Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt

Technische Universität München

Bioinformatic analysis of integrated *Arabidopsis thaliana* protein interaction networks

Stefan Altmann

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende:

Prof. Dr. Chris-Carolin Schön

Prüfende der Dissertation:

1. Prof. Dr. Claus Schwechheimer
2. Prof. Dr. Pascal Falter-Braun,
   Ludwig-Maximilians-Universität München
3. Prof. Dr. Thomas Rattei,
   Universität Wien

Die Dissertation wurde am 27.06.2019 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 18.10.2019 angenommen.

Abstract

Introduction. Protein-protein interactions are involved in virtually all biological processes in a cell. To identify protein-protein interactions, experimental high-throughput methods like the yeast two-hybrid system have been developed. The high number of identified interactions form large networks and only bioinformatic methods allow to infer global network properties and complex correlations between the network and the biological properties of the proteins.

In the following I describe the bioinformatic analysis of two *Arabidopsis thaliana* protein interaction networks derived from systematic yeast two-hybrid system screens: a network of proteins involved in phytohormone signaling, the PhyHormInteractome (PhI) and a network of effector proteins of three different pathogens interacting with *Arabidopsis thaliana* host proteins, the Plant-Pathogen Immune Network 2 (PPIN2).

Core signaling pathways of individual hormones and signal transduction by protein-protein interactions are well investigated, whereas signal integration of different hormone signaling pathways by protein-protein interactions has been elucidated only for a few cases. The aim of the analysis of the systematic PhI network is to analyze network structure, contact points of hormone signaling pathways in the network and correlation between network structure and biology.

It has been shown that effector proteins of two different pathogens converge on a common set of *Arabidopsis* host proteins. The aim was to integrate the effector-host protein interactions of a third pathogen, *Golovinomyces orontii*, into a previously elucidated effector-host network and to analyze convergence of all three pathogens. Moreover, the question was, which of the effector targets can be genetically validated and if effector binding left population genetic signatures in the host proteins, which show evidence for selection.

Results. The analysis of PhI revealed a scale-free topology with modular structures (communities). Seven of the identified communities showed an enrichment for a specific phytohormone pathway. The analysis of contact points between distinct hormone pathways revealed a dense interconnectedness between distinct pathways. Moreover, signal transduction by kinases and signal integration at transcription factors was found.

The examination of PPIN2 revealed both an intra- and interspecies convergence of effector proteins on common host proteins. The analysis of genetic validation experiments implies a high relevance of identified host proteins for pathogen infection. The integration of natural variation data from *Arabidopsis* ecotypes revealed that highly variable proteins preferentially interact with host proteins.

Discussion. The observed hierarchical topology of PhI and the identified hormone enriched communities show that network topology reflects biological processes. A strong interconnectedness of distinct phytohormone signaling pathways was observed, which is in contrast to low interconnectedness determined from literature-curated protein-protein interactions. The high number of direct pathway contacts indicate a strong reciprocal influence on each other rather than separated signaling pathways. In the course of this project, the signal transduction between distinct pathways could be shown in planta.

In PPIN2, the convergence analysis of effector proteins exposed a significant convergence of effector proteins from three different pathogens on common host proteins, despite their large evolutionary distance. Together with the results of the genetic validation of effector targets, the relevance of host proteins for infection could be shown. The integration of natural variation data from diverse ecotypes revealed a putative strategy of plants to overcome infection by pathogens.

Conclusions. The high number of protein interactions between phytohormone signaling pathways and the in planta validation of signal integration between pathways let conclude that phytohormone signaling pathways are strongly connected with each other. The analysis of PPIN2 showed a strong correlation of effector convergence and genetic validation of effector targets. This suggests, together with the identified variable genes interacting with effector targets that the observed network is the result of adaption in an evolutionary process.

Zusammenfassung

Einleitung. Protein-Protein-Interaktionen sind an nahezu allen biologischen Prozessen in einer Zelle beteiligt. Um Protein-Protein Interaktionen zu identifizieren, wurden experimentelle Hochdurchsatzmethoden wie das Hefe-Zwei-Hybrid-System entwickelt. Die große Anzahl an erhaltenen Interaktionen bilden große Netzwerke und nur mit bioinformatischen Methoden ist es möglich globale Netzwerkeigenschaften und komplexe Korrelationen zwischen dem Netzwerk und biologischen Eigenschaften der Proteine zu finden.

Im Folgenden beschreibe ich die bioinformatische Analyse von zwei *Arabidopsis thaliana* Protein-Interaktions-Netzwerken, die auf Grundlage von systematischen Hefe-Zwei-Hybrid-System Screens erstellt wurden: ein Netzwerk von Proteinen, die an der Phytohormon-Signalübertragung beteiligt sind, das PhyHormInteractom (PhI), und ein Netzwerk von Effektor-Proteinen dreier verschiedener Krankheitserreger, die mit *Arabidopsis thaliana* Wirts-Proteinen interagieren, das Pflanzen-Pathogen Immun Netzwerk 2 (PPIN2).

Die Hauptsignalwege einzelner Hormone und die Signalübertragung durch Protein-Protein Interaktionen sind gut untersucht, während die Signalintegration verschiedener Hormon-Signalwege durch Protein-Protein Interaktionen nur in wenigen Fällen aufgeklärt wurde. Ziel der Analyse des systematischen PhI-Netzwerks ist es, die Netzwerkstruktur zu analysieren, Kontaktpunkte verschiedener Hormonsignalwege zu identifizieren und die Korrelation zwischen dem Netzwerk und der darunterliegenden Biologie zu untersuchen.

Es wurde gezeigt, dass Effektor-Proteine von zwei verschiedenen Krankheitserregern auf einen gemeinsamen Satz von *Arabidopsis* Wirts-Proteinen konvergieren. Ziel war es, die Effektor-Wirt Protein Interaktionen eines dritten Erregers, *Golovinomyces orontii*, in ein zuvor aufgeklärtes Effektor-Wirt Netzwerk zu integrieren und die Konvergenz der Effektor-Proteine aller drei Erreger zu analysieren. Darüber hinaus stellte sich die Frage, welche der Effektor-Ziele genetisch validiert werden können und ob die Bindung der Effektoren populationsgenetische Signaturen in den Host-Proteinen hinterlassen haben, die Hinweise auf Selektion zeigen.

Ergebnisse. Die Analyse des PhI ergab eine skalenfreie Topologie mit modularen Strukturen (Communitys). Sieben der identifizierten Communitys zeigten eine Anreicherung für einen bestimmten Phytohormon-Signalweg. Die Analyse von Kontaktpunkten zwischen verschiedenen Hormonsignalwegen ergab eine dichte Vernetzung unterschiedlicher Signalwege. Darüber hinaus wurden Signalweiterleitung durch Kinasen und die Signalintegration bei Transkriptionsfaktoren gefunden.

Die Untersuchung von PPIN2 ergab sowohl eine Konvergenz von Effektor-Proteinen eines Pathogens als auch von mehreren Pathogenen auf gemeinsame Wirts-Proteine. Die Analyse von genetischen Validierungsexperimenten impliziert eine hohe Relevanz der identifizierten Wirts-Proteine für die Infektion mit Krankheitserregern. Die Integration von natürlicher Variation, die in *Arabidopsis* Ökotypen gefunden wurde, ergab, dass hochvariable Proteine bevorzugt mit Wirtsproteinen interagieren.

Diskussion. Die hierarchische Topologie des PhI und die identifizierten hormonangereicherten Communitys zeigen, dass die Netzwerktopologie biologische Prozesse widerspiegelt. Es wurde eine starke Vernetzung verschiedener Phytohormon-Signalwege beobachtet, die im Gegensatz zu einer niedrigen Vernetzung steht, die aus Literatur-kuratierten Protein Interaktionen bestimmt wurde. Die hohe Anzahl von direkten Kontakten zwischen den Signalwegen deutet auf eine starke gegenseitige Beeinflussung hin anstatt auf getrennte Signalwege. Im Rahmen dieses Projekts konnte die Signalübertragung zwischen verschiedenen Signalwegen in der Pflanze gezeigt werden.

Die Konvergenzanalyse in PPIN2 zeigte eine signifikante Konvergenz von Effektor-Proteinen von drei verschiedenen Pathogenen auf gemeinsame Wirts-Proteine, trotz der großen evolutionären Entfernung der Pathogene. Zusammen mit den Ergebnissen der genetischen Validierung von Wirts-Proteinen, die mit Effektor-Proteinen interagieren, konnte die Relevanz von Wirts-Proteinen für die Infektion gezeigt werden. Die Integration von natürlicher Variation in unterschiedlichen Ökotypen zeigte eine mögliche Strategie der Pflanzen, die Infektion durch Krankheitserreger zu überwinden.

Schlussfolgerungen. Die große Anzahl von Protein Interaktionen zwischen Phytohormon-Signalwegen und die Validierung der Signalintegration in Pflanzen lassen die Schlussfolgerung zu, dass die einzelnen Signalwege stark miteinander vernetzt sind. Die Analyse von PPIN2 zeigte eine starke Korrelation zwischen der Konvergenz von Effektoren und der genetischen Validierung von Effektor-Zielen. Zusammen mit den identifizierten variablen Proteinen, die mit Effektor-Zielen interagieren, lässt sich schlussfolgern, dass das beobachtete Netzwerk das Ergebnis eines evolutionären Prozesses ist.

Contents

List of Figures

List of Tables

Acronyms

Θ_W	Watterson's Θ
15mers	15 amino acid long sequence
AAP	Amino acid polymorphism
ABA	Abscisic acid
ABF	ABSCISIC ACID-RESPONSIVE ELEMENT-BINDING FACTOR
ABI	ABSCISIC ACID-INSENSITIVE
ACC	1-aminocyclopropane-carboxylic acid
ACS	1-aminocyclopropane-1-carboxylate synthase
AD	Activation domain
AFB	AUXIN SIGNALLING F-BOX PROTEIN
AHD	Arabidopsis Hormone Database
AHK	Arabidopsis histidine kinase
AHP	Arabidopsis histidine phosphotransfer protein
AI-1_{MAIN}	Arabidopsis Interactome 1 Main Screen
ARF	AUXIN RESPONSE FACTOR
ARR	Arabidopsis response regulator
ASA	Alpha and beta subunit of anthranilate synthase
AUX	Auxin
AUX/IAA	AUXIN/INDOLE-3-ACETIC ACID
BA	6-benzylamino purine
BAK1	BRI1-ASSOCIATED RECEPTOR KINASE1
BEE	BR enhanced expression
BiFC	Bimolecular fluorescence complementation
BIK1	BOTRYTIS-INDUCED KINASE 1
BIN2	BRASSINOSTEROID INSENSITIVE2
BKI1	BRASSINOSTEROID INSENSITIVE1 KINASE INHIBITOR
BL	Brassinolide
BR	Brassinosteroid
BRC1	BRANCHED1
BRI1	BRASSINOSTEROID-INSENSITIVE 1
BSK	BRASSINOSTEROIDSIGNALING KINASE
BSU1	BRI1-SUPPRESSOR1
BZR1	BRASSINAZOLE-RESISTANT 1
BZR2	BRASSINAZOLE-RESISTANT 2
CDG1	CONSTITUTIVE DIFFERENTIAL GROWTH1
CDK	Calcium-dependent kinase

CDPK	Cyclin-dependent kinase
CK	Cytokinin
Co-IP	Co-immunoprecipitation
COI1	CORONATINE-INSENSITIVE 1
CTR1	CONSTITUTIVE TRIPLE RESPONSE 1
D14	DWARF 14
D6PK	D6 PROTEIN KINASE
D_T	Tajima's D
DB	DNA binding domain
DELLA	DELLA domain-containing protein
DNA	Deoxyribonucleic acid
DRNL	DORNROSCHEN-LIKE
EBF	EIN3-BINDING F-BOX
edr	Enhanced disease resistance
eds	Enhanced disease susceptibility
EIL	EIN3-LIKE
EIN	ETHYLENE-INSENSITIVE
ER	Endoplasmatic reticulum
ET	Ethylene
ETI	Effector triggered immunity
ETP	ETHYLENE INSENSITIVE2-TARGETING PROTEIN
ETR	ETHYLENE RESPONSE
F_{st}	Wrigth's fixation index
FLS2	FLAGELLIN-SENSITIVE 2
FP	False positive
FPKM	Fragments per kilobase per million mapped reads
GA	Gibberellic acid
GA3OX1	Gibberellin 3-oxidase 1
GAI	GA-INSENSITIVE
GEO	Gene Expression Omnibus
GID	GIBBERELLIN INSENSITIVE DWARF
GO	Gene Ontology
Gor	Golovinomyces orontii
Hpa	Hyaloperonospora arabidopsidis
HSFA2	Heat shock factor A2
IAA	Indole-3-acetic acid
JA	Jasmonic acid
JA-Ile	jasmonoyl isoleucine
JAZ	JASMONATE ZIM-DOMAIN PROTEIN

KAI2	KARRIKIN-INSENSITIVE2
KAR	Karrikin
LD	Linkage disequilibrium
LRR	Leucine-rich repeat
MAMP	Microbial-associated molecular pattern
MAPK	Mitogen-activated kinase
MAX2	MORE AXILLARY BRANCHES2
mRNA	messenger RNA
n-chlor	Nuclear-encoded chloroplast-targeted
NB-LRR	Nucleotide-binding leucine rich repeat
NGS	Next Generation Sequencing
NINJA	NOVEL INTERACTOR OF JASMONATE ZIM-DOMAIN PROTEIN
NPR	NON-EXPRESSER OF PATHOGENESIS-RELATED GENES 1
nsSNP	Non-synonymous single nucleotide polymorphisms
OEC	*G. orontii* effector candidates
PAMP	Pathogen-associated molecular pattern
PCC	Pearson correlation coefficient
PCR	Polymerase chain reaction
PhI	PhyhormInteractome
PID	PINOID
PIF	PHYTOCHROME-INTERACTING FACTOR
PIN	PIN-FORMED
PP2C	Protein Phosphatases 2C
PPI	Protein-protein interaction
PPIN1	Plant-pathogen immune network 1
PPIN2	Plant-pathogen immune network 2
PR	PATHOGENESIS RELATED
PRR	Pattern recognition receptor
PRS	Positive reference set
Psy	Pseudomonas syringae
PTI	PAMP Triggered Immunity
PvA	PhyHormORFeome against AtORFeome
PvP	PhyHormORFeome against PhyHormORFeome
PWM	Position weight matrix
PYL	PYR-Like
PYR	Pyrabactin resistance
RCAR	Regulator component of ABA receptor
RFC3	Replication factor C subunit 3
RGA	REPRESSOR OF GA1-3
RGL	RGA-LIKE

RLK	Receptor-like kinases
RNA	Ribonucleic acid
RNA-Seq	RNA-Sequencing
RPKM	Reads per kilobase per million mapped reads
RR	Response regulator
RRS	Random reference set
SA	Salicylic acid
SAR	Systemic acquired resistance
SCF complex	Skp1-Cullin-F-box protein complex
SCL21	SCARECROW-like 21
SL	Strigolactone
SLY	SLEEPY
SMAX	SUPPRESSOR OF MAX2
SMXL	SUPPRESSOR OF MAX2-1-LIKE PROTEINS
SNP	Single Nucleotide Polymorphism
SnRK	SNF1-related kinase
SRA	Sequence Read Archive
sSNP	Synonymous single nucleotide polymorphisms
SSP_{PHO}	Search space PHO
TAA	Tryptophan aminotransferase
TAP-MS	Tandem affinity purification - mass spectrometry
TCP	TEOSINTE BRANCHED1, CYCLOIDEA, PCF1 and PCF2
TF	Transcription factor
TGA	TGA binding
TIR1	TRANSPORT-INHIBITOR RESPONSE 1
TP	True positive
TPL	TOPLES
TPM	Transcripts per million mapped reads
WRKY	WRKY domain containing protein
XRN4	Exoribonuclease 4
Y2H	Yeast two-hybrid system

1 Introduction

Plants are the source for food and material used to produce energy, medicine and many other goods. Moreover, they are an important economic factor in the production of wheat, corn and fruits for nutrition, but also for production of oil and as ingredients in other products like paper, clothing or beverages [1, 2]. Ongoing climate change will have a strong influence on crop production by decreasing production on average by 2.84 to 4.34 % per year [3]. To keep production stable, it is important that crops can cope with the stress factors caused by climate change, such as more or new pathogens or extreme climatic conditions. Therefore plants must be prepared for these stresses by breeding or genetics [4]. For this task it is important to understand plants on a molecular level: how they perceive environmental signals and integrate them into their growth regulatory decision network [5] and how pathogens and plants interact in order to infect a plant and defend against infection, respectively. To understand these complex mechanisms, the analysis of biological networks has emerged as a powerful approach [6].

Biotic and abiotic stresses activate different phytohormone signaling pathways in the plant. These signals activate divers biological processes, which allow plants to cope with stresses and regulate development. To gain insights into phytohormone signaling, I analyzed a protein-protein interaction network of proteins involved in phytohormone signaling pathways. A specific biotic stress is pathogen infestation of a plant. To understand how pathogen proteins interact with host proteins, a plant-pathogen protein interaction network is analyzed. For the analysis of both networks additional data from experiments and external databases were integrated into the network or analyzed together with the network.

In the first part of the introduction I give an overview about networks and properties of networks: methods to experimentally determine protein interaction maps, network topologies, and properties of networks. The second part deals with data from external sources, which can be integrated into networks. Here I will show gene ontology, gene expression data, and natural variation information. In the third part, phytohormone signaling pathways and signal transduction by protein-protein interactions are presented. The fourth part deals with host-pathogen protein interactions while infection of plants. At the end of the introduction, the objectives of the analyses of the networks are explained.

1.1 Networks

Many different entities like proteins, RNA, DNA, and small molecules e.g. phytohormones and enzymes are involved in the completion of a biological processes. These entities interact with each other in order to fulfill different tasks like signal transduction, regulation or metabolic functions. These interactions can be visualized dependent on the process as networks of protein-protein interactions, metabolic interactions, signaling networks, or transcription regulatory interactions [7].

1.1.1 Network mapping

For all enumerated types of interactions, a multitude of experimental methods has been developed for their detection. Especially, with the detection of more and more protein-protein interactions in the 1940s, the importance of physical interactions between proteins became apparent [8]. Since then a large number of experimental methods for both in vitro and in vivo protein-protein interaction detection has been developed. These methods can be divided into two groups: (i) methods that detect interactions between two proteins (binary interactions) and (ii) methods that detect complexes of several proteins. In vitro methods are e.g. tandem affinity purification - mass spectrometry (TAP-MS), co-immunoprecipitation (Co-IP), protein microarrays, protein-fragment complementation and others [9].

In vivo methods are e.g. bimolecular fluorescence complementation (BiFC) and the yeast two-hybrid system (Y2H), of which several variants exist, including high-throughput applications [10]. As both protein interaction networks which are analyzed in this thesis are derived by Y2H screens, this method is described here more detailed: In Y2H, two proteins of any organism can be tested for interaction in yeast. One protein (bait) is fused with a DNA-binding domain (DB) of a transcription factor, the second protein (prey) is fused to the activation domain (AD). If the prey and bait proteins interact, a functional transcription factor is reconstituted and a reporter gene is expressed, which signals the interaction between both tested proteins [10].

An adapted Y2H variant for high-throughput application has been used for example to generate the Arabidopsis Interactome 1 Main Screen (AI-1_{MAIN}), where 8,000 proteins were tested for binary interactions [11]. An empirical frame work for testing interaction mapping methods (see next section) has shown that this technique has a high specificity and generates a low number of false positive interactions [11, 12].

1.1.2 Quality assessment of mapping pipelines

Especially Y2H has been discussed to produce a lot of false positive interactions and a low coverage [13, 14], but over time limitations became apparent and could be partially overcome [14]. Only with the systematic investigation and comparison of different experimental methods an objective assessment of these methods became possible. A comparison of five different experimental methods revealed that Y2H is able to find about 25 % of the interactions from a positive reference set of known interactions, which a mean performance of the five tested methods [15]. Moreover, it could be shown that only a small fraction of the interactions are detected by all five methods and 25 % of the interactions are detected by only one method [15]. This partially overlapping set of detected interactions could be also observed for different Y2H versions [16].

To assess the quality of interactions produced by a specific interaction mapping pipeline, quality assessment framework was developed. By combining the results of the four parts of this framework the number of false positive and false negative interactions produced by the experimental method can be estimated [12]. In the following the four parts of quality assessment framework are described.

The **completeness** describes the number of tested protein pairs in relation to the defined number of proteins, which are planned to be tested. The completeness decreases for example if for Y2H some ORFs cannot be derived as AD- or DB-hybrid constructs [17, 12].

The **assay sensitivity** describes the percentage of interactions from a positive reference set, which can be identified with a specific interaction mapping pipeline [12]. Experimental observations show that assay sensitivity is usually between 20 - 35 % [15].

The **sampling sensitivity** describes the percentage of interactions which were found with a specific interaction mapping pipeline depending on the total number of interactions, which can be found by this specific method. Therefore, the screen must be repeated several times to determine the number of additional interactions, which can be identified with each additional repeat of the screen. This data can be used to model the total number of interactions, which can identified by a method, using an adjusted Michaelis-Menten equation. [12, 11]

These three measures can be used to estimate the **overall completion**, which describes the percentage of interactions, which were identified using a specific method in the previously defined set of proteins (search space). This can be used to estimate the total number of interactions, which are expected in the search space.

The fourth measure, the **precision** describes fraction of pairs, which can be also identified by different experimental method. Therefore the positive reference set and interactions from the mapping pipeline are tested with an additional interaction mapping pipeline. This provides information about the quality of the interactions, i.e. how much interactions can be reidentified with a different method [12, 16].

1.1.3 Network properties

The previously enumerated networks are biological networks, where nodes are defined entities like proteins and edges have a defined meaning e.g. a protein-protein interaction. Many other networks exist in real world and are subject of research in different research areas e.g. the internet (connections between routers), world wide web (links between webpages), science collaborations (co-authorships in publications), social networks (friendships), and graph theory in mathematics and physics [18]. In general, networks (graphs) are formed by nodes (vertices), which are connected by edges (links). An edge can be either directed or undirected. Directed edges denote for example the direction of a flow from a substrate to a product in a metabolic reaction or an action of one entity on another entity like a transcription factor, which regulates the expression of a gene. Undirected edges do not contain any direction information, like it is the case in protein-protein interactions [7].

Network measures

The nodes in a network can be characterized by different measures, which take into account the node's edges, the properties of its interacting nodes (neighbors), and its position in the network: The *degree (k)* indicates the number of edges, which are connected to a node. This is equivalent to the number of interactions with other nodes. In case of a directed network, each node has an out-degree k_{out} and in-degree k_{in}. The out-degree specifies the number of edges, which point to the neighbor node; the in-degree is the number of edges which point to the respective node (fig. 1.1 A) [19].

The interconnectedness of the direct interaction partners of a node is given by the *clustering coefficient (C)*. It measures how strong the interacting nodes are connected to each other. The clustering coefficient for node A (C_A) is described as $2n_A/k_A(k_A - 1)$, where n_A is the number of edges between the interacting nodes of A and k_A is the number of neighbors (fig. 1.1 B) [19].

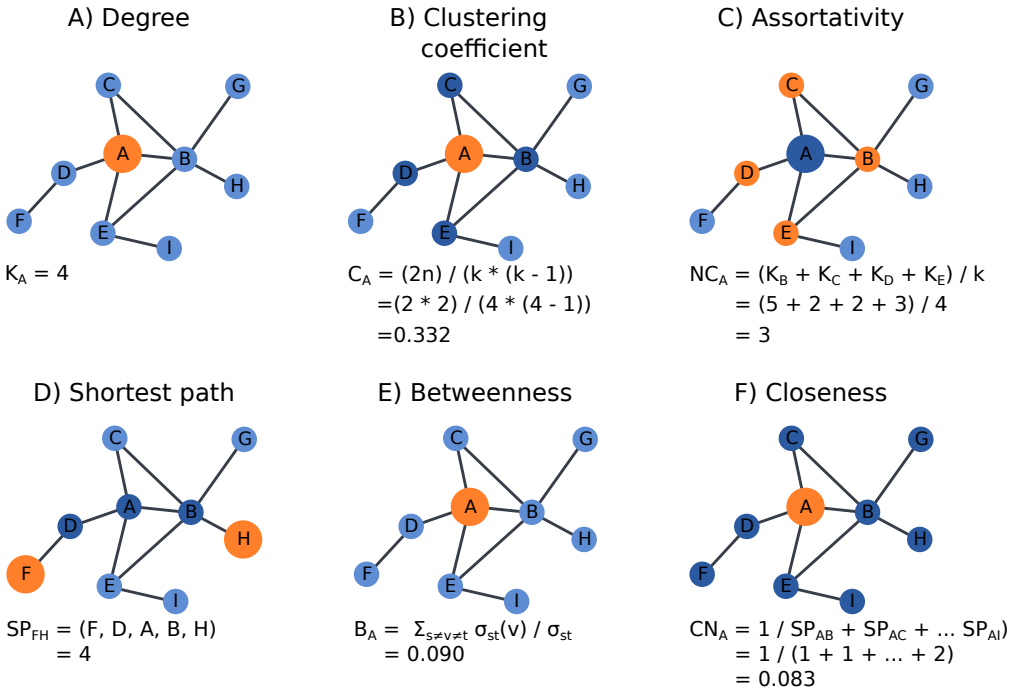


Figure 1.1: Selection of network measures adapted from [19]. For each measure the characterized node is marked in orange in the sample network. Below the network, the calculation and value of the respective measure is shown.

The *assortativity (NC)* is the mean degree of the neighbors of a node. The correlation of the degree and assortativity provides information, if the network is assortative or disassortative. A negative correlation suggests disassortativity, which means that nodes with a high degree (hubs) tend to interact with nodes with a low degree. A positive correlation indicates that hubs interact with hubs (fig. 1.1 C) [19].

The minimum distance between two nodes is described by the *shortest path (SP)*. It is the minimum number of nodes, which have to be passed following the edges from one node to another (fig. 1.1 D) [19].

The *betweenness (B)* measures the centrality of a node in the network. The betweenness centrality describes the fraction of shortest paths between all pairs of nodes, which pass a certain node. In terms of information flow, this would be the node where most information is transferred (fig. 1.1 E) [19].

Closeness (CN) also describes, how central a node is in a network, but in terms of lengths of shortest paths from the analyzed node to all other nodes in the network. The closeness is described by $CN(A) = 1/(\sum_{i=1}^{n} SP_{AX_n}/n)$, where A is the analyzed node and n is the number of nodes in the network without A (fig. 1.1 F) [20].

Network topology

The described network measures allow to characterize single nodes in terms of centrality and connectivity, but do not elucidate the overall connectivity of the nodes of a network. The overall connectivity can be described by different network topologies, which are characterized by specific distributions of degree and clustering coefficient. Three main

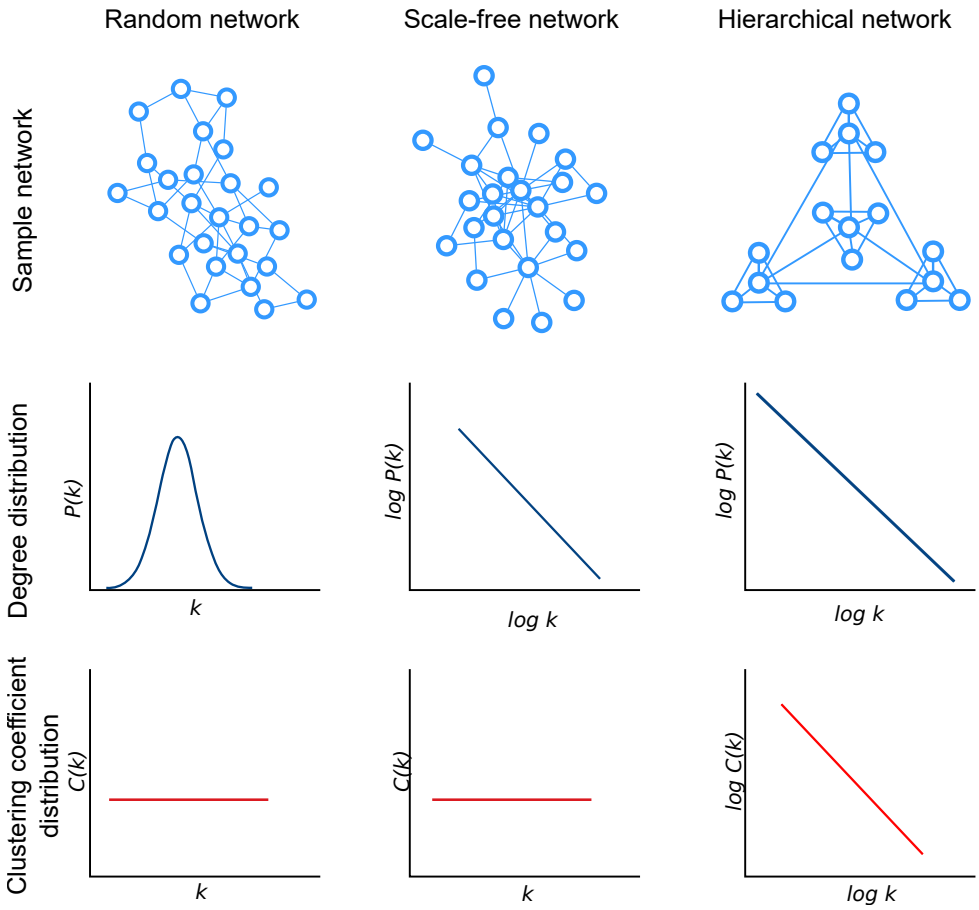


Figure 1.2: For each of the three commonly distinguished network topologies, the respective generic degree distribution and clustering coefficient distribution is shown. Figure adapted from [7]

topologies are distinguished in the field of molecular networks based on their characteristic distributions of degree and clustering coefficient.

In a *random network*, all nodes have a similar degree and they all have a similar interconnectedness. This results in a degree distribution, which follows a Poisson distribution. The clustering coefficient is independent of the degree of a node, which means neighbors always show a similar interconnectedness (fig. 1.2, random network) [7]. Random networks are mainly used in graph theory and do not reflect network structures observed in real world networks [18].

In a *scale-free network*, the degree distribution $P(k)$ follows a power law distribution, where the probability to observe a certain degree is proportional to $k^{-\gamma}$, where k is the degree and γ is the degree exponent determining the network properties. For $\gamma > 3$, hubs play no important role in a network, values of $\gamma < 2$ lead to a hub and spoke network and values in between lead to a hierarchy of hubs. In general, in this network a high number of nodes have a low degree and a very low number of nodes have a very high degree, which means that hubs are rare in the network (fig. 1.2, scale-free network) [7]. It is assumed that many real world networks like the internet, social networks including collaboration networks and airline networks have a scale-free structure [18].

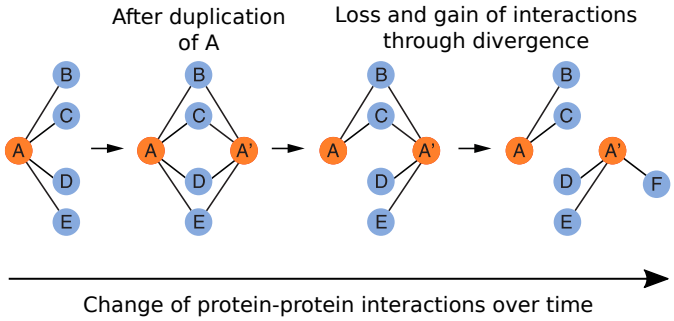


Figure 1.3: Schematic representation of interaction rewiring over time. After duplication of protein A, both homologous proteins A and A' share the same interaction partners. Over time, proteins diverge and lose / gain interactions resulting in a rewired network. Adapted from [11]

In a *hierarchical network*, not only the degree distribution follows a power law distribution, but also the clustering coefficient distribution. This implies the existence of highly connected subgraphs in the scale-free network, also called modules, cluster, or communities (fig. 1.2, hierarchical network) [7]. Most biological networks have a scale-free topology with hierarchical modularity, like protein-protein interaction networks, where most proteins have only a few interactions, a very low number of hubs exist and many molecules work together to achieve a distinct function. But also in metabolic and regulatory networks, a hierarchical topology is observed [7].

1.1.4 Network evolution

The hierarchical topology of biological networks is a result of evolutionary processes. The networks are derived from growth processes in which additional nodes are added to the network [7]. In case of protein-protein interaction networks, in the AI-1_{MAIN} network an evidence was found for the duplication-divergence model. This model assumes, that after duplication two paralogous proteins share their interaction partners, as they are identical. Over time the two paralogous proteins diverge and with increasing sequence dissimilarity they lose interactions, but can also gain new interactions (fig. 1.3 A). In AI-1_{MAIN}, an evidence was found that after whole genome duplication events, the paralogous proteins diverge fast and share after a short time period only 70 % of sequence identity and 40 % of interaction partners. Afterwards, the divergence increases slowly over time. Whole genome duplications are therefore major events for the evolution of networks [19]. This includes also phytohormone signaling pathways, which emerged over time. They share similar regulatory properties, but emerged at different time points in plant lineage [21] (see also section 1.3).

1.1.5 Network robustness

The evolutionarily derived scale-free structure also shows interesting properties in terms of robustness against node failures. An analysis of the scale-free network topology revealed that a random removal of nodes up to a certain degree does not destroy the network [22]. Up to a failure rate of 80 % of the nodes, the remaining 20 % still form a compact network connecting all remaining nodes [7, 22]. This elucidates the resilience of e.g. protein-protein interaction networks against failure of nodes by mutations [7]. On the other hand, if the hubs of a network are sequentially removed, at first the diameter

of the network increases rapidly, then the network desintegrates into isolated clusters [23] (fig. 1.4 A,C).

A random network, for comparison, shows similar behavior in both cases, the random removal and targeted removal of nodes. After removing a substantial number of nodes, the network desintegrates into tiny, non-communicating subgraphs [7] (fig. 1.4 B,C).

The importance of hubs was shown for the protein-protein interaction network of *Saccharomyces cerevisia*. The phenotypic consequence of a single gene deletion is affected to a large extent by the topological position of the protein. In systematic mutagenesis experiments, yeast showed a high tolerance to a substantial number of random protein deletions. But the targeted deletion of the proteins with phenotypic profiles and more than 15 interactions result in a lethality rate of 62 % [24].

1.1.6 Network communities

Besides the robustness, another inherent property of biological networks is the presence of functional modules, also called clusters or communities. These communities are highly interconnected groups of nodes within the network (fig. 1.5), which are indicated by a high clustering coefficient in determination of network topology [7].

Signatures of communities are present in many networks, not only biological networks: in social networks, communities represent grouping by interest or background; in citation networks, they e.g. represent related publications; in the web, they can represent webpages of similar topics [25]. For identification of communities, a large number of algorithms have been proposed. A selection of these algorithms is presented here to give an overview about different approaches.

The first approaches were based on **hierarchical clustering** of nodes based on weights, which indicate, which nodes are clustered together, e.g. the number of independent paths between two vertices (paths, which do not share any nodes except start and end node). The drawback of hierarchical clustering is the high number of isolated nodes, which are not assigned to a community [25].

An algorithm, which splits the network is the **edge betweenness algorithm**. Based on the edge betweenness, the network is split into communities. The edge betweenness is defined as the number of shortest paths between all nodes, which run through a certain edge. This represents the importance of an edge for information flow in the network. The edges with the highest betweenness are iteratively removed and the betweenness is recalculated until no edge remains. This algorithm is able to identify community structures in very different networks with high agreement of expectation, e.g. in Zachary's karate club study [26] to identify social structures, in a college football games network for identification of conferences (groups of football teams), in a collaboration network of a Santa Fe research institute to identify divisions, and in a food web of marine organisms to identify food chains [25].

Algorithms based on random walks are e.g. **walktrap** and **infomap**. Walktrap performs a large number of random walks with random start nodes. It is based on the idea, that short random walks tend to stay in the same community. Results from these short walks are used to clusters nodes into communities [27]. Infomap performs a very long random walk to analyze the information flow in the network. The description of the path is minimized by grouping nodes into modules [28].

A group of fast algorithms are greedy techniques, which are based on local optima, but do not necessarily find the global optimal solution. Greedy community identification

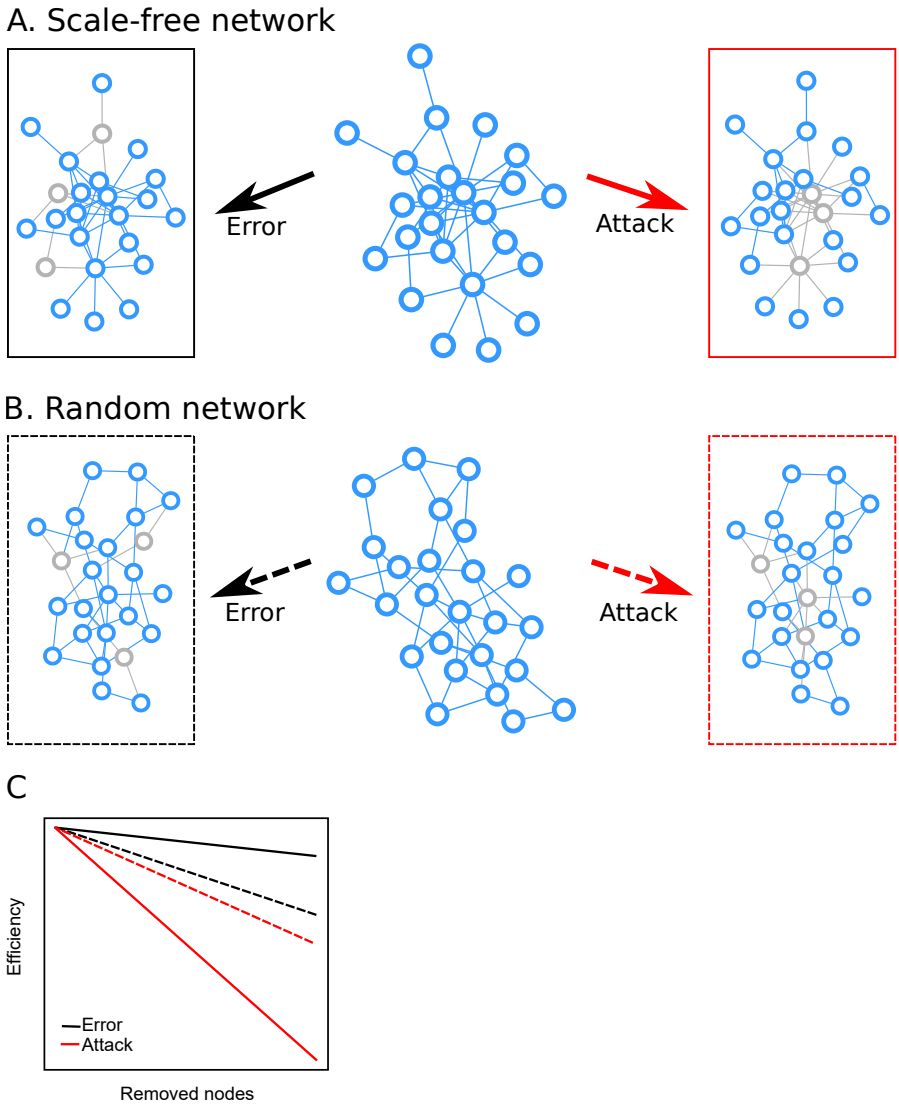


Figure 1.4: Robustness of different network topology to node removal by error or by attack. A) The random removal (error) of nodes in a scale-free network (solid lines) has only a low impact on the overall network structure. The targeted removal (attack) of central proteins and hubs destroys the networks and separates parts of the network. B) In the random network (dashed lines), both the random removal and targeted removal of nodes have a similar effect. C) Efficiency of the scale free network decreases very fast with number of removed node at targeted removal of nodes. Adapted from [6]

techniques are e.g. **fast greedy** [29] and **louvain** [30]. **Fast greedy** is a bottom-up method, which iteratively joins nodes / communities based on optimization of a modularity function. It tries to find the largest increase of modularity, which is described as the fraction of interactions within communities compared random expectation of interactions within a community [29].

Label propagation is an approach, where each node is assigned one of k labels. In an iterative manner, labels are updated by majority voting in the neighborhood of a vertex. Depending on the initial labeling of the nodes, different results are obtained. Therefore the algorithm has to be run a large number of times [31].

