Methods 131 (2017) 120–127

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Review of combinations of experimental and computational techniques to identify and understand genes involved in innate immunity and effector-triggered defence

Henrik U. Stotz [a,*], Rodrigo de Oliveira Almeida [b], Neil Davey [c], Volker Steuber [c], Guilherme T. Valente [b]

[a] School of Life and Medical Sciences, University of Hertfordshire, Hatfield AL10 9AB, UK
[b] Department of Bioprocess and Biotechnology, São Paulo State University (Unesp), School of Agriculture, Botucatu, Brazil
[c] Centre for Computer Science and Informatics Research, University of Hertfordshire, Hatfield AL10 9AB, UK

ARTICLE INFO

Article history:
Received 14 April 2017
Received in revised form 24 August 2017
Accepted 28 August 2017
Available online 1 September 2017

Keywords:
Breeding
Graph theory
Receptor-like protein
Systems biology

ABSTRACT

The innate immune system includes a first layer of defence that recognises conserved pathogen-associated molecular patterns that are essential for microbial fitness. Resistance (*R*) gene-based recognition of pathogen effectors, which function in modulation or avoidance of host immunity, activates a second layer of plant defence. In this review, experimental and computational techniques are considered to improve understanding of the plant immune system. Biocomputation contributes to discovery of the molecular genetic basis of host resistance against pathogens. Sequenced genomes have been used to identify *R* genes in plants. Resistance gene enrichment sequencing based on conserved protein domains has increased the number of *R* genes with nucleotide-binding site and leucine-rich repeat domains. Network analysis will contribute to an improved understanding of the innate immune system and identify novel genes for partial disease resistance. Machine learning algorithms are expected to become important in defining aspects of the immune system that are less well characterised, including identification of *R* genes that lack conserved protein domains.

Contents

1. Introduction

The innate immune system of plants and animals forms a first layer of defence against invading microbes and parasites [1]. Pattern recognition receptors (PRRs) recognise conserved pathogen-associated molecular patterns (PAMPs) that are essential for microbial fitness to activate this immune system [2]. Interactions between PRRs and PAMPs like flagellin and chitin stimulate mitogen-activated protein (MAP) kinase cascades that regulate transcription factors like WRKYs to elicit PAMP-triggered immunity (PTI) [3]. In addition to this innate immune response pathway, vertebrates have evolved an acquired immune system that includes T-cells, B-cells and antibodies [4]. This type of immunity is lacking from invertebrates and plants.

Adapted plant pathogens can overcome the first line of defence by producing effectors to dampen or evade plant immunity or otherwise manipulate the host. Pathogen effectors target different

* Corresponding author.
E-mail address: h.stotz@herts.ac.uk (H.U. Stotz).

http://dx.doi.org/10.1016/j.ymeth.2017.08.019
1046-2023/

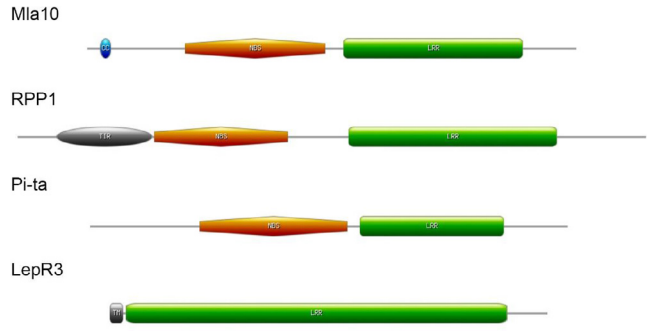

Fig. 1. Examples of the domain architecture of proteins encoded by resistance (*R*) genes. Coiled-coil (CC; blue), nucleotide-binding site (NBS; orange), leucine-rich repeat (LRR; green), Toll-interleukin receptor (TIR; grey oval) domains and a transmembrane region (TM; grey box) are shown. Lengths of proteins are drawn to scale. Mla10 confers resistance of barley against the powdery mildew pathogen *Blumeria graminis* f. sp. *hordei* with the effector Avr_{A10} [93]. RPP1 of *Arabidopsis thaliana* interacts directly with the effector ATR1 of *Hyaloperonospora arabidopsidis* [94]. The rice protein Pi-ta interacts with the AVR-Pita effector from the rice blast fungus *Magnaporthe grisea* [95]. *LepR3* encodes a receptor-like protein that confers resistance of oilseed rape against the phoma stem canker pathogen *Leptosphaeria maculans* [96]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cellular processes in host plants [5]. Some apoplastic effectors bind chitin to protect against host chitinases [6] or interfere with PAMP perception and signalling [7]. Other extracellular effectors inhibit defence-activating proteases [8,9]. Many intracellular effectors interfere with immune signalling [10]. As a result, effector-triggered susceptibility occurs [11]. In turn, plants have evolved receptors encoded by resistance (*R*) genes that interact with corresponding pathogen effectors or their host targets, resulting in effector-triggered immunity (ETI). Molecular identification of *R* genes was dependent on Flor's discovery of the gene-for-gene concept, meaning that host resistance and pathogen virulence are the result of genetic interactions between corresponding genes in the host and the pathogen [12,13]. Several decades after Flor's discovery the first *Avr* (coding for a pathogen effector) [14] and *R* genes [15] were cloned. The gene-for-gene concept has important implications for agriculture because resistance of a given crop cultivar depends on the presence of corresponding pathogen *Avr* genes. Pathogen races with mutated or deleted *Avr* genes are able to overcome crop resistance and therefore challenge agricultural production [16].

R genes encode either cytosolic or transmembrane receptors [17]. Cytosolic receptors contain nucleotide-binding site (NBS) and leucine-rich repeat (LRR) domains (Fig. 1). These domains may be preceded by N-terminal effector domains like coiled-coil (CC) or Toll-interleukin receptor (TIR) domains [18]. Exceptions to this rule are *Pto* and *PBS1* genes, which encode serine-threonine protein kinases [15,19]. Nevertheless, both kinases act in concert with *Prf* and *RPS5*, which encode NBS-LRR receptors, to initiate ETI against bacterial pathogens [20,21].

Compared to cytosolic receptors, information on *R* genes that encode transmembrane receptors is limited. Receptor-like proteins (RLPs) with an extracellular LRR (eLRR) domain, a transmembrane region and a cytosolic tail are involved in effector-triggered defence (ETD) of tomato (*Solanum lycopersicum*), oilseed rape (*Brassica napus*) and apple (*Malus domestica*) against extracellular fungal pathogens [17]. It is not clear to what extent transmembrane receptor kinases are involved in *R* gene-mediated resistance. For instance, the rice gene *Xa21* that confers resistance against *Xanthomonas oryzae* [22] was originally termed an *R* gene but is now considered a PRR; it recognises Ax21, a sulfated secreted peptide that serves as a signal in quorum sensing [23]. Similarly, *Xa3/Xa26*, which encodes a transmembrane receptor kinase, is referred to as a PRR providing broad-spectrum resistance against *X. oryzae* [24]. Adding to this controversy, the barley receptor kinase gene *Rpg1* is claimed to have the recognition specificity of an *R* gene, although the response pattern and timing is more reminiscent of PTI [25]. The corresponding eliciting molecule from spores of *Puccinia graminis* f. sp. *tritici* has not yet been isolated to classify it as an effector or a PAMP.

