Biol.* 28 (2012) 489–521. doi:10.1146/annurev-cellbio-092910-154055.
- [9] P.N. Dodds, J.P. Rathjen, Plant immunity: towards an integrated view of plant-pathogen interactions., *Nat. Rev. Genet.* 11 (2010) 539–548. doi:10.1038/nrg2812.
- [10] J.D.G. Jones, J.L. Dangl, The plant immune system., *Nature*. 444 (2006) 323–329. doi:10.1038/nature05286.
- [11] R.N. Pruitt, B. Schwessinger, A. Joe, N. Thomas, F. Liu, M. Albert, M.R. Robinson, L.J.G. Chan, D.D. Luu, H. Chen, O. Bahar, A. Daudi, D. De Vleeschauwer, D. Caddell, W. Zhang, X. Zhao, X. Li, J.L. Heazlewood, D. Ruan, D. Majumder, M. Chern, H. Kalbacher, S. Midha, P.B. Patil, R. V. Sonti, C.J. Petzold, C.C. Liu, J.S. Brodbelt, G. Felix, P.C. Ronald, The rice immune receptor XA21 recognizes a tyrosine-sulfated protein from a Gram-negative bacterium, *Sci. Adv.* 1 (2015) e1500245. doi:10.1126/sciadv.1500245.
- [12] C. Zipfel, S. Robatzek, L. Navarro, E.J. Oakeley, J.D.G. Jones, G. Felix, T. Boller, Bacterial disease

resistance in Arabidopsis through flagellin perception, *Nature*. 428 (2004) 764–767.

doi:10.1038/nature02485.

- [13] N. Hatsugai, D. Igarashi, K. Mase, Y. Lu, Y. Tsuda, S. Chakravarthy, H. Wei, J.W. Foley, A. Collmer, J. Glazebrook, F. Katagiri, A plant effector-triggered immunity signaling sector is inhibited by pattern-triggered immunity, *EMBO J.* 36 (2017) 2758–2769. doi:10.15252/embj.201796529.

Figure legends

Fig 1. Network topology, signal-convergence rules, and signal dynamics are important mechanistic information about a network. (A) The structure of an example signaling network. The shaded vertices show those with single input edges. (B) The network topology of (A). The shaded vertices in (A) are removed. (C-F) Networks with the same topology but different signal-convergence rules. For the sake of simplicity, Boolean networks, in which each vertex can take the value of either 1 ("on") or 0 ("off"). Vertices X and Y are input vertices, and vertices U and V are output vertices. Each row of the tables shows when vertices X and Y take particular values, what the output values at vertices U and V are. The logic operators of "AND" or "OR" are used as signal-convergence rules at two signal-converging vertices U and V. The value of 1 is shaded in the tables, so that it is easy to see that while the input patterns are the same in (C-F), the output patterns are different. (G-H) Dynamic response of network (D). The time unit is defined as the time for a signal to travel one edge length in the network. The initial values of the vertices are all 0 (Time=0). Then the value of input vertex X is changed to 1 permanently, and the value of input vertex Y is changed to 1 transiently (for 2 time units). The only difference between (G) and (H) is that the value change at vertex X was 1 time unit delayed in (H). As a result, the output at vertex V is qualitatively changed: while it is turned on transiently in (G), it never gets turned on in (H).

Fig 2. A conceptual diagram of the immune signaling network. (A) There are substantial networks upstream and downstream of the 4-sector network. Triggers a-d represent molecular patterns or pathogen effectors, which trigger signaling for PTI or ETI. The upstream network feeds multiple signals into the multiple input ports (shaded circles) of the 4-sector network. The four-sector network also feeds multiple signals into the downstream network. There are many outputs from the downstream network, which represent induced molecular defenses. Although we initially thought that the four

sectors form a bottleneck in the middle, later we discovered another sector, EMPIS, for ETI signaling. EMPIS was not detected earlier due to interactions between ETI and PTI signaling. There are early outputs and minor signals, which are independent of the four-sector network. They are represented by dotted arrows on the left in this panel but omitted in the subsequent panels. (B-C) Different triggers could generate different signal input patterns at the input ports of the four-sector network. This difference results in different signal input patterns to the downstream network and eventually results in different spectra of the defenses (Outputs). The differences in the signal input patterns to the next network and in the spectrum of the defense induced levels are represented by different patterns of the widths of the signal flow edges. (D) Combinations of triggers can be used for classification of pathogen types and generate specific signal input patterns. This idea is signified by illustration showing that the signal input patterns and the spectrum of the defenses in (D) are different from the additive effects of (B) and (C) for the signal input patterns and the spectrum of the defenses although the triggers for (D) (triggers a and c) are the union of triggers for (B) and (C).

Fig 3. Comparison between (A) an ordinary signaling network and (B) the immune signaling network. Particular challenges plants face in the immune signaling network are boxed in (B).

Fig 4. A conceptual diagram: two compensating immune signaling sectors can generate a deceptive selection landscape. When both JA (blue) and SA (green) signaling mediate immunity and they are antagonistic to each other (bottom), the selection landscape (red) resulting from the summation of JA- and SA-mediated immunity can conceal some immunity ("Concealed immunity") after adaptation of the pathogen strain (from "Original strain" to "Adapted strain").

Figures

Fig 1

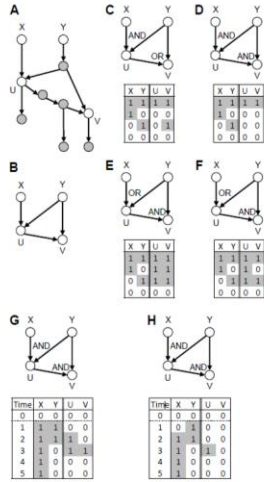


Fig 2

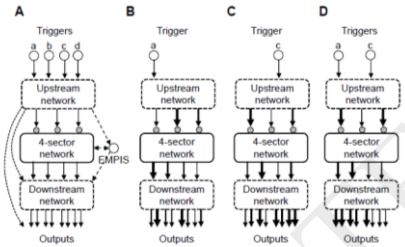


Fig 3

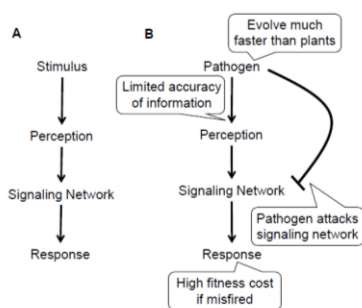


Fig 4