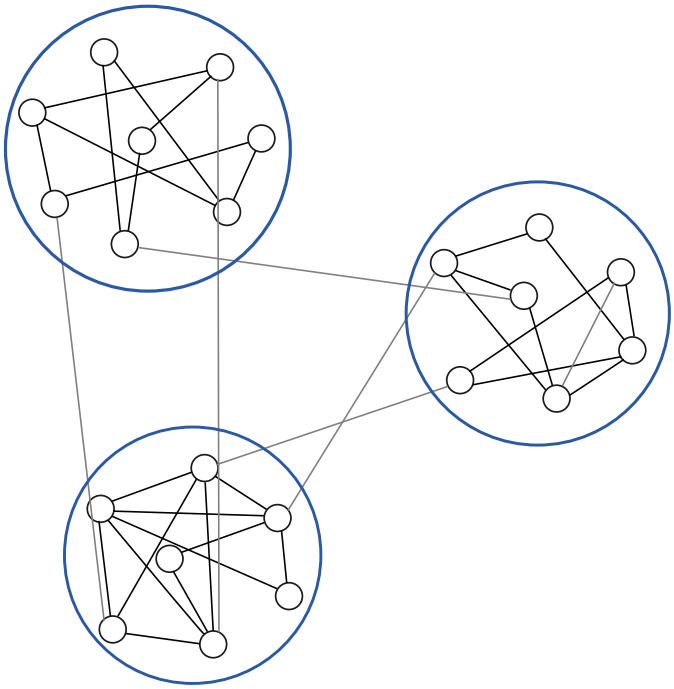


Figure 1.5: Schematic representation of a network with community structure. Blue circles show nodes, which form a community. Figure adapted from [25].

Only a selection of algorithms for identification of communities is presented here. Additional algorithms have been proposed for this task. Albeit, there is no general rule, which algorithm is the best. For each network, a suitable algorithm has to be identified on the basis of the network size and structure. It can be also helpful to compare clusterings of different algorithms to estimate suitability [32].

1.1.7 Association network

In networks often only a fraction of the nodes is characterized for their function or biological processes in which they take part. But even if they are annotated, it cannot be excluded that they are involved in additional processes. The interactions in the network can be exploited to transfer interactions between proteins by the "guilt-by-association" principle, which postulates that nodes with a similar interaction profile may have a similar function [33]. The interaction profile of a node is defined by the interactions and non-interactions with all other nodes in the network. For calculation of the interaction profile similarity, the number of common and unique interactions are compared and used to calculate a similarity value for a pair of nodes (fig. 1.6). Different measures are suitable to describe the similarity of the interaction profile, e.g. Jaccard index, Simpson index, Geometric index, Cosine index, Hypergeometric index and Pearson Correlation Coefficient (PCC). As biological networks are sparse, measures, which consider shared non-interaction partners are not suitable for this analysis [33].

All listed similarity measures except PCC are in a range of 0 to 1, PCC is in a range of -1 to 1. The example in figure 1.6, shows that the similarity measures score similarity differently. Additionally, a similarity can result in identical values, even if the overlap or the number of distinct interactions are different. The similarity measures have

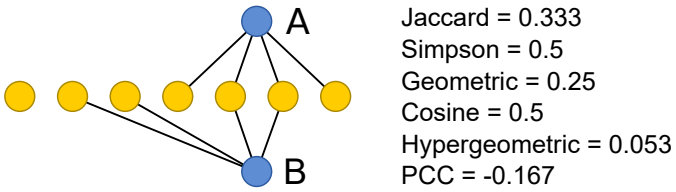


Figure 1.6: Example for interaction profile similarity. Node A and node B have two interactions in common and each node has two unique interactions. Similarity values of different similarity measures are shown right of network. (PCC: Pearson Correlation Coefficient). Adapted from [33].

strengths and weaknesses and have to be selected depending on the network structure and biological question [33].

Limitations of network approaches

Networks often look like a confusing collection of interactions, where it is difficult to identify structural details or functional groups (interaction hairballs). But networks are the result of a growth process and evolutionary adaption. Intrinsic structures and properties of biological networks can be elucidated by targeted analyses: networks have a scale-free topology with hierarchical modular structure and they are very robust against random failures of nodes, but vulnerable for targeted attack on hubs. The limitations of network approaches are the elucidation of the dynamics in terms of temporal and local occurrence of interactions, the characterization of modules for their functions and adaptation of the network to environmental stress conditions. Therefore additional data must be integrated.

1.2 Data integration

The integration of additional data can be used to further exploit networks and results from network analysis. However, the integration of data is not restricted to research on biological networks, but is an integral part of research in life sciences as the integration of multiple sources of information and data allows to better understand biological processes. With the possibility to generate large amounts of data using new sequencing techniques and other high-throughput methods, data integration becomes even more important [34, 35].

In case of integrated network analyses, diverse data can be integrated: to elucidate the functions of proteins of a certain community or identify pathways, gene ontology (GO) annotations can be integrated. To determine under which condition an interaction takes place, e.g. expression data can be used. The integration of information about polymorphisms in genes allow to explore evolutionary constraints acting on single genes. In the following these examples are further elucidated.

1.2.1 Gene Ontology annotations

The GO is a structured, precisely defined, common, and controlled vocabulary to describe the roles of genes and gene products in any organism. The GO contains terms, which are arranged in a hierarchical structure, where connections between the terms describe their

well-defined relationships. This forms a direct acyclic graph, which has three generic terms at the top: *biological process*, *molecular function* and *cellular component*. With each level below that, the description becomes more specific. The three top level terms divide the ontology into three categories: *Biological process* contains objectives, to which a gene or a gene product contributes. Each process consist of one or more functions, which often involve the chemical or physical transformation of an entity. The *molecular function* category describes the specific biochemical activity of a gene product. The *cellular component* category describes the place / organelle in the cell, where a gene product is active [36].

The GO database[1] can be used to search and visualize annotations of one or more gene products. For analysis, if a set of gene products share the same GO terms, the enrichment analysis is very common and a lot of tools for both enrichment analysis and result visualization have been developed. For both, analysis and visualization, online tools and plug-ins / packages for diverse programs / programming languages are available: Cytoscape [37] plug-ins: BiNGO [38], ClueGO [39]; online tools: FatiGO [40], FatiGO+ [41], DAVID [42], GOstat [43], GOrilla [44]; R packages: GSEABase [45], topGO [46] and visualization tools: REViGO [47]. In enrichment analysis, a set of gene products is tested for significantly overrepresented GO terms. Typically a set of genes is tested for enrichment against a background set using a hypergeometric or a binomial model. Enriched GO terms may suggest possible functions / biological processes of the gene products [44].

GO annotations are a resource for knowledge about specific genes collected from many publications and derived from diverse experiments. However GO annotations are not the proper resource, in case of systematic analysis of genes under specific conditions.

1.2.2 Expression data

Measuring expression of genes allow a genome-wide analysis of expression under diverse conditions in the complete organism or in single tissues. Differentially expressed genes under stress conditions, under treatment with various compounds, or at different time points compared to standard / start conditions can be identified.

The advent of the RNA-Seq technique revolutionized transcriptome profiling compared to other methods: RNA-Seq is not limited to detect transcripts that correspond to existing genomic sequence; it can reveal precise location of transcript boundaries, connectivity between exons, and sequence variation. RNA-Seq has a large dynamic range to measure expression levels, it can also measure expression level of very low and high expressed genes, where DNA microarrays lack sensitivity [48].

For measuring expression levels with RNA-Seq, the mRNA fraction of interest has to be extracted (fig. 1.7). The selected mRNA is fragmented and reverse transcribed into cDNA, which is subsequently sequenced with Next Generation Sequencing (NGS) technique. The resulting single end or paired end reads are aligned to a reference genome or transcriptome sequence, which is not a trivial task due to up to hundreds of millions of reads with sometimes ambiguous fits. From the degree of coverage of the reference sequence the expression levels of the respective genes / gene models are derived. The expression levels are represented as counts or using relative measures like RPKM (reads per kilobase per million mapped reads), FPKM (fragments per kilobase per million

[1] `http://www.geneontology.org`

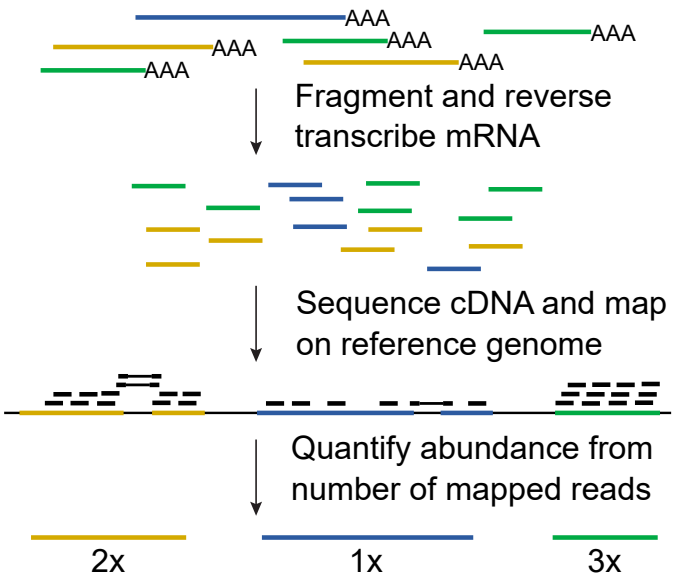


Figure 1.7: Overview about RNA-Seq pipeline. To identify expression of all genes under a certain condition or tissue, the mRNA is fragmented and reverse transcribed into cDNA. cDNA is fragmented and sequenced with Next Generation Sequencing (NGS) technique. Single end reads or paired end reads are aligned to the reference genome. From the number of mapped reads (coverage), the expression of each gene can be determined. Figure adapted from [49].

mapped reads), or TPM (transcripts per million), which are used for single-end read, paired-end reads or as measure normalized for gene length, respectively [49].

Resources for raw and already mapped RNA-Seq data are e.g. Gene Expression Omnibus (GEO)[2], Sequence Read Archive (SRA)[3], and ArrayExpress[4]. A large number of algorithms and software has been developed to analyze RNA-Seq data for differential expression [50].

The information about gene expression and differential gene expression can not only used to analyze single genes, but also binary protein complex abundance can be predicted. From mRNA levels, the protein abundance and hence the protein complex abundance can be modeled. In yeast, mRNA levels could explain a large portion of network dynamics in protein-protein interactions observed while diauxic shift [51].

1.2.3 Natural variation

Network dynamics on an evolutionary level can be investigated by including natural variation. The possibility to sequence large amounts of genomic sequences using NGS techniques also facilitated a deeper examination of variation in genomes. Far more polymorphisms can be detected with this technique compared to DNA microarrays. Mutations can be advantageous for an individual and increase its fitness, deleterious, if they decrease the fitness, or neutral, in case they have no effect on the fitness. The main interest in population genetics is to distinguish variations, which are caused by random genetic drift and neutral for the organism, from variations which are subject to selection

[2] www.ncbi.nlm.nih.gov/geo/
[3] www.ncbi.nlm.nih.gov/sra
[4] www.ebi.ac.uk/arrayexpress/

(caused by either advantageous or deleterious mutations). A distinction is made between different types of selection: *positive selection* is any type of selection, where mutations are advantageous and driven to fixation in a population, whereas *negative selection* or *purifying selection* is any type of selection, where mutations are deleterious and therefore removed from a population. A case, where more than one phenotype / allele is favored in a population is called *diversifying selection* or *balancing selection*. *Neutral selection* is present, if a mutation is neither favored nor discriminated, but comes to fixation by chance. The continued removal of deleterious mutations by negative selection is described by *background selection* [52, 53].

A lot of **selection measures** have been developed to identify alleles, which show evidence to have not evolved under neutral selection and to infer the type of selection from SNPs identified within a species or between species.

For detecting selection on the microevolutionary level, SNPs in alleles within a species are analyzed. For example, positive selection causes a beneficial allele to sweep to high prevalence or fixation rapidly within a population (population sweep). This leads to a reduction of genetic variation around the selected mutation. New mutations appear over time and lead to an excess of rare alleles. This can be detected by *frequency spectrum based methods* like Tajima's D, which compares the average number of nucleotide differences between pairs of sequences with the total number of segregating sites (SNPs) [54].

Tests, which include also sequences from other species detect selection on a macroevolutionary level. In these tests, like *McDonald-Kreitman Test* and d_N/d_S (Ka/Ks), the variation between species is included to determine the type of selection [54].

A third group of methods is based on the *linkage disequilibrium (LD)*. After a sweep, a selected allele has a strong linkage disequilibrium with its neighboring hitchhiker variants until it breaks down by recombination. The causal allele and its linked neighboring variants define a haplotype. LD based methods detect positive selection by looking for extended regions of strong LD compared to LD in the population. A region with a strong LD must have come to prevalence in relatively short time, otherwise recombination would have shortened the haplotype [54].

The fourth group of methods tests *differentiation between populations* of a species. The assumption of these tests is that in a particular environment different alleles are prevalent, because of different environmental pressures. If a selective pressure is acting on a certain locus between population, but not within populations, then the allele frequency among populations should be significantly different. A method, which compares allele frequencies within and between populations is e.g. *Wright's fixation index* (F_{st}) [54].

All tests can only suggest evidence for selection and different events affecting population structure like population expansion can be confounding. For evidence of selection, a combination of tests should be used [54, 55]. A lot of software packages and packages in R are available to analyze the large quantity of data for evidence of selection with different measures [56, 57].

NGS techniques and selection measures have already been deployed to gain insights into polymorphisms and adaption of *Arabidopsis thaliana* as it is a good model for identification of genes under selection: *Arabidopsis thaliana* is highly selfing and naturally exists as inbred lines. It is spread over Europe, Asia, Northern Africa and recently colonized North America [58]. It thrives in very diverse habitats, and is adapted to different temperature and precipitation conditions [59, 60, 61].

In the first phase of the 1001 Genomes Project[5], 80 accessions from eight regions throughout the species' native range were sequenced. It was detected, that recent species-wide selective sweeps are rare in *Arabidopsis* and deleterious mutations occur mainly in marginal populations [62].

For the final dataset of the 1001 Genomes Project, 1,135 natural inbred lines were re-sequenced to investigate the demographic history of *Arabidopsis*. More than 10 million biallelic SNPs and more than 1.4 million small-scale indels up to 40 bp were identified, which means on average one variant every 10 bp of the single copy genome. It was found, that pairwise differences between populations do not reflect the geography of the ecotypes, moreover relict groups and a complex pattern of colonization after last ice age was identified [58].

The sequencing of 180 lines from Sweden also uncovered signs of selection in ecotypes from a single region: selective sweeps were found in lines from northern Sweden and large structural variants indicate extreme strong selection [63]. More examples of local adaption are found in 173 lines from Sweden, which show strong adaption of flowering time to temperature [64].

A high number of polymorphisms is not only observed in genes related to flowering [65], but also in genes related to abiotic and biotic stress: Genes related to resistance against the pathogen *Hyaloperonospora arabidopsidis* show natural variation [66] and in genes involved in salt tolerance an excess of polymorphisms could be identified [67]. NGS technique allows to extend these analyses also to other plants like tomato, where also a high adaption to pathogen resistance was found [68].

The identified variations lead to visible phenotypes in the organism as a result of adaption to the environment. The variations affect various processes in the organism like regulatory processes by affecting non-coding regions [69] and have also implications for the functions of proteins, which among others affect protein-protein interactions.

On a longer evolutionary distance, the effect of variation on protein-protein interactions was shown for yeast proteins. A highly significant correlation between the number of interactions and the evolutionary distance of *S. cerevisiae* to either *Candida albicans* or *Schizosaccharomyces pombe* was observed [70]. This could be also shown in a shorter evolutionary distance for *Arabidopsis* proteins: the more distant two paralogous proteins are, the more divergent is the sequence and the lower is the number of common interaction partners [11]. The analysis of selective pressure acting on paralogous proteins, revealed a higher selective pressure on functional sites than on the rest of the protein [71]. Hub proteins, which are central of protein interaction networks, are observed to be under negative selection, whereas positive selection is found in peripheral regions of the network. This may be due to an increase of adaptive events in peripheral proteins [72]. Moreover the effect of deleterious variations in peripheral regions of the network is less severe than mutations affecting more central proteins [73, 74].

1.3 Phytohormone signaling

Phytohormones are small molecules, which together regulate virtually every aspect of growth, pattern formation, and reactions to biotic and abiotic stress of the plant. Phytohormone signals are perceived by receptor proteins and transmitted via PPIs. The signals of distinct hormones influence each other for a tight control of processes. So far,

[5] `http://1001genomes.org/`

a number of phytohormones and several compounds with hormone like effect have been found. Here I focus on the signaling pathways of abscisic acid (ABA), indole-3-acetic acid (IAA), the most abundant form of auxin (AUX), brassinosteroids (BRs), cytokinins (CKs), ethylene (ET), gibberellins (GAs), jasmonates (JAs), karrikin (KAR), salicylic acid (SA), and strigolactones (SLs) and give a brief overview about the core signaling pathways.

1.3.1 Signaling pathways

The hormone signaling machinery of these hormones emerged during evolution of land plants [21] and some signaling pathways share share similar processes for signal transduction, like ubiquitination by an SKP1, Cullin, F-box (SCF) containing complex, signaling by kinases (fig. 1.8), or localization of the receptor. All signaling pathways have in common that specific protein-protein interactions only take place in presence of the respective hormone, which leads in turn to signal transduction by altered protein-protein interactions and activation of transcription factors. These protein interactions form the core pathway, where the hormone signal is perceived by a receptor protein and forwarded by protein-protein interactions to the transcription factor. These protein interactions are only a small part of the interactome and the involved proteins influence also other processes by interacting with the proteins involved in these processes. In the following section the core pathways for the above listed phytohormones are briefly described.

Abscisic acid signaling is mainly associated with seed dormancy, drought responses and other growth processes [75]. In the absence of ABA, a clade of protein phosphatases 2C (PP2Cs) dephosphorylate SNF1-related kinases (SnRKs) 2.2, 2.3 and 2.6, which inactivates them and no signaling takes place [76]. In presence of ABA, the soluble receptors from the PYRABACTIN RESISTANCE (PYR)/PYR-LIKE (PYL)/REGULATORY COMPONENT OF ABA RECEPTOR (RCAR) family bind ABA, which facilitates binding to the PP2Cs ABI1 and ABI2 and inactivates them [77, 78]. SnRKs are activated by autophosphorylation and as a result SnRKs phosphorylates ABA-responsive transcription factors like ABSCISIC ACID-RESPONSIVE ELEMENT-BINDING FACTORs (ABFs) and ABSCISIC ACID-INSENSITIVE5 (ABI5) [79].

Brassinosteroids are involved e.g. in cell elongation, photomorphogenesis, seed germination, flowering, male fertility, senescence, and pathogen defense. The most active BR is brassinolide (BL), which is perceived by a family of three leucine-rich repeat (LRR) receptor like kinases, which are located at cell surface. At low BR levels, the main BR receptor BRASSINOSTEROID-INSENSITIVE 1 (BRI1) interacts with BRASSINOSTEROID INSENSITIVE1 KINASE INHIBITOR (BKI1), BRASSINOSTEROID-SIGNALING KINASE (BSK), and CONSTITUTIVE DIFFERENTIAL GROWTH1 (CDG1) and is inactive. The transcription factors BRASSINAZOLE-RESISTANT 1 and 2 (BZR1 and BZR2) are phosphorylated by BRASSINOSTEROID INSENSITIVE2 (BIN2) and retained by 14-3-3 proteins [21]. At high BR levels, BL binds to BRI1, which triggers interaction with the co-receptor BRI1-ASSOCIATED RECEPTOR KINASE1 (BAK1) [80]. This leads via phosphorylation of BSK1 and CDG1 to the phosphorylation of BRI1-SUPPRESSOR1 (BSU1) [81]. BSU1 phosphorylates BIN2, which inactivates it [82]. The transcription factors (TFs) BZR1 and BZR2 are in turn dephosphorylated by PP2A, which activate transcription of BR-responsive genes [82]

The **cytokinin** signaling pathway is involved in regulation of cell division, chloroplast development, leaf senescence, root and shoot growth, and branching. The CK recep-

tors ARABIDOPSIS HISTIDINE KINASE2 (AHK2), AHK3, and AHK4 are bound to endoplasmatic reticulum (ER) and the plasma membrane [83, 84, 85]. Upon binding to CK, AHKs autophosphorylate themselves and transfer the phosphate moiety afterwards to ARABIDOPSIS HISTIDINE PHOSPHOTRANSFER PROTEINs (AHPs) [86]. AHPs transport phosphate groups into the nucleus, where the B-type ARABIDOPSIS RESPONSE REGULATORS (ARR) is phosphorylated and can subsequently control expression of CK-dependent genes. The A-type ARRs, which are expressed upon CK signaling, act as negative feedback by contending about phosphate groups from AHPs [87].

Ethylene regulates fruit ripening, senescence, abscission, and response to biotic and abiotic stresses. ET is perceived via five ER membrane bound receptors: ETHYLENE RESPONSE 1 and 2 (ETR1 and ETR2), ETHYLENE-RESPONSE SENSOR 1 and 2 (ERS1 and ESR2) and ETHYLENE-INSENSITIVE 4 (EIN4) [88], which interact with CONSTITUTIVE TRIPLE RESPONSE 1 (CTR1) [89]. In the absence of ET, this kinase phosphorylates ETHYLENE INSENSITIVE2 (EIN2) [89], which in turn is degraded by the SCF complex with ETHYLENE INSENSITIVE2-TARGETING PROTEIN1/ 2 (ETP1/2) [90]. Additionally the TFs in the nucleus, EIN3 and EIN3-LIKE 1 to 3 (EIL1-3), are degraded in absence of ET by two F-Box proteins, EIN3-BINDING F-BOX 1 and 2 (EBF1 and EBF2) [91, 92]. In presence of ET, it is bound by the ET receptors, which prevent CTR1 from phosphorylating EIN2. As a consequence EIN2 is not degraded and its C-terminal part is be cleaved of and translocated to the nucleus, where it stabilizes the TFs EIN3 and EIL1-3 [89].

Salicylic acid is an important regulator of the immune response of plants and can induce systemic acquired resistance and hypersensitive response upon pathogen infection [93]. NON-EXPRESSER OF PATHOGENESIS-RELATED GENES 1 3 (NPR3) and NPR4 have been identified as SA receptors. NPR4 has a higher SA affinity than NPR3 and binds SA, if it is present at low concentrations. NPR4 subsequently inhibits degradation of NPR1, which accumulates and activates WRKY domain containing protein (WRKY) and TGA binding (TGA) transcription factors, which leads to basal resistance [93, 94]. Under high concentration of SA, it binds to NPR3. In this case NPR1 is degraded and allows programmed cell death [93, 94].

Gibberellic acid or also called **gibberellin** is involved in the regulation of germination and elongation growth in the plant [95]. In absence of GA, DELLA proteins (REPRESSOR OF GA1-3 (RGA), GA-INSENSITIVE (GAI) and RGA-LIKE 1 - 3 (RGL1 - RGL3)) bind to e.g. PHYTOCHROME-INTERACTING FACTOR (PIF) transcription factors and inhibit its activation. In presence of GA, GA binds to the receptor GIBBERELLIN INSENSITIVE DWARF1 (GID1) [95, 96]. This leads to a conformation change and promotes binding to DELLA proteins. This complex binds to the F-box proteins SLEEPY1 / 2 (SLY1/2). DELLA proteins become ubiquitinated and afterwards degraded. The derepression of e.g. PIFs activates expression of GA-dependent genes [97].

The recently discovered **karrikin** signaling pathway is involved in seed germination and early plant development. KARs are not endogenous compounds as the other nine phytohormones, but they are produced by burning sugar compounds, e.g. by wild fires. It is assumed that KARs mimic a so far unidentified compound, which could be similar to SL. The signaling pathways of SL and KAR are also very similar: under low KAR levels, SUPPRESSOR OF MAX2-1 (SMAX1) acts as transcription repressor. Under high levels of KAR, SMAX1 is bound by the co-receptor KARRIKIN-INSENSITIVE2 (KAI2). This complex interacts with MORE AXILLARY BRANCHES2 (MAX2) and

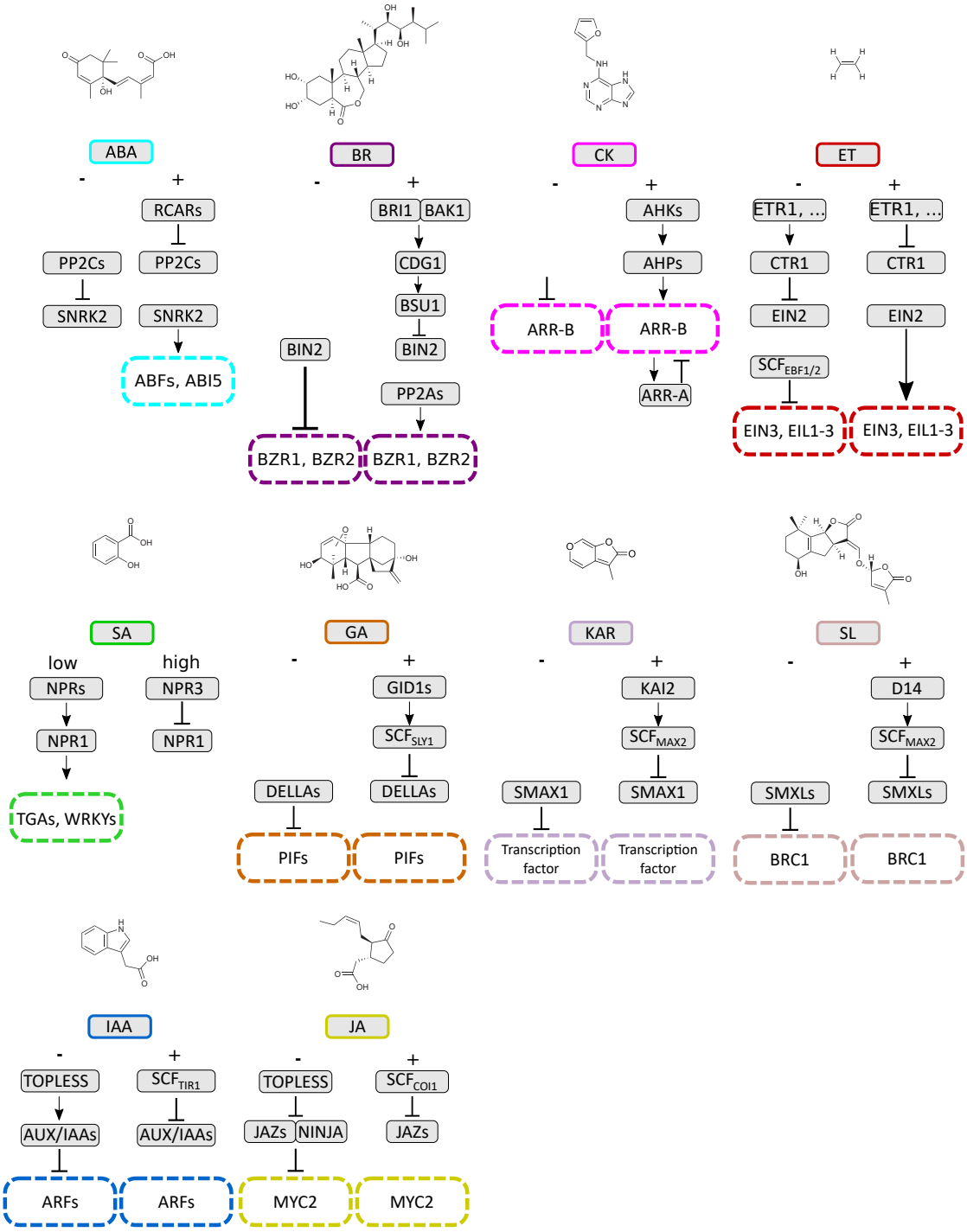


Figure 1.8: Hormone signaling pathways of ten phytohormones. For each pathway an exemplary protein or protein family is shown for receptors and intermediates. On top of each signaling pathway the structure of the respective hormone or one representative of the hormone family is shown. Pathways according to [21, 98]. Molecule structure of hormones except KAR adapted from [75]. KAR structure (KAR_1) from [99]. Arrows indicated activation or binding, blocking arrows indicate repression. "-" indicates low or no hormone present, "+" indicates high hormone presence. Transcription factors are in box with dashed border.

leads to ubiquitination by the formed SCF complex and degradation of SMAX1 by the 26S proteasome [99, 100].

Strigolactone is a recently identified phytohormone, involved in shoot branching, seed germination and seedling photomorphogenesis [101]. Its signaling pathway is similar to the GA pathway. In the absence of SL, SUPPRESSOR OF MAX2-1-LIKE PROTEINS (SMXLs) is likely to interact with TOPLESS (TPL) and represses BRANCHED1 (BRC1) [102]. In presence of SL, the receptor DWARF 14 (D14) perceives SL and interacts with MAX2, which leads to degradation of SMXLs. The transcription factor BRC1 is released, which leads to transcription of SL-dependent genes [101].

Indole-3-acetic acid, the most abundant form of bioactive **AUX**, is important in the regulation of plant growth and development, where it controls cell expansion and division [103]. Under low AUX levels, AUXIN/INDOLE-3-ACETIC ACID (AUX/IAA) proteins bind to the transcription factors AUXIN RESPONSE FACTORs (ARFs) and repress their activity by binding the transcriptional repressor TPL [104]. Under high AUX levels, AUX/IAAs bind to one of the six nuclear-localized F-box co-receptors for AUX called TRANSPORT-INHIBITOR RESPONSE 1 (TIR1) and AUXIN SIGNALLING F-BOX PROTEIN 1 to 5 (AFB1 - AFB5), which leads to ubiquitination by the SCF complex and subsequent degradation of AUX/IAAs and activation of IAA-dependent gene expression [105].

The bioactive molecule of **jasmonic acid** is conjugated to isoleucine (JA-Ile). The signaling pathway, very similar to the AUX signaling pathway, regulates defense against wounding and herbivores and is important for plant fertility and reproduction [106]. Under low JA levels, JASMONATE ZIM-DOMAIN PROTEINs (JAZs), NOVEL INTERACTOR OF JASMONATE ZIM-DOMAIN PROTEIN (NINJA), and TPL bind to transcription factors of the MYC-RELATED TRANSCRIPTIONAL ACTIVATOR (MYCs) family and repress their activation [107]. Under high JA levels, JAZ proteins are bound by the F-BOX protein and JA receptor CORONATINE-INSENSITIVE 1 (COI1). Like in the AUX pathway, JAZs are ubiquitinated and subsequently degraded. The lack of repression of MYC transcription factors activates JA-dependent gene expression [105].

1.3.2 Interactions of signaling pathways

The main phytohormone signaling pathways presented above illustrate that hormone signals are transmitted by protein-protein interactions and other mechanisms. This leads to activation or repression of hormone-dependent gene expression and other mechanisms like protein degradation. The different hormones are involved in the regulation of specific biological processes. Many of these processes are regulated by a concerted interaction of several phytohormone signaling pathways. In a number of studies, the activation of different phytohormone signaling pathways under the same condition and their reciprocal influence on each other could be elucidated.

Different hormone signaling pathways are activated as response to different stresses: The ABA signaling is mainly responsible for defense against abiotic stress, whereas SA, JA and ET are involved in response to biotic stresses. SA is activated in defense against biotrophic and hemi-biotrophic pathogens, JA and ET are activated in defense against necrotrophic and herbivorous insects. These four signaling pathways are known to interact for regulation of plant defense responses but also growth of the plant has to be adapted to stress conditions, which requires crosstalk with the growth promoting hormone signaling pathways of IAA, GA, and CK [108].

SA and JA have an antagonistic relationship, where SA can repress JA response to wounding. The suppression of JA response is likely mediated by NPR1 [109]. Also WRKY70 was found to repress the JA responsive gene *PDF1.2* [110]. SA signaling can also repress expression of *TIR1/AFB* genes, involved in AUX signaling and repress in this way AUX signaling [111]. A functional interaction between IAA and ET pathways has been observed in the regulation of root development [108]. Also JA and ET show a synergistic behavior. Both pathways induce / stabilize EIN3 in the process of root hair development [112] and are required to activate expression of ERF1 with subsequent activation of *PATHOGENESIS RELATED (PR)* genes [113]. Also CK and SA signaling act together in regulation of defense against hemi-biotrophic pathogens . The ABA and CK signaling pathways influence each other: the ABA signaling pathway can suppress CK synthesis and CK receptors can downregulate ABA signaling [108].

The evidence for crosstalk between hormone signaling pathways is mainly derived from phenotypes identified in mutant plants. Mutant plants, which show a phenotype in one hormone signaling pathway, also show a phenotype in a different hormone signaling pathway. From these experiments, a complex interaction between all phytohormone signaling pathways is assumed. However, the specific crosstalk between two pathways depends on the tissue, developmental stage, and environmental conditions [97].

The crosstalk between two pathways or the integration of signals from different signaling pathways can take place on different "levels": (i) regulation of hormone metabolism, (ii) hormone distribution, (iii) gene expression, (iv) control of key components of signaling pathways by other hormone signals [97], and (v) direct protein interactions.

The reciprocal **regulation of hormone metabolism** is e.g. found between AUX and ET signaling pathway: genes required for AUX biosynthesis (*alpha and beta subunit of anthranilate synthase* (*ASA1* and *ASB1*) and *tryptophan aminotransferase* (*TAA1*)) are under transcriptional control of ET. Consequently, treatment with ET increases AUX biosynthesis [114, 115, 116]. Rate-limiting genes involved in ET biosynthesis (*1-aminocyclopropane-1-carboxylate synthase* (*ACS*)) are under transcriptional control of the AUX signaling pathway [117]. Also JA production is influenced by AUX: in *arf6 arf8* mutants jasmonate levels are lower [118].

Crosstalk on the level of **hormone distribution** can be found in the root. During root development, AUX promotes lateral root initiation, while CK inhibits it by influencing expression of *PIN* genes, among others [119]. PIN proteins control the direction of AUX transport and establish an AUX gradient in the root [120]. Reduction of PIN expression disrupts formation of an AUX gradient and inhibits lateral root formation [119].

An interaction between two hormone signaling pathways on the level of **gene expression** is found e.g. between AUX and BR. A common set of genes are induced and repressed by both IAA and BR. A direct influence of BR on IAA-dependent gene expression was recently found: the BR-regulated kinase BIN2 directly regulates AUX response factor ARF2 by phosphorylation, which is an expression repressor. The phosphorylation inactivates ARF2 and activates IAA-dependent gene expression [121]. There was also found evidence for direct binding of the BR-regulated transcription factor BES1 to regulatory elements in the promoter region of GA biosynthesis genes and control their expression [122]. An overview of currently elucidated crosstalks on gene expression level is shown in figure 1.9.

So far only a low number of contacts between different phytohormone signaling pathways by direct protein-protein interactions were elucidated. One example is the above mentioned interaction between the BR-regulated kinase BIN2, which interacts with

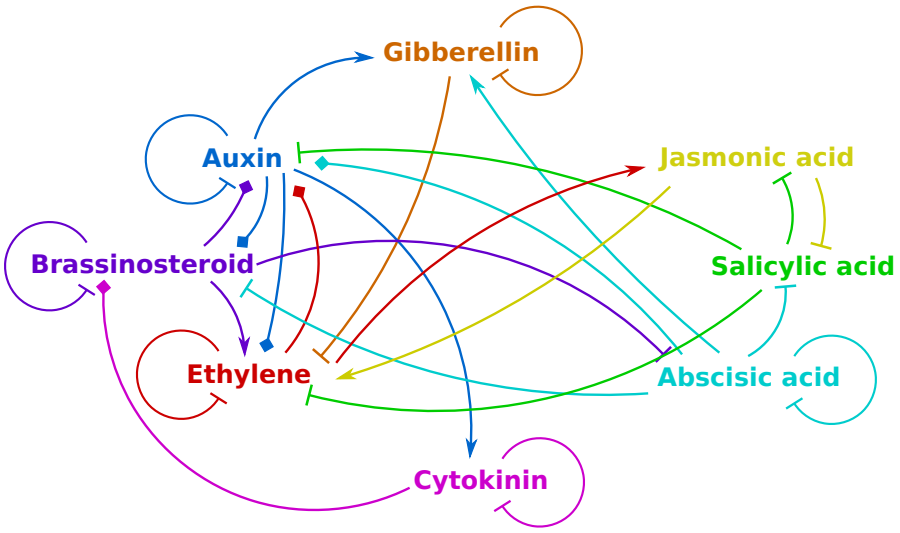


Figure 1.9: Regulation of hormone-dependent gene expression by crosstalk between hormone signaling pathways. Lines with arrowhead represent upregulation of hormone biosynthetic genes or downregulation of genes involved in hormone inactivation. Block arrows show opposite regulation. Diamond arrowheads represent ambiguous changes of gene expression. Figure adapted from [123] and extended with [124], [111], [108], [125], [126]

ARF2 involved in regulation of AUX-dependent gene expression [121]. A crosstalk between BR and GA was found to be mediated by direct interaction of BZR1 and RGA, a member of the DELLA protein family [127].

1.3.3 Signal transduction by phosphorylation

In some of the phytohormone core signaling pathways, the hormone signal is transmitted via phosphorylation of proteins by kinases. Moreover, kinases are involved in signal transduction in many biological processes like response to light, pathogen invasion, hormones, temperature, or nutrient deprivation. The signal is transmitted by kinases via transfer of the γ-phosphate from ATP to a serine, threonine or tyrosine of its substrate protein [128]. These phosphoamino acids diversify the chemical nature of protein surfaces. The phosphate group linked to one of the three amino acids can form hydrogen bonds or salt bridges. Phosphorylated proteins can be identified by phosphospecific-binding domains in other proteins and promote protein-protein interactions [129]. The phosphorylation is reversible, as the phosphate moieties can be removed by phosphoprotein phosphatases [128].

In *Arabidopsis thaliana*, a large number of genes encode kinases: so far 942 kinase domains in 940 proteins were identified. This corresponds to 3.4 % of all genes annotated in TAIR10. Of these 940 kinases, 561 are membrane-located receptor kinases and 381 soluble kinases. The membrane-bound kinases consist of large superfamily of receptor-like kinases (RLK) including transmembrane leucine-rich-repeat (LRR) receptor kinases, e.g. involved in sensing pathogen associated molecular patterns. The group of soluble kinases consists of 21 distinct kinase families, of which the largest families are mitogen-activated kinase cascades (MAPKs, MAP2Ks, MAP3Ks) involved in trans-

mission of extracellular stimuli, calcium-dependent kinases (CDPKs), cyclin-dependent kinases (CDKs), Snf-related kinases (SnRKs), and AGC kinases [130].

Thousands of in vivo phosphorylation sites have been identified, but for a lot of them the phosphorylating kinase is unknown [131, 132]. For some kinases / kinase families a consensus sequence around the phosphorylated site was identified. The consensus sequence defines a simplified motif, which specific amino acids occur most frequently up- and downstream of a phosphorylated site. These motifs differ strongly in size and complexity [133]. A phosphorylation motif of AGC kinases is e.g. TPRXpS(N/S), where p denotes the phosphorylated position [134].

1.4 Pathogen-host interactions

Plants are threatened by diverse pathogens, which use different strategies to use a host plant as basic food resource. Nematodes and aphids feed directly from plant cells using a stylet. Bacteria enter plants through pores and proliferate in the intercellular space (apoplast). Fungi and oomycetes grow hyphae between or on top of cells. They can additionally form haustoria in the plasma membrane of cells (fig. 1.10) [135, 136].

Plants developed a two-tiered strategy to defend against pathogens. All pathogens express small molecules, pathogen (or microbial)-associated molecular patterns (PAMPs or MAMPs), like flagellin from bacteria or chitin from fungi. PAMPs are recognized by the extracellular domain of pattern recognition receptors (PRRs). PRRs are divided in two classes: transmembrane receptor kinases and transmembrane receptor-like proteins lacking an internal signaling domain[135, 136]. *Arabidopsis thaliana* has 561 transmembrane receptor kinases [130] and 57 transmembrane receptor-like proteins [137].

After recognition of PAMPs by PRRs a PAMP Triggered Immunity (PTI) response is induced. An example for this process is the recognition of flagellin by the receptor FLAGELLIN-SENSITIVE 2 (FLS2), which interacts with the co-receptor BAK1, which leads to a transphosphorylation and subsequent activation of kinase signaling cascades (fig. 1.10). This leads to the transcription of at least 1,100 genes. Other PTI responses could be rapid influx of calcium ions, burst of active oxygen species, activation of mitogen-activated protein kinases (MAPKs), deposition of callosic cell wall appositions or localized cell death [138].