Based on the mentioned information, challenges and opportunities for computational analysis can be anticipated. Over 50 crop genomes have been sequenced, including staple crops maize, rice, wheat, barley, sorghum, potato, cassava, chickpea and soybean [26]. The degree of uncertainty associating specific proteins and their domains with *R* gene-mediated resistance makes it challenging to generate bioinformatic approaches to comprehensively search genomes for candidate *R* genes. Other challenges and limitations to *R* gene identification maybe related to problems in genome assembly and annotations caused by repetitive LRR sequences [27]. On the other hand, PTI is relatively well understood and computational network analysis may be used to systematically identify genes that are involved in innate defence response pathways and may contribute to partial resistance against pathogens.

For the remainder of this review, computational and experimental methods will be highlighted to improve the definition of the *R* gene complement in plant genomes. A detailed description of computational network analysis is given with its potential to assist in identification of genes that contribute to partial or quantitative disease resistance by combining information about gene and protein expressions, protein interactions, metabolomics and cellular signalling. An example of Machine learning is given to improve the definition of a specific class of *R* genes. This report closes with a consideration of advanced computational methods to incorporate mapped loci for identification of resistance genes. Such approaches will be an asset for plant breeding and crop protection.

2. Prediction of candidate *R* genes by homology- and motif-based searches

Sequenced genomes have been used to search for and functionally analyse all genes of a specific family. Homology searches using sequences of RLPs with eLRR domains, *CLV2* (*CLAVATA 2*), *TMM* (*TOO MANY MOUTHS*) and *Cf-9* (resistance gene against *Cladosporium fulvum*), have been used to detect 57 genes encoding RLPs in the genome of *Arabidopsis thaliana* [28]. Positive identification of RLPs was also based on the presence of a signal peptide, eLRR and transmembrane domains and a short cytoplasmic tail. Nevertheless, use of two characterised RLPs from *A. thaliana* (*CLV2* and *TMM*) and *Cf-9* from tomato may limit the power of detection of these types of proteins.

The MEME Suite web server [29] was used to discover putative NBS-LRR genes in the genome of *B. napus* [30]. Positive and negative training sets were used to identify discriminative motifs as in the case of the potato genome [31]. A total of 426 genes encoding NBS-LRR receptors with CC or TIR domains were identified in the *B. napus* genome [30].

A more extensive phylogenetic comparison of plant and animal proteins with NBS-LRR architectures was recently done [18]. The authors focused on the central STAND NTPase domains of R-proteins and animal NOD-like receptors. A hidden Markov model was used that was sensitive to NB-ARC (PF00931) and NACHT (PF05729) NTPases and other STAND NTPase subclades. The entire search probed 10,565,004 sequences and yielded 15,500 STAND NTPase domain hits. The result was further confined to generate a phylogeny with 964 STAND NTPase sequences. The presence of

this conserved domain in cytosolic R-proteins makes their definition and analysis relatively straightforward. By contrast, RLPs are ill defined containing a highly divergent eLRR domain, which makes research on these transmembrane receptors much more challenging.

Definition of *R* gene families in sequenced genomes can assist in cloning of resistance genes. However, only a subset of the RLP genes is expected to be involved in resistance against pathogens as others, including *CLV2* and *TMM*, are involved in development. Dual function in immunity and development may also not be excluded because the *Toll* gene of *Drosophila melanogaster*, encoding a transmembrane receptor with eLRR and cytosolic TIR domains, serves both functions [32].

3. Experimental approaches to expand the list of candidate *R* genes

R gene enrichment sequencing (RenSeq) was developed to increase the number of identified NBS-LRR genes [27]. This experimental approach overcomes the limitation of computational gene model predictions. The authors used the Agilent SureSelect Target Enrichment System to generate >48,000 120-mer biotinylated oligo-nucleotides based on 589 NBS-LRR sequences. Sequencing of the enriched library increased the identified NBS-LRR genes from 438 to 755 in the genome of potato (*Solanum tuberosum*). The newly identified NBS-LRR genes were used for SNP calling to narrow map locations of two *R* genes against the oomycete *Phytophthora infestans*. RenSeq has contributed to revision of gene models, correction of misassembled sequences, sequence gap closures and identification of deletions within NBS-LRR sequences. More recently, mutagenesis was combined with exome capture and bioinformatics (MutRenSeq) to clone two genes for resistance against the stem rust pathogen *Puccinia graminis* f. sp. *tritici* from hexaploid wheat [33].

Similar approaches ought to be applied to RLP genes to identify transmembrane receptors that interact with extracellular pathogen effectors. However, the lower conservation of eLRR domains relative to NBS-LRR sequences remains an obstacle. Another impediment is a more substantial lack of knowledge regarding transmembrane receptors that are encoded by *R* genes compared to *R* genes that code for intracellular NBS-LRR receptors.

4. Omics and network analyses to define the innate immune system of plants

Interactions between plants and microbes are not only controlled by signal exchange between partners but also influenced by a variety of environmental factors, including temperature and light [34]. These Interactions are dynamic and exceedingly complex. Detailed understanding of such interactions therefore requires omics and systems biology approaches.

The intention of systems biology is to obtain a holistic view of biology considering the relationships between all biotic and abiotic components [35]. Cells produce thousands of macromolecules with many interactions between them (Fig. 2A and B) generating transcriptional regulatory networks, signal transduction networks, protein-protein interaction networks (PPIs) and metabolic networks [36] (Fig. 2C-E). The gene regulatory networks (e.g. interactions between transcription factors and *cis*-regulatory elements) [37,38], protein-protein/DNA/RNA/enzyme/metabolite and miRNA-mRNA networks are physical interactions [37–40] (Fig. 2B and C). PPIs are most frequently represented by undirected edges [41] while other interactions are represented by directed edges [42] (Fig. 2D and E). In contrast, genetic interaction networks include functional consequences of two or more mutations in a cell

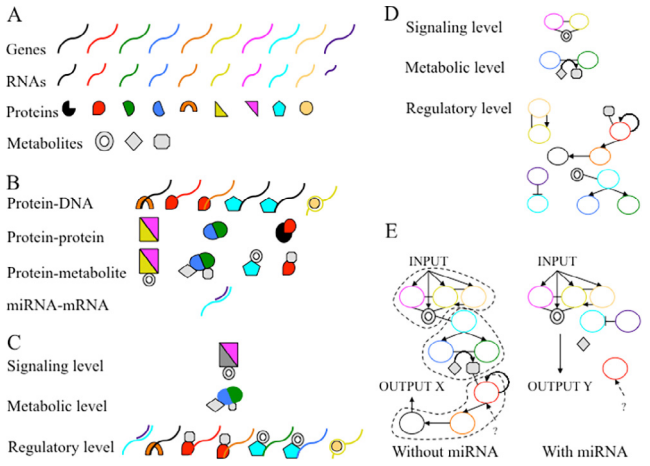

Fig. 2. Synthesis of systems biology at the molecular level. Here symbols and colours are representing genes, proteins, metabolites and graph nodes. A) Concepts of gene and protein expression and metabolism. All genes are coding sequences with exception of a non-coding RNA (in purple). Metabolites are shown as grey symbols. B) Different interactions observed in a cell. Transcription factors (dark orange, red or cyan symbol) are bound to promoters. The light orange symbol represents a histone that is involved in epigenetic regulation of a gene (in yellow). PPIs are also shown (yellow with pink, blue with green and black with red symbols). Four proteins bind or convert metabolites. One PPI (blue and green symbols) forms an enzyme complex that converts a substrate (diamond symbol) into a product (rounded box). The other proteins bind metabolites as part of a signalling process, e.g. binding of a metabolite or co-factor to a transcription factor. Only one microRNA-mRNA interaction is represented. C) Three different layers at the level of the molecular system. D) Graph visualization at the level of molecular systems. Directed edges (i.e. arrows) may indicate activators, repressors or metabolic conversions. Non-directed edges are also indicated. The light orange node (histone) is linked to the yellow node by two types of edges, a non-directed edge meaning physical interaction and an arrow meaning induction of gene expression (yellow symbol). E) Integrated view of sub-graphs. The left model results in an output X achieved in the absence of miRNA (purple node) from a signalling cascade that starts with a physical PPI (pink and yellow nodes) and a metabolite (open grey circle). The right model represents the same cascade but in presence of miRNA, which regulates translation of mRNA (cyan node). As a result, output X is not observed in the presence of miRNA because gene expression (blue and green nodes) is prevented in the absence of transcriptional activator (cyan node). The enzyme complex therefore does not form and metabolite conversion (diamond to rounded box) does not occur. This results in a lack of co-factor for another transcription factor (red node). Consequently, there is no positive feedback, resulting in absence of downstream gene expression (dark orange and black nodes). The final output Y is produced. Dashed circles illustrate the concept of communities. Question marks are unknown inducers of gene expression (red node). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

[42]. Whole interactions can be split into sub-systems [43] and communities [44] (Fig. 2B-E) using a hierarchical structure [43,45,46]. The systems dynamically change in response to different stimuli and evolution [43] with input (stimulus) influences on the system and outputs as the responses (Fig. 2E). Thus, systems biology is important (i) to generate predictions in a changing environment considering cellular responses to external or internal stimuli and (ii) to establish new hypotheses that can be experimentally tested [47,48].

Moreover, after completion of the first genome sequences the scientific community asked how this information is used to create organisms and how molecules work together to achieve complex tasks. Consequently, high throughput ‘omic’ technologies (e.g. transcriptomics, proteomics, metabolomics, lipidomics and fluxomes) have become the main source of data for systems biology at the molecular level [49]. Integration of multiple ‘omic’ datasets is important to better understand complex processes because they have complementary information, a good example of such integration being work on *M. genitalium* life cycle modelling [50].

Table 1
Descriptions of general molecular biology and computational approaches of original researches papers (reviews not included) about systems biology of innate immune systems in plants. The top 10 for each column are described in the bottom of table.

Molecular approaches[a]	Omics[a]	Global bioinform.[a]	Software/algorithms	Databases	Obs.[b]	Cite
BA; CHR; RNAi; Blot; qRT-PCR	–	–	Image J	TIGR Solanaceae Genomics Resource	1,2	[52]
BA; qRT-PCR	MAR	NPA	Image J; LEGG/EdLEGG/RepEdLEGG	Omnibus	–	[54]
–	ChIP-chip; MAR	DBC; ML; NA	BiNGO; Cerebral; Cytoscape; Enrichment Map; GSEA; MCL; Network Guided Forest; NetworkX; TileMap	AGRIS; BIOGRID; European Bioinformatics Institute ArrayExpress; IntAct; Omnibus; PubMed; SUBA3; TAIR; Weigel World Web	–	[55]
BA; qRT-PCR	MAR	DE	GeneSpring GX; Image-Pro Plus; MxPro; Robust Multiarray Average summarization	GenBank; NASCArrays	–	[58]
BA; qRT-PCR; METAB	MAR	CLU; NPA	ARANET Web tool; Analyst; GeneSpring GX; MAPMAN; VANTED	–	–	[59]
BA; qPCR	MAR	CLU; DE; NPA	AgilentAT6; BiNGO; GO-slim; R; WGCNA; biomaRt data-mining; pdInfoBuilder	Arabidopsis Information Resource; Arabidopsis Information Resource; AtGenExpress; Omnibus	–	[60]
–	MAR	DBC; NPA	ARACNE; BLAST; CLR; ClusterMaker; Clusterviz; Cytoscape; FAG-EC; MRNET; Markov Cluster Algorithm; POBO; R	DAVID; TAIR	–	[62]
BA; RT-PCR	MAR; RNA-Seq	CLU; DE; NA	ATCOECIS; AtRegNet; BiNGO; CLR; CORNET; Cytoscape; DESeq; Image J; LeMoNe; ModuleViewer; R; TwixTrix; Web MicroRNA Designer; edgeR	AGRIS; AraCyc; AraNet; ArrayExpress; NASCArray; Omnibus; PLACE; PLAZA; Plant TFDB; TAIR	–	[64]
BA; RT-PCR	MAR	NPA	BLASTP; INPARANOID	BIND; BIOGRID; IntAct; InterPro; KEGG; MINT; NCBI; TAIR; WormNet; YeastNet	–	[66]
BA	MAR	CLU; DE; NPA	Cytoscape; Expander; Genomic Research Multiple Experiment Viewer; R; ROBIN; TANGO	AGRIS AtTFDB; ArrayExpress; DATF; Omnibus; PlantsPP	–	[67]
–	MAR	CLU; GE; NPA	BioLayoutExpress 3D; GOstats; Graphviz; Markov Cluster Algorithm; R; arrayQualityMetrics; gcRMA	ArrayExpress; AtGenExpress; TAIR	–	[68]
–	MAR	CLU; GE; NA	–	TAIR; Omnibus	–	[69]
–	MAR	CLU; GE; NPA	AtRegNet; Cytoscape; FANDOM; InferGene	AtGenExpress; NASCArrays; TAIR	–	[70]
MS; qRT-PCR	–	NMA	–	–	1	[71]
Blot	RNA-Seq	DE	BEDTools; Generic GO Term Finder; R; TopHat	Omnibus; GenBank	–	[72]
BA; Blot; CHR; qRT-PCR	MAR	DE; HCLU	Affymetrix GeneChip Operating; Genesis; Genevestigator; R	Omnibus	–	[73]
Blot; qRT-PCR	MAR	CLU; DE	Cluster; NimbleGen; R; Treeview	Omnibus	–	[74]

Top 10

Freq. Item[c]	Freq. Item[c]	Freq. Item[c]	Freq. Item[c]	Freq. Item[c]
9 BA	14 MAR	8 NPA	8 R	9 Omnibus
7 qRT-PCR	2 RNA-Seq	8 CLU	5 Cytoscape	7 TAIR
4 Blot	1 ChIP-chip	7 DE	3 Image J	3 AtGenExpress
2 RT-PCR		3 NA	3 BiNGO	3 ArrayExpress
2 CHR		3 GE	2 Markov Cluster Algorithm	2 NASCArrays
1 qPCR		2 DBC	2 GeneSpring GX	2 IntAct
1 RNAi		1 NMA	2 CLR	2 GenBank
1 MS		1 ML	2 AtRegNet	2 BIOGRID
1 METAB		1 HCLU	1 pdInfoBuilder	2 Arabidopsis Information Resource
			1 gcRMA	2 AGRIS

[a] Abbreviations: BA, bioassay; CHR, chromatography; CLU, clustering; DBC; database creation; HCLU, hierarchical clustering; MAR, microarray; METAB, metabolome; MS, mass spectrometry; DE, differential expression analysis; GE, gene expression; ML, machine learning; NA; network analysis; NMA; networks modelling/analysis; NPA; networks prediction/analysis;
[b] Obs. 1, not global discussions concerning networks; Obs. 2, analysis on *Solanum tuberosum*.
[c] The numbers before the abbreviations in the top 10 description relates to the number of papers using them.