To suppress PTI, some pathogens are able to deliver their effector proteins into the host cell. Bacteria encode about 20-30 effector proteins, which are directly injected in the host cell via a type-III secretion system. Effector proteins target host proteins involved in PTI in order to suppress it. The injected effector proteins are recognized by nucleotide-binding leucine rich repeat (NB-LRR) proteins, which leads to the effector triggered immunity (ETI) [135].

Three models have been proposed, how effector proteins are recognized by NB-LRR genes: (i) direct interaction of NB-LRR proteins with effector proteins, (ii) effector proteins modify a protein and this modification in turn is recognized by the NB-LRR protein, and (iii) the effector protein binds to the host protein and this complex is recognized by the NB-LRR protein. PTI also activates kinase signaling, which leads to transcription of defense genes [135].

Important downstream responses of ETI and PTI are the activation of the hormone signaling pathways of SA, JA, and ET. These pathways are important regulators of the expression of defense genes. SA is involved in the resistance to biotrophic pathogens, whereas JA-ET are involved in resistance to necrotrophic pathogens. Recently it has

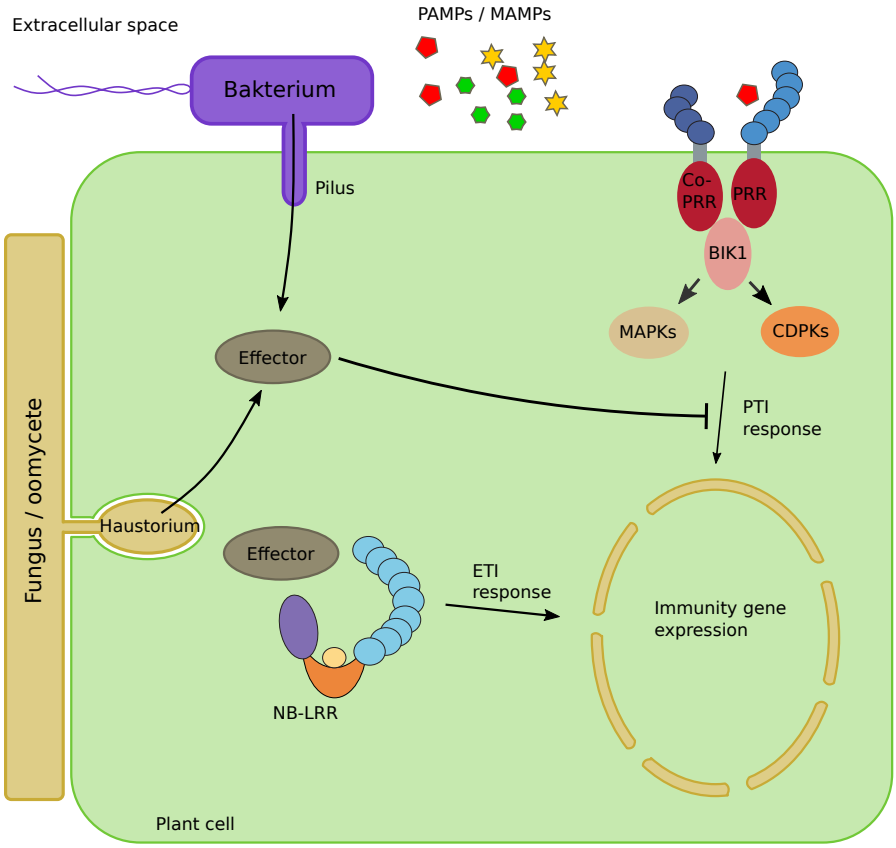


Figure 1.10: Bacteria, fungi and oomycetes colonize plants and propagate in the extracellular spaces of plant tissues. Fungi and oomycetes extend hyphae into this space and can also grow haustoria, which are specialized feeding structures. All pathogens, including nematodes and aphids release small molecules, so called pathogen (or microbial)-associated molecular patterns (PAMPs or MAMPs) into the extracellular space, which are perceived by the extracellular domain of pattern recognition receptors (PRRs). In general, PRRs interact with co-receptors like BRASSINOSTEROID INSENSITIVE 1-ASSOCIATED KINASE 1 (BAK1) and elicit PAMP-triggered immunity (PTI) with its intracellular domain. After transphosphorylation of the complex, signal can be mediated by different kinases like BOTRYTIS-INDUCED KINASE 1 (BIK1), by mitogen-activated protein kinases (MAPKs), or calcium-dependent protein kinases (CDPKs). Bacteria deliver effector proteins via a type III secretion system into the cell; fungi and oomycetes deliver effector proteins via haustoria or other structures into the cell. These effector proteins directly interact with host proteins where they can suppress PTI. Intracellular nucleotide-binding leucine-rich-repeat (NB-LRR) proteins can recognize effector proteins and induce effector triggered immunity (ETI). Figure adapted from [135].

been found, that all three pathways act synergistically in PTI to amplify the immune response [135]. Additionally the IAA pathway can be repressed by PTI [138].

Also pathogens influence the host's hormone signaling pathway: *P. syringae* produces coronatine, a JA mimic, which suppresses SA mediated defense responses. The fungal pathogen *Gibberella fujikuroi* produces GA, which leads to 'foolish seedling' disease. Other pathogens are able to produce CK to retard senescence of leafs [138].

So far only a low number of effector proteins have been systematically analyzed for interactions with proteins from the respective host organism. For plants, effector proteins

of the two pathogens *Hyaloperonospora arabidopsidis* and *Pseudomonas syringae* have been tested for interaction with proteins from *Arabidopsis thaliana* [139].

1.5 Objectives

Using experimental high-throughput methods for protein interaction mapping, millions of protein pairs can be tested for interaction. The high number of identified interactions form large networks and only bioinformatic methods allow to infer global network properties and complex correlations between the network and the biological properties of the proteins. The aim of this thesis was to analyze two protein-protein interaction maps derived from a high-throughput Y2H interaction mapping pipeline and integrate diverse data sources and data from other high-throughput experiments. Below the specific aims for each network are described.

The aim of the analysis of the systematic PhI network was to analyze the network structure of the protein interaction map and correlate the network structure with the underlying biology. In the network, direct and indirect contact points of distinct hormone signaling pathways should be systematically analyzed to determine the extent of signal transduction between distinct hormone pathways. Moreover, the examination of the direction of signal flow should address the issue of missing directions of signaling in a protein-protein interaction network. Therefore, possible kinase-substrate interactions and protein-protein interactions of TFs were examined. As hormone signaling pathways are important for integration of environmental signals into growth decisions, the investigation of genetic signatures of phytohormone signaling proteins addresses the question where adaption to stresses might take place in the genome.

In a host-pathogen network has been shown that effector proteins of two different pathogens converge on a common set of Arabidopsis host proteins. The aim was to integrate the effector-host protein interactions of a third evolutionarily distant pathogen, *Golovinomyces orontii*, into a previously elucidated effector-host network of interactions between *Arabidopsis* and the two pathogens *Hyaloperonospora arabidopsidis* and *Pseudomonas syringae*. With this integrated network, the question should be answered, if the previously observed interspecies convergence can be also found with these three evolutionarily distant pathogens and if there also exists convergence of effector proteins of a single pathogen. Moreover, the question is, which of the effector targets can be genetically validated and if effector binding left population genetic signatures in the host proteins, which show evidence for selection.

2 Results
Phytohormone signal transduction

Phytohormones are small signaling molecules, which are involved in regulation of many biological processes and integration of internal and external stimuli into growth decisions of the plant. Phytohormone signals are usually transmitted via protein-protein interactions from a receptor protein to a transcription factor, which controls hormone-dependent gene transcription.

An unbiased overview about transduction of hormone signals via protein-protein interactions can be gained using a high throughput interaction mapping pipeline to screen a set of proteins related to phytohormones that is as complete as possible.

In the following sections the generation of the PhyhormInteractome (PhI), a protein-protein interaction map for phytohormone related proteins is presented. Subsequently extensive analyses are applied on this network to elucidate the structure of the network and signaling transduction within and between hormone signaling pathways.

2.1 Network mapping

Mapping of protein-protein interactions by high-throughput interaction mapping methods is a crucial step to obtain an as comprehensive as possible, unbiased, and high-quality interaction map. Therefore, selection of loci, cloning of ORFs, interaction mapping using a high-throughput Y2H pipeline, bioinformatic analyses of intermediate stages, and final compilation of the protein interaction map are described in the following. The single steps in this pipeline are partially conducted by colleagues from the group of Pascal Falter-Braun, but results of these steps are essential for later analyses. Therefore results of each step of the pipeline are described. Subsequently, the quality of the Y2H mapping pipeline is assessed and used to estimate the overall completion of the derived protein interaction map.

2.1.1 Search space definition and ORFeome assembly

To generate a systematic and comprehensive protein interaction map of proteins involved in phytohormone signaling, a set of proteins is required, which comprises all proteins, which are involved in phytohormone signaling or likely involved in phytohormone signaling. This set of proteins is called *search space*. This set of proteins was determined from the Arabidopsis Hormone Database (AHD) 2.0 [140], which covers 615 *Arabidopsis thaliana* proteins having genetic evidence from mutant or transgenically overexpressing or silencing *Arabidopsis* lines to be involved in one or several phytohormone signaling pathways for the hormones ABA, AUX, BR, CK, ET, GA, JA, and SA. SL and KAR are not curated in this database. The search space consists of (i) all hormone-related genes with genetic evidence listed in the AHD 2.0 [140], (ii) genes, which are members of a gene family or transcription factor family, which is significantly enriched in the set

of genes having genetic evidence in AHD 2.0 (see table A.1), and (iii) genes proposed by experts. This initial search space comprised 1227 loci, which is about 4.6 % of all *Arabidopsis thaliana* loci. This set of loci is referred as search space PHO (SSP_{PHO}). The described initial composition of the search space was conducted by Pascal Falter-Braun.

The defined list of 1,227 in the search space had to be derived as physical ORFs for the Y2H screen. Out of this list, 688 ORFs were already available as Y2H constructs in the AtORFeome2 collection, which comprises more than 12,000 ORFs (fig. 2.1). The remaining ORFs were derived from ABRC stock center or had to be cloned from cDNA: 163 ORFs could be ordered as Gateway clones, 113 ORFs as plasmid clones from ABRC, and 253 ORFs had to be cloned from cDNA.

For isolation genes from mRNA, the tissue or combinations of tissues containing the highest concentration of the respective mRNA should be used. Therefore the concentrations of mRNA for the respective 253 ORFs were determined from the dataset *E-TABM-17 - Transcription profiling by array of organism parts from different strains of Arabidopsis* [141]. For 217 genes, the three tissues with highest expression were successfully determined; for the remaining genes, no expression data are available in this dataset (see also methods 5.2, table A.2).

Subsequent collection of plant material, and cloning of required ORFs from cDNA was conducted by Melina Altmann.

Overall 1,216 out of 1,227 ORFs could be derived as Y2H constructs. Compilation and cloning of ORFs was conducted by Melina Altmann from the group of Pascal Falter-Braun.

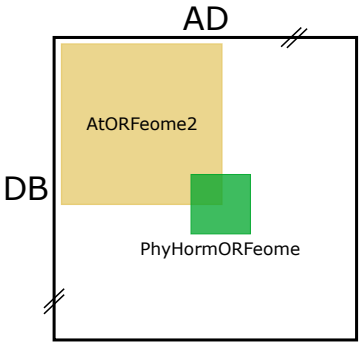


Figure 2.1: Search space defined by PhyHormORFeome and AtORFeome2. AtORFeome2 comprises of ORFs of around 12,000 loci out of 27k loci of *A. thalina*; PhyHormORFeome contains ORFs from about 1,200 loci related to phytohormone signaling. Loci were defined in the search space PHO (SSP_{PHO}). Both ORFeomes have an overlap of 688 loci.

2.1.2 Y2H interaction mapping

To identify physical binary interactions between proteins in the PhyHormORFeome collection, an established high-throughput Y2H interaction mapping pipeline was used. This pipeline was previously applied to identify protein-protein interactions in a collection of about 8000 proteins in the AtORFeome which yielded an interaction map consisting of 5664 interactions between 2661 proteins [11], protein-protein interactions between proteins in AtORFeome and a collection of ORFs from pathogens *Hyaloperonospora arabidospsidis* and *Pseudomonas syringae* [139], and protein-protein interactions between AtORFeome and proteins from oomycete *Golovinomyces orontii* [142].

This Y2H mapping pipeline has been extensively tested and compared to other methods in [143, 17, 12, 144, 145, 16]. Experimental steps in this pipeline were conducted by Melina Altmann.

This Y2H screening pipeline consists of four steps: Primary screen, secondary phenotyping, candidate interaction identification, and verification (see also 5.3). In the first step, pools consisting of 188 clones containing a vector with an ORF plus activation domain (AD) are tested for interaction with single clones containing a vector with an ORF and a DNA binding domain (DB). All growing yeast colonies are tested for autoactivating DB clones in the second phenotyping step to keep only yeast colonies with real interacting protein pairs. These two steps were repeated three times for a higher saturation of the number of detectable interactions. As the primary screen is conducted for time and cost reasons with pools of AD clones, interacting protein pairs have to be identified by sequencing of polymerase chain reaction (PCR) products of genes in AD and DB vectors of interacting proteins. For cost reasons, short unique DNA sequences (barcode sequences) were added to PCR product while PCR reaction, which allow pooling and subsequently a combined sequencing of 96 AD- and DB-ORFs with next generation sequencing (NGS) technique.

From results of barcode NGS the interacting protein pairs could be identified (see section 5.3). Unexpectedly, for each AD-DB pair, multiple loci for the respective AD and DB could be identified as candidate interaction partners for each positive colony. As all candidate interaction pairs are to be tested multiple times independently in the verification step, there is no need to identify a candidate interaction partner with high certainty as long as the interacting pair is contained in the list of pairs to be verified.

The question was now, how many of the identified loci per AD-DB pair should be used to generate candidate interactions. To determine this threshold, all combinations of up to three AD- and three DB-loci per tested yeast colony which had to a read frequency of at least 10 % of the reads were selected. These loci were used to generate all possible AD-DB combinations for the respective clone, which are referred as candidate interactions. These candidate interaction were tested against literature interactions derived from IntAct [146] and BioGRID [147] and a random set of 1000 protein pairs to determine, if loci identified with a low number of reads are likely to produce candidate pairs containing real interactions or not.

In all tested combinations, literature interactions were contained (fig. 2.2 A). Whereas in random protein combinations almost no literature interactions could be identified (fig. 2.2 B). This led to the conclusion that loci identified with a low number of reads are more likely to be interaction partner in a real interaction than random pairs. Therefore, all possible combinations up to the third most frequent locus with a frequency $\geq 10\%$ were used to generate candidate interactions pairs.

From the alignment results derived from Bowtie 2 using up to the third best hit with a minimum of 10 % aligned sequences, 833 unique candidate interactions from 314 loci were generated.

To connect expected hormone signaling pathways to the existing protein interaction maps AI-1_{MAIN} and AI-1_{REPEAT}, the clone collection PhyHormORFeome (AD pools) was additionally one time screened against the AtORFeome (DBs) clone collection. 2475 candidate interactions between 1408 proteins were identified after barcode NGS.

As both clone collections, AtORFeome and PhyHormORFeome, have 690 clones in common, an overlap of the candidate interactions of PvP screen and PvA screen is expected. We observed 167 shared candidate interactions between both screens.

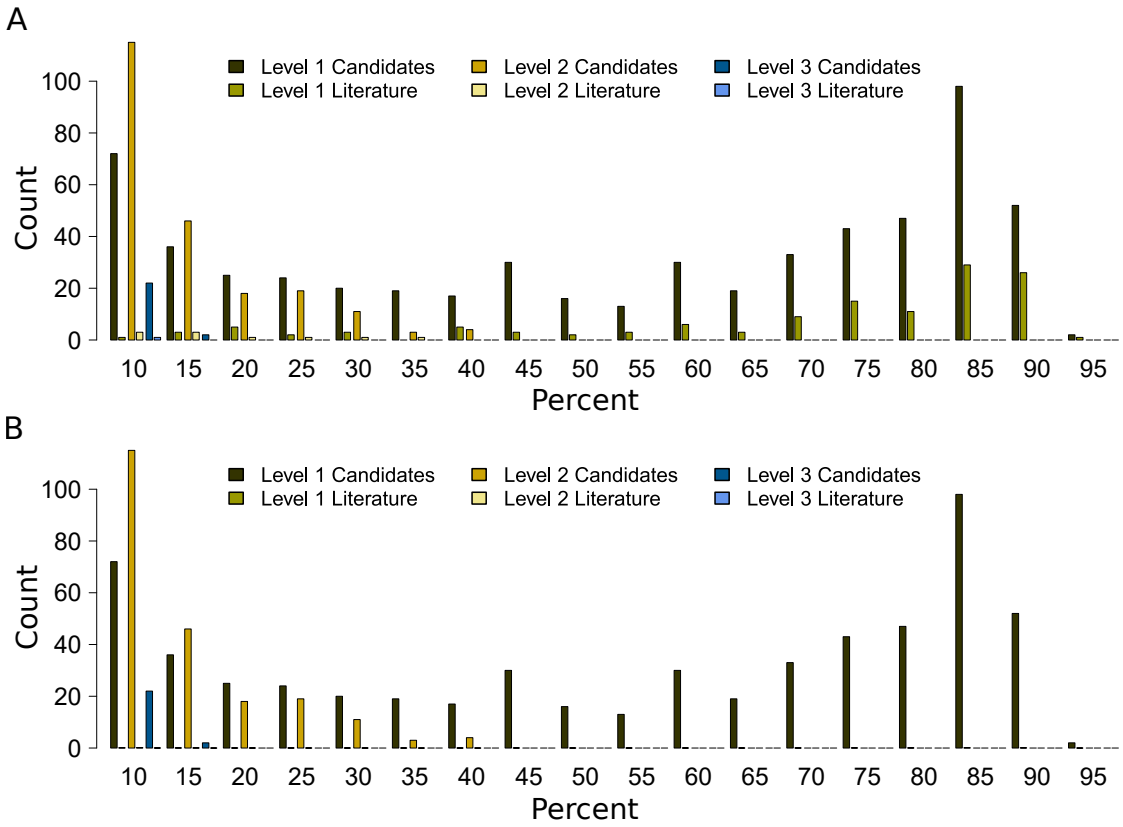


Figure 2.2: A) Candidate interactions contain literature interactions. Candidate interactions generated from the best hit ("Level 1") of NGS sequencing results contain literature interactions, even if they were identified only with a low number of sequences. But also candidate interactions, which contain one or two interaction partners from the second or third best hit of NGS results contain already known literature interactions. x-axis: percentage of aligned sequences, y-axis: number of interactions. B) Random candidate interactions drawn from search space SSP_{PHO}. In 1000 times randomly drawn interactions with the same number of interactions are almost no literature interactions contained.

Clones for all candidates interactions were subsequently newly picked from the ORFeomes and five times independently verified for interaction and autoactivation. Only interactions which showed three times a stronger growth on control plates compared to selective plates were treated as real interaction (see fig. 5.2, Step 4 Verification).

After 5-fold independent verification of candidate interactions for the screen of the PhyHormORFeome against the PhyHormORFeome (PvP), 475 interactions between 251 proteins were identified (fig. 2.3). This protein interaction map is referred as PhyhormInteractome (PhI). For the screen PhyHormORFeome against AtORFeome (PvA), 698 interactions between 581 proteins were assembled after 4 fold independent verification (fig. A.1). This protein interaction map is referred as $PhyhormInteractome_{out}$ (PhI_{out}). In PhI_{out}, 192 out of 581 proteins are in the search space SSP_{PHO} and interact with 389 proteins not in search space SSP_{PHO}. Both networks have 87 interactions in common. A combined network of both Y2H screens contains 1086 interactions between 696 proteins (fig. A.2), of which 299 proteins are part of the defined search space

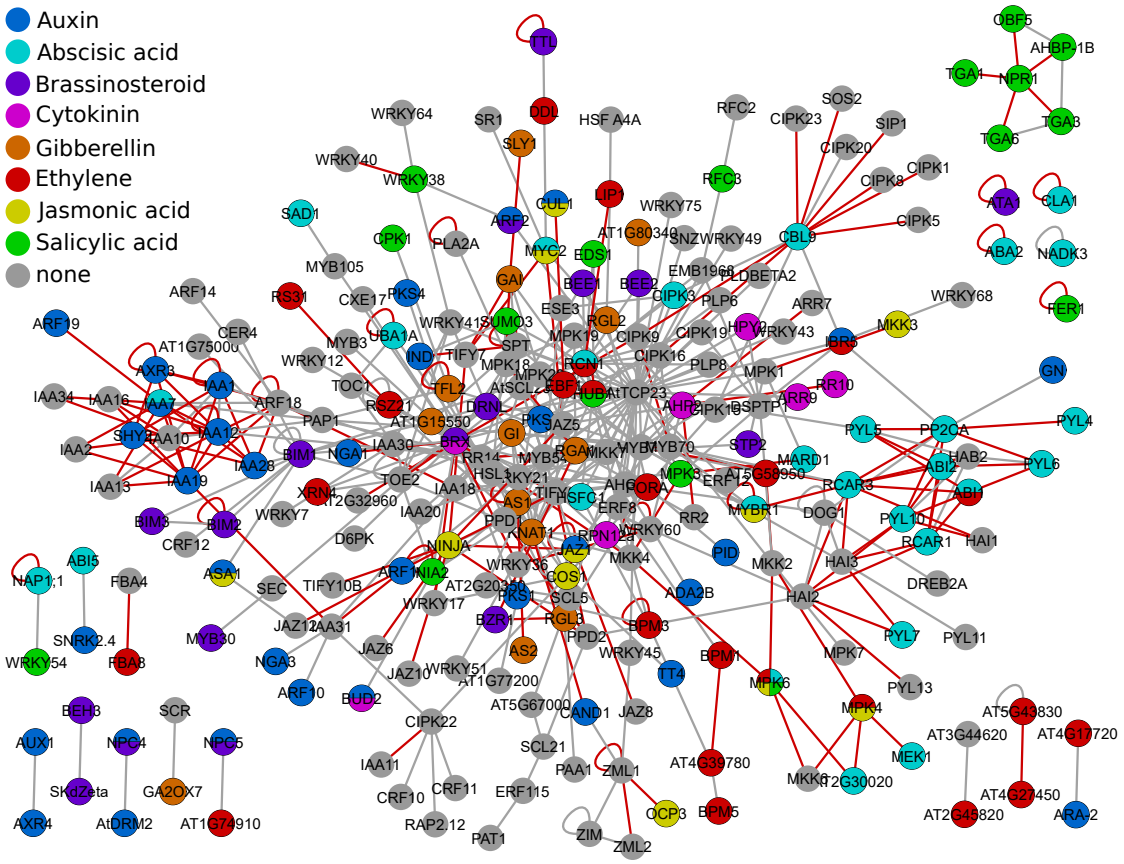


Figure 2.3: Network representation of PhI network resulting from PvP Y2H screen. Proteins are annotated with phytohormone annotations from AHD 2.0 restricted to annotations derived from genetic evidence. Red edges show recovered interactions curated in the IntAct or BioGRID databases. Network was visualized with Cytoscape [37].

SSP_{PHO} for PhI. The combined protein interaction map of PhI and PhI_{out} is referred as $PhyhormInteractome_{extended}$ (PhI_{ext}).

In PhI_{out}, 95 loci are contained, which are also contained in PhI. These 95 proteins interact with another 6 proteins in SSP_{PHO}, but not in PhI, and 281 proteins not in SSP_{PHO}. The 88 proteins, contained in search space SSP_{PHO} but not in PhI, show exclusively interactions with 163 proteins not in search space SSP_{PHO}. Only 18 out of these 163 proteins have a GO annotation related to phytohormone signaling pathways. Overall 377 proteins without any hormone annotation are contained in PhI_{ext}, of which 314 where not contained in our search space SSP_{PHO}.

The results of PhI_{out} show that (i) the overlap between the two ORFeomes, PhyHorm-ORFeome and AtORFeome, is reflected by the interactions: both interactions where both interaction partner are in SSP_{PHO} and interactions where only one interaction partner is SSP_{PHO} were found. (ii) Some proteins in SSP_{PHO} only interact with proteins not SSP_{PHO}, which could indicate that they are not involved in phytohormone signaling. (iii) Proteins involved in phytohormone signaling interact with a large number of proteins involved in various processes, which leads to the assumption that a lot of biological processes are influenced phytohormone signaling pathways.

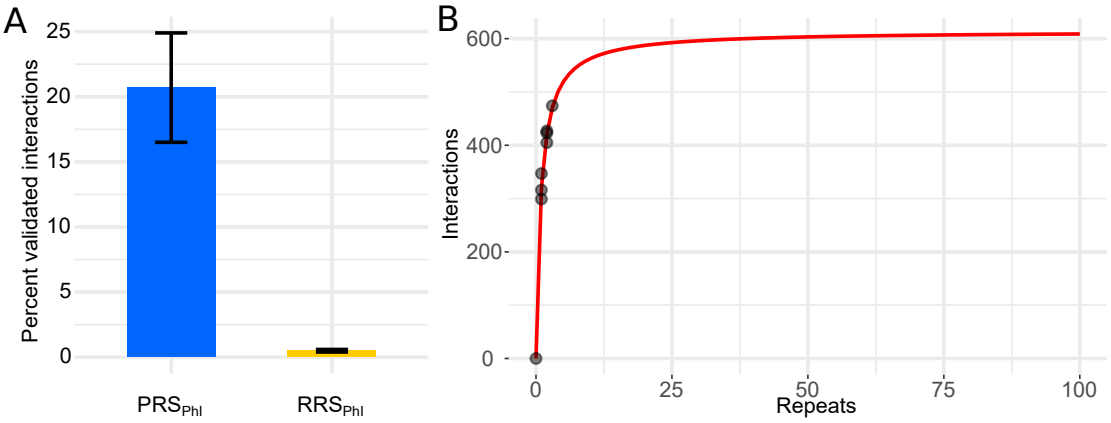


Figure 2.4: Y2H assay sensitivity and sampling sensitivity in PvP. A) Assay sensitivity in PvP Y2H screen. PRS: Positive Reference Set, RRS: Random Reference Set B) Sampling sensitivity in Y2H mapping pipeline. Michaelis-Menten curves is fitted through number of interaction found by number of repeats in three repeats of the PvP screen. Black dots indicate number of interactions in three repeats and combinations of three repeats, respectively.

2.1.3 Quality assessment

An important criteria for the reliability of experimentally determined protein-protein interactions derived from a high-throughput interaction mapping pipeline is the quality and coverage of the data produced by the mapping pipeline. To validate our Y2H interaction mapping pipeline in terms of completeness, assay sensitivity, sampling sensitivity, and precision, we used an empirical approach for quality assessment as described in [12].

The completeness of the screen can be calculated from the number of AD and DB clones in the PhyHormORFeome compared to the number of defined loci in the search space SSP_{PHO}. Overall 1201 ORFs could be cloned, of which 1181 were available in both vectors, 15 ORFs could be cloned in AD vectors only and 5 ORFs could be cloned in DB vectors only. Compared to the defined search space SSP_{PHO} of 1227 loci, a completeness of 94.2 % was obtained.

With each interaction mapping pipeline a fraction of all existing interactions can be identified. With each repeat a decreasing number of new interactions can be detected until a saturation of the maximum detectable interactions is reached. The fraction of detected interactions can be estimated by calculating the sampling sensitivity (see section 5.4.3). Therefore an adjusted Michaelis-Menten equation is used to fit a curve through the number of identified interactions by the number of repeats of the screen [12, 11]. The fitted Michaelis-Menten curve reaches its saturation at 614 ± 37 interactions. This is the estimated number of interactions, which can be detected with our Y2H pipeline using the described search space SSP_{PHO}. Compared to 475 interactions in PvP screen, we reached a sampling sensitivity of $77.6\,\% \pm 4.7\,\%$ (see fig. 2.4 B).

The assay sensitivity provides information about the number of true positive and false positive interactions identified by the mapping pipeline. Therefore the interaction mapping pipeline was tested against 92 interactions from a positive reference set (section 5.4.4), consisting of manually recurated *Arabidopsis* protein-protein interactions and 95 random interactions sampled from search space SSP_{PHO} as negative set. From the positive reference set, 19 interactions were successfully verified, of which 6 could be

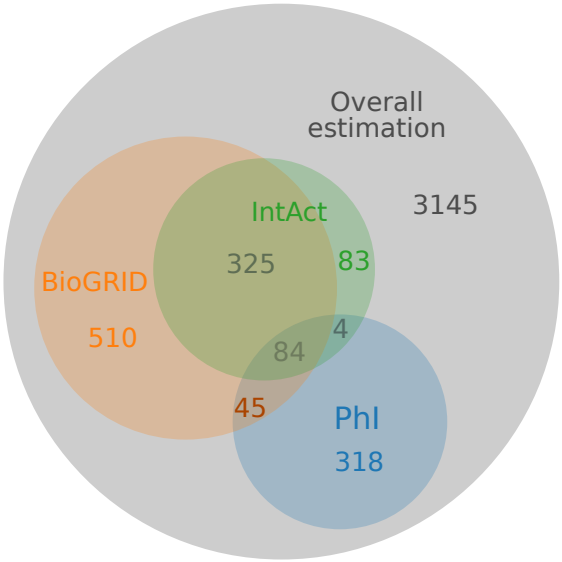


Figure 2.5: Number of interactions, which overlap between the different sets of binary interactions, in relation to the estimated number of interactions expected in the search space SSP_{PHO}. Not shown are interaction in network PhI_{out}.

verified in both directions. From the random reference set, 0 interactions could be verified. This means, we reached an assay sensitivity of 20.7 % ± 4.2 % and an assay specificity of 100 %. The overall sensitivity, which is a combination of assay sensitivity and sampling sensitivity, reaches 16.0 % ± 6.3 %, which is comparable to the overall sensitivity observed in the AI-1_{MAIN} Y2H screen [11]. Including the completeness, the overall completion [16] is 15.1 % ± 6.3 %. This means that around 15.1 % of all binary interactions in the search space SSP_{PHO} were found. This overlap of around 15 % is therefore also expected for a randomly selected subset of all possible binary interactions in the search space. The observation that PhI has an overlap of 17.9 % and 13.4 % with the network map derived from binary interactions of IntAct and BioGRID restricted to search space SSP_{PHO} (fig. 2.5) fits therefore to the calculated overall completion, respectively.

On the basis of the overall completion of 15.1 % ± 6.3 % in PhI, which corresponds to 475 interaction, the total number of binary interactions in the search space SSP_{PHO} can be estimated to about 3145 interactions. Integrating all binary interactions from PhI and interactions in search space from PhI_{ext}, BioGRID, and IntAct, 1605 interactions are present. This means a coverage of about 51.7 % of the search space SSP_{PHO} (fig. 2.5).

2.2 Network topology

Most biological networks, including protein-protein interactions networks, have the same properties resulting from a growth process. They share a scale-free topology with hierarchical modularity, presence of hubs, disassortativity, topological robustness, and groups of highly interconnected nodes, which often carry out specific cellular functions. These common properties are likely to be the result of gene duplication events [7].

To analyze the network topology of PhI, I calculated degree distribution and clustering coefficient and compared it to the respective distributions of the systematic protein interaction networks AI-1_{MAIN} [11] and two literature curated networks from IntAct [146] (fig. 2.7 A) and BioGRID [147] (fig. 2.7 B) restricted to our search space $\mathrm{SSP}_{\mathrm{PHO}}$ and binary interactions. The distributions of these two measures are used to distinguish random network structure, scale free network structure and hierarchical network structure.

We hypothesized that a phytohormone related protein-protein interaction map has the properties observed for other biological networks, which means that both the degree distribution and clustering coefficient follow the power law distribution.

The calculation of the degree distribution and clustering coefficient showed the expected results for our systematic phytohormone network and also AI-1_{MAIN}. Both, degree distribution and clustering coefficient, followed a power law distribution, which indicates a hierarchical topology [7]. In both literature derived interaction maps, degree distribution followed a power law distribution. The clustering coefficient $C(k)$ in the network derived from BioGRID was independent of k and showed a uniform distribution and in IntAct $C(k)$ increased with increasing k (fig. 2.6).

The results showed that PhI and AI-1_{MAIN} have a hierarchical network structure, which implies the presence of modules. The network derived from BioGRID showed for both measures a distribution, which indicates a scale free network structure. The network derived from IntAct did not fit to any of the three network topologies; it had an unusual high connectivity of the neighborhood or nodes with a high degree.

The results let conclude that PhI network has the topology, which is expected for a biological network, whereas both LCI networks did not.

The question was how the different network topologies for interactions in the same search space can be explained. The observed network topologies of the LCI networks could be a result from experimental biases, where hypothesis driven small or meso scale experiments lead to a dense connectivity between intensely studied proteins. This could explain the high number of interactions observed between proteins involved in the same hormone signaling pathway (fig. 2.7 A,B). Another reason could be a bias in curation of interactions from literature, where interactions had for any reason not been included in the databases. Our systematic experimental approach, on the other hand, is able to detect interactions between protein pairs, which were so far not in the focus for a possible functional relationship and would not be tested in a hypothesis driven experiment. This lead to a bias free connectivity between proteins.

2.3 Communities in networks

Proteins, which are involved in the same biological process, protein complexes, signaling cascades, or transcriptional regulatory circuits often show strong connectivity and form modular structures in networks [7, 148]. These functional modules are often referred to as communities. In a network, a community is characterized by the higher number of interactions between nodes, which belong to the same community, than to the rest of the network [25]. That strongly interconnected proteins are often part of the same biological process has been shown e.g. in AI-1_{MAIN} [11]. As the hierarchical network topology of PhI implies the existence of communities and proteins involved in the same process often form communities, the assumption was that the PhI network contains functional communities for single phytohormone signaling pathways.

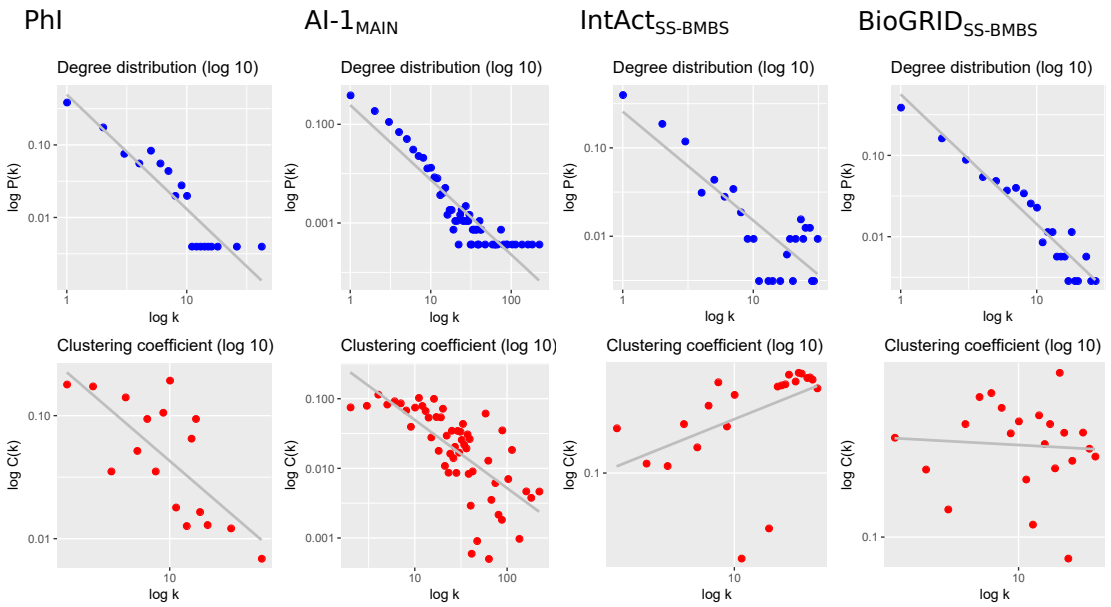


Figure 2.6: Degree distribution (upper row, blue dots) and clustering coefficient distribution (lower row, red dots) of four networks: PhI, AI-1_{MAIN}, literature curated interactions in IntAct and BioGRID restricted to binary interactions in search space SSP_{PHO}, respectively.

For identification of communities, the edge betweenness algorithm described by Girvan and Newman in [25] has been used. This algorithm iteratively removes the edge with the highest betweenness, i.e. the edge which is used from most shortest paths in the network, from a network and recalculates the edge betweenness until no edges remain. For detection of communities, the implementation of this algorithm in the R package igraph [149] was used. The outcome was compared against other community detection algorithms which showed similar results (see appendix table 5.2). The identified community structure of PhI was compared against the community structure of the three networks AI-1_{MAIN}, literature curated binary interactions restricted to search space SSP_{PHO} contained in $\text{IntAct}_{\text{SS-BMBS}}$ and $\text{BioGRID}_{\text{SS-BMBS}}$.

To test, if the identified communities are functionally relevant in phytohormone signaling, communities in all four networks were tested for enrichment with the annotation for a specific phytohormone signaling pathway using Fisher's exact test. Using hormone annotations based on genetic evidence from AHD 2.0 [140], in PhI seven communities are enriched in functions in one of these seven phytohormones: ABA, IAA, BR, CK, ET, JA, SA (fig. 2.8). To test significance of the number of enriched communities, I tested the enrichment of communities in 1000 degree preserved randomized networks. In PhI only up to three enriched communities were contained in randomized networks (Figure 2.8 B). This shows the non-random existence of phytohormone related functional modules.

Literature curated interaction maps and AI-1_{MAIN} were also analyzed for communities and enrichment of hormone annotations within communities. In AI-1_{MAIN} (A.4) 5 enriched communities were identified in the complete network and 3 in the largest connected subgraph. In both LCI networks, $\text{IntAct}_{\text{SS-BMBS}}$ and $\text{BioGRID}_{\text{SS-BMBS}}$, 13 and 10 enriched communities (fig. 2.9 A, A.3 A) were identified, respectively. In the largest connected subgraph of $\text{IntAct}_{\text{SS-BMBS}}$, eight hormone enriched communities for

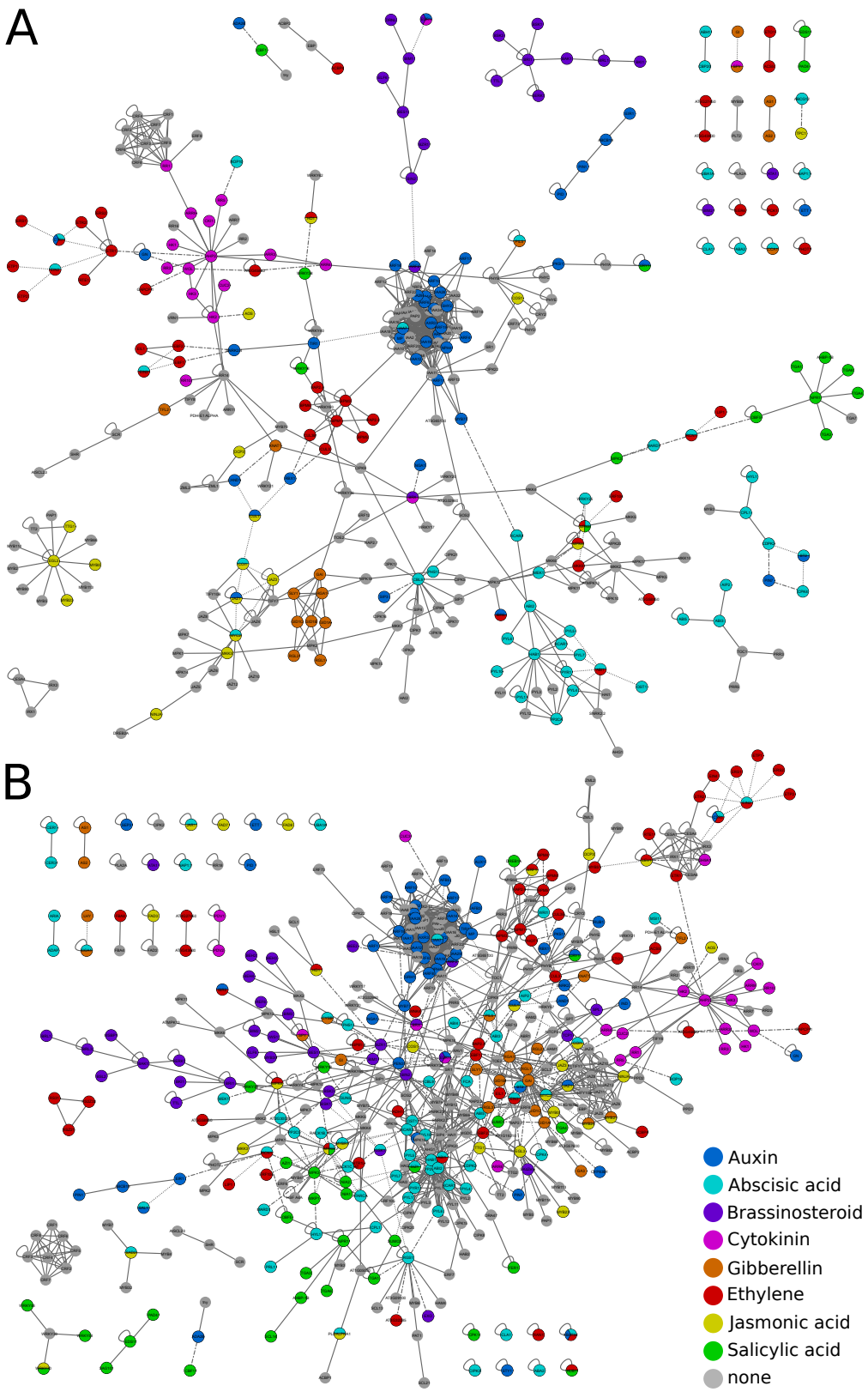


Figure 2.7: Network representation of binary protein-protein interactions curated from literature in A) IntAct database and B) BioGRID database contained in search space SSP_{PHO} and annotated with phytohormone annotations from AHD 2.0.