Phytohormones, including salicylic acid (SA), jasmonic acid (JA), and ethylene (ET), are involved in plant immunity against pathogens. SA signalling contributes to immunity against biotrophic and hemibiotrophic pathogens, which require living host cells for colonisation. Conversely, JA/ET signalling contributes to defence against necrotrophic pathogens that kill host cells during invasion [51]. The expression analysis of specific genes of *Solanum tuberosum* demonstrated that sometimes both SA and JA hormones are produced and act together on induction of defence responses [52]. In *A. thaliana* the SA, JA and ET can create synergistic or antagonistic networks to regulate plant immunity against pathogens [53] (Table 1).

Expression profiles can be used to construct a network model for host immunity against pathogens. A weak regulatory relationship was detected in signalling of the immunity network after challenging *A. thaliana* with *Pseudomonas syringae* containing the effector *AvrRpt2* [54] (Table 1); using 22 known immune signalling components representing different sectors, such as MAP kinases, nitric oxide (NO), reactive oxygen species (ROS), callose and phytohormone sectors, regulatory relationships were modelled based on similarities with mRNA expression profiles of *A. thaliana* mutants deficient in one of these network components. This resulted in a static model with extensive negative regulatory relationships between signalling sectors during ETI in response to *AvrRpt2*. Importantly, predictions from these models were experimentally validated, highlighting the significance of this modelling approach to generate testable hypotheses.

Machine learning approach (Network Guided Forests) associated with systems biology was used to analyse transcription factors in *Arabidopsis thaliana* combining protein-DNA interactions, PPIs,

chromatin modification and co-expression data at multiscale-comparisons of PTI and ETI. The results suggest coordination among network modules to maintain robustness of the immune response [55] (Table 1). Other results reinforce that PTI and ETI network acts synergistically [56,57].

Integration of biotic and abiotic stimuli for gene regulation is another important aspect. Plant immunity can be affected by environmental factors [34], suggesting that optimized responses of plants are dependent on integration of complex information. However, it is not completely understood how plants integrate this complex information into a response. Complex transcriptional responses of *A. thaliana* to single and multiple stresses were assessed [58–60] (Table 1) and it was concluded that multiple stresses activated a combination of specific transcriptional responses that contributed to stress tolerance in plants [61].

Modelling can be used to predict new functions for genes. Using several datasets, networks can be constructed to link genes and use a guilt-by-association strategy to predict gene function. Different strategies of network modelling have been used to study plant defence responses, including co-expression networks [62], functional association networks [63], static regulatory networks using genome-wide datasets [64], dynamic regulatory networks using time series data [54] and regulation and organising principles of networks using multiple regression models or machine-learning approaches [55]. The co-expression concept can be extended using functional association networks and incorporating multiple large-scale data sets, e.g. transcriptome data, experimental information on PPIs, protein sequence and gene-gene association data from organisms, thus enhancing predictive ability [65]. Using this approach, a remarkable network was generated able to predict the role of novel genes in seed pigmentation, drought tolerance and lateral root formation of *A. thaliana* [66]. In another study, high-scoring regulatory proteins (such as transcription factors, kinases and phosphatases) were fused into a co-expression network, enhancing correct prediction of phenotypes of known stress regulators from 36% to 62% [67]. Co-expression datasets of *A. thaliana* were used to show that biotic stress responses and hormone treatments induced several network modules and that hormones modified a module related to defence regulation, while effector proteins from *P. syringae* repressed two other modules [68]. Within these modules, genes of unknown function potentially contribute to plant defence responses. A large collection of expression datasets on multiple tissues, treatments and mutant genotypes was used to infer regulatory relationships between *A. thaliana* transcription factors and their target genes, generating a model that can be used to investigate the plasticity of the transcriptional network [69,70] (Table 1).

Taken together, the papers using different molecular biology, omics and systems biology approaches report on (i) the roles of signalling sectors in innate immune responses, (ii) the importance of multifactorial analysis for system modelling, (iii) the role of network modules, the influence of gene, protein, metabolite, abiotic factors and interaction on biological systems, (iv) new gene/protein candidates for further experimentation, (v) antagonistic or synergistic actions on PTI and ETI and (vi) the robustness of immune systems [51–75]; sometimes databases and software were created [55,62]. The papers share common features such as use of qRT-PCR (to search for candidates or validation), microarray data (RNA-Seq has not yet become the most used technique for transcriptomics), network predictions, clustering and differential expression analysis, use of R packages, Cytoscape, BiNGO and large databases such as Omnibus and TAIR; it shows that the community deal with different technologies, theories, tools and approaches to better understand biological phenomena. Many authors prefer to make their own network predictions (mainly co-expression networks) than to use previously established ones, which suggests a large interest in finding new edges/nodes important for biological systems (Table 1). Many omics data are available for plants such as *Arabidopsis thaliana* (TAIR database) [76], *Brassica napus* (BRAD database) [77], *Solanum lycopersicum* (SOL Genomics Network) [78], *Oryza sativa* (Rice DB database) [79], *Triticum aestivum* (GrainGenes database) [80] and *Zea mays* (MaizeGDB database) [81]. However, usually innate immune system studies using omics/systems biology have only been done in *Arabidopsis* (Table 1) and not yet in crop species.

The innate immune system of plants is complex. As with all other complex phenotypes, systems biology may be used to better understand biological processes because a holistic view can provide important additional insights. To our knowledge, systems biology of innate plant immunity is not thoroughly investigated. Omics and systems biology can be combined to study the innate immune system to generate high-throughput data and provide information about interactions and changes in response to different stimuli. Such studies are expected to generate new concepts for research on plant innate immunity.

5. Machine learning to identify *R* genes lacking conserved domains

Machine learning is a large topic covering aspects of how a machine can modify its response(s) according to its previous experiences [82]. In bioinformatics the aspect of automated machine learning that is most often used is that of data analysis. In this case it is usual that a large quantity of data is present and the machine is programmed to look for patterns in the data that can lead to a better understanding of its structure, which in turn may facilitate the interpretation of any new data from the same domain.

In machine learning there is a distinction made between so-called supervised and unsupervised learning, although this distinction does not necessarily place problems into two non-overlapping classes. In both cases the goal is to gain an understanding of any structure implicitly present within the data, which may in turn lead to novel interpretations of new (previously unanalysed data).

In general, the data used in such analyses consists of a set of multidimensional vectors, where each vector represents a single data item. The vectors are normally of the same arity (size). For example a vector may represent a sequence of DNA according to some measured features of such a sequence. Given such a set of feature vectors, known as an unlabelled data set, it is possible to visualise the data in 2 dimensions, for example by using a principal component projection, or to look for natural groups in the data. This is often done using standard statistical methods such as agglomerative clustering. More recently it has been proposed that techniques from machine learning such as Self Organising Maps [83] or Neural Gas [84] can be used. It should also be noted that Non-Negative matrix decomposition is a technique with a firm mathematical footing that is becoming increasingly applied to data analysis and mining [85].

However when each item of data, as represented by its feature vector, has an associated label, such as "possible *R* gene" then the data is said to be labelled and so-called supervised learning method can be used. Here an algorithm is used to find decision boundaries between two or more classes. The aim of this is not just to characterize the classes of the existing data, but to be able to extrapolate the likely class of previously unclassified data. We have used this technique in another related bioinformatics project [86]. There are many machine learning tools that can be used, such as Support Vector Machines, Random Forests and Bayesian methods. But in all cases the process is not straightforward, beset by problems around overfitting and underfitting the data [87]. In general a good model of the data can be learnt if there is a good quantity

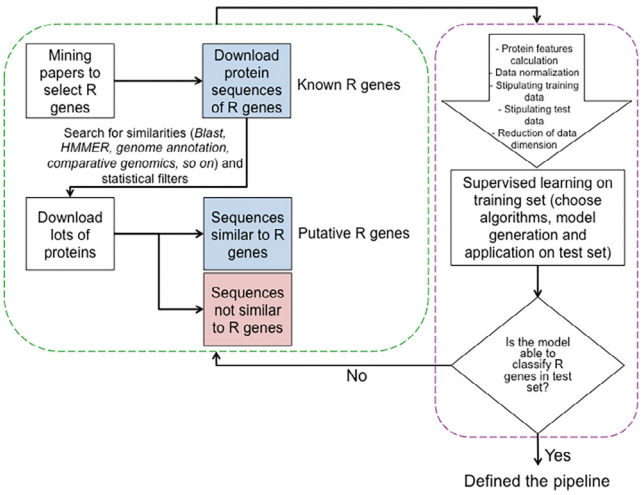

Fig. 3. Machine learning pipeline to create a model to classify *R* genes. Green rounded box, database preparation; purple rounded box, machine learning pipeline; blue boxes, positive examples; red box, negative examples. The final question of this pipeline is whether the model is able to classify *R* genes of the test set (the unlabelled instances). If the classification is not satisfactory (due to overfitting, low true positive rate or other reasons), pipeline improvement is induced based on new data acquisition for subsequent machine learning procedures. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of high quality data: that is data without missing values, consistent labels and with all classes being well represented. If this is the case then high quality predictions can be made. In fact, the creation of a model to classify *R* genes using supervised learning can be related to problems of data quantity (lack of large amounts of positive data (the *R* genes) while lots of negative data are easily found), which may be solved using many bioinformatics tools and databases (Fig. 3). Machine learning algorithms have recently been developed to expand on the identification of candidate *R* genes [88,89], but they have not extensively been tested or applied to sequenced plant genome datasets.

Whilst it is often easy enough to identify positive examples of the class of interest (e.g. disease-related genes, regulatory nucleotide sequences), it is sometimes difficult to identify negative examples. This may be due to uncertainty or simply lack of data. Nonetheless for a trainable model to be successful it is usually the case that both classes (positive and negative) are well sampled. This will be the case if "near missies" are well represented as such cases are particularly useful in determining high quality decision boundaries. An extended discussion of this issue has been published [86].

6. Concluding remarks and outlook

Until now, molecular biological and 'omic' strategies have been used to investigate many biological problems. As these technologies generate large amounts of data, comparative genomics and systems biology are state-of-art to generate global snapshots and show useful biological information. However, patterns in genomics and systems biology are not easily understood, and machine learning is an interesting tool to clarify these patterns. All methods described here aim to generate models useful to explain phenomena, create new hypotheses to be tested experimentally or provide information for technological improvements. Different approaches have provided results that converge to generate knowledge in different areas (Fig. 4). Some areas depend on direct data from another area (e.g. bioinformatics requires molecular biology or omics), in spite of some relationships not being direct. Studies on

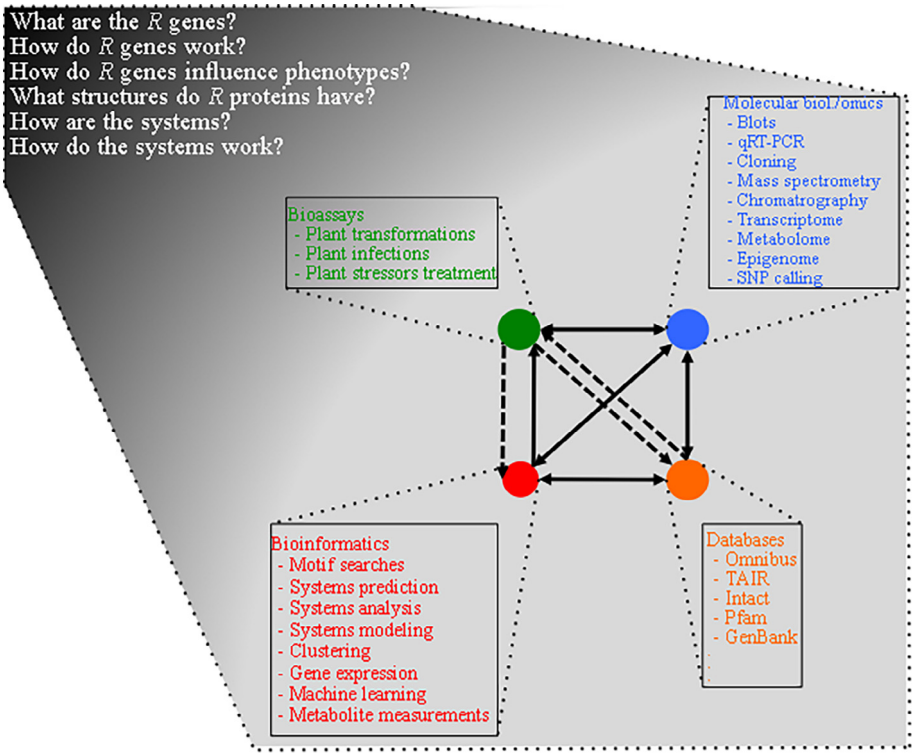

Fig. 4. Overview of methods and questions that motivated studies on plant innate immunity. Arrows indicate relationships between bioassays (e.g. phenotypic analysis of transgenic plants, responses to stress treatments), molecular biology/omics, bioinformatics and public databases. Dashed arrows mean that results can be used indirectly to fill knowledge gaps in the target node/area. For instance, bioassays generate results that can contribute to databases, however, most of data in the latter are derived from bioinformatics or molecular biology/omics. On the other hand, bioassays experiments can be designed based on direct information from databases, bioinformatics and previous molecular biology/omics information. Blot include immunoblots and northern blots.

plant innate immune systems involve overlapping strategies, tools, data and areas.