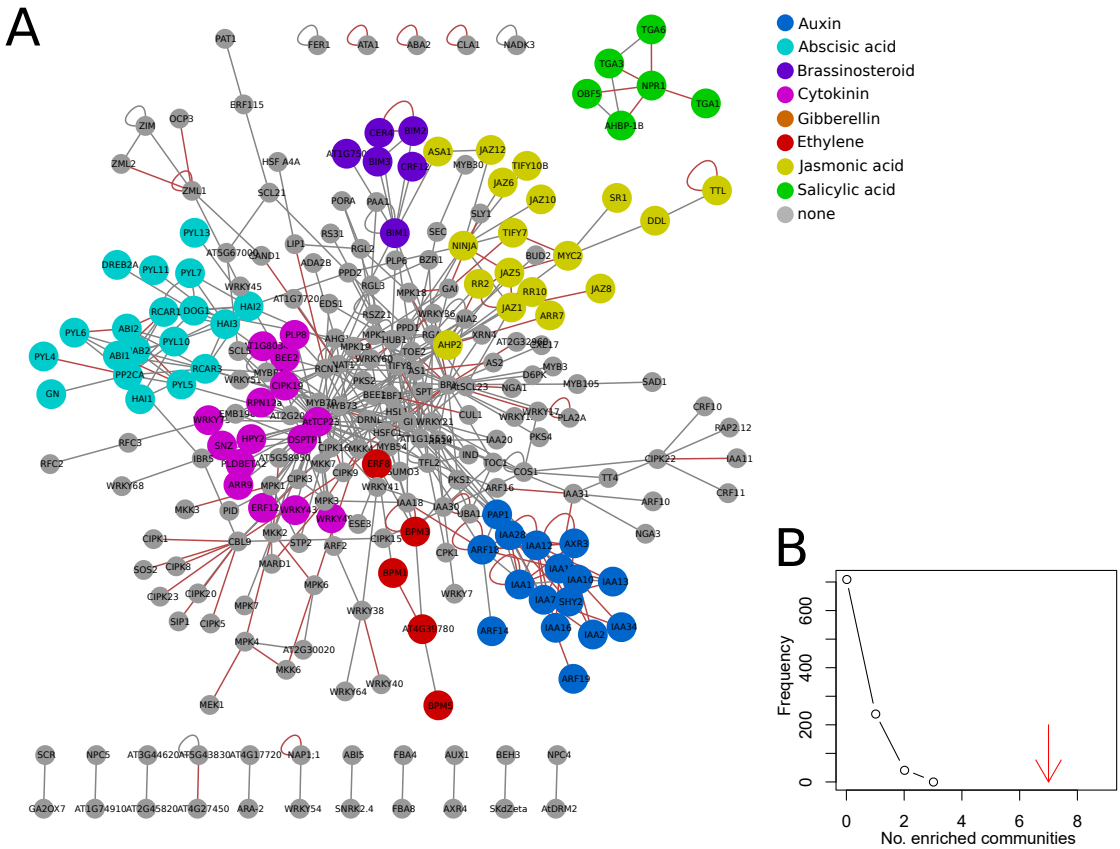


Figure 2.8: A) PhI network colored by hormone enriched communities. ET: community number 12, IAA: 4, BR: 8, CK: 16, JA: 7, ABA: 3, SA: 15. B) Number of enriched communities in PhI using AHD genetic annotations is marked with a red arrow. Distribution shows number of enriched communities in 1000 randomized networks.

seven hormone signaling pathways were identified, whereby two separate communities were enriched with ET. Five small subgraphs, which are also single communities, were enriched with five different hormone signaling pathways. One of them was enriched for JA, which was not observed in the largest connected subgraph (fig. 2.9) and four communities enriched in IAA, ABA, BR, and SA. In the largest connected subgraph of $\text{BioGRID}_{\text{SS-BMBS}}$, one enriched community for each out of the eight tested phytohormones was contained. Additionally two enriched communities for ET and SA in the smaller subgraphs were found.

The community detection showed that the functional modules implied by the hierarchical structure could be identified and assigned to a biological function. Moreover, the significant number of hormone enriched communities showed that topological substructures of the network are corresponding to biological functions and do not occur by accident. The low number of phytohormone enriched communities identified in $\text{AI-1}_{\text{MAIN}}$ is probably the result of the low number of hormone related protein contained in this network. The search space used for $\text{AI-1}_{\text{MAIN}}$ contained only half of the defined search space SSP_{PHO} of PhyHormORFeome. Therefore only a low number of interactions in search space SS_{PHO} is contained in $\text{AI-1}_{\text{MAIN}}$ (106 interactions). The community detection and enrichment analysis in LCI networks showed that in both networks the identified

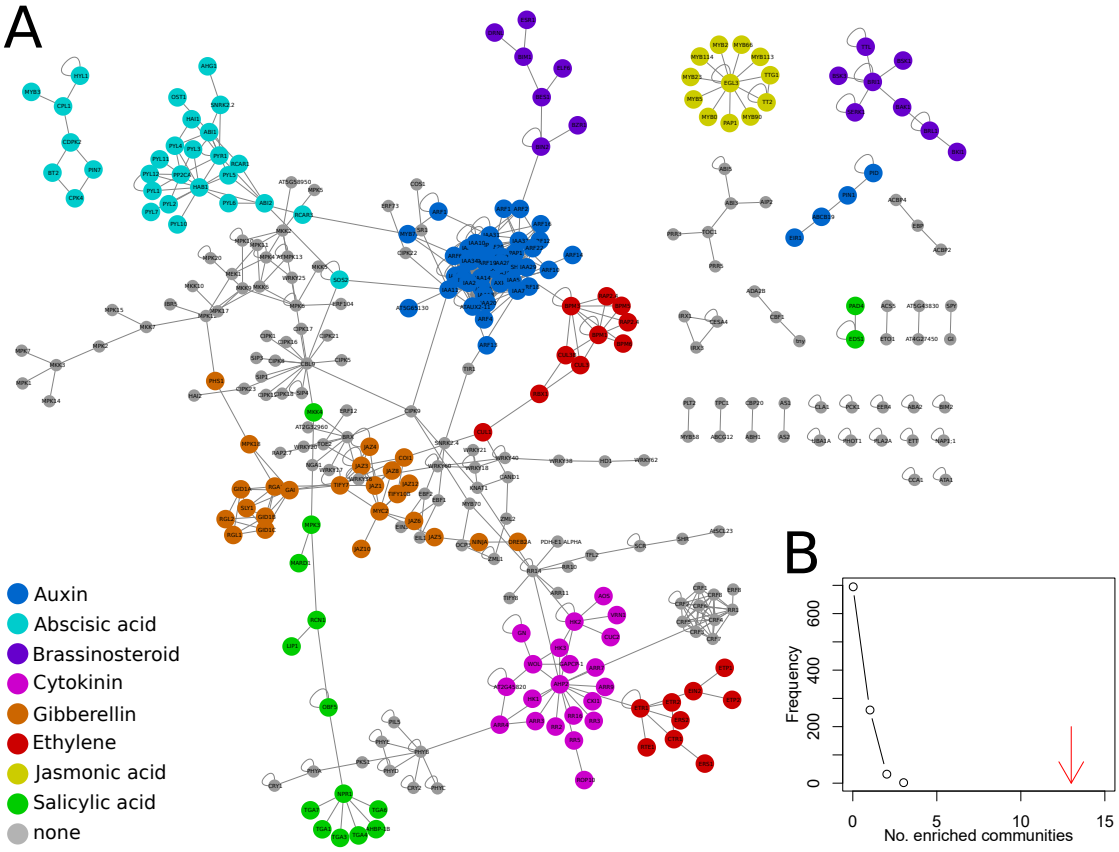


Figure 2.9: A) IntAct network colored by hormone enriched communities. B) Number of enriched communities in IntAct using AHD genetic annotations is marked with a red arrow. Distribution shows number of enriched communities in 1000 randomized networks.

communities from topological properties also correspond to biological function. A number of communities were enriched with proteins related to a phytohormone signaling pathway. In contrast to PhI, for some phytohormone signaling pathways not only one but multiple communities were enriched. This could either imply missing interactions due to hypothesis driven experiments [17] or different biological processes regulated by the same phytohormone are regulated by separated unconnected pathways.

The results from all analyzed networks let conclude, that network topology reflects modularity of biological functions. The significant higher number of enriched communities in real networks compared to degree preserved randomized networks showed that correlation between functional and topological modularity is a property of protein-protein interaction networks.

Community network

The network structure and community detection algorithm are crucial for identification of communities. The number and size of communities and the connections between communities depend on these two parameters. The community detection algorithm has only a low influence on the detected communities for the above examined networks (see appendix table 5.2). Different network structures are therefore reflected by different

community structures. These community structures can be represented as a community network, where a node represents all proteins, which belong to the same community and edges represent interactions between proteins belonging to two different communities [150]. This simplified, coarse-grained representation is a higher level of abstraction of the underlying PPIs. The obtained community network can be used for analyses and comparisons to gain insights into PPI networks.

For the analysis of community networks, the largest connected subgraph of a network was used to assemble the respective community network. The community networks were generated from the results obtained from the edge betweenness algorithm. The comparison of the community networks of the two systematic networks, PhI and AI-1_{MAIN} with both literature curated interaction maps reveal differences in the community network structure and community network properties. In the largest connected subgraph of PhI (218 nodes), 23 communities were identified, where each community consists on average of 9.5 nodes. The mean degree of a community is 6.1 for 23 communities and the clustering coefficient is 0.41. In the largest connected subgraph of the systematic network AI-1_{MAIN} a similar mean degree of 6.8 for 98 communities and a lower clustering coefficient of 0.19 is observed. For both binary literature curated interaction maps restricted to search space $\mathrm{SSP}_{\mathrm{PHO}}$ a lower mean degree for communities was observed: Communities from $\mathrm{IntAct}_{\text{SS-BMBS}}$ have a mean degree of 2.3 and 3.3 from $\mathrm{BioGRID}_{\text{SS-BMBS}}$. Clustering coefficient of $\mathrm{BioGRID}_{\text{SS-BMBS}}$ (12 communities in 294 nodes) reaches 0.38, whereas clustering coefficient of $\mathrm{IntAct}_{\text{SS-BMBS}}$ (18 communities in 253 nodes) is only 0.06 (fig. 2.10 E, F). The mean degree of a community in the systematic Y2H community networks is around two-fold higher than in the LCI community networks. This shows the stronger connectivity between communities, which fits to the hierarchical network structure of the systematic networks.

All community networks have a negative assortativity, which means that all networks are disassortative. In these networks, a large number of low connected nodes have an interaction with nodes having a high degree. This is a property of biological networks like protein-protein interaction maps [151]. This property is more pronounced in both systematic Y2H derived networks (fig. 2.10 G).

The clustering coefficient of a network represents how strong a network is connected. A low value near zero means a low connectivity between nodes; a value of 1 means every node is connected with each other. Both, PhI and BioGRID, show a relative strong connectivity between communities compared to IntAct and AI-1_{MAIN}. The difference between these two networks lies in the number of interactions between communities, which is derived from the number of protein-protein interactions between communities: PhI has most times more than one interaction between two communities, whereas BioGRID has only one interaction between two communities. AI-1_{MAIN} has a weak connectivity mainly due to a high number of communities which interact with only one community. IntAct has the lowest clustering coefficient, which shows the weak connectivity between communities (see width of edges in fig. 2.10 A-D).

PhI and AI-1_{MAIN} show a community network (fig. 2.10 A, B), which is expected for biological networks [7], whereas both literature curated networks show network properties and a network structure, which are uncommon for protein-protein interaction networks (fig. 2.10 C, D). The community network of PhI illustrates the high number of modules and their strong interconnectivity compared to the low modularity in LCI community networks. This leads to the conclusions that PPI networks are more modular

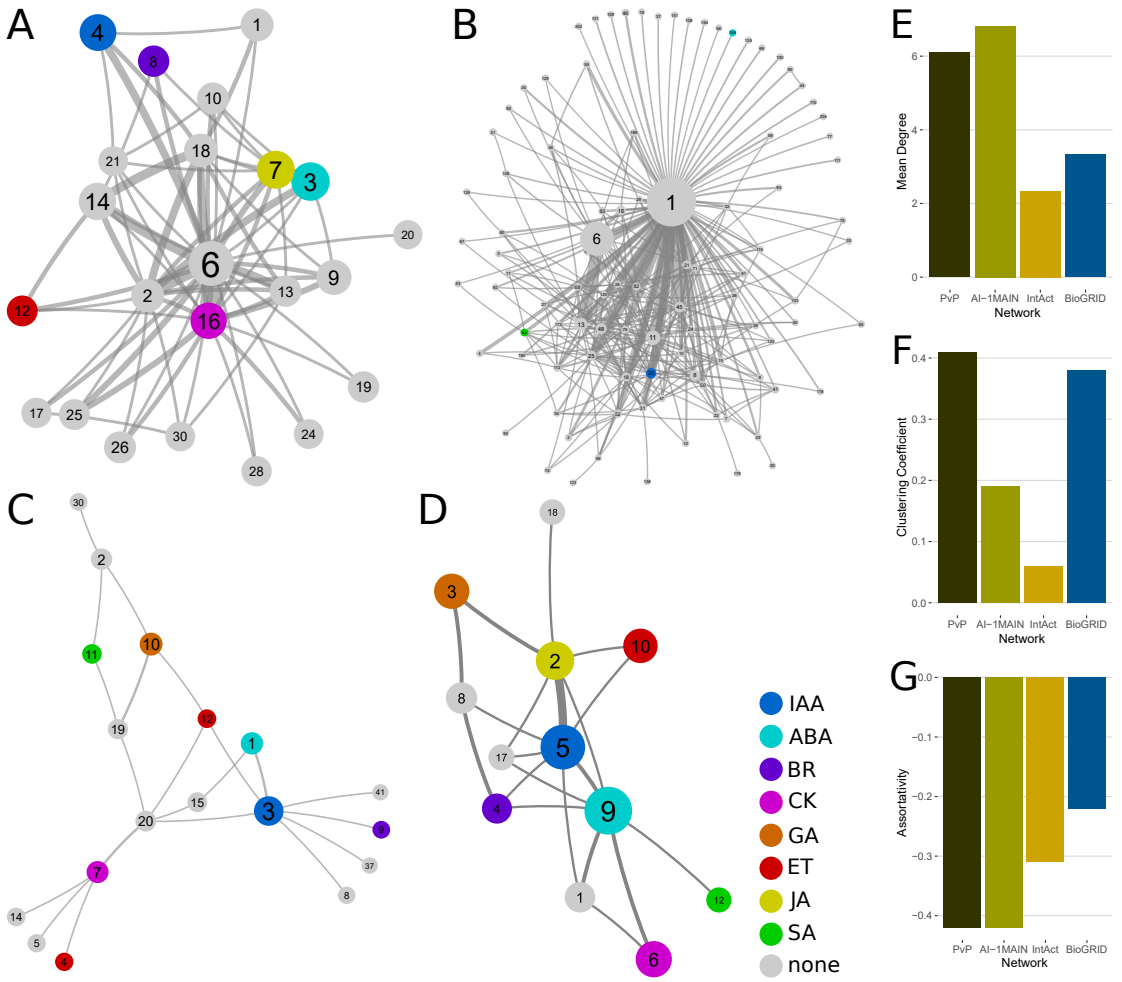


Figure 2.10: A-D) Networks represent interactions between communities. Nodes represent a community. Size of the node is relative to the number of proteins in this community. Edges are interactions between proteins in two different communities. Edge width is relative to the number of interactions between two communities. Number in the node is number of the community. A) PhI B) AI-1_{MAIN} C) $IntAct_{SS\text{-}BMBS}$ D) $BioGRID_{SS\text{-}BMBS}$ E-G) Network properties of community interaction networks. E) Mean degree of communities. F) Clustering coefficient of communities. G) Assortativity of communities.

and biological processes influence each other more than expected from analyses of LCI networks.

2.4 GO enrichment of communities

The community enrichment analysis based on AHD 2.0 annotations revealed that a number of communities in PhI are enriched with a hormone annotation. This showed that topological properties of networks reflect biological functions. To elucidate biological functions of other communities, a suitable method is to additionally integrate GO annotations into the network. Proteins in the same community in PhI can be analyzed for common biological process annotations. This helps to interpret the common function of proteins with strong connectivity and gives rise to functionality of proteins with no

or low annotation depth in GO. Additionally the enrichment found with annotations from AHD 2.0 can be confirmed and verified with gene ontology annotations, which are independently curated from GO annotations. Therefore all communities in PhI were tested for enriched GO terms and analyzed for their function in biological processes (see methods in section 5.8).

In the largest connected subgraph of PhI, 24 communities were identified and also the second largest subgraph consisting of 6 proteins forms one community. These communities were tested for GO enrichment; communities in smaller unconnected subgraphs were not tested for enrichment. Only 4 out of these 25 communities showed no enrichment in any GO term, whereas 21 were enriched in one or more GO annotations. All hormone enriched communities from previous analysis except the BR-community (community 8) also showed an enrichment in one or more GO annotations related to the respective hormone. In the following paragraph a selection of GO term enriched communities is briefly described for their respective biological functions.

A detailed analysis of community 8 shows a GO enrichment of *gametophyte development*. It has been shown that BRs are regulating anther and pollen development in *Arabidopsis* [152], which supports the BR annotation of this community identified in the previous analysis. Community 3, which is enriched for ABA using AHD annotations, has an enrichment in GO terms related to response to heat, water deprivation and salinity, and stomatal movement. The influence of ABA signaling on response to these environmental stresses has been extensively analyzed and reviewed in [153, 154]. The IAA community 4 has an enrichment for the GO term *response to auxin*, but also for root related GO terms. The role of AUX in root growth and lateral root formation has been well examined [155]. The JA community 7 is one of largest identified communities consisting of 18 nodes. Three JA related GO terms are enriched: *jasmonic acid biosynthetic process*, *jasmonic acid mediated signaling pathway*, and *response to jasmonic acid*. Besides the JA related terms, several biological processes like *response to wounding*, *response to fungus*, and *regulation of growth* are enriched, where JA is known to be involved [156]. The SA community 15 consists of six proteins which are connected via the SA binding protein NPR1 [157]. It has also two GO terms enriched, which are related to SA and several GO terms related to defense and immunity. SA is known to be involved in systemic acquired resistance (SAR) upon infection with pathogens [158]. The CK community 16 consists of 15 proteins, which are mainly connected by TCP23. GO enrichment is dominated by various GO terms related to metabolic processes. Beneath *response to cytokinin*, cell growth and meristem development is enriched, which can be also found in literature related to CK [159]. The largest community, number 6, is central of the community network and has a connection to every other community. This community is not enriched in any hormone annotation, but contains proteins from each phytohormone signaling pathway except CK. The most frequent hormone is GA with five occurrences. The most enriched GO terms are *cotyledon development*, *embryonic pattern specification*, and *meristem initiation*. Two communities (25 and 26) were identified, which are enriched in MAP Kinase cascades and contain mainly MAP kinases or MAPK kinases. Both communities have no direct interaction, but both interact with the same three communities: 2, 6, and 16. Both MAP Kinase communities have an enrichment of common GO terms, e.g. *stomatal complex development*, *plant ovule development*, *negative regulation of programmed cell death*, and *carpel development*. Both communities have general defense and stress response in common, but to different specific stresses. Community 25 is enriched in osmotic and oxidative stress; community 26 is enriched in

response to cold and *defense response to fungus and bacterium*. MAP kinases in distinct communities could therefore be in charge for response to distinct stresses. The enriched GO terms of all other communities can be found in appendix in tables A.12 - A.32.

Biological functions arise from interactions among components (proteins, DNA, RNA, and small molecules) in the cell. These components form functional modules to carry out specific cellular functions. Therefore, to understand biological phenomena and diseases, it is crucial to understand the functional modules [160, 161, 162]. Both, hormone enrichment analysis and GO enrichment analysis of topological communities revealed that topological communities correspond to hormone signaling pathways and biological processes. Moreover, GO enrichment analysis confirmed hormone enriched communities, which were identified using AHD annotations. The combination of hormone annotations and GO annotations allowed to infer which biological processes are regulated by which hormone signaling pathways.

2.5 Hormone pathway interconnectedness

Community analysis of PhI revealed a high connectivity between communities compared to networks derived from literature curated interactions. In contrast, literature curated interaction networks showed much lower connectivity between communities. Network structure and connectivity of communities in PhI gave rise to the hypothesis that all phytohormone signaling pathways are strongly interconnected, which was unexpected from analysis of LCI networks. Therefore the capabilities of signal transduction between phytohormone signaling pathways should be examined. To systematically answer the strength of interconnectedness of phytohormone signaling pathways, the number direct contacts and the distance between two pathways should be systematically determined.

To measure the distance within and between phytohormone signaling pathways, the mean shortest path length between all protein pairs whose annotation belong to a particular phytohormone combination was determined, whereby only shortest paths were taken into account, which connect to pathway A and B without crossing additional proteins involved in pathway A or B (see methods section 5.9). Also the number of direct interactions between proteins from each phytohormone combination and the number of proteins annotated with a certain phytohormone was examined. These three properties were determined for all four previously described networks (PhI, AI-1_{MAIN}, $\mathrm{IntAct}_{\mathrm{SS\text{-}BMBS}}$, and $\mathrm{BioGRID}_{\mathrm{SS\text{-}BMBS}}$) using hormone annotations derived from AHD 2.0.

A comparison of the number of proteins in the network annotated with a certain hormone showed that all four networks have a comparable number of proteins annotated for the respective phytohormone signaling pathway. A similar number of proteins are annotated with IAA, ABA, and ET. A lower number of proteins are annotated with the remaining hormones (Figure 2.11 A-D, upper panel).

The networks differ strongly in the number of direct protein-protein interactions within and between the phytohormone signaling pathways. PhI has at most 40 (ABA) and on average 17 direct interactions within a signaling pathway. Between hormone signaling pathways, the number of direct interactions are on average 3.6 in a range of 0 (GA-JA, SA-JA) to 15 (ABA-ET) (Figure 2.11 A, lower panel, size of the circles).

In Arabidopsis Interactome 1 Main Screen (AI-1_{MAIN}), which contains only a low number of hormone related proteins (about 50 % of PhyHormORFeome search space $\mathrm{SSP}_{\mathrm{PHO}}$), a very low number of direct interactions within and between hormone signaling pathways was found: a maximum of 12 and on average 2.8 interactions within a signaling

pathway. Direct interactions between hormone signaling pathways were found for 7 out of 28 combinations with a maximum of four interactions and on average 0.5 interactions (Figure 2.11 B, lower panel, size of the circles).

Literature curated interaction maps showed a much higher number of interactions within hormone signaling pathways compared to PhI. In IntAct on average 52 interactions in a range from 14 (SA) to 168 (IAA) interactions are found and in BioGRID on average 46.5 interactions in a range from 14 (GA) to 166 (IAA). Between hormone signaling pathways, in IntAct on average 3.2 interactions exist, which is less compared to PhI. For 10 combinations no interactions are found, but for ABA-IAA 22 and ABA-ET and 19 interactions exist, respectively. In BioGRID only for two hormone combinations (ET-GA, CK-SA) no direct interactions exist. On average 6 interactions between different hormone signaling pathway are present (Figure 2.11 C, D, lower panel, size of the circles).

Even if there is no direct interaction between two proteins belonging to different phytohormone signaling pathways, a signal can be transmitted by proteins not yet annotated with the respective hormone. To take this into account the direct shortest path length between two proteins with the respective hormone annotations was determined. Only shortest paths were taken into account, which do not contain a protein annotated with the same phytohormone as the two proteins which shortest path length was determined. This avoids biases due to expanded protein families and single distant proteins (see methods section 5.9).

In PhI, the length of shortest paths is between 2.2 and 4.2 (see distribution inside fig. 2.11 A, lower panel) with a mean shortest path length of 3.5. The shortest path lengths were observed for proteins within the CK and within the GA pathway. The third shortest path length was found between the CK and GA annotated proteins. This could be a hint for strong interconnectedness of these two signaling pathways.

In the second systematic network, AI-1_{MAIN}, also a narrow distribution of the mean shortest path lengths between 4 and 5.8 is observed similar to the distribution in PhI. The shortest path was observed for protein interactions within the GA signaling pathway.

The broadest distribution of the shortest path lengths was observed in the LCI network derived from IntAct interactions. The shortest path lengths were in range from 1.5 to 10.2 (see fig. 2.11 C, lower panel) with a mean of 6.4. Within the hormone signaling pathways of IAA, BR, CK, and GA the shortest path lengths are found.

In the BioGRID network there was a stronger interconnectedness between phytohormone signaling pathways than in the IntAct network, but lower than in PhI. The shortest paths are between 2.9 and 6.6 with a mean of 5.2. Strongest interconnectedness is found within IAA, BR, CK, JA, and ABA (see fig. 2.11 D, lower panel).

The hierarchical network structure implies a strong interconnection between functional modules of a network [7]. This strong interconnectedness was reflected by the narrow distribution of the mean shortest path lengths between the hormone signaling pathways. The narrow distribution, also observed in AI-1_{MAIN}, was in contrast to the wide distribution observed in LCI networks. The strong differences of distances between different signaling pathways and within pathways may be due to hypothesis driven biases inherent in LCI networks [17].

The high number of direct interactions and the short average length of the shortest paths to the next annotated protein shows the strong connectivity between different hormone signaling pathways. This leads to the conclusion that there are no separated

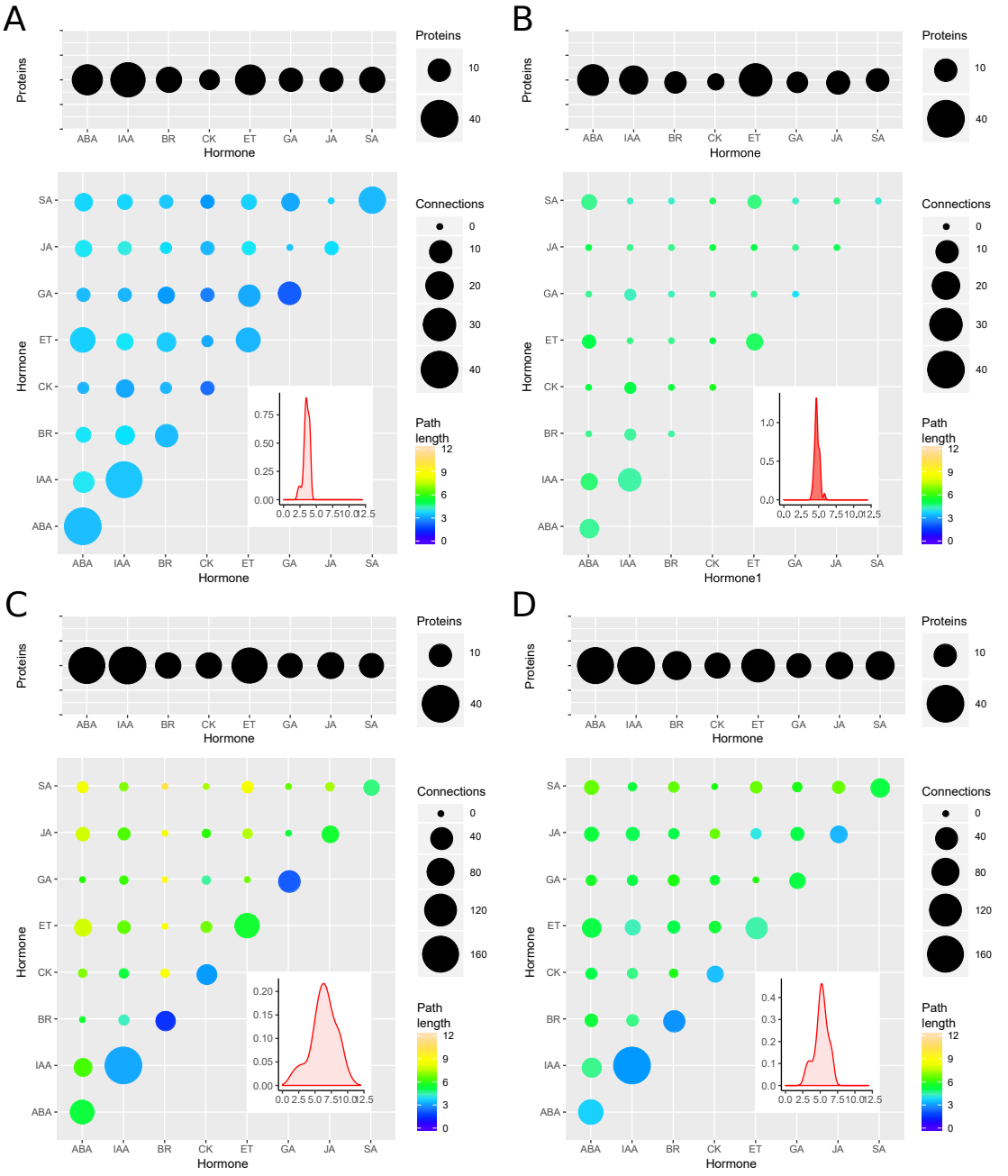


Figure 2.11: Hormone interconnectedness in protein interaction network maps of A) PhI, B) AI-1_{MAIN}, C) $\text{IntAct}_{\text{SS-BMBS}}$, D) $\text{BioGRID}_{\text{SS-BMBS}}$ using hormone annotations of AHD 2.0 with genetic evidence. For each network map, two panels are shown: The upper panel contains the number of proteins annotated with the respective hormone. The lower panel shows the number of connections between hormones and the average shortest path length. The embedded figure in the lower panel shows distribution of shortest path lengths.

signaling pathways per phytohormone with linear connections, but rather a dense signaling network.

2.6 Hormone annotation inference

So far more than 27,000 genes and in most cases several additional gene models per gene are annotated in the *Arabidopsis thaliana* genome. Many functions could be elucidated for a lot of the genes in numerous experiments, but the annotations are still incomplete. Predictions can help to generate hypothesis about possible functions of gene products, which can be tested in experiments. The basis for these predictions is that gene products with related functions tend to share properties such as genetic or physical interactions. This property was e.g. exploited to construct a network from genetic interactions between genes, which reveals relations between biological processes and the inherent functional organization of the cell in yeast [163]. This principle is known as *guilt by association* [33].

The aim of this analysis is to infer hypotheses from network topology about possible phytohormone signaling pathways in which proteins are involved. The shared property used to infer hypotheses are shared interactions with other proteins, the so called interaction profile. From the interaction profiles of a all nodes in a network an association network is calculated [33]. In this network, two proteins are connected, if their interaction profile has a similarity above a certain threshold. This interaction similarity can be used to transfer annotations between connected proteins [33].

Similarity of interactions between all pairs of nodes in PhI was calculated using the geometric index, which corresponds to the product of the proportion of shared nodes [33]. A clustered matrix with all pairwise similarity values is shown in figure A.5.

To generate an association network from similarity values, the similarity value at 1 % of most similar interactions was selected as cutoff. The derived similarity cutoff value was 0.2, which was also used in previous experiments as threshold [163, 164]. This threshold yielded 683 association edges between 191 proteins (see fig. 2.12).

For phytohormone annotation of proteins in the association network, annotations from AHD 2.0 and GO annotations (without annotations derived from evidence code IEA) were used. Out of 191 proteins in the association network, 62 have no hormone annotation, neither in AHD 2.0 nor in GO. The interactions in the association network can be categorized in five types: The association network contains 98 interactions where both proteins have the same hormone annotation (type I) and 123 interactions between proteins with no annotation (type II). Between proteins with overlapping annotations (type III), 47 interactions are contained in this network. This means that both proteins have one or more hormone annotations in common, but also one or more hormone annotations are unique in one or both proteins. Between protein pairs with and without hormone annotation, 298 similar interaction profiles are found (type IV). For proteins with strictly different hormone annotations (type V), 117 interactions are contained in the network.

Interactions of type III and type V are predictions, which can be used to experimentally determine, if one or both proteins are involved in additional phytohormone signaling pathways of the respective interacting protein. Interactions of type IV can be used to determine, if the non-annotated protein is involved in the phytohormone signaling pathway(s) of the interacting protein. With these candidates the respective hormone signaling pathways can be extended to additional proteins. It is important to keep in

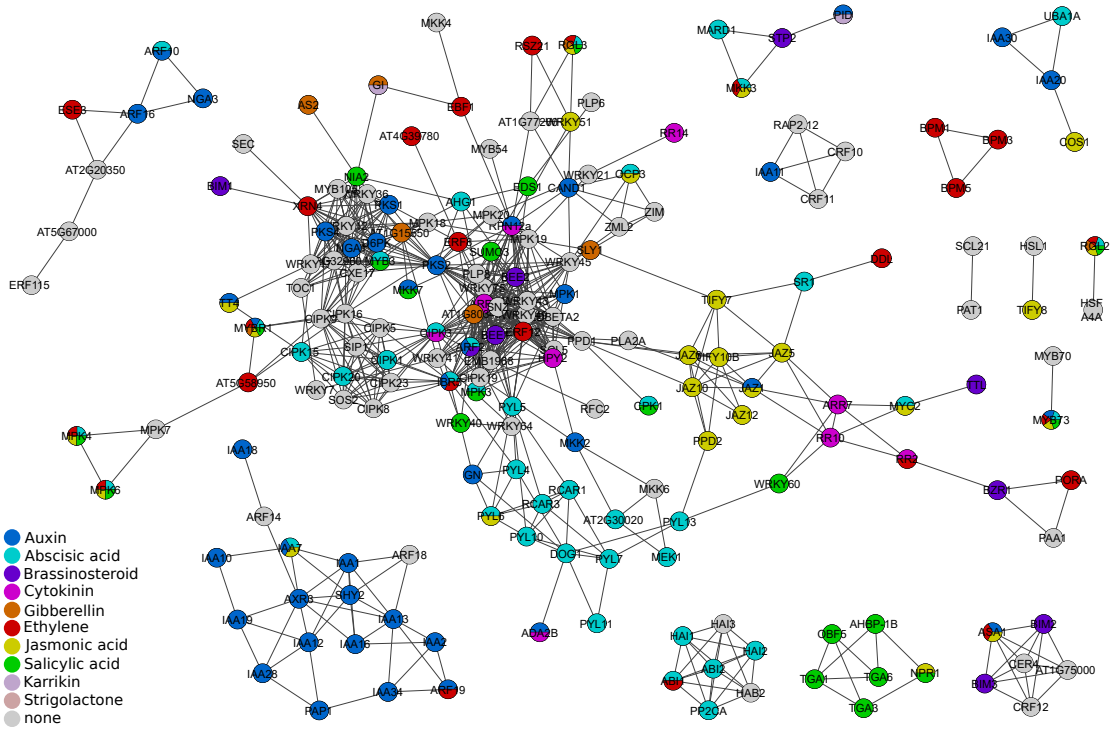


Figure 2.12: Association network of PhI interactions. Hormone annotations from AHD 2.0 and GO combined. Edges show similar interaction profile.

mind that proteins with a similar interaction profile in most cases do not physically interact, but can interact.

2.7 Phytohormone signal integration and points of crosstalk

Signal integration of different phytohormone signaling pathways or the reciprocal influence between phytohormone signaling pathways (crosstalk) to precisely control development, breeding, and response to stresses has been found in *Arabidopsis* mainly on the gene network level [123]. But only a small number of crosstalks induced by protein-protein interactions are described so far and only a low number of crosstalk candidates have been described to date (see section 2.5). One example for the influence of one signaling pathway to another is the interaction between JAZ1 and RGA, which is a member of the DELLA protein family. JAZ proteins function as transcriptional repressors of the JA signaling response and DELLA proteins are important regulators of GA signaling and inhibit expression of GA-induced genes in absence of GA. In Y2H and pull-down experiments an interaction between JAZ1 and RGA was proven. Pull-down experiments indicate that RGA and MYC2 compete for binding to JAZ1, which suppresses response to JA [165, 166].

We hypothesized that new phytohormone crosstalks achieved by protein-protein interactions can be identified in plant experiments. Therefore, knock-out plants of proteins assumed to be involved in a certain phytohormone pathway are expected to show a similar phenotype like the interacting protein, which is known to be in the examined phytohormone signaling pathway.

For identification of contact points between hormone signaling pathways mediated by protein-protein interactions, all protein-protein interactions between distinctly annotated proteins were identified. Additionally linear interactions between three proteins, where two differently annotated proteins are connected by a non-annotated protein, were extracted from PhI_{ext}. In this analysis, combined hormone annotation from AHD 2.0 and GO were used. Identified candidate crosstalks were divided up into two different categories: category I contains crosstalks between proteins, which are annotated strictly different. Crosstalks in category II are between proteins, which are differently annotated but have one or more hormone annotations in common.

2.7.1 Pairwise crosstalks

In PhI overall 173 candidate crosstalks of category I and 232 candidate crosstalks of category II were identified (fig. 2.13 C), whereby each distinct hormone combination of a single protein-protein interaction was counted as a crosstalk candidate between two different hormone signaling pathways. Crosstalk candidates of category I were found in 82 interactions between 79 proteins (fig. 2.13 A, C) and of category II in 77 interactions between 88 proteins (category II), respectively. Including PhI_{out} interactions, the number of candidate crosstalks of category I increases to 217 and 295 in category II, respectively (fig. A.6 C, table A.37). Including PhI_{out}, candidate crosstalks of category I are derived from 104 interactions between 103 proteins and candidate crosstalks of category II from 97 interaction between 108 proteins. Although the network PhI_{ext} is more than twice the size (1086 interactions between 696 proteins) compared to PhI (475 interactions between 251 proteins), the number of candidate crosstalks increases in both categories by around 20 %. This is expected, because 384 out of 581 proteins in PhI_{out} are not in the search space SSP_{PHO} and only 34 out of the 384 proteins have hormone related annotations in GO. The remaining proteins, so far uncharacterized for being involved in phytohormone signaling, may help to elucidate how phytohormones can influence which biological processes by protein-protein interactions.

A detailed analysis of crosstalk candidates of category I and II between phytohormone signaling pathways revealed crosstalk candidates between all signaling pathways except for KAR and SL, where only a low number of crosstalks and no crosstalks were found (fig. 2.13 B, tables A.33 - A.36). The high number of crosstalks in category II between ABA - IAA and ABA - ET pathways is unexpected, because in the community analysis no interaction between the ABA and ET community was observed and only a low number of interactions between the ABA and IAA communities in literature curated networks was observed. It is also notable that BR annotated proteins never share annotations with ET, GA and JA (only category I crosstalks). This means that the proteins involved in BR signaling pathway are not directly involved also in ET, GA or JA signaling pathways. This can imply either a lack of annotations or a low interlink between this pathways. Only three proteins are contained in PhI with KAR annotation, but nevertheless connections to all hormone signaling pathways except BR, CK and SL were found. No interactions were found in PhI for proteins involved in SL signaling pathway, therefore crosstalks for this phytohormone are not present in this network.