Genetic mapping of crop improvement traits include traditional linkage analysis in segregating populations, genome-wide association studies and other approaches [90]. Despite the power of genomic breeding methods, breeders still have to contend with long lists of candidate genes. To narrow these lists of genes, additional information about molecular networks and phenotypes needs to be considered [91]. Machine learning approaches are expected to be able to increase the power of detection of key genes involved in important agronomic traits, including resistance against infectious pathogens and parasites [92].

Acknowledgments

This work was in part supported by an ERA-NET for Coordinating Action in Plant Sciences Grant [BB/N005112/1] to HUS. We are also grateful to the financial support of Santander Universities UK.

References

[1] C. Zipfel, Pattern-recognition receptors in plant innate immunity, Curr. Opin. Immunol. 20 (2008) 10–16.
[2] C.A. Janeway, Approaching the asymptote? Evolution and revolution in immunology, Cold Spring Harbor Symposium of Quantitative Biology 54 (1989) 1–13.
[3] G. Tena, M. Boudsocq, J. Sheen, Protein kinase signaling networks in plant innate immunity, Curr. Opin. Plant Biol. 14 (2011) 519–529.
[4] D.T. Fearon, R.M. Locksley, The instructive role of innate immunity in the acquired immune response, Science 272 (1996) 50–53.
[5] S. Asai, K. Shirasu, Plant cells under siege: plant immune system versus pathogen effectors, Curr. Opin. Plant Biol. 28 (2015) 1–8.
[6] H.A. van den Burg, S.J. Harrison, M.H. Joosten, J. Vervoort, P.J. de Wit, *Cladosporium fulvum* Avr4 protects fungal cell walls against hydrolysis by plant chitinases accumulating during infection, Mol. Plant Microbe Interact. 19 (2006) 1420–1430.
[7] R. de Jonge, H.P. van Esse, A. Kombrink, T. Shinya, Y. Desaki, R. Bours, et al., Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants, Science 329 (2010) 953–955.
[8] H.C. Rooney, J.W. Van't Klooster, R.A. van der Hoorn, M.H. Joosten, J.D. Jones, P. J. de Wit, Cladosporium Avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance, Science 308 (2005) 1783–1786.
[9] G. Doehlemann, C. Hemetsberger, Apoplastic immunity and its suppression by filamentous plant pathogens, New Phytol. 198 (2013) 1001–1016.
[10] A.P. Macho, C. Zipfel, Targeting of plant pattern recognition receptor-triggered immunity by bacterial type-III secretion system effectors, Curr. Opin. Microbiol. 23 (2015) 14–22.
[11] J.D.G. Jones, J.L. Dangl, The plant immune system, Nature 444 (2006) 323–329.
[12] H.H. Flor, Inheritance of reaction to rust in flax, J. Agric. Res. 74 (1947) 241–262.
[13] H.H. Flor, Genetics of pathogenicity in *Melampsora lini*, J. Agric. Res. 73 (1946) 335–357.
[14] B.J. Staskawicz, D. Dahlbeck, N.T. Keen, Cloned avirulence gene of *Pseudomonas syringae* pv. *glycinea* determines race-specific incompatibility on *Glycine max* (L.) Merr, Proc. Natl. Acad. Sci. U.S.A. 81 (1984) 6024–6028.
[15] G.B. Martin, S.H. Brommonschenkel, J. Chunwongse, A. Frary, M.W. Ganal, R. Spivey, et al., Map-based cloning of a protein kinase gene conferring disease resistance in tomato, Science 262 (1993) 1432–1436.
[16] M. Figueroa, N.M. Upadhyaya, J. Sperschneider, R.F. Park, L.J. Szabo, B. Steffenson, et al., Changing the game: using integrative genomics to probe virulence mechanisms of the stem rust pathogen *Puccinia graminis* f. sp. *tritici*, Front. Plant Sci. 7 (2016) 205.
[17] H.U. Stotz, G.K. Mitrousia, P.J. de Wit, B.D.L. Fitt, Effector-triggered defence against apoplastic fungal pathogens, Trends Plant Sci. 19 (2014) 491–500.
[18] J.M. Urbach, F.M. Ausubel, The NBS-LRR architectures of plant R-proteins and metazoan NLRs evolved in independent events, Proc. Natl. Acad. Sci. U.S.A. 114 (2017) 1063–1068.
[19] M.R. Swiderski, R.W. Innes, The Arabidopsis *PBS1* resistance gene encodes a member of a novel protein kinase subfamily, Plant J. 26 (2001) 101–112.
[20] V. Ntoukakis, I.M. Saur, B. Conlan, J.P. Rathjen, The changing of the guard: the Pto/Prf receptor complex of tomato and pathogen recognition, Curr. Opin. Plant Biol. 20 (2014) 69–74.
[21] B. Day, S.Y. He, Battling immune kinases in plants, Cell Host & Microbe 7 (2010) 259–261.
[22] W.Y. Song, G.L. Wang, L.L. Chen, H.S. Kim, L.Y. Pi, T. Holsten, et al., A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*, Science 270 (1995) 1804–1806.
[23] S.W. Han, M. Sriariyanun, S.W. Lee, M. Sharma, O. Bahar, Z. Bower, et al., Small protein-mediated quorum sensing in a Gram-negative bacterium, PloS One 6 (2011) e29192.
[24] C.J. Park, M.Y. Song, C.Y. Kim, J.S. Jeon, P.C. Ronald, Rice BiP3 regulates immunity mediated by the PRRs XA3 and XA21 but not immunity mediated by the NB-LRR protein, Pi5, Biochem. Biophys. Res. Commun. 448 (2014) 70–75.
[25] J. Nirmala, T. Drader, X. Chen, B. Steffenson, A. Kleinhofs, Stem rust spores elicit rapid RPG1 phosphorylation, Mol. Plant-microbe Interact.: MPMI 23 (2010) 1635–1642.
[26] J.F. Wendel, S.A. Jackson, B.C. Meyers, R.A. Wing, Evolution of plant genome architecture, Genome Biol. 17 (2016) 37.
[27] F. Jupe, K. Witek, W. Verweij, J. Sliwka, L. Pritchard, G.J. Etherington, et al., Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations, Plant J. 76 (2013) 530–544.
[28] G. Wang, U. Ellendorff, B. Kemp, J.W. Mansfield, A. Forsyth, K. Mitchell, et al., A genome-wide functional investigation into the roles of receptor-like proteins in Arabidopsis, Plant Physiol. 147 (2008) 503–517.
[29] T.L. Bailey, M. Boden, F.A. Buske, M. Frith, C.E. Grant, L. Clementi, et al., MEME SUITE: Tools for motif discovery and searching, Nucleic Acids Res. 37 (2009) W202–W208.
[30] B. Chalhoub, F. Denoeud, S. Liu, I.A. Parkin, H. Tang, X. Wang, et al., Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome, Science 345 (2014) 950–953.
[31] F. Jupe, L. Pritchard, G.J. Etherington, K. Mackenzie, P.J. Cock, F. Wright, et al., Identification and localisation of the NB-LRR gene family within the potato genome, BMC Genomics 13 (2012) 75.
[32] B. Lemaitre, E. Nicolas, L. Michaut, J.-M. Reichhart, J.A. Hoffmann, The dorsoventral regulatory gene cassette *spatzle/Toll/cactus* controls the potent antifungal response in Drosophila adults, Cell 86 (1996) 973–983.
[33] B. Steuernagel, S.K. Periyannan, I. Hernandez-Pinzon, K. Witek, M.N. Rouse, G. Yu, et al., Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture, Nat. Biotechnol. 34 (2016) 652–655.
[34] J. Hua, Modulation of plant immunity by light, circadian rhythm, and temperature, Curr. Opin. Plant Biol. 16 (2013) 406–413.
[35] H. Kitano, Foundations of Systems Biology, first ed., The MIT Press, Cambridge, 2001.
[36] B.H. Junker, Networks in biology, in: B.H. Junker, F. Schreiber (Eds.), Analysis of biological networks, John Wiley & Sons, Hoboken, 2008, pp. 3–14.
[37] M. Banf, S.Y. Rhee, Computational inference of gene regulatory networks: Approaches, limitations and opportunities, Biochim. Biophys. Acta-Gene Regul. Mech. 2017 (1860) 41–52.
[38] D.C. Tian, Q.Q. Gu, J. Ma, Identifying gene regulatory network rewiring using latent differential graphical models, Nucleic Acids Res. 44 (2016) e140.
[39] R. Guigó, The coding and the non-coding transcriptome, in: M. Walhout, M. Vidal, J. Dekker (Eds.), Handbook of Systems Biology: Concepts and Insigths, Academic Press, Cambridge, 2012, pp. 27–41.
[40] T. Ideker, N.J. Krogan, Differential network biology, Mol. Syst. Biol. 8 (2012) 567.
[41] A.-R. Carvunis, F.P. Roth, M.A. Calderwood, M.E. Cusick, G. Superti-Furga, M. Vidal, Interactome networks, in: M. Walhout, M. Vidal, J. Dekker (Eds.), Handbook of Systems Biology: Concepts and Insigths, Academic Press, Cambridge, 2012, pp. 45–63.
[42] M.L. Bulyk, M.A.J. Walhout, Gene regulatory networks, in: M. Walhout, M. Vidal, J. Dekker (Eds.), Handbook of Systems Biology: Concepts and Insigths, Academic Press, Cambridge, 2012, pp. 65–88.
[43] A. Frolova, M. Obolenska, Integrative Approaches for Data Analysis in Systems Biology: Current Advances, Applied Physics and Engineering (Ysf), 2016 II International Young Scientists Forum (2016) 194–198.
[44] R. Steuer, G.Z. López, Global network properties, in: B.H. Junker, F. Schreiber (Eds.), Analysis of Biological Networks, John Wiley & Sons, Hoboken, 2008, pp. 31–63.
[45] C.E. Riera, C. Merkwirth, C.D. De Magalhaes, A. Dillin, Signaling networks determining life span, in: R.D. Kornberg (Ed.), Annual Review of Biochemistry, Vol 85, Annual Reviews, Palo Alto, 2016, pp. 35–64.
[46] H.Y. Yu, M. Gerstein, Genomic analysis of the hierarchical structure of regulatory networks, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 14724–14731.
[47] M. Altaf-Ul-Amin, F.M. Afendi, S.K. Kiboi, S. Kanaya, Systems biology in the context of big data and networks, Biomed. Res. Int. (2014) 428570.
[48] H.W. Engl, C. Flamm, P. Kugler, J. Lu, S. Muller, P. Schuster, Inverse problems in systems biology, Inverse Probl. 25 (2009) 123014.
[49] B. Karahalil, Overview of systems biology and omics technologies, Curr. Med. Chem. 23 (2016) 4221–4230.
[50] J.R. Karr, J.C. Sanghvi, D.N. Macklin, M.V. Gutschow, J.M. Jacobs, B. Bolival, et al., A whole-cell computational model predicts phenotype from genotype, Cell 150 (2012) 389–401.
[51] J. Glazebrook, Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens, Annual Review of Phytopathology, Vol 43, Annual Reviews, Palo Alto, 2005, pp. 205-227.
[52] V.A. Halim, S. Altmann, D. Ellinger, L. Eschen-Lippold, O. Miersch, D. Scheel, et al., PAMP-induced defense responses in potato require both salicylic acid and jasmonic acid, Plant J. 57 (2009) 230–242.
[53] C.M.J. Pieterse, A. Leon-Reyes, S. Van der Ent, S.C.M. Van Wees, Networking by small-molecule hormones in plant immunity, Nat. Chem. Biol. 5 (2009) 308–316.