The combination of many studies elucidated that JAZ proteins serve as signaling hubs and affect not only the JA signaling pathway, but multiple signaling pathways by protein-protein interactions [166]. Via protein-protein interactions of JAZ proteins, the JA, GA, IAA and ET pathways are influencing each other [166]. The number of protein-

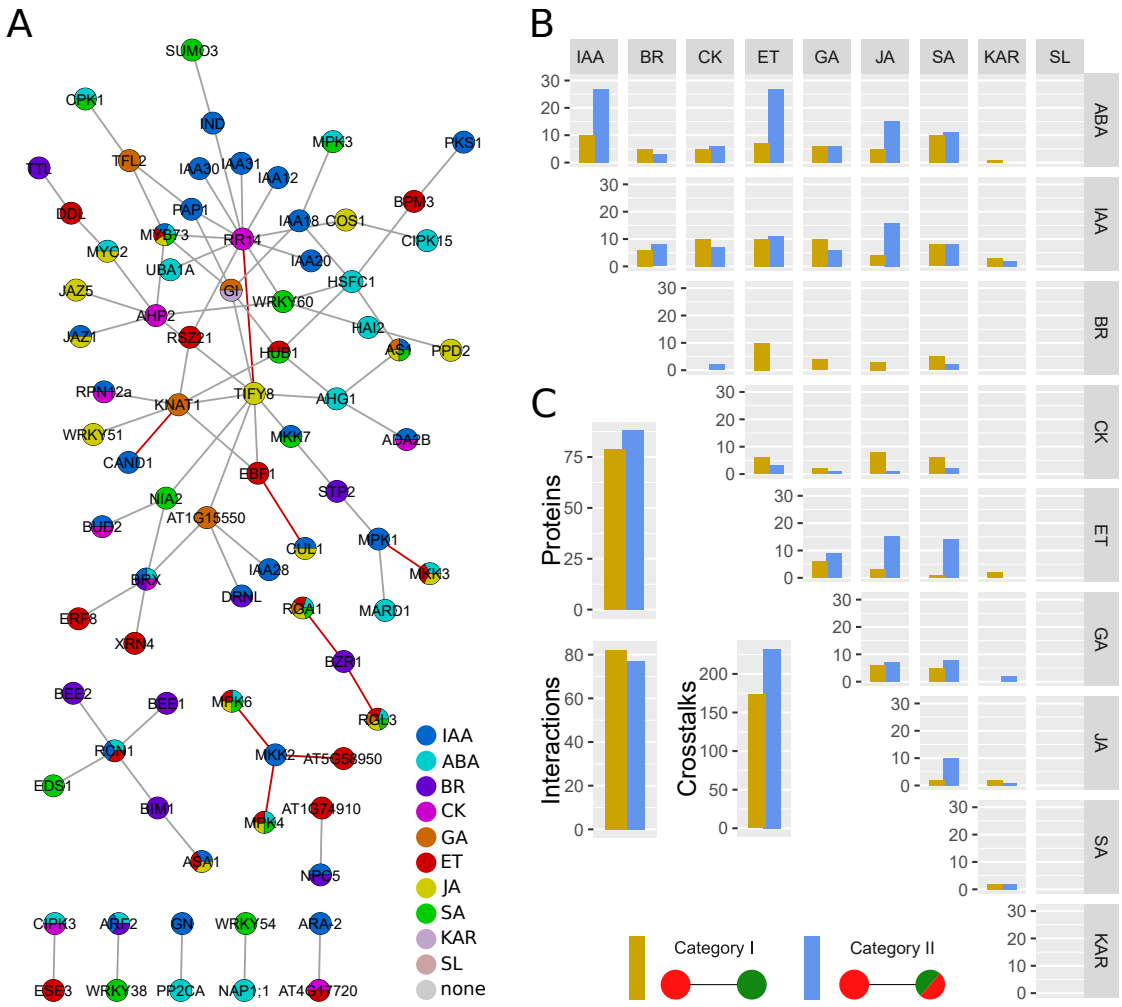


Figure 2.13: PhI pairwise crosstalks A) Network representation of pairwise crosstalks of category I in PhI B) Number of pairwise crosstalks category I (orange) and category II (blue) in PhI for individual hormone combinations. C) Number of proteins, interactions and crosstalks per category.

protein interactions of proteins involved in different phytohormone signaling pathways indicates that the reciprocal influence of phytohormone signaling pathways on each other takes place on a much larger scale. A lot of PPIs between proteins of different signaling pathways were found which could enable crosstalk between the pathways.

The elucidated PPIs indicate that crosstalk via protein interactions is not restricted to particular cases, but could be a common mechanism for signal transduction between pathways. An experimental validation could work out to which degree signals are transmitted by the identify crosstalk candidates.

Although it is known, that crosstalk between phytohormone signaling pathways are essential for developmental processes [97], defense and environmental adaption to stresses [108], only a few crosstalks mediated by protein-protein interactions are described. One of the few cases, for example, is the interaction between the BR regulated BIN2 and ARF2 involved in AUX signaling pathway. Phosphorylation of ARF2 by the kinase BIN2 increases expression of AUX-induced genes [121]. Another example is the interaction

between BZR1 and RGA, where both proteins serve as negative regulator of both the BR and GA signaling pathway [127]. An antagonistic crosstalk between the JA and ET signaling pathway was identified through the interaction of MYC2 and EIN3 [167].

The high number of crosstalk candidates found in PhI and PhI_{ext}, indicate, that signal transduction between phytohormone signaling pathways is a common mechanism.

2.7.2 Three node crosstalks

In the previous section, crosstalk candidates were identified by interactions between proteins, which are involved in distinct phytohormone signaling pathways. However, not all proteins have been analyzed for their role in phytohormone signaling and have therefore no annotation. Nevertheless, these proteins could be also involved in phytohormone signal transduction. To identify potential signal transduction by non-annotated proteins, linear interactions between three proteins, where two hormone annotated proteins are connected by a non-annotated protein (connector), were determined. This interactions are referred as *three node crosstalks*.

In PhI, 27 connector proteins were identified, which together connect 72 proteins involved in distinct phytohormone signaling pathways (fig. 2.14 A). Between two and six proteins are linked by a connector. TOE2 (AT560120) and TCP23 (AT1G35560) are outlier, as they connect a higher number of proteins. TOE2 connects eight proteins and TCP23 has 25 interactions partners with a phytohormone annotation. More than 80 % of annotated proteins interact with one or two connector proteins. Two proteins, MYB73 (AT4G37260) and BRX (AT1G31880, NLM9) have interactions with 8 and 10 connector proteins, respectively. Both proteins have multiple hormone annotations. Considering only three node interactions of category I (strictly distinct hormone annotations of connected proteins), the resulting subnetwork of PhI consists of 93 interactions between 76 proteins, which contains 509 candidate crosstalks. The high number of crosstalk candidates results from the number of phytohormone pathway interaction combinations, which can be derived from the network. For each connector protein the number of candidate crosstalks can be calculated by $\frac{(n-1)*n}{2}$, where n is the number of interactions * hormone annotations (see also methods 5.11). The subnetwork consisting of category II crosstalks (partially overlapping hormone annotations) consists of 100 interactions between 79 proteins and results in 734 crosstalk candidates (fig. 2.14 C). The number of crosstalks candidates between distinct hormone combinations is similar to distribution in pairwise crosstalk (compare figs. 2.14 B and 2.13 B).

In the combined network PhI_{ext}, many proteins not in search space SSP_{PHO} and without hormone annotations are included, which interact with proteins involved in phytohormone signaling pathways and act as connector proteins. 109 connector proteins are contained in the combined network, which connect 220 phytohormone related proteins (fig. A.7, table A.38). This is a threefold increase in connector proteins and hormone related proteins. Besides TOE2 and TCP23, two additional highly connected connector proteins show up: HSL1 (AT4G32010) and ARF18 (AT3G61830), which connect 11 and 9 hormone related proteins. Although ARF18 is known as *auxin response factor 18* and all interaction partners have annotations related to AUX and additionally other hormone signaling pathways, it has no experimentally validated GO annotation related to AUX.

For experimental validation of crosstalk candidates, three node interactions from a combined network PhI_{ext} can be used, but systematic analyses were restricted to PhI to avoid biases caused by different Y2H screening depth and screened ORFeomes.

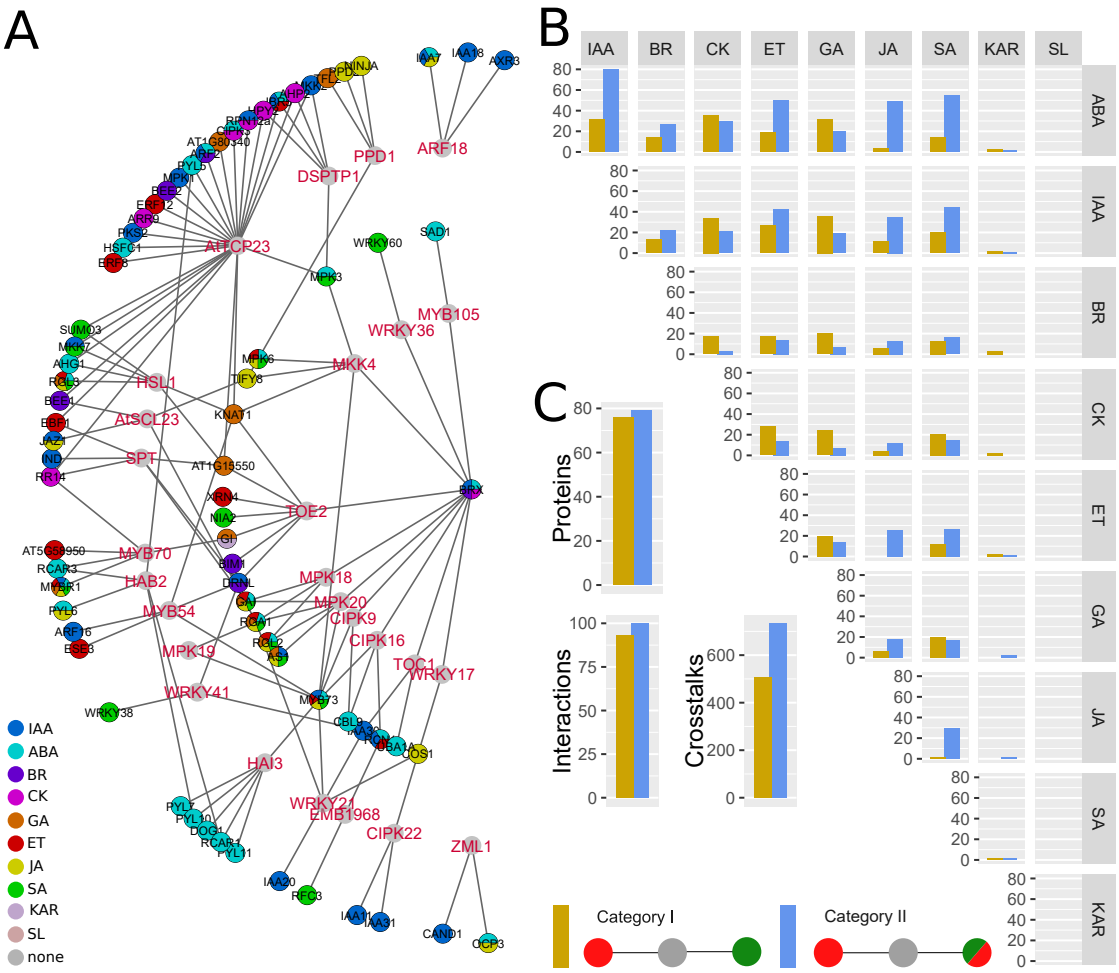


Figure 2.14: A) PhI three node crosstalks category I and II. B) PhI number of three crosstalks category I (orange) and category II (blue). C) Number of proteins and interactions in a network consisting of category I and II crosstalks candidates.

In PhI, 20 out of 27 connectors have at least one crosstalk of category I and not only crosstalks of category II. For these 20 connectors the crosstalk profiles were analyzed, which means for each connector protein the connected phytohormone pathways were enumerated (fig. 2.15 A). These connector proteins can link different phytohormone signaling pathways to regulate morphogenesis and development of a plant. Depending on the number of interactions, many pathways or only specific pathways are interconnected. The connectors WRKY transcription factors 17, 21, 36, and 41 connect SA, JA, and IAA to several other pathways, respectively. TCP23 interconnects all phytohormone pathways except SL, KAR and JA despite the high number of interactions. Also TOE2, the protein with the second most interactions has no connection to the JA signaling pathway.

SPT has six interactions with GAI, EBF1, DRNL, GA3OX1, IND and RR14 (fig. 2.15 B), which allows to form 15 different three node protein interaction pattern, where SPT may function as connector of two proteins or of two phytohormone signaling pathways, respectively. Two interacting proteins, DRNL and GAI, are involved in multiple hormone signaling pathways and therefore increase the number of possible hormone sig-

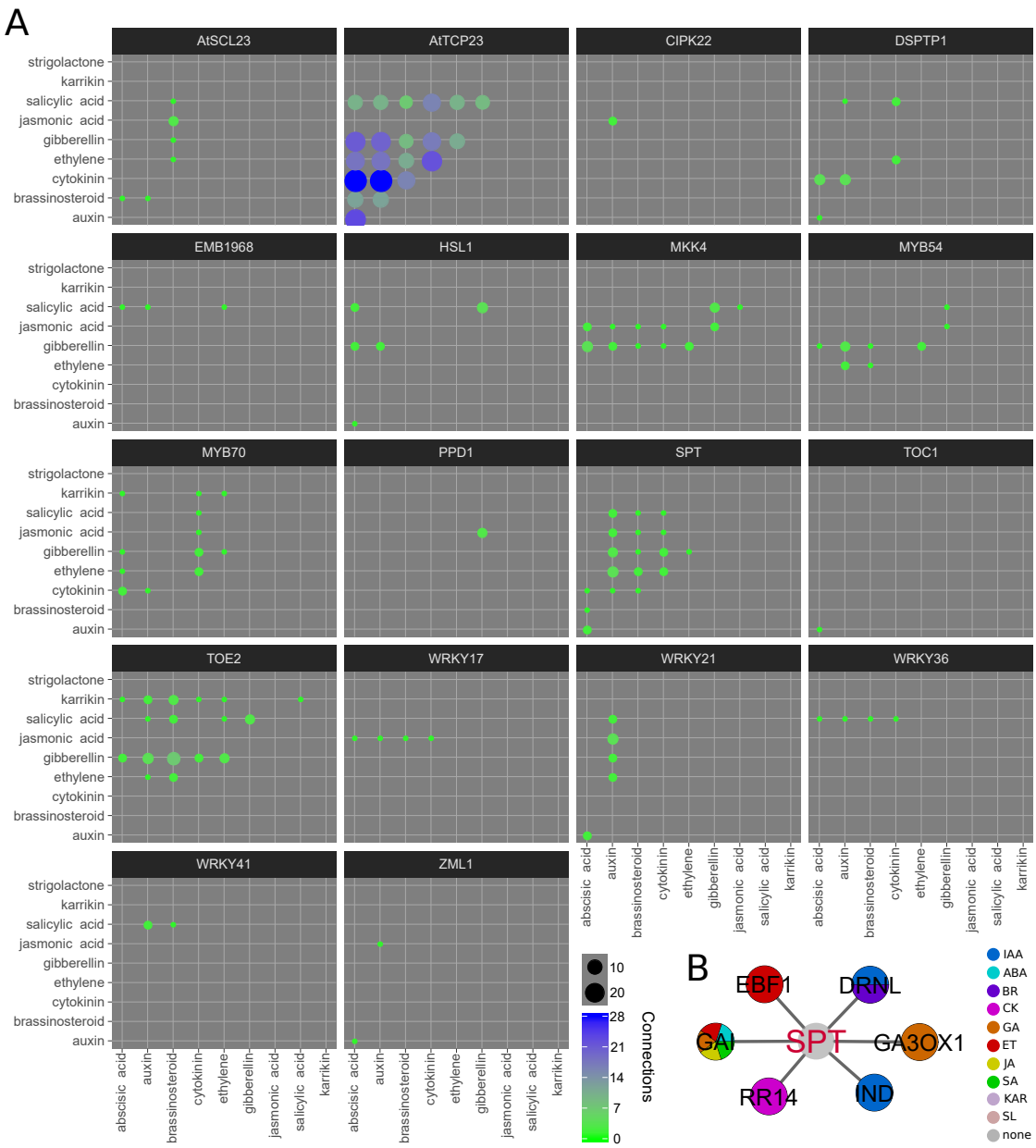


Figure 2.15: A) Hormone profiles of connector proteins show the number of hormone combinations connected by the respective connector protein. B) Network representation of SPT as connector protein with its hormone annotated interacting proteins. SPT connects proteins involved in seven distinct hormone signaling pathways. Resulting hormone profile is in row three, column three of panel A.

naling pathway connections up to a total of 18 combinations of interconnections between distinct signaling pathways. SPT could therefore act as signal transmitter between two phytohormone signaling pathways, if SPT and the two respective genes are expressed and their resulting proteins are available for binding. This means that genes should be involved in the same biological process and/or the respective genes should be expressed under the same condition and in the same tissue. This is analyzed in the following paragraph.

SPT mutant plants have abnormal, unfused carpels [168] and show a reduced seed dormancy [169]. It has been shown that a conserved acidic domain in SPT is essential for carpel function [170] and SPT and IND have a joint function in gynoecium and fruit development [171]. SPT also integrates time of day and temperature signaling to control vegetative growth rate [172]. In accordance to SPT's functions, RNA-Seq expression data [173] show a high expression in carpel and dark-grown seedlings. Nine physical interactions of SPT are currently contained in IntAct and BioGRID, where one interaction partner is IND and two other interaction partners are the DELLA proteins RGA1 and RGL2. The already known interaction partner of SPT, IND, has also a high expression in carpel [173]. This shows that both proteins are expressed in the same tissue. So far, for EBF1 no function in carpel development is described. EBF1 is important for fine-tuned ET response and is highly expressed in most tissues, including carpel [173]. Response regulator 14 (RR14) is known to be involved in CK-activated signaling pathway [174]. A direct interaction with the DELLA protein GAI has been shown in [175]. No function in carpel development has been shown so far, but also a high expression in RNA-Seq data in carpel is observed [173]. For GAI, a member of the DELLA protein family, 149 physical interactions are collected in IntAct and BioGRID database, but so far no physical interaction with SPT is described. Nevertheless, in [176] a cross-regulation of SPT by DELLA proteins was shown: SPT protein levels are low, if DELLAs are abundant and both, SPT and DELLAs restrain growth of the Arabidopsis seedling. RNA-Seq expression data show a high expression of GAI in most conditions and tissues except pollen. Gibberellin 3-oxidase 1 (GA3OX1) is known to be involved in later steps of the gibberellic acid biosynthetic pathway. No direct interaction with SPT is known so far, but a negative regulation by the DELLA protein GAI has been found [177]. A high expression of GA3OX1 is observed in RNA-Seq expression data of dark-grown seedlings and in root. DORNROSCHEN-LIKE (DRNL) has so far no described interaction with SPT or the other SPT interaction partners. Studies showed that drnl mutants are affected in floral organ outgrowth [178] and that DRNL might play a critical role in stamen emergence [179]. RNA-Seq expression data show an overall very low expression of DNRL in all tissues and conditions.

SPT and almost all interaction partners of SPT have an experimentally proven role in floral organs or show a high expression in the same tissue and under the same conditions in RNA-Seq experiments. This means that gene products, i.e. proteins, of SPT and its interaction partners are physically present, which is the prerequisite for a PPI. This suggests that SPT can interact with several of its interaction partner and can operate as connector for different signaling pathways in floral organs.

But also other connector proteins are known to play a role in different processes, which may indicate to mediate signaling between different pathways: TOE2 is a transcription factor involved in repression of flowering [180] and also involved in response to pathogen [142]. TCP23 is ubiquitously expressed and is involved in flowering time control and leaf and root development [181].

The example of SPT and its interaction partners and other connector protein shows that the identified connector proteins are likely to play an important role in linking different phytohormone signaling pathways.

2.8 Condition-dependent crosstalks

Y2H experiments are able to detect a part of all possible pairwise interactions in a given set of proteins (see also section 2.1.3). This set of protein-protein interactions represents a static map of interactions of the examined organisms, but it has been shown that protein interactions dynamically rewire under altered conditions [51, 148, 182, 183, 184]. Additionally, protein-protein interactions are detected in the Y2H system independently of the origin organisms. Therefore, the condition, the time point, the organelle, and the tissue of the specific interaction is undefined. So far no comprehensive proteome level data are available for *Arabidopsis thaliana* to get context for protein interactions. But it was shown that context can be predicted from RNA-Seq expression data: for protein-protein interactions in diauxic shift in yeast, prediction of context was possible with a high probability [51], albeit some additional factors can influence both translation of mRNA to protein and protein-protein interactions. Due to degradation or cell-to-cell movement, the mRNA level can differ from the protein level. Also post-translational modifications or additional molecules can influence activity or conformation of a protein, which influences binding activity. In the diauxic shift study was also observed that fully connected subnetworks centered around hub proteins form the dynamic parts in the protein-protein interaction network [51]. These hub proteins are called *date hubs* in contrast to *party hubs*. Date hubs interact with a lot of proteins at different time points, whereas party hubs interact with a lot of proteins at the same time point [185].

The PhI contains several hub proteins, where TCP23 has by far the most interactions. Moreover, in the previous analysis of three node crosstalks (section 2.7.2) could be shown, that TCP23 could potentially connect different phytohormone signaling pathways. The hypothesis was that TCP23 acts as a date hub and therefore interacts only with a portion of elucidated interaction partners in a specific condition, which could be either a tissue or an environmental condition. Furthermore, the question was which pathways are possibly connected in a specific condition.

To answer these questions, we calculated a score for protein-protein interactions as described in Celaj et al, 2007 [51]: Under the assumption that mRNA levels derived from RNA-Seq experiments reflect abundance of the respective proteins with a certain degree of uncertainty, a score for each protein-protein interaction (interaction score) was modeled for different conditions (see section 5.12) based on mass action by Gurudutta Panda. For an interaction score ≥ 2, we assumed that the interaction can take place in the respective condition because of sufficient presence of mRNA and hence protein of both interacting proteins.

This interaction score was used to analyze protein interactions of TCP23: In PhI, TCP23 interacts with 42 proteins and for the interactions with each of the interaction partners at least one condition is observed with an interaction score ≥ 2. Twenty five of the interacting proteins have an annotation for a hormone signaling pathway. The specific conditions under which the interactions of a specific interacting protein could take place are referred as condition profile.

The clustering of interaction scores of TCP23 and its 25 interacting proteins with hormone annotations shows a rough clustering into four clusters of proteins which have

an interaction score above 2 in similar conditions (fig. 2.16). It stands out, that all clusters contain differently hormone annotated proteins and proteins with the same annotation do not directly cluster together. For example, EBF1 and ERF8, both involved in ET signaling, are more similar to ARF2 (IAA, ABA, BR) and MPK3 (ABA, JA), respectively. The two homologous proteins BEE1 and BEE2 are together in one cluster, but the condition profile of BEE2 is more similar to the condition profile of the AUX annotated protein PKS2 than to BEE1.

The dynamic of interactions observed in yeast [51] can be also observed for the interactions of TCP23. Specific interactions take place only under specific conditions, which qualifies TCP23 as date hub. The clustering shows that not proteins involved in the same phytohormone signaling pathway have the most similar interaction profile with TCP23, but proteins involved in different pathways. This leads to the conclusion that likely the signals of different pathways merge on TCP23 or the pathways are connected by TCP23.

Interaction profiles of paralogous proteins BEE1 and BEE2

The interaction of TCP23 with the two paralogous proteins BEE1 and BEE2 led to the question to which degree the conditions overlap in which they interact with TCP23. This could also give a hint about the functional divergence of these proteins under the assumption of the duplication - divergence model proposed in [11] (see also section 1.1.4 and fig. 1.3).

BEE1 (BR enhanced expression 1, AT1G18400) and BEE2 (BR enhanced expression 2, AT4G36540) are both involved in the BR signaling pathway. For the third homologous protein BEE3 (BR enhanced expression 3, AT1G73830), no interaction was found in our Y2H screen. These three proteins are members of the basic helix-loop-helix transcription factor family, which consists of 162 genes [186, 187, 188]. To elucidate common and distinct processes of BEE1 and BEE2, condition profiles of both BEE proteins are compared to condition profiles of remaining TCP23 interacting proteins. This shows under which condition a certain BEE protein and another TCP23 interactor is likely interacting with TCP23 (table 2.1), e.g. AHG1 and BEE1 have and interaction score ≥ 2 with TCP23 in two conditions (first row in table 2.1) and BEE2 shares no common condition with AHG1.

The analysis shows that BEE1 and BEE2 have only rarely the same condition profile under the same condition with the same TCP23 interaction partner (table 2.1). In detail, BEE1 and BEE2 have one condition in common, where both paralogous proteins and the four proteins ARF2, ARR9, GA3OX1, and MKK7 show interactions with TCP23 at the same time. This occurs under treatment with elevated CO_2 (table 2.1, column 24). Only with ERF8, HSFC1, and PKS2 both BEE1 and BEE2 show additionally an interaction with TCP23 in the three same conditions. There are also proteins, which share conditions exclusively with only one of the two BEE proteins: AHG1, GA3OX2, ERF12, KNAT1, PYL5, and SUMO3 have only conditions with BEE1 in common, whereas BEE2 has no exclusive conditions with other proteins. But compared to BEE1, BEE2 and other proteins show in more common conditions an interactions with TCP23 than BEE1.

The analysis shows that conditions of BEE1 and BEE2 overlap only in a low number of conditions. In most of the conditions either BEE1 or BEE2 and a third protein can interact with TCP23 due to a sufficiently high edge score. Although BEE1 and BEE2 are both involved in the same processes, an increasing number of processes is

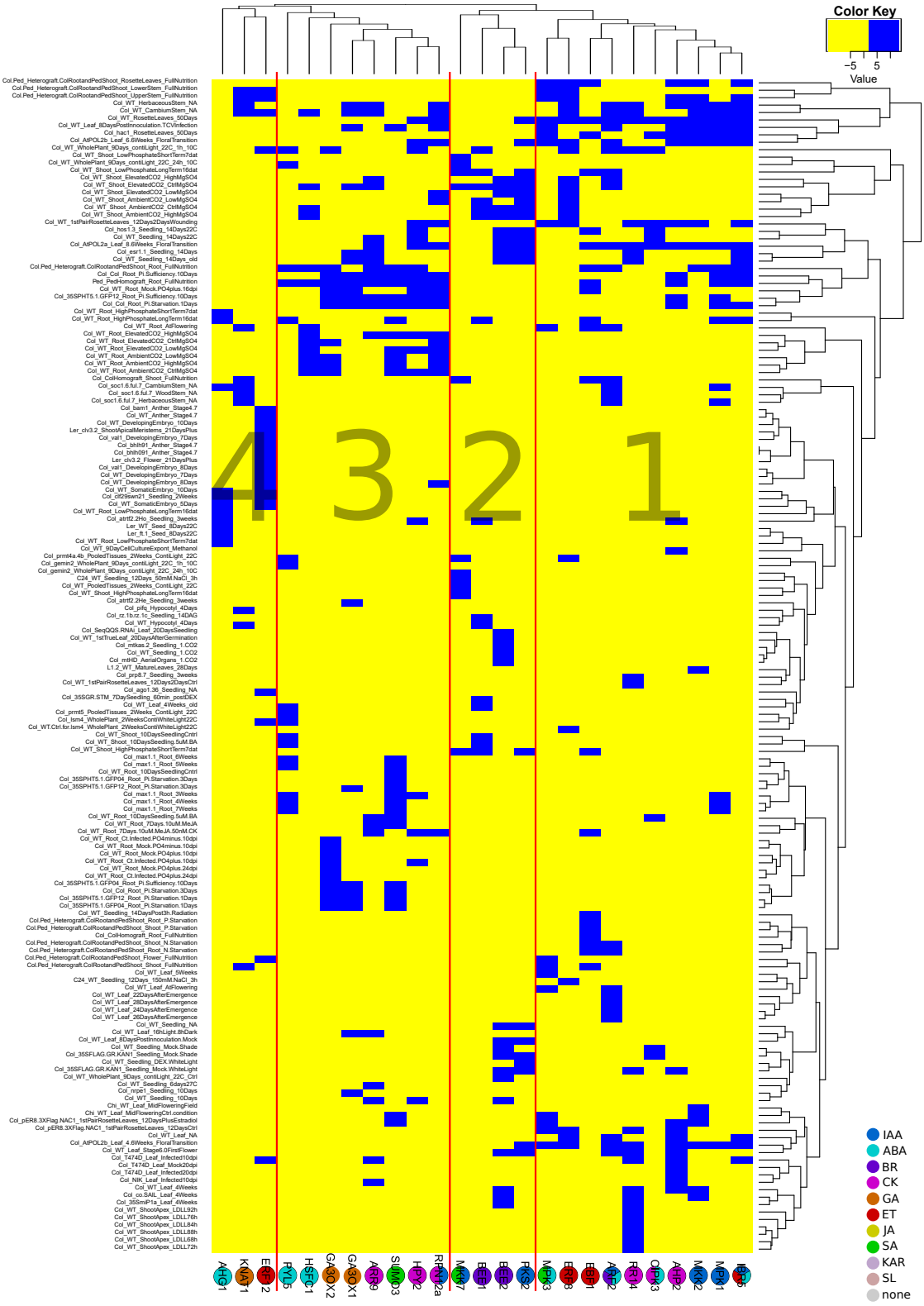


Figure 2.16: Condition-dependent interactions of TCP23 with hormone annotated proteins. Interaction score ≥ 2 (blue) means interaction of the respective protein with TCP23 under the respective condition. Numbers denote cluster of interactions.

elucidated, where only one of the paralogs plays a role: Phenotyping of triple knock-out plants of BEE1, BEE2 and BEE3 showed shorter hypocotyls and smaller floral organs. As response to ABA was reduced in *BEE1* overexpressing plants, a function as signaling intermediates of BEE proteins was suggested [189]. In [190] the common function of BEE1, BEE2 and BEE3 in shade avoidance was shown. Besides the common and redundant functionality of these three BEE proteins, many processes were identified in which only one or two BEE proteins play a role. Only the expression of BEE1 is dependent on MPK4 [191] and only BEE1 is underexpressed in knock-out plants of the E3 SUMO Ligase SIZ1 [192] and in *cie1* knock-out plants [193]. BEE1 and its closest homolog BEE3 have an important role in pollen growth in the reproductive tract [194] and only BEE1 interacts with CES in order to promote BR synthesis [195]. Unique roles of BEE2 are less abundant described than of BEE1: BEE2 is shown to be downregulated under drought stress [196] and also under treatment of plants with pathogen associated molecules elf18, flg22 or chitin [197]. This correlates with the observation that plants overexpressing BEE2 show a partially reduced immunity [197]. An interaction with PAR1 was only found with BEE2, which may attenuate AUX responses [198].

Interaction of the paralogous proteins BEE1 and BEE2 allowed to search for divergence on the basis of the edge score under different conditions. The analysis of the conditions under which they interact with TCP23 at the same time than another proteins involved in phytohormone signaling revealed a strong divergence in the expression of the two paralogous proteins.

2.9 Identification of signal transduction by potential kinase - substrate interactions

Post-translational modifications of proteins are an important mechanism of a cell to respond to external and internal signals of changing conditions and to control biological processes. Methylation, acetylation, ubiquitylation, sumoylation, and phosphorylation among others are possible post-translational modifications of specific amino acid residues in proteins. The most prominent modification is the phosphorylation, where one phosphate moiety is added to a specific serine, threonine, or tyrosine amino acid (phosphorylation site) of a substrate protein by a kinase [199].

In *Arabidopsis thaliana* 940 kinases based on 942 kinase domains were identified, which is 3.4 % of the 27,416 gene models annotated in TAIR10. The kinases are separated into two clades: the membrane located kinases and the soluble kinases, which comprise 561 and 381 kinases, respectively [130]. In comparison, the human kinome comprises 518 putative kinases, which is about 1.7 % of all human genes [200]. The large number of kinases results mainly from genome duplication events, where expanded kinase families have roles in plant-specific processes and contributes to adaptive evolution [201].

Phosphorylation sites are usually surrounded by a conserved amino acid sequence (phosphorylation motif), which often contributes to kinase-substrate recognition [133]. Although phosphorylation motifs for some kinases have been elucidated, no comprehensive collection for *Arabidopsis thaliana* is available.

Kinases do not only interact with proteins in order to phosphorylate them, but these interactions can also build a scaffold together with the kinase in order to specify substrate specificity or allosterically control the activity of a kinase. Some kinases must

Table 2.1: Concurrent interaction of BEE1 and BEE2 and another TCP23 interacting protein with TCP23

	Col_WT_Hypocotyl_4Days	Col_AtPOL2a_Leaf_8.6Weeks_FloralTransition	Col_WT_Leaf_Stage6.0FirstFlower	Col_WT_Leaf_4Weeks	Col_co.SAIL_Leaf_4Weeks	Col_35SmiP1a_Leaf_4Weeks	Col_WT_Leaf_8DaysPostInnoculation.Mock	Col_WT_Leaf_4Weeks_old	Col_WT_WholePlant_9Days_contiLight_22C_1h_10C	Col_WT_Root_HighPhosphateLongTerm16dat	Col_atrtf2.2Ho_Seedling_3weeks	Col_WT_Seedling_14Days_old	Col_esr1.1_Seedling_14Days	Col_WT_Seedling_10Days	Col_35SFLAG.GR.KAN1_Seedling_Mock.WhiteLight	Col_WT_Seedling_Mock.Shade	Col_35SFLAG.GR.KAN1_Seedling_Mock.Shade	Col_WT_Seedling_NA	Col_WT_Seedling_14Days22C	Col_hos1.3_Seedling_14Days22C	Col_WT_Shoot_AmbientCO2_CtrlMgSO4	Col_WT_Shoot_AmbientCO2_HighMgSO4	Col_WT_Shoot_AmbientCO2_LowMgSO4	Col_WT_Shoot_ElevatedCO2_CtrlMgSO4	Col_WT_Shoot_ElevatedCO2_HighMgSO4	Col_WT_Shoot_ElevatedCO2_LowMgSO4	Col_WT_Shoot_10DaysSeedlingCutrl	Col_WT_Shoot_10DaysSeedling.5uM.BA	Col_WT_Shoot_HighPhosphateShortTerm7dat	Col_WT_Shoot_LowPhosphateLongTerm16dat
AHG1:BEE1										✓	✓																			
AHP2:BEE1											✓																			
AHP2:BEE2		✓	✓	✓											✓															
ARF2:BEE1									✓															✓						✓
ARF2:BEE2		✓	✓																					✓	✓					
ARR9:BEE1																								✓						
ARR9:BEE2		✓										✓	✓	✓					✓					✓	✓					
GA3OX2:BEE1									✓																					
CIPK3:BEE1									✓																					
CIPK3:BEE2		✓														✓	✓		✓	✓										
EBF1:BEE1										✓																			✓	✓
EBF1:BEE2		✓																		✓										
ERF12:BEE1									✓																					
ERF8:BEE1									✓												✓	✓	✓	✓						
ERF8:BEE2																					✓	✓		✓	✓	✓				
GA3OX1:BEE1																								✓						
GA3OX1:BEE2												✓	✓											✓						
HPY2:BEE1									✓		✓																			
HPY2:BEE2		✓												✓					✓	✓										
HSFC1:BEE1																					✓	✓		✓						
HSFC1:BEE2																					✓	✓		✓						
IBR5:BEE1										✓																				
IBR5:BEE2		✓										✓	✓																	
KNAT1:BEE1	✓																													
MKK2:BEE2		✓		✓	✓																									
MKK7:BEE1																								✓					✓	✓
MKK7:BEE2																								✓						
MPK1:BEE1										✓																				
MPK1:BEE2		✓																												
PKS2:BEE1																					✓	✓	✓	✓					✓	✓
PKS2:BEE2		✓	✓				✓					✓	✓		✓		✓	✓	✓	✓	✓	✓		✓	✓	✓				
PYL5:BEE1								✓	✓	✓																	✓	✓		
RPN12a:BEE1																							✓							
RPN12a:BEE2		✓																								✓				
RR14:BEE1									✓																					
RR14:BEE2		✓	✓	✓	✓	✓						✓	✓		✓															
SUMO3:BEE1										✓																				

be activated by phosphorylation by other kinases like in the well known MAP kinase cascade [133].

Since phytohormone signals are also transmitted by phosphorylation, the question arose, where in the PhI hormone signals are transmitted by kinase - substrate interactions. Therefore the aim was to identify phosphorylation motifs in kinase interacting proteins in PhI to predict potential kinase - substrate interactions.

To identify possible kinase - substrate interactions in PhI, all kinases in PhI were identified and classified for their kinase family they belong to. For each represented kinase family, possible phosphorylation motifs were determined from a set of 15 amino acid long sequences with a phosphorylated residue in the center (15mers), which was derived

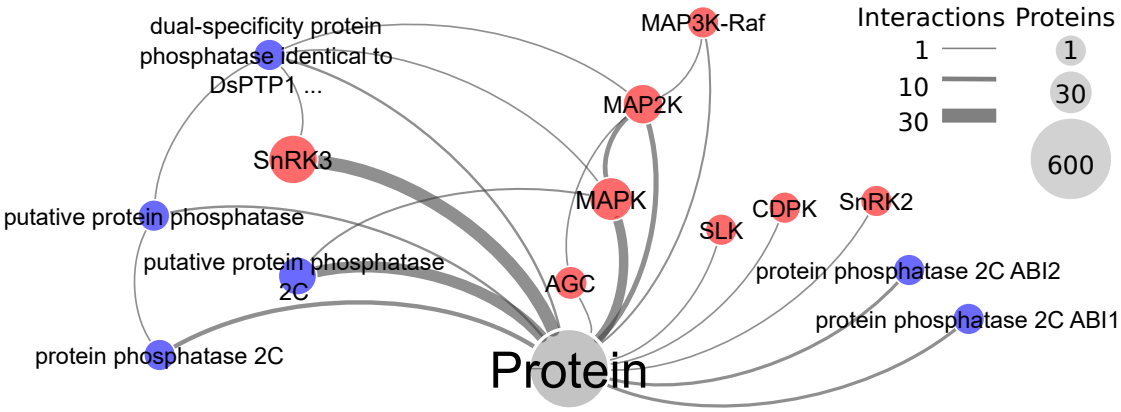


Figure 2.17: Kinase and phosphatase family interactions in PhI network. Size of vertices show number of proteins in each family, thickness of edges show number of interactions between families. Red nodes are kinase families, blue nodes are phosphatase families, gray node contains all proteins which belong to neither kinases nor phosphatases.

from substrates of the respective kinase family. The 15mers were extracted from a list of short phosphorylated peptide sequences provided by Waltraud Schulze by personal communication. For each kinase family, several motifs were determined. The number of motifs for each kinase family was estimated by a clustering approach using an elbow plot and principal component analysis (fig. 2.18). Subsequently, these motifs were used to determine possible phosphorylation sites and substrates of proteins in PhI_{ext}, which interact with kinases from the respective family. The significance of identified phosphorylation sites were tested against the complete Arabidopsis proteome. To test this approach, motifs were also tested on known substrates of the respective kinase family derived from PhosPhAt [202, 203, 204].

In PhI 34 kinases from 8 families and 10 phosphatases from 6 groups were identified (table A.40), which show mainly interactions with proteins not belonging to kinase or phosphatase families. Some interactions between kinase and phosphatases were observed, where most of the interactions are found between MAP2Ks and MAPKs (fig. 2.17 and table A.39). Including interactions from PhI_{out}, seven additional kinases from AURORA, CDK, CKL, RLCK 3, RLCK 9, and soluble kinases family and two additional phosphatases could be identified.

Identification of phosphorylation sites in positive control

PIN1 is a known substrate of AGC kinases and is used as positive control for this approach. PIN1 is used to demonstrate the identification of motifs from known substrates and in particular the identification of PIN1 as substrate of the AGC kinase family and of phosphorylation sites in the amino acid sequence of PIN1. PIN-FORMED (PIN) proteins are membrane bound AUX efflux transporters. The AUX transport is activated by phosphorylation of PIN proteins by D6 PROTEIN KINASE (D6PK) and PINOID (PID)/WAG kinases of the Arabidopsis AGCVIII kinase family. Four phosphorylated positions in the first intracellular hydrophylic loop of PIN1 [205] have been identified so far: S231, S252, and S290 are phosphorylated by PID and S271 by D6PK [134, 206].