[54] M. Sato, K. Tsuda, L. Wang, J. Coller, Y. Watanabe, J. Glazebrook, et al., Network modeling reveals prevalent negative regulatory relationships between signaling sectors in Arabidopsis immune signaling, PLoS Pathog. 6 (2010) 15.
[55] X.B. Dong, Z.H. Jiang, Y.L. Peng, Z.D. Zhang, Revealing shared and distinct gene network organization in Arabidopsis immune responses by integrative analysis, Plant Physiol. 167 (2015) 1186–1203.
[56] K. Tsuda, F. Katagiri, Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity, Curr. Opin. Plant Biol. 13 (2010) 459–465.
[57] O. Windram, K.J. Denby, Modelling signaling networks underlying plant defence, Curr. Opin. Plant Biol. 27 (2015) 165–171.
[58] N.J. Atkinson, C.J. Lilley, P.E. Urwin, Identification of genes involved in the response of Arabidopsis to simultaneous biotic and abiotic stresses, Plant Physiol. 162 (2013) 2028–2041.
[59] C.M. Prasch, U. Sonnewald, Simultaneous application of heat, drought, and virus to Arabidopsis plants reveals significant shifts in signaling networks, Plant Physiol. 162 (2013) 1849–1866.
[60] S. Rasmussen, P. Barah, M.C. Suarez-Rodriguez, S. Bressendorff, P. Friis, P. Costantino, et al., Transcriptome responses to combinations of stresses in Arabidopsis, Plant Physiol. 161 (2013) 1783–1794.
[61] A. Mine, M. Sato, K. Tsuda, Toward a systems understanding of plant-microbe interactions, Front. Plant Sci. 5 (2014) 423.
[62] J.P. Tully, A.E. Hill, H.M.R. Ahmed, R. Whitley, A. Skjellum, M.S. Mukhtar, Expression-based network biology identifies immune-related functional modules involved in plant defense, BMC Genomics 15 (2014) 421.
[63] T. Lee, H. Kim, I. Lee, Network-assisted crop systems genetics: network inference and integrative analysis, Curr. Opin. Plant Biol. 24 (2015) 61–70.
[64] V. Vermeirssen, I. De Clercq, T. Van Parys, F. Van Breusegem, Y. Van de Peer, Arabidopsis ensemble reverse-engineered gene regulatory network discloses interconnected transcription factors in oxidative stress, Plant Cell 26 (2014) 4656–4679.
[65] O. Windram, C.A. Penfold, K.J. Denby, Network modeling to understand plant immunity, in: N.K. VanAlfen (Ed.), Annual Review of Phytopathology, Vol. 52, Annual Reviews, Palo Alto, 2014, pp. 93–111.
[66] I. Lee, B. Ambaru, P. Thakkar, E.M. Marcotte, S.Y. Rhee, Rational association of genes with traits using a genome-scale gene network for *Arabidopsis thaliana*, Nat. Biotechnol. 28 (2010) 149-U14.
[67] V. Ransbotyn, E. Yeger-Lotem, O. Basha, T. Acuna, C. Verduyn, M. Gordon, et al., A combination of gene expression ranking and co-expression network analysis increases discovery rate in large-scale mutant screens for novel *Arabidopsis thaliana* abiotic stress genes, Plant Biotechnol. J. 13 (2015) 501–513.
[68] S.S. Ma, H.J. Bohnert, S.P. Dinesh-Kumar, AtGGM2014, an Arabidopsis gene co-expression network for functional studies, Sci. China-Life Sci. 58 (2015) 276–286.
[69] J. Carrera, S.F. Elena, Computational design of host transcription-factors sets whose misregulation mimics the transcriptomic effect of viral infections, Sci. Rep. 2 (2012) 1006.
[70] J. Carrera, G. Rodrigo, A. Jaramillo, S.F. Elena, Reverse-engineering the Arabidopsis thaliana transcriptional network under changing environmental conditions, Genome Biol. 10 (2009) R96.
[71] Y. Kim, K. Tsuda, D. Igarashi, R.A. Hillmer, H. Sakakibara, C.L. Myers, et al., Mechanisms underlying robustness and tunability in a plant immune signaling network, Cell Host & Microbe 15 (2014) 84–94.
[72] T. Maekawa, B. Kracher, S. Vernaldi, E.V.L. van Themaat, P. Schulze-Lefert, Conservation of NLR-triggered immunity across plant lineages, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 20119–20123.
[73] N. Tintor, A. Ross, K. Kanehara, K. Yamada, L. Fan, B. Kemmerling, et al., Layered pattern receptor signaling via ethylene and endogenous elicitor peptides during Arabidopsis immunity to bacterial infection, Proc. Natl. Acad. Sci. U.S.A. 110 (2013) 6211–6216.
[74] K. Tsuda, A. Mine, G. Bethke, D. Igarashi, C.J. Botanga, Y. Tsuda, et al., Dual regulation of gene expression mediated by extended MAPK activation and salicylic acid contributes to robust innate immunity in *Arabidopsis thaliana*, PLoS Genet. 9 (2013) e1004015.
[75] A.C. Vlot, D.A. Dempsey, D.F. Klessig, Salicylic acid, a multifaceted hormone to combat disease, Ann. Rev. Phytopathol. 47 (2009) 177–206.
[76] E. Huala, A.W. Dickerman, M. Garcia-Hernandez, D. Weems, L. Reiser, F. LaFond, et al., The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant, Nucleic Acids Res. 29 (2001) 102–105.
[77] F. Cheng, S. Liu, J. Wu, L. Fang, S. Sun, B. Liu, et al., BRAD, the genetics and genomics database for Brassica plants, BMC Plant Biol. 11 (2011) 136.
[78] N. Fernandez-Pozo, N. Menda, J.D. Edwards, S. Saha, I.Y. Tecle, S.R. Strickler, et al., The Sol Genomics Network (SGN)–from genotype to phenotype to breeding, Nucleic Acids Res. 43 (2015) D1036–D1041.
[79] R. Narsai, J. Devenish, I. Castleden, K. Narsai, L. Xu, H. Shou, et al., Rice DB: an Oryza Information Portal linking annotation, subcellular location, function, expression, regulation, and evolutionary information for rice and Arabidopsis, Plant J. 76 (2013) 1057–1073.
[80] H. O'Sullivan, GrainGenes, Methods Mol. Biol. 406 (2007) 301–314.
[81] C.M. Andorf, E.K. Cannon, J.L. Portwood 2nd, J.M. Gardiner, L.C. Harper, M.L. Schaeffer, et al., MaizeGDB update: new tools, data and interface for the maize model organism database, Nucleic Acids Res. 44 (2016) D1195–D1201.
[82] P. Domingos, A few useful things to know about machine learning, Commun. ACM 55 (2012) 78–87.
[83] T. Kohonen, Self-organized formation of topologically correct feature maps, Biol. Cybern. 43 (1982) 59–69.
[84] B. Fritzke, A growing neural gas network learns topologies, in: G. Tesauro, D.S. Touretzky, T.K. Leen (Eds.), Advances in Neural Information Processing Systems, MIT Press, Cambridge MA, 1995.
[85] D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (1999) 788–791.
[86] F. Rezwan, Y. Sun, N. Davey, R. Adams, A.G. Rust, M. Robinson, Using Varying Negative Examples to Improve Computational Predictions of Transcription Factor Binding Sites., Engineering Applications of Neural Networks, Springer, Berlin, 2012. pp. 234-243.
[87] V.N. Vapnik, V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
[88] S.K. Kushwaha, P. Chauhan, K. Hedlund, D. Ahren, NBSPred: a support vector machine-based high-throughput pipeline for plant resistance protein NBSLRR prediction, Bioinformatics 32 (2016) 1223–1225.
[89] T. Pal, V. Jaiswal, R.S. Chauhan, DRPPP: A machine learning based tool for prediction of disease resistance proteins in plants, Comput. Biol. Med. 78 (2016) 42–48.
[90] R.J. Snowdon, F.L.I. Iniguez Luy, Potential to improve oilseed rape and canola breeding in the genomics era, Plant Breeding 131 (2012) 351–360.
[91] X. Wang, N. Gulbahce, H. Yu, Network-based methods for human disease gene prediction, Brief Funct. Genomics 10 (2011) 280–293.
[92] E. Glaab, Using prior knowledge from cellular pathways and molecular networks for diagnostic specimen classification, Briefings Bioinf. (2015).
[93] S. Bai, J. Liu, C. Chang, L. Zhang, T. Maekawa, Q. Wang, et al., Structure-function analysis of barley NLR immune receptor MLA10 reveals its cell compartment specific activity in cell death and disease resistance, PLoS Pathog. 8 (2012) e1002752.
[94] S. Chou, K.V. Krasileva, J.M. Holton, A.D. Steinbrenner, T. Alber, B.J. Staskawicz, Hyaloperonospora arabidopsidis ATR1 effector is a repeat protein with distributed recognition surfaces, Proc. Natl. Acad. Sci. U.S.A. 108 (2011) 13323–13328.
[95] Y. Jia, S.A. McAdams, G.T. Bryan, H.P. Hershey, B. Valent, Direct interaction of resistance gene and avirulence gene products confers rice blast resistance, EMBO J. 19 (2000) 4004–4014.
[96] N.J. Larkan, D.J. Lydiate, I.A. Parkin, M.N. Nelson, D.J. Epp, W.A. Cowling, et al., The *Brassica napus* blackleg resistance gene *LepR3* encodes a receptor-like protein triggered by the *Leptosphaeria maculans* effector AVRLM1, New Phytologist 197 (2013) 595–605.