For identification of phosphorylation motifs, 15mers are clustered by sequence similarity. Clustering of 15mers of proteins phosphorylated by AGC kinases show a good

separation at a partition into four clusters in both elbow plot (fig. 2.18 A) and principle component analysis (fig. 2.18 B). From each of these four sets of 15mers a position weight matrix was derived, which contains the frequency of each amino acid at each of the 15 positions of clustered 15mers. This position weight matrix (PWM) was used in turn to identify potential phosphorylation sites in proteins. Sequence logo representations of the four position weight matrices are shown in figure 2.18 C-F. Considering the most abundant amino acids per position, restricted to positions reaching a minimum bit score of 1, the motif TxRxpS[NS]x[EG] (p denotes phosphorylated position, x denotes an arbitrary amino acid) can be derived from PWM 4.1 (fig. 2.18 C), RxxpS from PWM 4.3 (fig. 2.18 E) and pSPRR from PWM 4.4 (fig. 2.18 F). PWM 4.2 has a very low variation and reaches a bit score of 1 at each position, which results in the motif PMKQHVRp[ST][VP]LGLQL[KL]. Two similar phosphorylation motifs have been described for AGC kinases: TPRxpS[NS] has been found as PIN-specific motifs for kinases from AGCVIIIa group [134, 207], RxRxxS/T motif has been found as preferred phosphorylation motif of animal AGC kinase Akt [208]. This shows that phosphorylation motifs of a kinase family can be extracted from known substrates of the respective family.

After extraction of phosphorylation motifs, all four position weight matrices have been used to identify possible phosphorylation sites in PIN1 and to test, if they fit significantly better on PIN1 than on other *Arabidopsis* proteins (background). Therefore a motif score for each 15mer around a phosphorylatable amino acid is calculated, which is the product of the frequencies of the respective amino acid at the respective position relative to the phosphorylated amino acid. A test of PWM 4.1 on PIN1 shows that this PWM fits better on PIN1 than on proteins in background distribution. The observed P-value of PIN1 is 0.0002, which shows a high probability for PIN1 to be phosphorylated by an AGC kinase (fig. 2.19 A). A detailed analysis of the identified phosphorylated position in PIN1 reveals, that a serine residue at position 252 (S252) best fits to the PWM. This position has been found to be phosphorylated by AGCVIIa kinases together with S290, S231 and S271, which also have a higher motif score than other residues in PIN1 (fig. 2.19 B) [134, 206]. The results let conclude, that it is possible to clearly identify PIN1 as a target of AGC kinases using the position weight matrices and moreover it is possible to identify phosphorylated positions in the protein.

Identification of substrates in PhI_{ext}

The described approach was used to identify possible kinase substrate interactions in the combined network PhI_{ext}. On the one hand, these interactions help to identify points of signal integration by phosphorylation, on the other hand a direction of the signal flow can be inferred. In the combined network, 27 potential phosphorylation targets for kinases from the eight families cyclin-dependent kinase (CDK), calcium-dependent protein kinase (CDPK), mitogen-activated protein kinase (MAPK), mitogen-activated protein kinase kinase (MAP2K) , mitogen-activated protein kinase kinase kinase (MAP3K), Shaggy-like kinases (SLK), SNF1-related kinase 2 (SnRK2), and SNF1-related kinase 3 (SnRK3) were identified (table 2.2). Additionally 10 out of 13 already in literature described substrates of the respective kinase families were correctly reidentified (table A.43). These 27 possible substrates have 47 interactions with 30 kinases, whereby a protein can have both roles, substrate and kinase e.g. in case of MAPKs and MAP2Ks, and are counted here twice as they represent distinct signaling events.

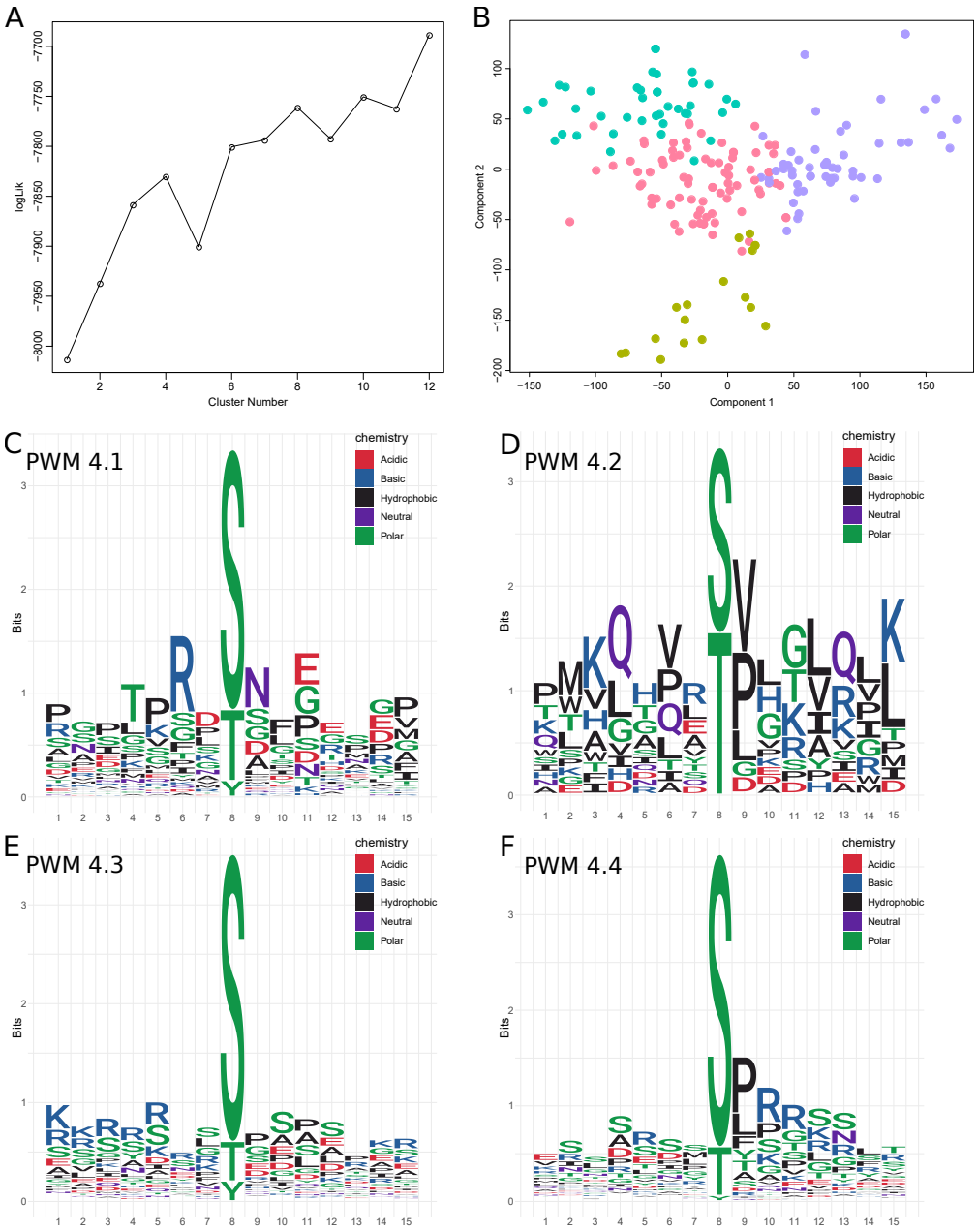


Figure 2.18: AGC kinase family clustering results. A) Elbow plot of AGC 15mer clustering. B) Principal component analysis of AGC 15mers. Colors show assignment of 15mers to one of four clusters. C-F) Sequence logo representation of PWMs. Central positions is phosphorylated amino acid. Seven positions upstream and seven positions downstream are considered. Y-axis shown information content as bit score. C) Phosphorylation motif derived from cluster 1. D) Phosphorylation motif of cluster 2. E) Phosphorylation motif of cluster 3. F) Phosphorylation motif of cluster 4.

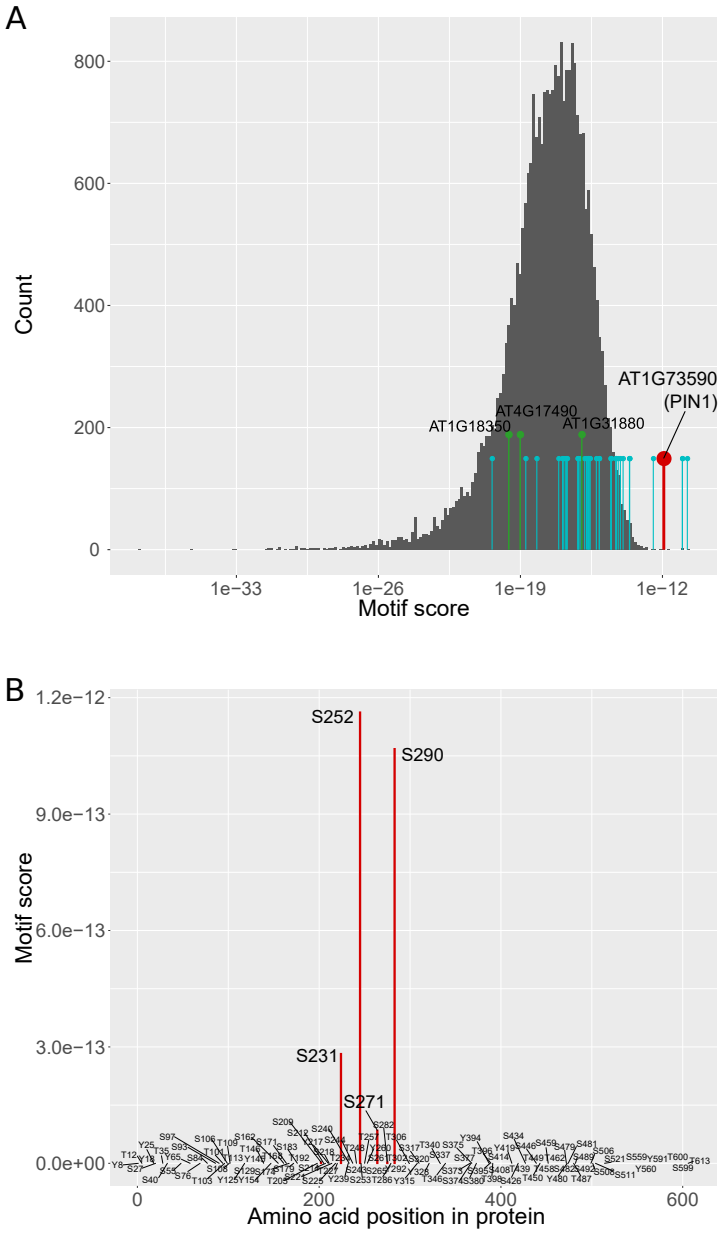


Figure 2.19: Motif scores using PWM derived from AGC 15mers in cluster 4.1. A) Gray distribution shows best score of PWM derived from 15mers in cluster 1 on all 27,416 protein sequences of primary gene models in TAIR10. Blue bars show best score on proteins known to be phosphorylated by AGC kinases. Red bar shows best score in PIN1. Green bars shown proteins interacting with AGC kinases in PhI_{ext}. B) Motif scores in PIN1 for each phosphorylatable position, which contains either amino acid serine (S), threonine (T) or tyrosine (Y). Character and number shows position and amino acid in the protein sequence. Red bar shows the motif score. No bar means motif score is virtually zero and does not fit. PWM fits on four positions with a motif score greater zero: S231, S252, S271, and S290.

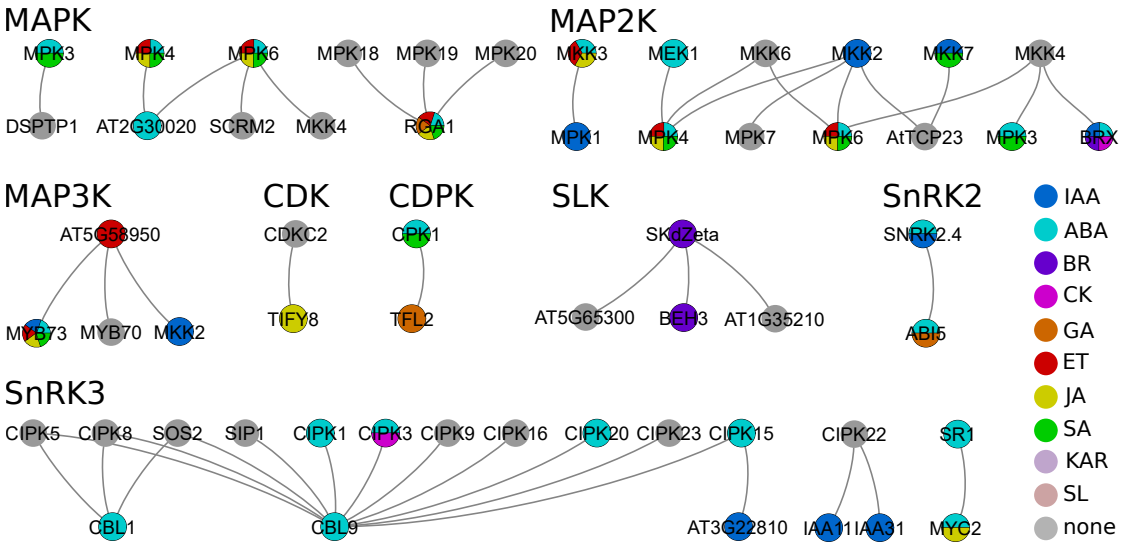


Figure 2.20: Possible kinase - substrate interactions contained in PhI_{ext} grouped by kinase families. For each kinase family the respective kinases are in the upper row and possible substrates are in lower row. Nodes are colored for their respective hormone annotation in AHD 2.0 and GO.

The analysis of putative kinase - substrate pairs showed that both kinases and possible substrates had a strong variation in the number of hormone annotations. MAP kinases, MAPK kinases, MAPKK kinases and / or their respective possible substrates had hormone annotations for up to five hormone signaling pathways (fig. 2.20). Taken together, MAPK and MAP3K had hormone annotations for five different hormone pathways and MAP2K had annotations for seven different hormone signaling pathways (fig. 2.20). Proteins involved in CDK, CDPK, SLK, SnRK2, and SnRK3 signaling pathways had only up to two hormone annotations and taken together annotations for up to three hormone signaling pathways (fig. 2.20). This indicated that a larger number of phytohormone signals are integrated in the MAPK cascade than in the other kinase interactions.

Nevertheless, crosstalks of category I or category II are contained in the interactions of all kinase families except in the SLK and CDK family (see section 5.11). Taken together, six crosstalks of category I between differently annotated proteins and seven crosstalks of category II with overlapping phytohormone annotations are contained.

With the approach of extracting position weight matrices from known substrates and using it for identification of new substrates, both known and new substrates could be predicted. This allowed to infer a direction of the signal flow in the protein interaction network. Moreover, a signal integration from different pathways especially in MAPK cascade was observed.

2.10 Signal integration at transcription factor level

All phytohormone signaling pathways in plants have receptor proteins, which sense the respective phytohormone molecules. These signals are often transmitted to transcription factors, which control expression of target genes [21]. The BR signaling pathway, for example, contains the receptor kinases BRI1, which senses BR and activates a signaling pathway, which can activate a small family of TFs. These TFs regulate expression of

Table 2.2: Protein-protein interactions between proteins, which belong to a kinase family and proteins, where a PWM of the respective kinase family fits significantly better than on background. Table shows kinase family, interacting protein locus, best P-value of comparison against background, and if phosphorylation event between kinase family and locus has been described in literature (known) or not (new).

Kinase	Locus	P-Value	Status
CDK	AT4G32570	4,03E-02	new
CDPK	AT5G17690	2,65E-02	new
MAPK	AT2G01570	3,78E-02	new
MAPK	AT3G23610	3,57E-03	new
MAPK	AT2G30020	1,60E-03	new
MAPK	AT1G51660	7,73E-03	known
MAPK	AT1G12860	2,50E-02	new
MAP2K	AT1G35560	2,69E-02	new
MAP2K	AT4G01370	7,30E-05	known
MAP2K	AT1G31880	3,14E-03	new
MAP2K	AT3G45640	3,65E-05	known
MAP2K	AT1G10210	1,82E-04	known
MAP2K	AT2G43790	4,00E-05	known
MAP2K	AT2G18170	2,92E-04	known
MAP3K	AT2G23290	3,80E-02	new
MAP3K	AT4G37260	4,67E-03	new
MAP3K	AT4G29810	3,92E-02	known
SLK	AT4G18890	1,70E-02	new
SLK	AT5G65300	1,38E-02	new
SLK	AT1G35210	1,91E-02	new
SnRK2	AT2G36270	2,55E-04	known
SnRK3	AT5G47100	7,30E-05	known
SnRK3	AT3G17600	8,06E-03	new
SnRK3	AT4G28640	1,01E-02	new
SnRK3	AT1G32640	5,69E-03	new
SnRK3	AT4G17615	3,65E-05	known
SnRK3	AT3G22810	4,07E-02	new

more than thousand genes [209]. The TFs BZR1 and BZR2 in BR signaling pathway interact with DELLAs proteins from GA signaling path, which have an inhibitory effect on transcription factor activity [210]. In this case, phytohormone signal integration takes place on transcription factor level, where repressor proteins influence transcription factor activity. To identify these signal integration events, PhI_{ext} was screened for interactions of proteins with known repressor function and transcription factors.

The combined network PhI_{ext} was screened for interactions between 71 repressor proteins (table A.44) and transcription factors. In the combined network, 140 transcription factors were identified, whereby 44 repressor proteins are also transcription factors with DNA binding activity. Out of 140 transcription factors, 22 have 42 interactions with 23 repressor proteins (fig. 2.21). Seven of these transcription factors have interactions with

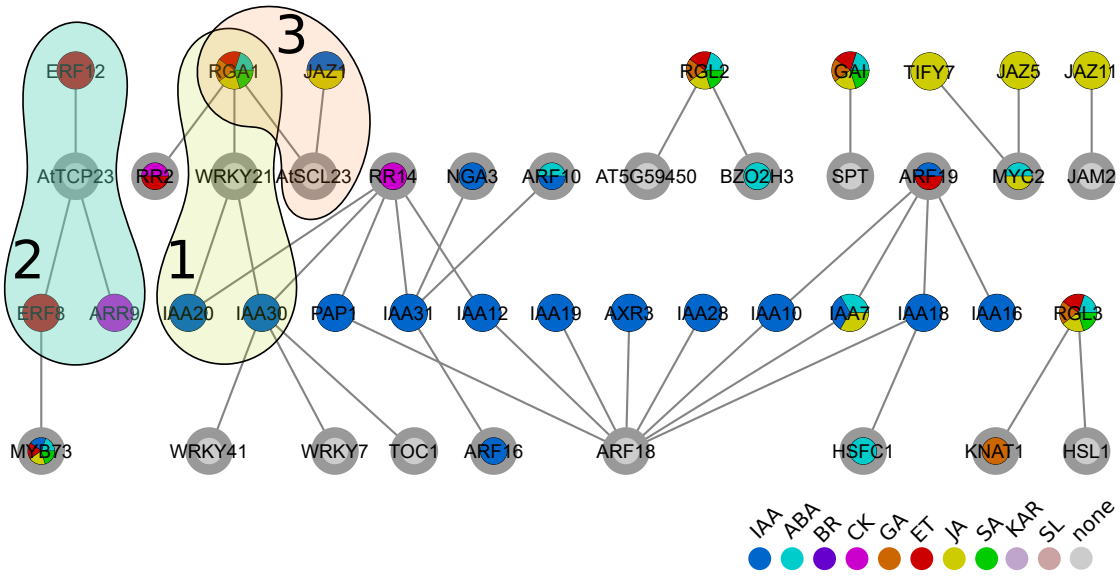


Figure 2.21: Hormone signaling repressor protein - transcription factor interactions. Transcription factors are marked with a grey border. Colors show combined hormone annotations from AHD 2.0 and GO.

multiple repressor proteins and also some IAA repressor proteins interact with multiple transcription factors.

Three transcription factors TCP23, WRKY21, and SCL23 were so far not associated with any phytohormone signaling pathway, but each of these three transcriptions factors is interacting with repressors from different phytohormone signal pathways. WRKY21 is interacting with two IAAs from the AUX pathway and RGA1 involved in multiple pathways (fig. 2.21, interaction group 1), TCP23 is interacting with two ERFs from ET signaling pathway and ARR9 from CK signaling pathway (fig. 2.21, interaction group 2) and SCL23 has interactions with RGA1 and JAZ1, which is involved in AUX and JA signaling pathway (fig. 2.21, interaction group 3). There are also cases where a transcription factor associated with a hormone signaling pathway interact with a repressor from a different hormone signaling pathway. This is for example the case for RR14 involved in CK signaling pathway, which has multiple interactions with repressors from AUX signaling pathway.

This analysis showed that signal integration can take place at transcription factors. In the identified cases, TFs interact with repressor proteins involved in distinct hormone signaling pathways, which could indicate that different hormone pathways control the expression of genes. The binding of repressor proteins to TFs could be a possible mechanism to regulate hormone dependent gene expression by different hormone pathway.

2.11 Natural variation in PhI

Arabidopsis thaliana occurs in Europe, Asia and Northern Africa and has recently colonized Northern America [58]. It thrives in many diverse habitats with different temperature and precipitation conditions, on different grounds and from sea level up to 4250m. Therefore *Arabidopsis* is suitable to analyze adaptive traits [59]. Many genetic loci could

be linked to adaptation to local environment of the respective ecotype [60, 61]. These loci often show deviations from neutral selection in allele frequency. These adaptive polymorphisms in a population can be used to identify genes under natural selection. Therefore single nucleotide polymorphisms (SNPs) are compared between ecotypes to identify their frequency. The frequencies of SNPs within a gene are exploited by neutrality tests to determine the type of selection. To test a gene for neutral selection, different neutrality tests can be used, e.g. Tajima's D, Fu and Li's D*, Fu and Li's F*, and others [52]. Alleles should be analyzed with multiple tests of neutrality to avoid misinterpretation of a single test [55]. To identify genes under positive or balancing selection involved in phytohormone signaling, coding sequences derived from 1135 accessions [58] were tested with four tests to identify deviations from neutral selection: Tajima's D [211], Fu & Li's F* and D* [212], and Ramos' R_2 [213], which are all based on the allele frequency spectrum to determine type of selection. Tajima's D test compares estimates of the number of segregating sites and the mean pair-wise difference between sequences. Fu and Li's F* compares the number of singletons to the mean pair-wise difference between sequences. Fu and Li's D* compares the number of singletons to the total number of nucleotide variants of the sequence [214]. Ramos' R_2 is based on the comparison of the difference between the number of singletons per sequence and the average number of nucleotide differences [214] and can be used to cross check Tajima's D results [215].

A strong positive Tajima's D indicates balancing selection, where multiple alleles exist in the population, whereas a strong negative Tajima's D indicate accumulation of SNPs in each single allele (singletons). The first case can be a result of either a population contraction or weak bottleneck; the second case can be observed in case of population expansion or strong bottleneck [216].

The interpretation of Fu and Li D* statistic is similar to Tajima's D, where a negative value also indicates an excess of singletons and a positive value indicates a lack of singletons. The Fu and Li F* has a similar behavior than Fu and Li D*. Both tests indicate therefore the same population scenarios than Tajima's D.

A low value in Ramos' R_2 test indicates an expanding population and a high number of singletons, similar to Tajima's D. Ramos' R_2 is more sensitive for this event in a small population than Tajima's D [213].

For each selection measure, as cut-off for selecting significantly deviating genes from neutral selection, the two-sided 95 % confidence interval from all 27,416 genes in *Arabidopsis thaliana* was selected, similar as described in [217, 218, 219]. Whereby this is the confidence interval for neutral mutations and all values lower than the 2.5 percentile and greater than the 97.5 percentile are considered as significant. Genes with values outside the two-sided 95 % confidence interval are assumed to be under non-neutral selection.

In PhI, ten genes have a Tajima's D value not in the two-sided 95 % confidence interval (table 2.3) defined by all *Arabidopsis thaliana* genes. All genes except CLA1 also have a value outside 95 % confidence interval in one of the other selection measures. All other genes have values outside the confidence interval only in one selection measure or in two related selection measures like Fu & Li's F* and D*. For further analysis, all genes, which are not in 95 % confidence interval in Tajima's D and one additional method are assumed to show evidence for deviation from neutral selection (deviating genes).

These nine genes were further analyzed in terms of network properties of the respective proteins in PhI. The mean values of degree, clustering coefficient (transitivity), betweenness, and closeness of the nine deviating genes were compared against mean values derived from 10,000 degree preserved randomized networks or networks with randomly

Table 2.3: All loci of PhI, which have a value outside the 95 % confidence interval in one of the selection measures. Only values not in the confidence interval are shown. The upper and lower value of the confidence interval is shown in the last two lines.

Locus	Symbol	Tajima's D	Fu and Li's D*	Fu and Li's F*	Ramos R2
AT1G25490	RCN1	-2.48			0.00
AT1G54490	XRN4	-2.54			0.00
AT1G77470	RFC3	-2.44			0.00
AT2G01760	RR14	0.86			0.08
AT2G04890	SCL21	-2.54	1.88	-8.59	
AT2G25000	WRKY60	-2.48			0.00
AT3G23610	DSPTP1	0.59			0.08
AT4G01026	PYL7	1.09			0.09
AT4G15560	CLA1	-2.49			
AT5G56580	MKK6	1.32			0.09
AT1G07430	HAI2		-9.65	-7.04	
AT1G32640	MYC2		0.59	-7.37	
AT4G18880	HSF A4A		-9.30	-7.05	
AT4G30080	ARF16		-9.69	-7.18	
AT5G62000	ARF2		-9.81	-7.00	
AT5G25190	ESE3			-0.61	
AT1G77200					0.00
AT2G45820					0.00
AT2G46130	WRKY43				0.00
AT3G45640	MPK3				0.00
AT3G61860	RS31				0.00
AT5G19000	BPM1				0.00
AT5G65210	TGA1				0.00
lower		-2.42	-8.85	-6.85	0.006
upper		0.48	0.05	-0.69	0.08

permutated vertex ids, respectively. Both degree, clustering coefficient (transitivity) and betweenness of deviating genes showed no significant deviation from random expectation. A significant deviation from random expectation was found for closeness with an observed p-value of 0.0009 (fig. 2.22). The mean closeness of all nine deviating genes is significantly lower than expected by chance. As closeness is defined by the inverse of the average length of the shortest paths to all other vertices in the graph, this means that deviating genes are less central in the network and / or less connected than expected by chance. This fits to the observation that hubs in protein-protein interaction network are more conserved [24, 220] than proteins at the periphery of the network [72].

Four (RR14, DSPTP1, PYL7, MKK6) out of the nine deviant genes showed a positive Tajima's D value, the remaining five genes (RCN1, XRN4, RFC3, SCL21, WRKY60) had a negative value. A positive value is a signal for balancing selection, which means that different alleles exist in the population. A negative value means that each single ecotype has accumulated private SNPs, which are hardly shared between ecotypes [52]. All deviant genes have been subject of examinations, which allows to identify their biomolecular functions and common features from literature.

MKK6 is required together with MPK13 for lateral root formation [221] and there is evidence for interaction with Pseudomonas syringae protein HopF2, which can inhibit pathogen associated molecular pattern (PAMP)-triggered immunity [222]. PYL7, also known as RCAR2, involved in ABA signaling is highly expressed in guard cells and

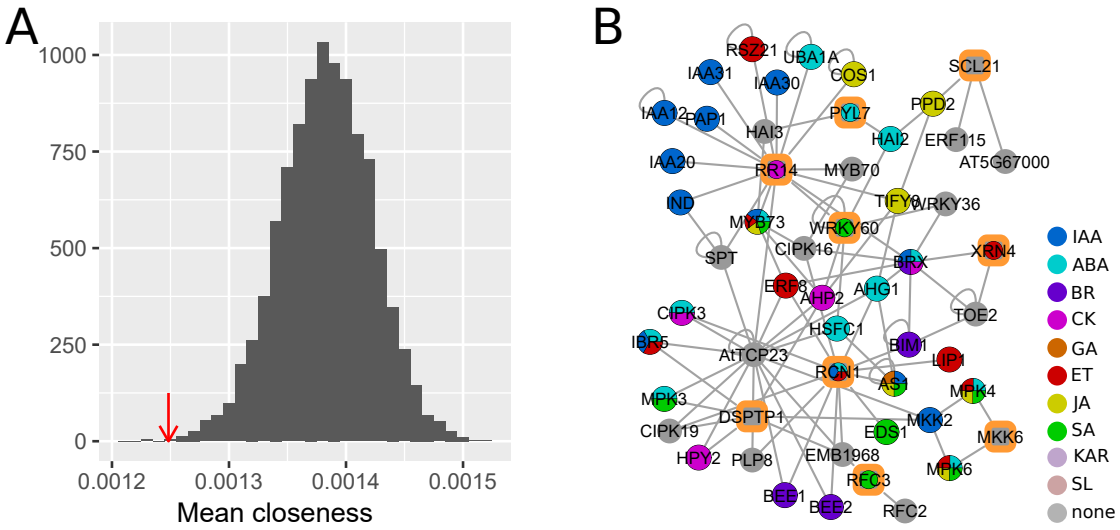


Figure 2.22: A) Distribution of Closeness mean values of genes with significant Tajima's D value in 10,000 degree preserved randomized network. Red arrow shows closeness mean value in real network. B) The nine deviant genes are marked with an orange border. Network shown is a subnetwork of PhI, including deviant genes and their first neighbors.

seedlings together with other PYR-PYL/RCAR members [223]. It is involved in closing of guard cells on drought sensing and PYL7 overexpressing plants are more resistant to drought [224]. Compared to other members of PYR-PYL/RCAR family, PYL7 displays an opposite gene expression profile with increased expression in response to ABA, salt, osmotic, or drought stress [225]. DSPTP1, a dual-specificity phosphatase was shown to dephosphorylate MPK4 [226] and interacts in the PhI network also with MPK3 and MKK2. DSPTP1 acts as regulator of ABA accumulation and is a negative regulator in osmotic stress signaling during seed germination and seedling establishment [227]. RR14 is one of 11 members of the type-B ARR family of which seven are suggested to be associated with CK signal transduction. The specific role of RR14 could not be elucidated so far, but complex interactions with other hormones including AUX, ET and GA are assumed [228]. RR14 is the only type B family member repressed by *Pseudomonas syringae pv. tomato DC3000*. Repression appears to be dependent on type III secretion system effectors [229].

RCN1, is a protein phosphatase 2A subunit and functions as a positive regulator of the PP2A holoenzyme. RCN1 is involved in root gravitropism and basipetal AUX transport in root as antagonist of PID [230] and in regulating postembryonic root development and stress response [231]. It acts as a link between glucose and the BR-signaling pathway [232], is involved in phototropism by dephosphorylating PKS4 [233], regulates BR signaling by dephosphorylation of BRI1 [234], and functions in methyl jasmonate signaling and signal crosstalk between methyl jasmonate and ABA [235]. rcn1-1 mutants are impaired in stomatal closure [235] and in dephosphorylation of Phot2, which results in enhanced blue light response [236].

XRN4 (exoribonuclease 4) decays specific mRNAs of proteins involved in nucleic acid binding, DNA or RNA binding, and nuclear-encoded chloroplast-targeted (n-chlor) proteins. Moreover, it has an influence on expression of proteins involved in response to abiotic or biotic stimulus [237]. Response to biotic stress was shown in xrn4 mutant

lines, which showed decreased susceptibility against Turnip mosaic virus [238]. The response to abiotic stress was shown in heat shock experiments: xrn4 mutants exposed to short-term severe heat stress showed increased survival rate compared to wild-type plants [239], whereas xrn4 null mutants exposed to moderately high temperature over a long period were almost unable to survive [239]. This could be explained by a reduction in the degradation of heat shock factor A2 (HSFA2) and ethylene response factor 1 (ERF1) mRNA in xrn4 mutants [239].

RFC3 (replication factor C subunit 3) is involved in negative regulation of systemic acquired resistance in Arabidopsis. The knock-out mutant plant rfc3-1 is hypersensitive to SA and more resistant to Hpa Noco2 compared to wild-type plants [240, 241].

SCL21 (SCARECROW-like 21) is a transcription factor of the GRAS protein family, whose members play diverse roles in plant development [242]. Additionally SCL21 is together with PAT1 a positive regulator of phytochrome A signal transduction for high-irradiance responses [243] and target for nematode secretory peptides, which bind to SCL21 and SCL6 and stimulates root growth [244].

WRKY60 (WRKY DNA-binding protein 60) and the two structurally related proteins WRKY18 and WRKY40 form homocomplexes and heterocomplexes. Triple and double mutants of theses WRKY proteins were substantially more resistant to *P. syringae* but more susceptible to *B. cinerea* than wild-type plants. On the other hand single mutant plants of WRKY60 were more susceptible to both *P. syringae* and *B. cinerea* [245]. WRKY18 and WRKY60 have a positive effect on plant ABA sensitivity and can inhibit seed germination and root growth. Both proteins also enhance sensitivity to salt and osmotic stress [246] and were found to interact with the promoters of ABI4 and ABI5 [247].

Both, the four genes with a positive Tajima's D and the five genes with a negative Tajima's D have been found to be involved in important developmental processes of the plants and most interestingly in response to biotic and abiotic stress responses. A link between natural variation with increased variability in genes associated with adaption to biotic and abiotic stresses has been already been found in plants [58, 248, 249] and also in human [221].

Enhanced disease susceptibility or enhanced disease resistance phenotypes are mostly identified in knock-out lines, which lead to severe phenotypes. These phenotypes are normally not observed in nature. Natural occurring SNPs can lead to more subtle phenotypes, which nevertheless can increase fitness without severe side effects. The variations in the nine deviating genes lead probably to a changed functionality of the genes, which increases resistance to various stresses . Many genes are involved in response to biotic and abiotic stresses, but the identified nine genes are likely the points in the protein-protein interaction network, where a change of phytohormone signal transduction might take place in order to adapt to different environmental challenges.

3 Results
Host-pathogen networks

Plants are threatened by various abiotic and biotic stresses. One of these biotic stresses are microbial pathogens, from which some are able to inject bacterial virulence factors, termed effector proteins, into the host cell. These effector proteins are able to influence the host's immune system with the aim to suppress defense responses. On the other hand, if effector proteins or evolutionary conserved pathogen (or microbial) associated patterns are recognized by host proteins, immune reactions are activated. [250, 136]

To elucidate the binding of effector proteins to host proteins and identify the manipulated host processes, a systematic Y2H screen of effector proteins from the oomycete *Hyaloperonospora arabidopsidis* and the bacteria *Pseudomonas syringae* against a collection of immune system proteins and about 8,000 other proteins from *Arabidopsis thaliana* was conducted 2011 by Mukthar et al. [139]. The analysis of the resulting protein-protein interaction maps revealed that both pathogens interact with a limited set of proteins. It could be shown that there is a significant overlap in targeted host proteins despite the large evolutionary distance of the examined pathogens [139].

To further examine host-pathogen protein-protein interactions, effector proteins of a third evolutionarily distant pathogen, the fungi *Golovinomyces orontii*, were screened for interactions with *Arabidopsis* proteins. Results were integrated with interactions from *Hyaloperonospora arabidopsidis* and *Pseudomonas syringae*. Target proteins of all three pathogens were examined for disease phenotypes in infections assays in planta. Results have been published in [142]. Here I present bioinformatic analyses of protein-protein interactions and of phenotyping experiments.

3.1 Effector-host interaction map

For the Y2H screen, 84 *G. orontii* effector candidates (OECs) were successfully cloned, whereby 15 out of 84 OECs showed auto-activation and were excluded from Y2H screen. The remaining 69 OECs were screened against the ORFeome collection AtORFeome2,

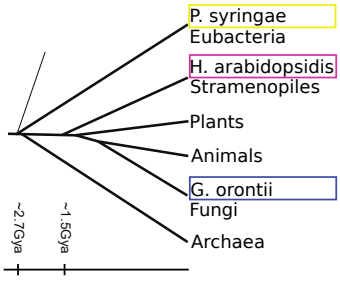


Figure 3.1: Evolutionary distance of *G. orontii* to *P. syringa* is about 2.7 billion years and to *H. arabidopsidis* about 1.5 billion years.

Table 3.1: Number of effector proteins (Effectors), interacting host proteins (Host interactors), interacting host proteins in AI-1_{MAIN} and number of interactions between effector proteins and host proteins for pathogens examined in PPIN1 [139], *Pseudomonas syringae* (Psy) and *Hyaloperonospora arabidopsidis* (Hpa) and pathogen analyzed in this project, *Golovinomyces orontii* (Gor). Results are shown for different sets of host search spaces: host proteins in AtORFeome1 (8k space), host proteins in AtORFeome2 (12k space), and host proteins in AtORFeome1 and immune proteins (PPIN1).

		Psy	Hpa	Gor	Sum
8k Space	Effectors	30	52	41	123
	Host interactors	56	118	45	178
	Host interactors AI-1_{MAIN}	47	107	42	155
	Effector - host protein interactions	99	230	93	421
12k Space	Effectors	-	-	46	46
	Host interactors	-	-	60	60
	Effector - host protein interactions	-	-	122	122
PPIN1	Effectors	30	53	-	83
	Host interactors	61	122	-	165
	Effector - host protein interactions	106	234	-	340

which contains about 12,000 ORFs from *Arabidopsis thaliana*. For interaction mapping, a Y2H pipeline with a very low false discovery rate [12, 145, 16] was used. This pipeline was also used previously for mapping 8,000 *Arabidopsis thaliana* ORFs from clone collection AtORFeome1 resulting in the Arabidopsis Interactome 1 (AI-1) [11] and for screening effector proteins of *Hyaloperonospora arabidopsidis* and *Pseudomonas syringae* against the ORF collection AtORFeome1 and a set of immune proteins and protein fragments. These interactions form the Plant-Pathogen Immune Network 1 (PPIN1) [139]. In the Y2H screen of OECs against AtORFeome2, 122 interactions between 46 effector proteins and 60 host proteins were identified. In this network map, an effector protein interacts on average with 2.3 host proteins and 16 host proteins interact with multiple effector proteins.

The questions were, which proteins and which biological processes are targeted by the effector proteins of the fungus and moreover which of the targets overlap with the targets of the previously analyzed pathogens *Hyaloperonospora arabidopsidis* and *Pseudomonas syringae*. This means that both, the convergence of all pathogens on host proteins and of each pathogen separately had to be analyzed to identify overlapping host targets. To answer, which processes are targeted, a GO analysis of the respective host targets is required. Moreover, the questions arose, which of the targets are relevant in immunity of the plant against the respective pathogen and if effector binding left population genetic signatures in the host proteins, which show evidence for selection. To determine the relevance of proteins for immunity, the results of phenotyping experiments of T-DNA lines were analyzed for their relevance in disease susceptibility and resistance. The natural variation of proteins in *Arabidopsis* ecotypes was used to examine, how binding of host proteins by effector proteins could be prevented.

To systematically analyze, which host proteins are targeted by *G. orontii* and if its effector proteins also converge on common host proteins, an integrated network of effector host proteins of all three pathogens with identical parameters in terms of host search

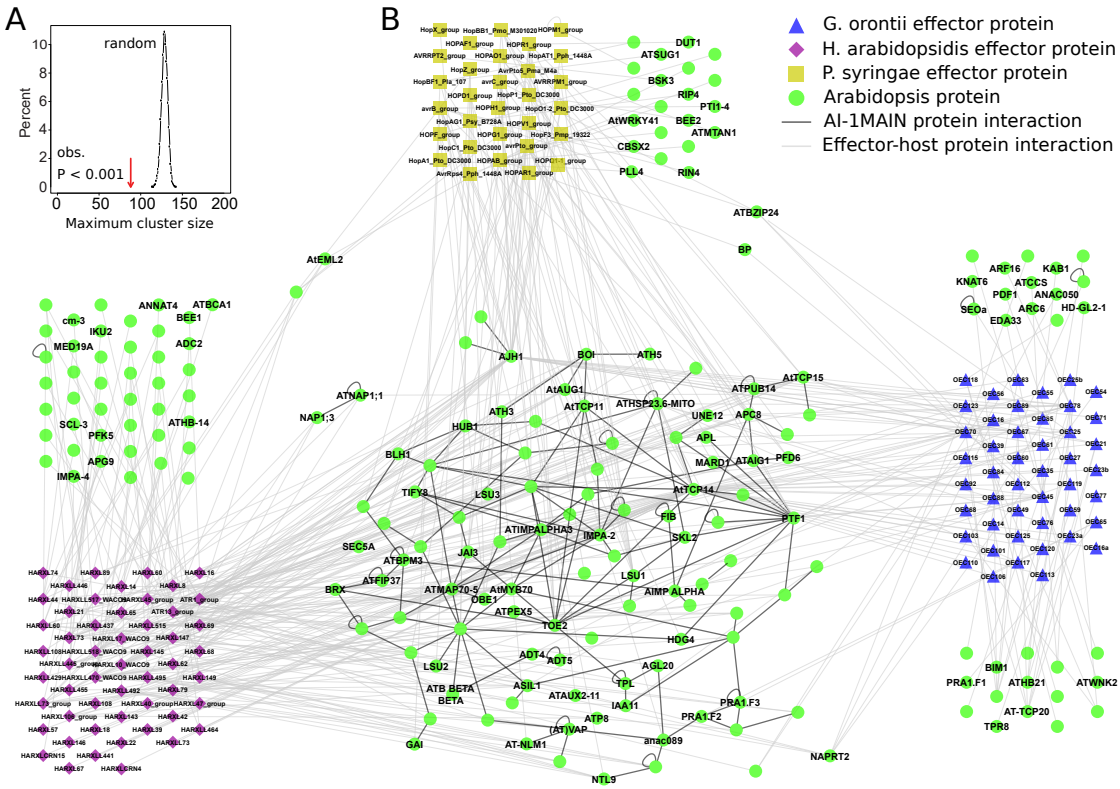


Figure 3.2: A) Size distribution of largest connected subgraph in 10,000 degree preserved randomly rewired AI-1_{MAIN} networks compared to observed size (red arrow) shows that host proteins are less connected than expected by chance. B) Merged effector - host protein interactions from *G. orontii* effectors (blue), *H. arabidopsidis* effectors (violet), and *P. syringae* (yellow). Effector - host interactions of *H. arabidopsidis* and *P. syringae* are from [139]. Interactions between host proteins are from AI-1_{MAIN} [11]. Network visualization using Cytoscape [37].

space had to be generated. Therefore only interactions between effectors and host proteins, which are in AtORFeome1 (8k space) were included for systematic analysis (table 3.1, 8k Space). Additionally, direct interactions between host proteins were added from the systematic network AI-1_{MAIN} to form the network Plant-Pathogen Immune Network 2 (PPIN$2_{8\mathrm{k_sys}}$). This network consists of 421 effector - host protein interactions between 123 effector proteins and 178 host proteins and 162 interactions between host proteins. For all subsequent statistical analyses, this network was used.

Including interactions from AI-1_{MAIN}, a subgraph consisting of 88 directly connected host proteins was identified. A comparison against 10,000 degree preserved randomized AI-1_{MAIN} networks show that the host proteins are less connected than expected chance (fig. 3.2 A). This could indicate that effector proteins do not target a specific part of the network, but attack different network areas with different functions.

3.2 GO analysis of host proteins

Effector proteins target a range of host proteins in order to suppress immune reaction of the plant and to ensure appropriate conditions. To answer the questions, which biological processes in the host are targeted by all and specific pathogens, respectively, GO terms

of host targets were analyzed for enrichment relative to all genes in AtORFeome1 for each pathogen separately and for all pathogens together.

The most enriched terms for both, host targets of each pathogen separately and host targets of all pathogens combined, were "regulation of defense response" and "regulation of response to stimulus" (fig. 3.3). Host targets of all pathogens shared also general GO terms related to regulation of several biological processes (fig. 3.3 A). Where the host targets of both, *H. arabidopsidis* and *G. orontii*, were enriched in "response to auxin stimulus" (fig. 3.3 B, D), host targets of *P. syringae* were enriched in "salicylic acid mediated signaling pathway" (fig. 3.3 C). SA and AUX are important for susceptibility and resistance to pathogens [111, 251]. Host proteins of the fungi *G. orontii* were additionally enriched in "copper ion transport" (fig. 3.3 D). This is of interest, as copper might be involved in regulation of defense response [252] and proteins involved copper translocation are also involved in pathogen defense [253]. All GO enrichment results can be found in table S3 in [142].

GO analysis showed that all pathogens have in common host targets related to response to stimulus and defense response, but target different phytohormone signaling pathways. *H. arabidopsidis* and *G. orontii* target the AUX signaling pathway and *P. syringae* targets the SA signaling pathway. Only the fungus *G. orontii* targets an additional assumed defense mechanism of plants, the copper ion transport.

3.3 Convergence of effector proteins on common host proteins

In the protein interaction network of *H. arabidopsidis* and *P. syringae* a convergence of effector proteins of both pathogens on common host proteins was observed. From the integration of effector - host interactions of three pathogens, the question arose, if there is a convergence of two pathogens and of all three pathogens on common host proteins. Moreover, I wanted to know, if there is also a convergence of effector proteins of a single pathogen on host proteins (intraspecies convergence).

3.3.1 Interspecies convergence

In PPIN1, a convergence of the effectors of both pathogens, Hpa and Psy, on a common set of host proteins was observed (interspecies convergence) [139]. The analysis of the fungal pathogen Gor allows to examine, if an interspecies convergence exists despite the larger evolutionary distance. In PPIN2, there are 24 host proteins on which two pathogens each converge and 9 host proteins on which all three pathogens converge (fig. 3.4 E). The observed overlap between Gor and Hpa is larger than the overlap between Hpa and Psy, which fits to their respective evolutionary distances (compare fig. 3.1). The interspecies convergence was also compared against simulations with a degree preserved randomized network for all pairwise pathogen combinations and all three pathogens. For all pathogen combinations a significantly higher number of proteins are targeted by two or more pathogens compared to random simulations (fig. 3.4 F-I). This shows that all three pathogens bind a specific set of proteins.

3.3.2 Intraspecies convergence

To analyze intraspecies convergence, I evaluated, if the Gor effector protein convergence is significant. Therefore, the number of targeted host proteins was compared against a

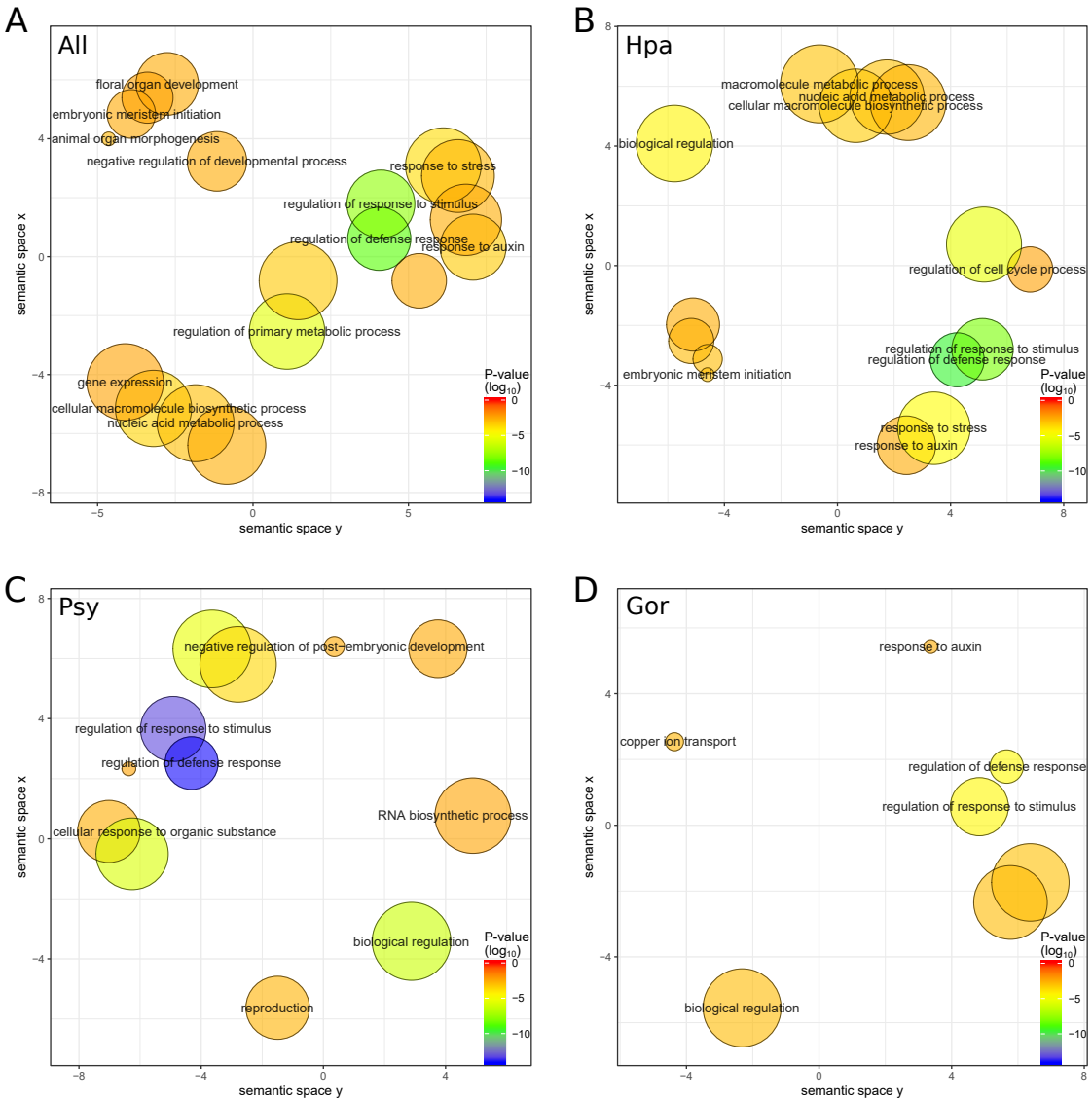


Figure 3.3: Enriched gene ontology terms in A) all host targets, B) host targets of *H. arabidopsidis* (Hpa), C) *P. syringae* (Psy) and D) *G. orontii* (Gor) effector proteins. GO terms are grouped by semantic similarity. Size of circles show number of proteins annotated with the respective GO term.

simulation of Gor effector proteins randomly interacting with AI-1_{MAIN} proteins. The distribution of the number of host proteins derived from 10,000 simulations was compared against the real number of host proteins, which are targeted by Gor effector proteins. In case of random binding, the real number of host proteins is expected to be inside the distribution derived from simulations; in case of convergent binding the real number of host proteins should be significantly lower than observed in simulations (fig. 3.4 A). The simulation shows that in random expectation more than 80 host proteins are targeted by Gor effector proteins, which is significantly higher than the experimentally observed 45 host proteins (fig. 3.4 B). This analysis was also conducted for effectors of Hpa and Psy and for both pathogens a significant intraspecies converge was found, too (fig. 3.4 C, D).

3.3.3 Conclusion

The enumeration of effector proteins binding to a host protein shows that all proteins, which are bound by three pathogens are also subject of intraspecies convergence of one pathogen and 16 out of 23 proteins bound by effectors of two pathogens are also subject of intraspecies convergence of one of the pathogens (fig. A.25 B). This shows the importance of these host proteins for infection by the examined pathogens.

Both, intra- and interspecies convergence was observed in PPIN2, which gives rise for the hypothesis, that all three pathogens attack a core functionality of the plant, which is important for the pathogens' successful infection. Additionally each pathogen specifically attacked proteins involved in biological processes, which support infection by the single pathogen. The strong and highly significant convergence suggests that there is acting natural selection for influencing specific biological processes of the host organism.

3.4 Analysis of genetic validation of effector targets

A Y2H screen detects protein-protein interactions without the context in which an interactions takes place. The analyzed effector - host protein interactions are assumed to take place either to influence the host's immune system and facilitate infection or the plant proteins bind to effectors in order to avoid infection. But also other reasons for interaction are seem reasonable, like influencing biological processes to adjust environment for the pathogen.

3.4.1 Infection phenotypes

To evaluate which host proteins, identified in the Y2H assay, also show genetic support to be involved in immunity or play a role while infection by a pathogen, infection assays with *Arabidopsis* mutant lines were performed. Overall, 124 effector interacting host proteins were tested in available T-DNA insertion lines for altered infection phenotypes [254]. In total, 179 T-DNA lines were confirmed for homozygosity and T-DNA insertion into the gene of interest. These validated lines were phenotyped in infection assays using *G. orontii* isolate MPIPZ [255] (virulent on Col-0 [256, 257]), *P. syringae* strain *P. syringae pv. tomato* DC3000 and three *H. arabidopsidis* isolates: Emwa1 and Emoy2 (avirulent on Col-0) and Noco2 (virulent on Col-0 [258]). These three Hpa isolates allowed to detect both enhanced disease susceptibility (eds) and enhanced disease resistance (edr) phenotypes. Analysis of T-DNA lines and phenotyping assays were conducted by Petra Epple, Kristin Wiley, Nathan McDonald, M. Shahid Mukhtar, Jeffery L. Dangl, Ralf Weßling, Sabine Haigis, Paul Schulze-Lefert and Ralph Panstruga.

Infestation of T-DNA lines with each inoculated pathogen was compared against infestation of the respective Col-0 plants. Log_2 fold changes were calculated to determine the phenotype (edr or eds); p-values were calculated to determine if the infestation is significantly lower of higher in the T-DNA line compared to Col-0. A significant different infestation compared to Col-0 could be observed in 63 out of 124 tested effector interactors, so 51 % of host proteins could be genetically validated (fig. 3.5 A, B, fig. A.25, table S4 in [142]). These 63 proteins will be referred as effector targets. In T-DNA lines of 25 effector targets, an enhanced disease susceptibility was observed compared to Col-0. This implies a role of these effector targets in immune defense of the host. T-DNA lines for 21 out of the 63 effector targets exclusively showed an enhanced dis-

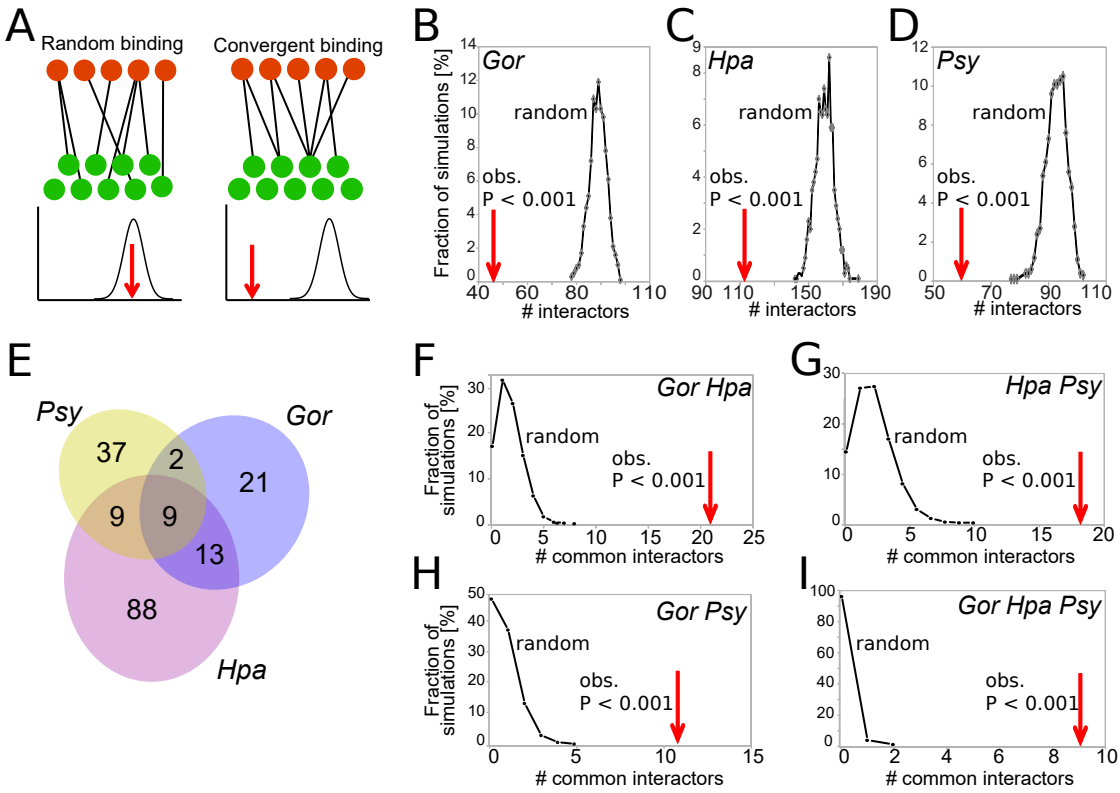


Figure 3.4: Inter- and intraspecies converge of effector proteins. A) Schematic illustration of intraspecies convergence. Left panel: If effector proteins (red) randomly bind to host proteins (green), observed number of targeted host proteins (red arrow) should be inside the distribution of the randomly expected number of host proteins. In case of convergence to a specific set of host proteins (right panel), the real number of host proteins should be smaller than the distribution of the number of randomly expected host proteins. B) Distribution of the number of randomly expected host proteins interacting with *Gor* effector proteins and the observed number of interacting host proteins (red arrow). C) As in B, but for *Hpa* effector proteins. D) As in B, but for *Psy* effector proteins. E) Venn diagram, which show the number of host proteins interacting with effector proteins from one, two, or all three pathogens. F-I) Interspecies convergence of effector proteins on common host proteins. F) Interspecies convergence of host proteins from *Gor* and *Hpa*. Distribution shows number of common host proteins by random expectation; red arrows shows real number of common host proteins. G) As in F, but for *Hpa* and *Psy* effector proteins. H) As in F, but for *Gor* and *Psy* effector proteins. I) As in F, but for *Gor*, *Hpa* and *Psy* effector proteins.

ease resistance compared to Col-0. This is mainly observed in infection assays with the virulant *Hpa* isolate Noco2. A possible reason for edr could be a role of effector targets either in preparation of proper growth conditions for the pathogens or immune response signaling, where the pathogen fails to suppress the plant's immune system. For 17 effector targets, both phenotypes were observed for the same effector target at treatment with different pathogens or different *Hpa* isolates. Noticeable are here three TCP transcription factors, TCP13, TCP14, and TCP19, which show an eds phenotype with the biotrophic pathgens *Hpa* and *Gor*, and an edr phenotype with the hemibiotrophic *Psy*.

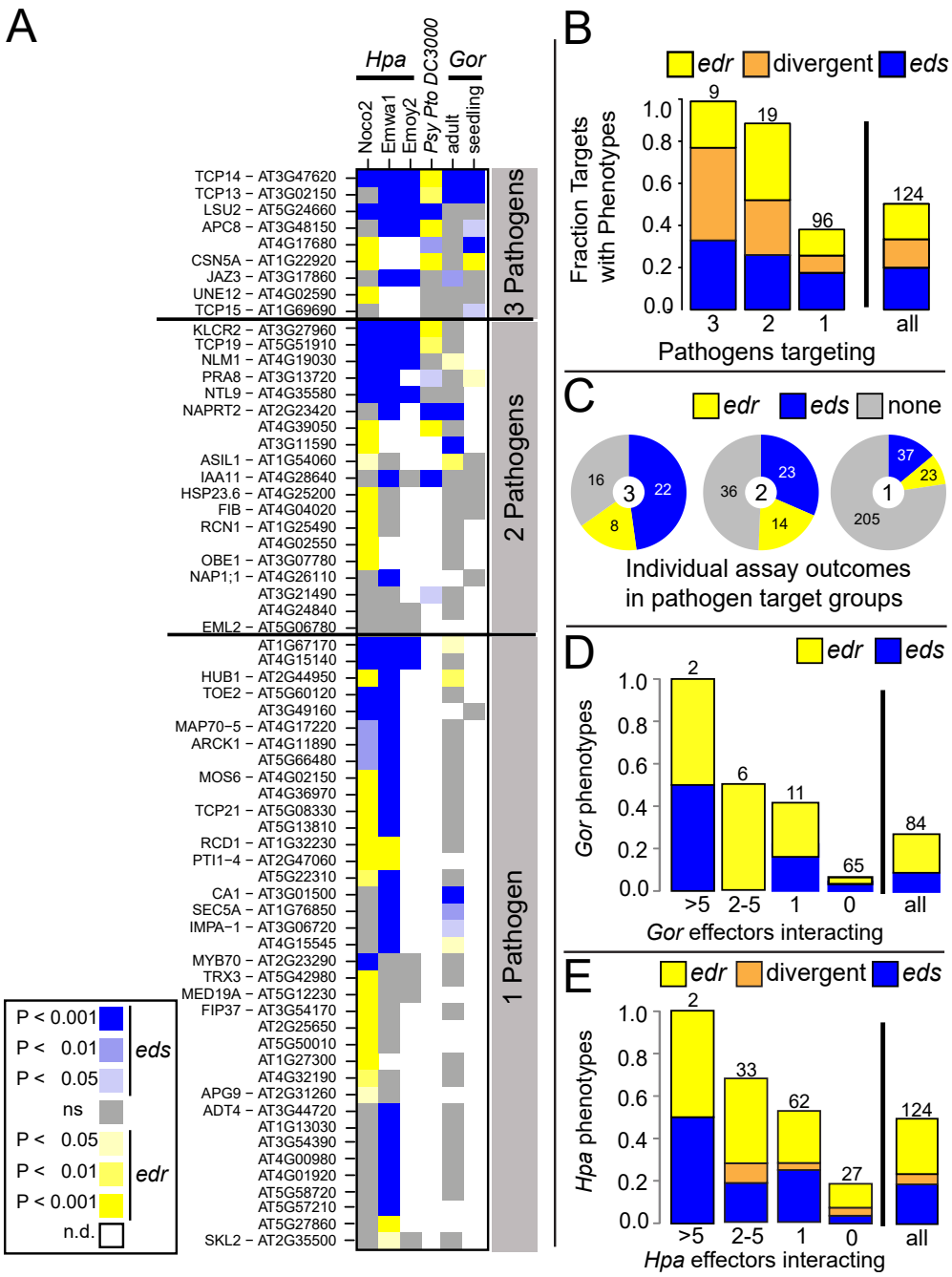


Figure 3.5: Phenotyping results of effector targets in T-DNA lines. A) Significant phenotyping results of T-DNA lines of 63 effector targets under treatment with the respective pathogen / growth stage. Results are grouped by number of pathogens interacting with the respective protein in the Y2H assay and sorted by number of phenotypes. B) Percent of T-DNA lines, which show either an edr, eds, or both phenotypes for effector targets interacting with the respective number of pathogens in Y2H assay. "all" show results without regard of number of interacting pathogens. Number above the bar is number of proteins in this group in Y2H assay. C) For each effector group (targeted by 3, 2, or 1 pathogen), the number of assay outcomes (edr, eds, no phenotype) are shown. D) Fraction of edr and eds phenotypes for effector target interacting with >5, 2-5, 1, or 0 *Gor* effector proteins compared to all tested T-DNA lines. Number at the top of the bar show number of proteins in the respective group. E) As in D, but for effector targets of the pathogen *Hpa*.

3.4.2 Correlation of effector convergence and phenotype density

In the systematic $PPIN2_{8k\_sys}$ a non-random intra- and interspecies convergence on host proteins was observed. This suggests, that these proteins are biologically relevant for pathogen infection. I hypothesized that this relevance might be reflected in the phenotypic characterization of the respective T-DNA lines, which were genetically tested in infection assays.

To test this hypothesis, the analyzed T-DNA lines were grouped by the number of pathogens targeting the respective protein in $PPIN2_{8k\_sys}$ and identified the lines' combined phenotype (fig. 3.5 B). In T-DNA lines of proteins targeted by 3 pathogens, all lines showed a phenotype. The number of observed phenotypes is decreasing with the number of targeting pathogens. This shows the positive correlation between number of observed phenotypes and number of targeting pathogens. To exclude that this correlation results from a deeper phenotypic interrogation in lines targeted by more pathogens, the number of observed phenotypes per number of assays was examined (fig. 3.5 C). Therefore for the fraction of eds and edr phenotypes in all assays for proteins interacting with three, two or one pathogen was analyzed. This analysis shows that the fraction of observed phenotypes is increasing with the increasing number of targeting pathogens (fig. 3.5 C) and does not depend on the number of assays.

To analyze, if this correlation also exists for proteins, which are subject for intraspecies convergence, assays were grouped for proteins targeted by >5, 2-5, 1 or 0 pathogen targets by effector proteins of *Gor* and *Hpa* (fig. 3.5 D, E), respectively. In both cases the fraction of proteins, for which an edr or eds phenotype could be observed was higher for proteins with a higher number of targeting effector proteins.

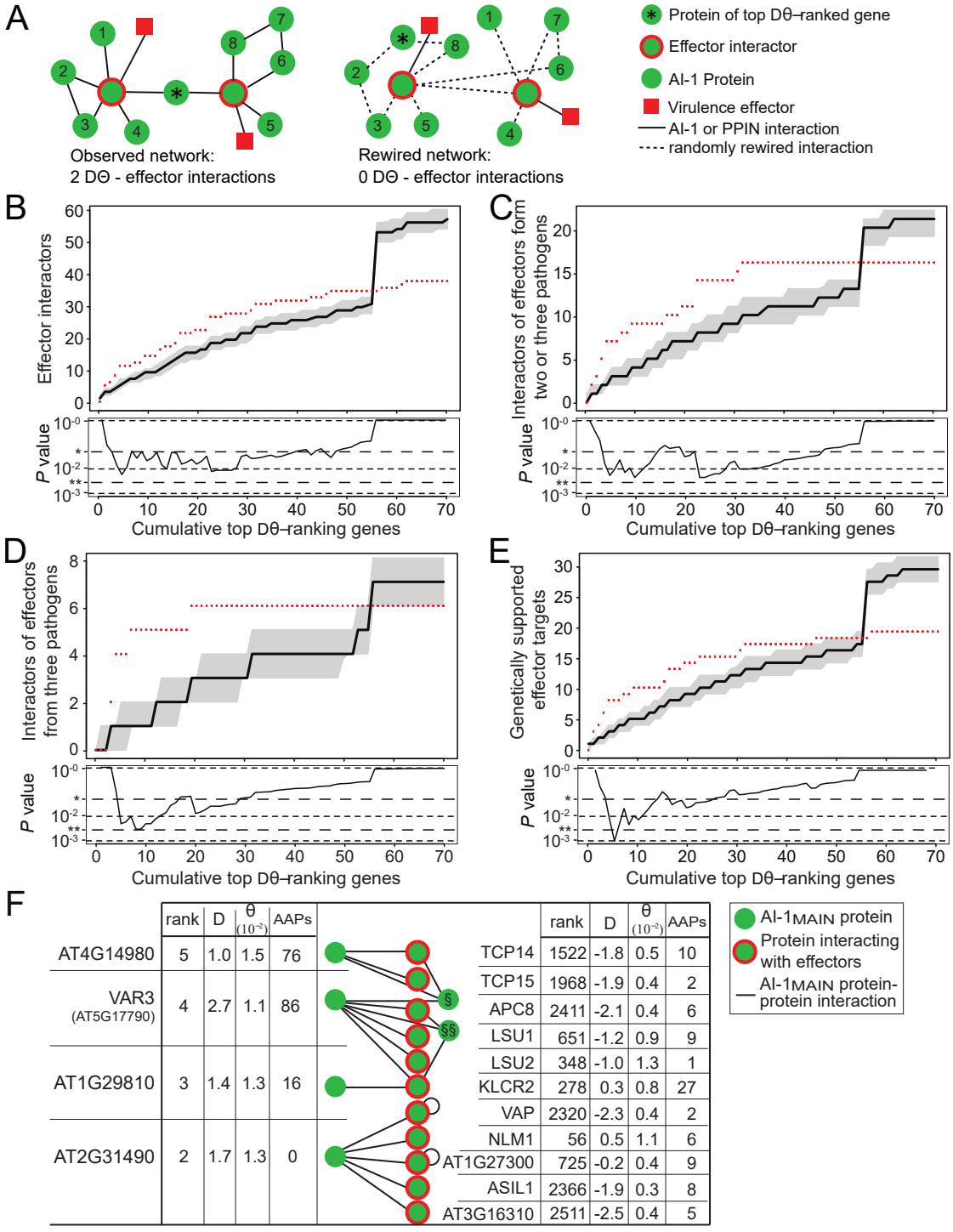
3.5 Analysis of natural variation in host network

One possibility for effector targets to avoid interaction with pathogen effector could be to change amino acid sequence at the binding site. These mutations can be detected by testing genes for deviation from neutral selection. The analysis of natural variation was supported by Stefan R. Henz, Kai Christian Bader, Klaus F. X. Mayer, Detlew Weigel, Jeffery L. Dangl and Pascal Braun.

For analysis of natural variation, the genomes of 80 natural accessions of *Arabidospsis thaliana* were used. These 80 accessions were collected in eight regions over Europe and Asia, where they adapted to diverse environmental conditions. These accessions were sequenced as part of the 1001 genomes project[1] and mapped to the reference genome of the accession Col-0 [62]. For genomic sequences of all genes in $AI\text{-}1_{MAIN}$ two measures were calculated to assess deviations from neutral selection: Tajima's D (D_T) and Watterson's Θ (Θ_W) to determine allele frequencies deviation and scaled mutation rate, respectively [211, 259]. As both measures assess deviation from neutral selection based on different sequence properties, gene ranking in both measures deviate from each other. Therefore we ranked genes based on their relative position in the ranked list of each measure (DΘ-ranking) (see table S7 in [142]).

To identify variation in the protein sequences of the 80 natural accessions plus Col-0, a consensus codon sequence of coding sequences of the 81 accessions was constructed. The consensus codon sequence was used to determine the consensus protein sequences and to identify non-synonymous single nucleotide polymorphisms (nsSNPs). The protein

[1] `http://1001genomes.org/`



	rank	D	θ (10^{-2})	AAPs
AT4G14980	5	1.0	1.5	76
VAR3 (AT5G17790)	4	2.7	1.1	86
AT1G29810	3	1.4	1.3	16
AT2G31490	2	1.7	1.3	0

	rank	D	θ (10^{-2})	AAPs
TCP14	1522	-1.8	0.5	10
TCP15	1968	-1.9	0.4	2
APC8	2411	-2.1	0.4	6
LSU1	651	-1.2	0.9	9
LSU2	348	-1.0	1.3	1
KLCR2	278	0.3	0.8	27
VAP	2320	-2.3	0.4	2
NLM1	56	0.5	1.1	6
AT1G27300	725	-0.2	0.4	9
ASIL1	2366	-1.9	0.3	8
AT3G16310	2511	-2.5	0.4	5

Figure 3.6: Proteins with high natural variability interact with effector targets. A) Schematic illustration of the analysis of interactions between variable proteins and effector targets. Left network: Number of effector targets interacting in AI-1_{MAIN} with top DΘ ranking proteins. Right network: number of interactions between top DΘ ranking proteins with effector targets after degree-preserved random rewiring of AI-1_{MAIN}. (Continued on next page)

Figure 3.6: (Continued from previous page.) B) Cumulative number of top ranking DΘ proteins interacting with effector interacting proteins in AI-1_{MAIN} compared to observed number in 1,000 randomly rewired networks. Y-axis shows number of effector interactors interacting with the number of top DΘ ranking genes shown on x-axis. Red dotted line represents real data determined from AI-1_{MAIN}, black line shows mean value from 1,000 degree-preserved randomly rewired AI-1_{MAIN} networks, gray area shows value between 25^{th} and 75^{th} percentile. The lower panel provides the experimentally determined p-value (* 0.05; ** 0.005) for each number of considered top ranking AAP genes. C) As in B, but number of interacting effector proteins, targeted by at least two pathogens are determined per number of top ranking AAP genes. Strong increase of effectors in simulation at x = 56 stems from the high-degree protein NLM1 (AT4G19030). The high number of rewired interactions for this protein increase number of interacting effector targets in all categories. C) As in B, but number of interacting effector proteins, targeted by three pathogens are determined per number of top ranking AAP genes. E) As in B, but number of interacting effector proteins having an immune phenotype in infection assays are determined per number of top ranking AAP genes. F) Subnetwork of highest ranking DΘ proteins - effector target interactions. Among the 13 interaction partners of the five highest ranking DΘ proteins are 11 effector targets, including the five most targeted proteins. Tables show combined rank, D_T, Θ_W and amino acid polymorphisms (AAP). Non-effector targets: AT1G51580 (§), ZPF7(§§).

---

sequences of all accessions were compared against the consensus protein sequence to identify altered amino acids in the protein sequences of 2,653 AI-1_{MAIN} proteins (see table S7 in [142]).

The first hypothesis was that direct effector targets show evidence for balancing selection to avoid binding with effector proteins. This is the case, if host proteins have a high positive Tajima's D value. Four groups of host proteins were tested for significant higher Tajima's D value compared to random expectation: all effector interactors, effector interactors with an interspecies convergence of 2 or 3 pathogens, effector interactors with an interspecies convergence of all three pathogens, and effector interactor showing an immune phenotype in infection assays. All four groups showed no deviation from random expectation (fig. A.26 A). A possible explanation for this observation might be the fact, that most of the effector interactors have a high number of interactions in the AI-1_{MAIN} network and therefore a high number of amino acid changes at binding sites could have a severe impact on many interactions with other host proteins. The following hypothesis was, that the effect of pathogen binding is detectable in the network neighborhood of effector binding proteins. Therefore the mean Tajima's D value of all interaction partner of effector binding proteins was compared against expected distribution in randomized network. In this analysis, no deviation from random expectation could be detected for any of the four groups of effector binding proteins (fig. A.26 B).

As many proteins interacting with effector binding protein are not involved in biological processes of immunity or pathogen defense, the analysis was focused on proteins showing a high variation in the DΘ-ranking. We analyzed, if high ranking proteins are preferentially interacting with host proteins which interact with effector proteins. For each cut-off of number of highest ranking genes, the number of effector targets were counted, which interact with these high-ranking genes in AI-1_{MAIN} (fig. 3.6 A, left network). The derived number of interacting effector targets was compared against the number of interacting effector targets in 1,000 degree-preserved networks (fig. 3.6 A, right network). This analysis was conducted for the previously defined groups: all effector interactors, effector interactors with an interspecies convergence of 2 or 3 pathogens,

effector interactors with an interspecies convergence of all three pathogens, and effector interactors showing an immune phenotype in infection assays. The derived values for the real interactions from AI-1_{MAIN} are shown as red dotted line in figures 3.6 B-E. The mean number of expected effector interactors from 1,000 degree-preserved randomized networks is shown as black line, gray area shows values between the 25^{th} and 75^{th} percentile. In the lower panel of figures 3.6 B-E is shown the experimentally derived p-value for each cut-off. Up to a certain cut-off, always a significant p-value and a higher number of interactions with effector interactors than expected by random is observed. A ranking of variable proteins using amino acid polymorphisms (AAP) shows a similar result and supports the finding, that highly variable proteins tend to interact with proteins interacting with effector proteins.

The network of four out of the five highest ranking variable proteins shows interactions with eleven effector interacting proteins and two non-effector interacting proteins (fig. 3.6 F). The effector interacting proteins seem to be in contrast of the high-ranking proteins to be under purifying selection, with low ranks in the DΘ ranking.

4 Discussion

The main focus of this thesis was to analyze systematic protein interaction maps derived from Y2H high-throughput experiments with bioinformatic methods, to gain insights into their structural properties, and correlate them with underlying biology. This was achieved by analyzing network properties of interaction maps and integration of external knowledge to identify and contextualize inherent network structures. In the scope of this thesis, I analyzed the protein interaction maps derived from Y2H screens from two different projects. The systematic analysis of phytohormone signal transduction by protein-protein interactions in *Arabidopsis thaliana* and the analysis of host-pathogen protein interactions of effector proteins of three pathogens interacting with *Arabidopsis thaliana* host proteins.

4.1 Phytohormone signal transduction and signal integration

To elucidate phytohormone signal transduction and signal integration between different signaling pathways by PPIs, a collection of more than 1,200 selected ORFs, the PhyHormORFeome, were analyzed for binary PPIs in a Y2H interaction mapping pipeline. Additionally, the PhyHormORFeome was screened against a the AtORFeome2, a clone collection with more than 12,000 different ORFs from *Arabidopsis thaliana*, to connect phytohormone signaling pathways with other biological processes. In the scope of this thesis, the interaction mapping pipeline was assessed for its quality and the resulting protein interaction map was integrated with additional data and bioinformatically analyzed to gain insights into phytohormone signaling.

4.1.1 Y2H mapping pipeline produced a high quality interaction map

For the analysis of the Y2H interaction mapping pipeline, an empirical quality assessment frame work was used, where three measures of the performance of the Y2H interaction mapping pipeline were determined: completeness, sampling sensitivity, and assay sensitivity. The completeness defines the percentage of interactions, which were screen compared to the initially defined search space SSP_{PHO}. The sampling sensitivity determines the percentage of identified interactions compared to the theoretically possible number of interactions, which can be identified with the used screening pipeline. For the assay sensitivity a set of known and random interactions were tested to determine the percentage of true positive (TP) and false positive (FP) interactions, which are expected.

More than 98 % of the defined search space SSP_{PHO} of phytohormone related proteins could be cloned for the PhyHormORFeome collection and thus a completeness of 94.6 % was achieved. Sampling sensitivity was estimated to 77.6 % from the number of interactions derived in three repeats of the screen. The screen of the PRS and the RRS with the Y2H mapping pipeline showed an assay sensitivity of 20.7 % and no false positive interaction was found. Together, this results in an overall completion of 15.1 % $\pm$ 6.3 %. From

the two screens with this mapping pipeline, PhyHormORFeome against PhyHormORFeome and PhyHormORFeome against AtORFeome2, the two protein interaction maps PhI, consisting of 475 interactions between 251 proteins and PhI_{out} consisting of 1086 interactions between 696 proteins were derived.

The quality assessment of the interaction mapping pipeline revealed a high quality of the pipeline. The assay sensitivity is lower compared to sensitivity observed in screen of AI-1_{MAIN} with about 36.4 % [11], but it is in the range 17 - 20 %, what was observed in [12]. Similar to our pipeline, in [12] also a low false positive rate of < 0.5 % was found. The screen of AI-1_{MAIN} reached a sampling sensitivity of around 36.5 % with two repeats of the screen [11] and for a screen of a human ORFeome, a sampling sensitivity of 45 % - 53 % for one screen depending on the Y2H configuration, was estimated and approximately six screens would be needed for 90 % saturation [12]. Compared to these screen we reached a high sampling sensitivity of 77.6 % with three repeats and around 58 % for one screen. The overall completion for our defined search space was compared to the overlap with literature curated interactions in the same search space derived from IntAct [146] and BioGRID [147]. The observed overlap was 17.9 % and 13.4 %, respectively. This confirms the estimation of 15.1 % of the overall completion.

The quality assessment shows both that the number of identified interactions is in the range, what is expected and the reliability of the interactions is high. Moreover, the systematic Y2H screen yields unbiased interactions and does not suffer from biases as observed in hypothesis driven experiments [17].

4.1.2 The PhyhormInteractome has a hierarchical network structure

To reveal the inherent topology of the PhI, the distribution of the degree and the clustering coefficient was determined.

Both, the degree distribution and the clustering coefficient distribution followed a power law distribution, which let infer a hierarchical topology. This topology was also found in AI-1_{MAIN}, whereas LCI networks from BioGRID and IntAct showed a scale free and an unusual network structure, respectively.

The PhI corresponds to expectation of a biological network, most of which have a hierarchical network structure [7].

This let conclude that the structure of the PhI does better fit to biological networks, than currently available literature curated networks.

4.1.3 Communities in PhyhormInteractome correspond to biological processes

In the hierarchical network topology, groups of nodes with a stronger connection among themselves than to the rest of the network are inherent. These groups, called communities, were identified and tested for their enrichment in the annotation of a certain phytohormone using AHD 2.0 annotations. Additionally, the communities were tested for enrichment of GO annotations and a coarse-grained community network was generated and examined.

In PhI, for seven out of eight phytohormones one enriched community was identified, which are significantly more communities than expected by chance (fig. 2.8). In both literature curated interaction maps from IntAct and BioGRID, for some phytohormones multiple communities were detected (fig. 2.9).

The gene ontology analysis using GO annotation from TAIR10 confirmed the hormone enrichment for six out of seven hormone enriched communities of the preceding analysis. All seven hormone enriched communities are enriched in GO terms which are related to the respective hormone signaling pathways, like response to specific stresses. Other communities were enriched in pathways involved in developmental processes and response to biotic and abiotic stresses.

The coarse-grained visualization of communities as a community network elaborates differences between a systematic and a literature curated protein interaction network. Community networks of both systematic networks show a higher degree of community nodes and are more disassortative compared to literature curated networks.

It has been described, that cellular functions, such as signal transmission, are conducted by interacting molecules, which form communities [160]. Therefore, the identification of hormone enriched communities confirms that the network structure reflects the underlying biology of hormone signal transduction pathways. Moreover, the GO enrichment analysis confirmed hormone enrichment and other functions of hormone signaling pathways, e.g. response to heat, water deprivation and salinity, and stomatal movement in ABA [153, 154], root growth and lateral root formation in AUX pathway [155], or *response to wounding*, *response to fungus*, and *regulation of growth* in JA pathway [156].

The coarse grained community networks were analyzed for their network properties. The community networks reflect the properties of the protein interaction networks. The community network of PhI is disassortative, what is expected for biological networks [151].

Community analysis and subsequent enrichment analysis with two orthogonal datasets, AHD 2.0 and GO, confirmed that the PhI network reflects the underlying biology. The community network visualization showed the strong connectivity of distinct signaling pathways compared to LCI networks.

4.1.4 Hormone pathways are strongly interconnected

The analysis of the network topology, the communities, and the community networks revealed strong differences between the four compared protein interaction networks PhI, AI-1_{MAIN}, $\mathrm{IntAct}_{\mathrm{SS\text{-}BMBS}}$, and $\mathrm{BioGRID}_{\mathrm{SS\text{-}BMBS}}$. To systematically analyze the connections within and between hormone signaling pathways, three properties were examined: number of proteins per phytohormone signaling pathway, the number of protein interactions, and average shortest path length between and within phytohormone signaling pathways.

All four networks have a comparable number of proteins in the respective hormone signaling pathways, which avoids a bias in comparing number of interactions and shortest path lengths between the networks. Both systematic networks (PhI and AI-1_{MAIN}) have around 4.7 (PhI) and 5.6 (AI-1_{MAIN}) times more interactions within a hormone signaling pathway than between hormone signaling pathways, whereas both literature curated networks have 7.7 (BioGRID) and 16.3 (IntAct) times more interactions within pathways than between pathways. The main difference between the networks becomes visible on closer inspection of the connectivity in terms of shortest path lengths. While PhI has a range of mean shortest path lengths of 2.2 - 4.2 and AI-1_{MAIN} a range of 4 - 5.8 within and between phytohormone signaling pathways, the range is much wider in the literature curated networks: IntAct has a range of 1.5 - 10.2 and BioGRID has a range of 2.9 - 6.6.

Literature curated networks are in line with literature, that signal integration does mainly not take place by direct contacts of protein-protein interactions [123]. In the PhI network, on the other hand, we find a lot of direct interactions of proteins from distinct hormone signaling pathways and an overall short distance between pathways.

These results show that phytohormone signaling pathways are much stronger connected than expected by analysis of literature curated interactions (fig. 2.11) and described examples in literature.

4.1.5 Hormone annotations can be inferred by interaction similarity

The testing of proteins for interactions can be performed in Y2H in high-throughput in a reasonable time, but the involvement of a protein in a specific phytohormone signaling pathway can be proven mainly by experiments with plants. This means a much higher effort and can be scaled to high-throughput with difficulties only. In this case, hypotheses from predictions helps to narrow down the number of experiments, which have to be conducted.

Calculating similarities of interaction profiles of proteins resulted in an association network, where two proteins are connected by an edge, if they have similar interactions. The association network derived from the PhI network shows a high number of interactions between proteins involved in the same phytohormone signaling pathway (fig. 2.12), which reflects protein interactions in strongly interconnected communities. Besides these interactions, 164 interactions with partially overlapping and 298 interactions with different or no annotations have been identified. This is an immense reduction of tests compared to theoretically 251 * 10 tests for a complete characterization of all 251 proteins in PhI in all ten examined phytohormones, whereby different experimental methods / parameters for one phytohormone are not yet taken into account.

In the association network, JAZ1 (AT1G19180) and MYC2 (AT1G32640) are found to have a similar interaction profile like ARR7 (AT1G19050). ARR7 is annotated with CK, whereas JAZ1 and MYC2 are not. Within the scope of this project we aimed to validate the two proteins JAZ1 and MYC2 for their additional role in CK signaling pathway. This was tested in planta using knock-out lines of JAZ1 and MYC2 under treatment with BA. In this experiment a lower anthocyanin content compared to wildtype plants is expected. In both knock-out lines under treatment with BA, a lower anthocyanin content per g fresh weight compared to wild type was observed. This shows evidence for JAZ1 and MYC2 to be also involved in CK signaling pathway.

Although these are only two examples where a predicted annotation was successfully verified, it shows that predictions and experiments can complement each other.

4.1.6 Phytohormone signaling crosstalks candidates could be validated

The systematic analyses of PhI, network topology, communities, and interconnectedness, revealed both the existence of functional modules with a strong connectivity and interconnectedness and a high number of connections between them. Moreover, the interconnectedness analysis supported the hypothesis that all phytohormone signaling pathways are strongly interconnected and signals can be potentially transmitted via a very low number of protein interactions between distinct signaling pathways. Therefore, systematically all interactions, where a crosstalk could take place were determined in PhI and PhI_{ext}, where a crosstalk is defined as possible signal transduction between two distinct hormone signaling pathways. Additionally, all potential points of crosstalk

were determined, where the crosstalk is mediated by a connector protein, which is itself not involved in a phytohormone signaling pathway. Crosstalks were classified into two categories: category I, if the interacting or connected proteins are involved in distinct signaling pathways and category II, if the interacting or connected proteins have partially overlapping hormone annotations.

In PhI, overall 173 candidate crosstalks of category I and 232 candidate crosstalks of category II were identified, in the extended network PhI_{ext} 217 crosstalks in category I and 295 crosstalks in category II were found. Possible crosstalks, which are mediated by a connector proteins were found for 27 connector proteins in PhI: 509 possible crosstalks in category I and 734 possible crosstalks in category II.

The high number of identified crosstalk candidates is in contrast to the low number of crosstalks by protein-protein interactions described in literature. A selection of crosstalks described in literature are the interaction between the BR regulated BIN2 and ARF2 involved in AUX signaling pathway [121], the interaction between BZR1 and RGA, where both proteins serve as negative regulator of both the BR and GA signaling pathway [127] and an antagonistic crosstalk between the JA and ET signaling pathway through the interaction of MYC2 and EIN3 [167].

In the scope of this project, a small set of from pairwise crosstalk candidates was selected to experimentally validate, if the interaction between two proteins leads to integration of signals from different phytohormone signaling pathways. In case of signal integration both proteins should affect the same biological process and lead in case of gene knock-out plants to similar phenotypes under treatment with the same phytohormone.

From crosstalks candidates identified in PhI, candidates to test for phytohormone signal integration were selected depending on expression of interacting proteins in seedling stage and common expression in same tissue. The second criterion for selection of a crosstalk protein pair was that the insertion in the T-DNA line is placed in the first exon. These two preconditions were met for proteins involved in 17 crosstalk candidates from PhI and for 2 crosstalk candidates from two specific Y2H screens in the search space SSP_{PHO} (fig. 4.1). All knock out lines were tested in the respective hormone assays together with control and wild type lines (table 4.1).

Phenotyping assays have been conducted for following phytohormones: BR, ABA, IAA, GA, JA, SA, ET and CK. In the BR assay two new proteins, DDL and RCN1, were successfully tested for presence of a root elongation phenotype so far only observed in their BR annotated interaction partners.

In the ABA germination assay, WRKY54 and AHP2 showed in one repeat a phenotype compared to wildtype, whereas DDL did not. For a significant result, more repeats will be conducted. BEE1, AS1, EDS1, BEE2, BIM1, JAZ3, and MYB77 are also candidates for ABA signaling pathway and will be tested in a root elongation assay.

Eleven proteins are candidates for an additional role in AUX signaling pathway: BPM3, BEE1, NIA2, COS1, RCAR1, EDS1, AHP2, CBL9, BEE2, BIM1, and PP2CA. So far in one repeat of this assay no phenotype could be demonstrated for these genes, which means additional adjustment of the AUX phenotyping assay is needed.

All four candidates of the GA signaling pathway were tested for hypocotyl elongation under PAC treatment. HUB1 and NIA2 showed a significantly shorter root under treatment compared to wildtype, whereas RCN1 and JAZ3 did not.

The candidate genes (RCN1, DDL, NIA2, AHP2, TT4) and control (COI1) tested in a root elongation assay under treatment with JA did not show a phenotype under the

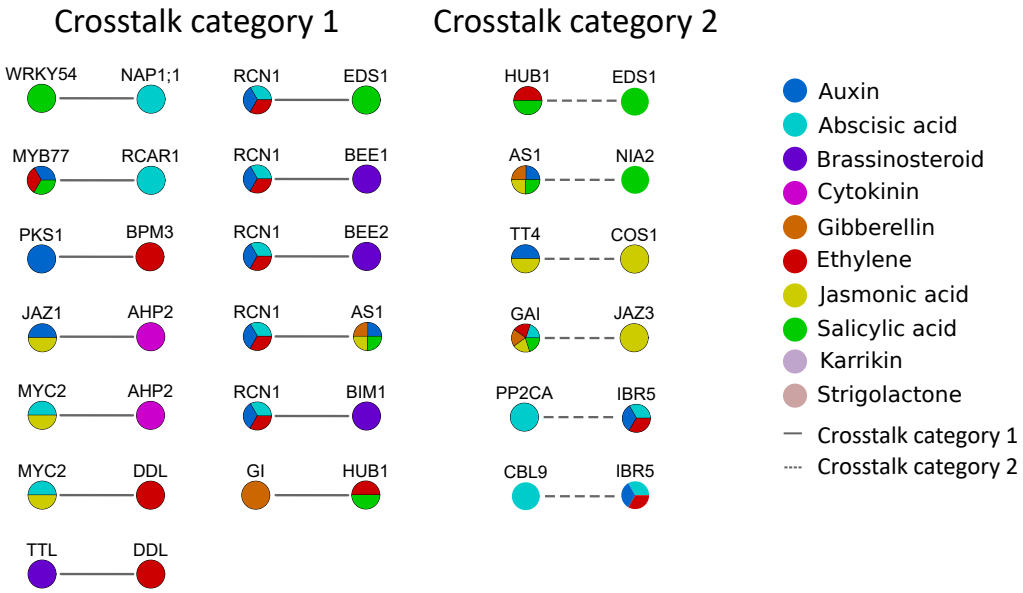


Figure 4.1: Selected Y2H interaction pairs for the seedling assays. Proteins involved in these pairwise interactions were partially tested to be involved in the phytohormone signaling pathway(s) of its interaction partner.

tested concentration. Therefore additional adjustments of the assay are needed for this phytohormone signaling pathway.

SA signaling pathway candidates were tested using an infection assay with *Pseudomonas syringae pv. tomato*. All three candidates tested (HUB1, RCN1, and NIA2) showed in tested knock out lines in at least one repeat a significantly higher susceptibility, which supports their involvement in SA signaling pathway.

Involvement in ET signaling pathway was tested by measuring apical hook elevation, radial hypocotyl swelling and shorter hypocotyl length, also known as triple response [260]. Five (GI, PKS1, EDS1, CBL9, and TTL) out of 10 tested knock out lines showed a significant different response to treatment with 1-aminocyclopropane-carboxylic acid (ACC) in one out of these three phenotypes of the triple response.

Involvement in CK signaling pathway was tested with a root elongation assay under treatment with 6-benzylamino purine (BA) and additional measurement of anthocyanin content. In the first repeat of this assay for both candidates (JAZ1 and MYC2) the expected phenotype was observed.

From all tested knock-out lines for phenotypes under treatment with phytohormones, 14 times the respective phenotype could be observed and 8 times no phenotype was observed. This preliminary result shows that in more than 60 % of tested gene knock-out - phytohormone combinations, a phytohormone induced a phenotype associated with the respective interacting protein. As knocked-out genes were so far annotated to have a role in distinct phytohormone signaling pathways, our phenotyping experiments suggest an influence by additional phytohormones on the currently known hormone regulated biological process.

The identified number of crosstalk candidates imply that crosstalks mediated by protein-protein interactions are more common than expected from literature. Moreover, the high number of verified crosstalks by protein-protein interactions in planta,

Table 4.1: Phenotyping results of T-DNA lines included assay for phytohormone signal integration. Black dots show known annotation for each protein; red dots mark the hormones in which line has to be tested. All lines marked with a "x" are used in a particular assay as positive control, a "+" indicates the wild type line tested in the respective assay. Table reproduced from Melina Altmann's thesis.

Locus ID	Symbol	Annotation	ABA	BR	IAA	GA	JA	SA	ET	CK
AT3G29350	AHP2	CK	●		●		●			●
AT2G37630	AS1	GA,IAA,JA,SA	●		●	●	●	●	●	
AT1G18400	BEE1	BR	●	●	●				●	
AT4G36540	BEE2	BR	●	●	●				●	
AT5G08130	BIM1	BR	●	●	●				●	
AT2G39760	BPM3	ET			●				●	
AT5G47100	CBL9	ABA	●		●				●	
AT2G44050	COS1	JA			●		●			
AT3G20550	DDL	ET	●	●			●		●	
AT3G48090	EDS1	SA	●		●			●	●	
AT1G14920	GAI	ABA,ET,GA,JA,SA	●			●	●	●	●	
AT1G22770	GI	GA				●		●	●	
AT2G44950	HUB1	SA,ET				●		●	●	
AT2G04550	IBR5	ABA,IAA, ET	●		●				●	
AT1G19180	JAZ1	IAA, JA			●		● x			●
AT3G17860	JAZ3	JA	●			●	● x	●		
AT3G50060	MYB77	ET, IAA, SA	●		●			●	●	
AT1G32640	MYC2	ABA, JA	●				●		●	●
AT4G26110	NAP1;1	ABA	●					●		
AT1G37130	NIA2	SA			●	●	●	●		
AT2G02950	PKS1	IAA			●				●	
AT3G11410	PP2CA	ABA			●			●	●	
AT1G01360	RCAR1	ABA	●		●			●		
AT1G25490	RCN1	ABA, ET, IAA	●	●	●	●	●	●	●	
AT5G13930	TT4	IAA,JA			●		●			
AT5G58220	TTL	BR		●					●	
AT2G40750	WRKY54	SA	●					●		
AT4G26080	ABI1	ABA	x							
AT3G24650	ABI3	ABA	x							
AT3G20770	EIN3	ET							x	
AT2G01570	RGA	GA				x				
	RGA/GAI	GA				x				
AT3G11540	SPY	CK								x
Ecotype	Col-0		+	+	+	+	+	+	+	+
Ecotype	Ler-0	0				+				

indicate that crosstalk between hormone signaling pathways by protein-protein interactions is a common mechanism. The identified three node crosstalk candidates also show a large number of possible signal transduction events between hormone signaling pathways, although these candidates remain to be validated in planta.

4.1.7 Condition-dependent crosstalks

Protein-protein interactions are identified in Y2H interaction mapping independently of the conditions under which they occur in the organism. To contextualize interactions, an interaction score from RNA-Seq data was calculated, where the mRNA level serves as a substitute for the protein level. For all interactions in PhI at least one condition were

identified, where the PPI likely takes place. This increases confidence for protein-protein interactions derived from our high-throughput Y2H interaction mapping pipeline.

Interaction scores were additionally used to analyze interactions between the hub protein TCP23 and its interacting proteins. TCP23 interacts with 25 proteins, which are involved in one or more phytohormone signaling pathways. The interaction scores allow to identify conditions, under which TCP23 is likely to interact with two proteins involved in different phytohormone signaling pathways and mediate a crosstalk between them.

Clustering of the interactions between TCP23 and its 25 annotated interaction partners over 157 conditions revealed four groups of interaction partners, which interact with TCP23 under the same conditions (fig. 2.16). Two of the TCP23 interaction partners are the two homologous proteins BEE1 and BEE2 (Brassinosteroid Enhanced Expression). The analysis of conditions, where a protein interact with TCP23 under same condition than BEE1 and BEE2, respectively, elucidates a complex pattern of interactions: Both homologs interact with TCP23 under the same conditions than ARF2, ARR9, GA3OX1, and MKK7, but only BEE1 interacts with TCP23 in the same conditions than AHG1, GA3OX2, ERF12, KNAT1, PYL5, and SUMO3. In other conditions where ERF8, HSFC1, and PKS2 interact with TCP23, either BEE1 or BEE2 also interact with TCP23 (table 2.1).

Initially, both BEE1 and BEE2 were considered as redundant proteins [189] and still common functions are found [190], but later on specific functions for each BEE protein were identified [191, 192, 193, 194].

The analysis supports literature that BEE1 and BEE2 have both common and distinct functionality dependent on the environmental condition. This is in line with the hypothesis, that homologous genes derived by duplication events loose and gain interactions over time, which leads to loss and gain of functions [19, 11].

4.1.8 Signal transduction by kinase - substrate interactions

As protein-protein interaction networks are undirected networks, no direction of the transmitted phytohormone signal can be derived. This means, from an interaction can't be deduced, if a proteins transmits or receives a hormone signal. In case of kinase - substrate interactions the direction is obvious: the kinase transmits a phosphate moiety on its substrate. To exploit this knowledge, all kinases in PhI_{ext} and their interacting proteins were identified and analyzed for a potential kinase - substrate interaction. Therefore, potential phosphorylation motifs per kinase family were inferred from phosphorylated 15mers of substrates. These phosphorylation motifs were tested for fit on kinase interacting proteins in PhI_{ext} and on known substrates.

An analysis of the test on PIN1, known to be phosphorylated by AGCVIII kinases revealed that all four known phosphorylation sites could be identified. In PhI_{ext}, 41 kinases belonging to 13 kinase families were identified. For nine kinase families, phosphorylation motifs could be inferred from phosphorylated 15mers. Motifs were tested in total on 74 kinase interacting proteins. On 27 proteins, one of the motifs of the respective kinase family matched significantly better than on the background consisting of all 27,416 *Arabidopsis thaliana* proteins (table 2.2). The 74 kinase interacting proteins contained 13 already known substrates from five different kinase families. From these 13 substrates, ten substrates from five different families could be reidentified. The 27 potential substrates have 47 interactions with 30 kinases, where a protein can have

both roles, substrate and kinase e.g. in case of MAPKs and MAP2Ks. In the identified potential kinase - substrate interactions a different number of integrated phytohormone pathways were observed: CDK, CDPK, SLK, SnRK2, and SnRK3 signaling pathways integrate hormone signals from up to two different phytohormones. A large number of phytohormone signals are integrated in the MAPK cascade (fig. 2.20).

Kinases are involved in most of the hormone signaling pathways, either as receptors or in transmitting the signal. For example CK is perceived by membrane-associated kinases, which transfer a phosphate to AHP, which phosphorylated in turn ARR proteins and activate expression of CK dependent signaling. [75, 97, 21].

As mainly the core signaling pathways have been elucidate so far, additional kinase - substrate interactions in hormone signal transduction are likely. However, experimental validations of the predicted interactions are needed to verify phosphorylation and identify involved hormone signaling pathway.

4.1.9 Signal integration at transcription factor level

Another potential point of signal integration is regulation of expression by different phytohormone signaling pathways. Each phytohormone signaling pathway starts with a receptor and then the signal is often transmitted to a transcription factor, which regulates the expression of more than 1,000 genes. Therefore literature was searched for proteins known to have a repressing function in any phytohormone signaling pathway. The PhI_{ext} was searched for interactions between 71 identified repressor proteins and 140 transcription factors contained in PhI_{ext}.

Forty two interactions between 22 transcription factors and 23 repressor proteins were found. Three of these transcription factors interact with repressors involved distinct phytohormone signaling pathways: TCP23, WRKY21, and SCL23. Eight interactions between four out of five DELLA proteins with seven transcription factors were found. This is of interest, as each of the four DELLA proteins is involved in five different phytohormone signaling pathways.

The repressor-TF interactions could have a strong influence on the expression of many gens. These interactions could explain the crosstalk of different hormone pathways, which take place on gene regulatory level [123].

Transcription factors are the last layer in a signaling cascade and can activate the expression of thousands of genes. At this layer, the interaction of transcription factors and repressors involved in distinct hormone signaling pathways concertedly control expression of genes. For a comprehensive protein interaction map of transcription factors and repressor proteins, a Y2H screen between 71 repressors and about 2000 transcription factors was executed by Melina Altmann. Based on the result of this Y2H screen, an adapted Y1H screen was used to determine expression regulation of transcription factors by repressor proteins. In a systematic screen conducted by Julius Palme, repressors from six phytohormone signaling pathways were tested against eight transcription factors. It could be shown that the repressor proteins can have an enhancing or repression effect on the transcription.

4.1.10 Natural variation indicates adaptation to environment

Phytohormone signaling is a way how plants incorporate signals derived from the environment and adapt its growth to the habitat. As *Arabidopsis* occurs in Europe, Asia and Northern Africa and thrives in diverse habitats, it is suitable to analyze adaption

to the environment on genomic level. Therefore coding sequences of 1135 ecotypes were analyzed for deviation from neutral selection using different selection measures.

Nine genes in PhI showed deviation from neutral selection in different selection measures. All nine genes are involved in response to different stresses: They are involved in coping with drought stress, salt and osmotic stress or they are involved in pathogen infection. Interestingly the genes under strong balancing or positive selection are less central in the network as expected by chance.

So far, several traits have been identified in *Arabidopsis* which are subject to adaption: It could be shown in swedish line, that that flowering time adapted to temperature [64], and polymorphisms in flowering related genes has been observed earlier [65]. Moreover, genes related in resistance to the pathogen *Hyaloperonospora arabidopsidis* show natural variation [66] and in genes involved in salt tolerance an excess of polymorphisms could be identified [67]. Genes which are subject to non-neutral selection are in both literature and in PhI involved in response to biotic and abiotic stresses. This makes is likely that the genes identified in PhI are under selection due to adaption to specific environmental conditions.

The identified genes under non-neutral selection are involved in phytohormone signal transduction and response to stresses. An analysis of the ecotype and its phenotype could possibly link alleles to specific traits.

4.2 *Arabidopsis* host protein - *Gor* effector protein interactions

To identify protein-protein interactions between *Gor* pathogen effector proteins and *Arabidopsis* host proteins, effectors were tested for interaction with host proteins in a Y2H high-throughput mapping pipeline. Obtained interactions were integrated in a protein interaction map of interactions between *Arabidopsis* and the two pathogens *H. arabidopsidis* and *P. syringae* from [139] to elucidate host targets of evolutionary distant pathogens. Host targets of all three pathogens were genetically validated and correlated with the protein interaction map.

4.2.1 High-quality effector-host interactome

To obtain a basis for systematic analysis of effector - host protein interactions between effector proteins of three pathogens and *Arabidopsis* host proteins, interaction maps of all three pathogens were restricted to a common search space of *Arabidopsis* proteins. Therefore, a systematic host - effector protein interaction map was compiled including effector - host interactions for all three pathogens, restricted to a common subset of around 8,000 host proteins. Additionally direct interactions between host proteins were integrated from AI-1_{MAIN} [11] (fig. 3.2 B).

The resulting protein interaction map contains effector - host protein interactions from three evolutionary distinct pathogens (fig. 3.1) with biotrophic (*Hpa*, *Gor*) and hemibiotrophic (*Psy*) lifestyle.

Due to the use of the same high-throughput Y2H interaction mapping pipeline, for testing proteins for interactions, all three networks have been generated under the same conditions. Therefore no bias due to the mapping pipeline is expected, which otherwise could be expected [16]. The used interaction mapping pipeline is the same than used for AI-1_{MAIN}, which yielded a high sensitivity and specificity [11].

The integrated interaction map is a high-quality, bias free interaction map of effector - host protein interactions, which is suitable for systematic bioinformatic analysis.

4.2.2 GO enrichment analysis of host proteins reveals differences between pathogens

To determine the biological processes which are influenced by effector proteins, host proteins of each pathogen were analyzed for enrichment of GO terms.

The analysis of enriched GO terms showed that all three pathogens bind to proteins annotated with "regulation of defense response" and "regulation of response to stimulus". A difference can be found by the phytohormone signaling pathway, which is preferentially targeted by the pathogens: Both biotrophic pathogens, *Hpa* and *Gor*, bind preferentially to proteins involved in "response to auxin stimulus", whereas the effectors of the hemibiotrophic pathogen *Psy* bind to host proteins enriched in "salicylic acid mediated signaling pathway". Host proteins of the fungi *G. orontii* were additionally enriched in "copper ion transport".

Although both, AUX and SA, phytohormone signaling pathways are important for immunity [111, 251], pathogens with different lifestyle focus on different signaling pathways. The AUX signaling pathway, influenced by *Hpa* and *Gor*, is known be targeted by pathogens. An increased AUX level has a negative effect on the plant's immunity [261]. For the hemibiotrophic pathogen *Pst DC3000*, which has a biotrophic lifestyle in early stages, but a necrotrophic lifestyle in later stages, it seems to be important to control the SA signaling pathway. On the one hand a high SA level can lead to an increased resistance to Pst DC3000 [262] and on the other hand some necrotrophic pathogens manipulate the SA signaling pathway to promote the infection of the plant [263]. The "copper ion transport" process targeted by *Gor* effectors is of interest, as copper might be involved in regulation of defense response [252] and proteins involved in copper translocation are also involved in pathogen defense [253].

GO enrichment analysis revealed that pathogens with different lifestyle bind to proteins involved in different phytohormone signaling pathways. Although this has been already described, it is confirmed by the GO analysis and moreover the GO enrichment analysis, shows that significant number of proteins involved the respective hormone pathways are targeted.

4.2.3 Effector proteins converge on common host proteins

In the protein interaction map of *Hpa* and *Psy* interactions with *Arabidopsis* proteins, a convergence of effector proteins from both pathogens on common host proteins was shown [139]. This analysis was extended to effector proteins of all three pathogens and also conducted for effectors of a single pathogen.

The analysis of the extended protein interaction network $\text{PPIN2}_{\text{8k\_sys}}$ revealed both a statistically significant intraspecies and interspecies convergence of the effector proteins of all three pathogens.

This shows that convergent evolution [264] also exists on pathogen - host protein interactions.

The extent of convergence observed in PPIN2 lets assume a high pressure for interaction with specific host proteins, as these proteins a presumably important for successful infection. This may buffer the manipulation of certain biological processes in the plant even in case of loss of an effector protein or selection in the plant against a specific

effector protein. Another hypothesis for convergence could be the delivery and action of convergent effector proteins at different time points [138] to repeatedly influence the same process.

4.2.4 Genetic validation of host proteins

The genetic validation of host proteins in infection assays with T-DNA lines of host proteins successfully demonstrated the altered susceptibility or resistance compared to wild type for more than half of the host proteins. The results of the genetic validation were analyzed for correlation with results from convergence analysis.

A correlation between the interspecies convergence and the number of observed phenotypes was detected. This correlation was shown to be independent of the number of tests per T-DNA line. Moreover, a correlation between the number of effectors per pathogen and the number of observed phenotypes was found.

Convergence analysis and phenotyping assays revealed that TCP14 is the most targeted host protein, which also showed highly significant disease phenotypes in each of the six assays for all three pathogens. TCP14 is interacting with 23 *Gor* effectors, 25 *Hpa* effectors and four *Psy* effector proteins. The family members TCP13, TCP15, and TCP19 are also targeted by multiple pathogens and showed disease phenotypes in infection assays. This suggest that these members of the TCP transcription factor family play an important role while pathogen infection, moreover as their importance for plant immunity has been shown recently [265].

To independently validate the convergence of effector proteins on host targets, TCP14 was tested for co-localization with 11 *Hpa* effector, 19 *Gor* effectors and three *Psy* effectors by colleagues. Therefore a TCP14-RFP fusion gene was overexpressed in *Nicotiana benthamiana* were it localized to sub-nuclear foci. If TCP14-RFP was expressed together with an effector protein, which did not interact in Y2H assay, no co-localization of both proteins could be observed. But if TCP14-RFP was co-expressed with Y2H interacting effector proteins, it re-localized into sub-nuclear foci 64 % of *Hpa* effectors, 74 % of *Gor* effectors and one of three *Psy* effectors. These results were confirmed in co-immunoprecipitation experiments for single effectors (see also fig. 4 in [142]).

The correlation between convergence and number of phenotypes supports the hypothesis that there is evolutionary pressure to bind specific host proteins, which might be important for successful infection. The colocalization assay confirmed the physical binding of effector and host proteins and the importance of TCP14 as effector target.

4.2.5 Natural variation in host target protein network

A high variability of targeted host proteins was expected due to an assumed evolutionary pressure to overcome binding by effector proteins. To test this hypothesis, a combined score of the two neutrality tests, Tajima's D and Watterson's Θ were calculated and the number of amino acid polymorphisms in the host proteins were analyzed.

This hypothesis could not be confirmed, moreover, on the contrary, a conservation of these proteins was found. Many of the effector interacting host proteins are hubs and have a lot of interactions. A high variability of these proteins might affect many biological processes in the plant and could lead to severe phenotypes. This is in line with the observation, that for hub proteins a negative correlation between degree and variation rate was observed [266]. Further analysis of natural variation revealed that

the most variable host proteins significantly more often interact with effector interacting proteins, than expected by chance.

The analyses of the plant pathogen immune network revealed insights into possible strategies of pathogens for successful infestation. Additionally a strategy of plants to overcome an infection was found. On the one hand, pathogen effectors converge on host proteins and on the other hand host proteins interact with highly variable proteins, which might function as guards of the host proteins.

4.3 Conclusion

The aim in both presented projects was to bioinformatically analyze protein-protein interaction networks and integrate additional information into the network to elucidate biological meaning of the findings, understand in the context of available information and gain new knowledge about biological processes.

The PhI showed a strong interconnectedness by protein-protein interactions between phytohormone signaling pathways, instead of separated pathways. The large number of crosstalk points between distinct pathways and their in planta validation draw a picture of a hormone signaling network, where distinct pathways influence each other at multiple points. Even these results make signaling more complex, it may help to understand how plants integrate environmental signals and which processes are affected by specific stresses.

The PPIN2 revealed a convergence of effectors of three evolutionarily distant pathogens, *Golovinomyces orontii*, *Hyloperonospora arabidopsidis* and *Pseudomonas syringae*, on common host proteins of *Arabidopsis thaliana*. The genetic validation of host proteins showed their importance for immunity and infection and revealed a set of core immune proteins.

The results of the analyses of the protein interactions maps in this thesis elucidated the strong interconnectedness of hormone signal transduction and the convergent binding of effector proteins. Together, this knowledge can support the community to find ways how crops can be enabled to deal with stresses caused by global warming.

5 Methods
Phytohormone signal transduction

5.1 Definition of search space SSP_{PHO}

The search space defines the set of proteins, which should be searched for binary interactions by the Y2H interaction mapping pipeline described in [267]. The search space was defined by Pascal Falter-Braun using a collection of genes provided in the Arabidopsis Hormone Database 2.0 [140] having genetic evidence to be involved in the signaling pathway of one or multiple of these phytohormones: ABA, AUX (IAA), BR, CK, ET, GA, JA, and SA (`http://ahd.cbi.pku.edu.cn/`). For other known hormones (like KAR and SL) and signaling molecules (like NO, glucose and the "branching hormone" produced by the MAX pathway [123]) no information is available about genetic evidence in this database and no other systematic collection is available so far.

All loci with genetic evidence (mutant or transgenic study) were selected. Loci of a gene family, were tested for enrichment with Fisher's exact test. If a gene family showed a significant enrichment, all members of the gene family were included in the search space. Eleven loci were included following the suggestions of colleagues. This resulted in a list of 1227 loci (see table A.3). This list of loci is referred as search space SSP_{PHO}.

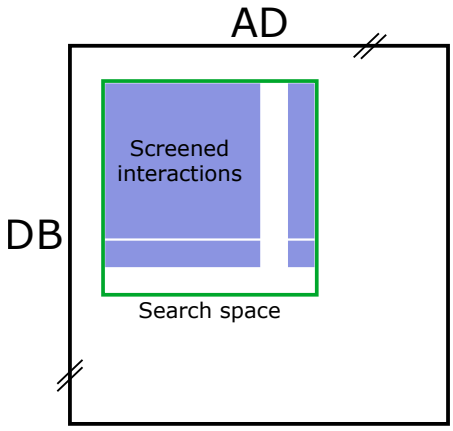


Figure 5.1: In the defined search space SSP_{PHO} (green box), which is a subset of all *Arabidopsis thaliana* ORFs (black box), all AD-hybrid constructs were screened against all DB-hybrid constructs, which could be derived for the PhyHormORFeome (blue area). Some ORFs could not be derived as AD- or DB-hybrid constructs or both. These could not be screened for interactions (white area within green box).

5.2 Generation of PhyHormORFeome

The PhyHormORFeome denotes the physical collection of ORFs defined in the search space. 688 loci of the search space SSP_{PHO} were already contained in the AtORFeome2 collection of *Arabidopsis thaliana* clones, which is located in our lab.

163 loci were available as ABRC Gateway(TM) clones and 113 loci as ABRC template plasmids. The remaining loci had to be isolated from *Arabidopsis thaliana* cDNA.

For isolation of genes from cDNA, the tissue or combinations of tissues containing the highest concentration of the respective mRNA should be used. The concentrations of mRNA were initially visually processed by Karl Kugler from the dataset *E-TABM-17 - Transcription profiling by array of organism parts from different strains of Arabidopsis* measured with *A-AFFY-2 - Affymetrix GeneChip Arabidopsis Genome [ATH1-121501]* provided on the EMBL-EBI ArrayExpress platform as part of [141]. This dataset contains the transcript abundance of 17 tissues / tissue combinations.

For extraction of genes from plant material, the three tissues with highest expression for each gene were extracted. The best tissues were afterwards used for making of cDNA from the respective genes. Subsequent collection of plant material, and cloning was conducted by Melina Altmann.

1207 ORFs could be retrieved as AD and DB, 16 ORFs could be retrieved only as AD and 5 only as DB. Overall ORFs for 1216 loci were retrieved.

5.3 Y2H screening pipeline

The Y2H screening pipeline consists of four steps: Primary screen, secondary phenotyping, candidate interaction identification, and verification (fig. 5.2).

Steps 1, 2 and 4 were executed by Melina Altmann. At the end of step 2, a PCR using barcode primers having a unique *barcode* nucleotide sequence per well and clone type (AD, DB) has been executed. While PCR reaction, a unique nucleotide sequence per well is attached to the PCR product. The PCR product has been sent for NGS to SeqWell (Boston). The candidate interaction pairs were extracted from received de-multiplexed NGS reads stored in separate files in FASTQ format.

FASTQ files contain a quality score for each base in each read. This quality score should be taken into account for sequence alignment. This means that bases identified with a poor quality have a lower impact on the alignment. A suitable program for this task is Bowtie2 [268], which can be downloaded from `http://bowtie-bio.sourceforge.net/bowtie2/index.shtml`.

Bowtie2 requires an index file containing the reference sequences, against which the sequencing results are aligned. This index file is built using the command bowtie2-build.bat -f [reference sequences fasta file] [index file name]. The coding sequences of all primary gene models listed in TAIR10 were included in the used index file (see listing 5.1).

```
1 C:\bowtie2-2.2.9> bowtie2-build.bat
2       -f TAIR10_cds_20110103_representative_gene_model_updated.fa
3       ath_cds
```

Listing 5.1: Bowtie2 command to build an index file named ath_cds using Arabidopsis thaliana coding sequences of representative gene models.

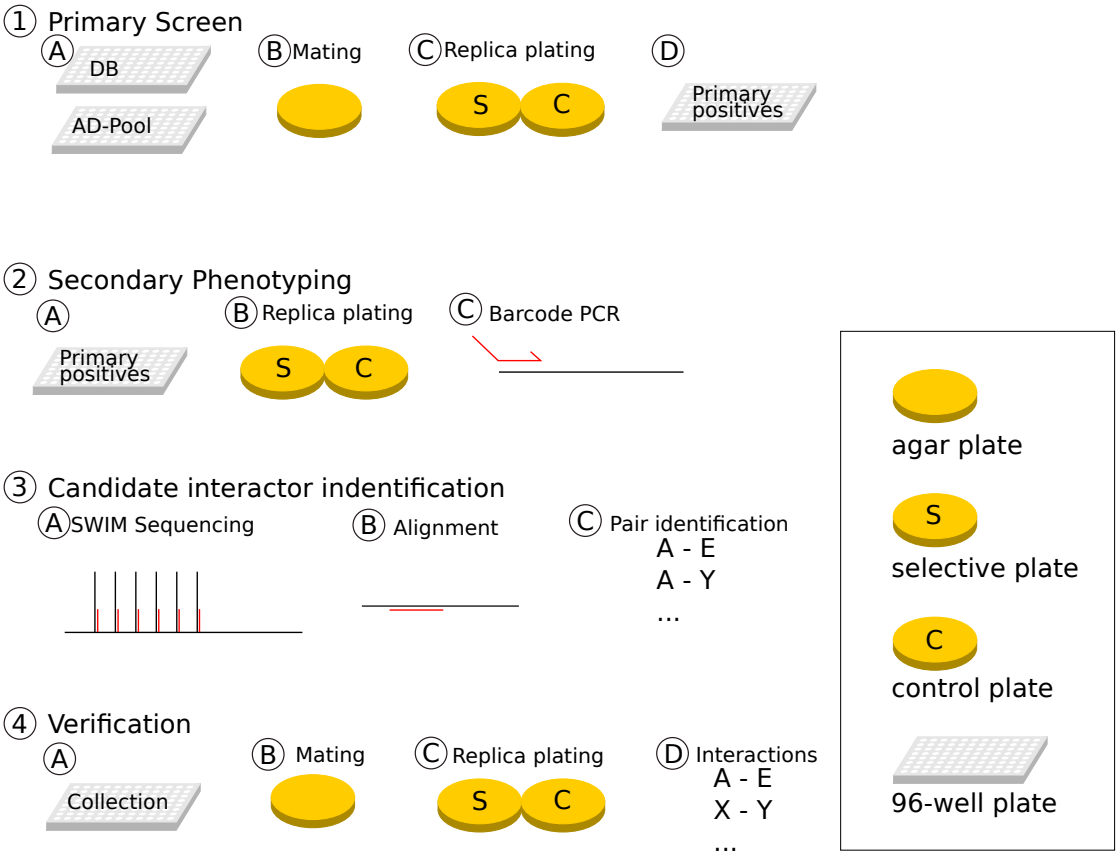


Figure 5.2: Our Y2H interaction mapping pipeline consists of four steps: 1. Primary screen: Single DB clones are mated with pools containing 188 AD clones. Positive colonies are selected from replica plating plates. 2. Secondary Phenotyping: primary positive colonies are tested for auto activating DB clones. Non-autoactivating colonies are prepared for sequencing by barcode PCR for subsequent identification. 3. Candidate interactor identification: Barcoded PCR products are sequenced using NGS. NGS sequences are aligned against reference sequences of tested proteins to identify interacting AD and DB clone. From alignment results interacting candidate protein pairs are identified. 4. Verification: All candidate pairs are verified four times independently to test for proteins for interaction.

Fastq-files can be aligned one by one to the reference sequences in the index file using the command *bowtie2.bat -x index_file_name –local -S out_file_name -U input_file_name*. Additional alignment parameters for bowtie2.bat have to be set according to the input data and alignment parameters have to be adjusted. An extended description of the parameters can be found on Bowtie2's homepage: `http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml`. Since many files had to be processed, they were handled with a custom Perl script iterating over all FASTQ files (see listing 5.2).

```
1 #!/usr/local/bin/perl
2
3 use strict;
4 use warnings;
5 # catch input argument
6 ($#ARGV == 0) or die "Usage: $0 [directory]\n";
7 # load package
```

```
use File::Find;

# find all files in all subdirectories of the given directory and send
# them to subfunction
find ( \&send_to_bowtie2, @ARGV );
# subfunction to send fastq file to bowtie2
sub send_to_bowtie2{
    # check, if files start with "AD" or "DB" and end with "fastq"
    if($_ =~ '[AD,DB].*fastq$'){
        print "$_\n"; # print file name
        print $File::Find::name."\n"; # print path and file name
        my $outfile = $File::Find::name; # out file name
        $outfile =~ s/fastq/sam/g; # replace fastq by sam
        print $outfile."\n"; # print out file name

        # build string for starting bowtie2, adjust bowtie2 folder
        my $bowtie_string = "C:/bowtie2-2.2.9/bowtie2.bat
            -x C:/bowtie2-2.2.9/ath_cds --nofw --no-head --no-sq
            --local -S $outfile -U $File::Find::name";
        print $bowtie_string."\n"; # print bowtie2 string
        system($bowtie_string); # start bowtie2
    }
}
```

Listing 5.2: Iteration over all FASTQ files in all subdirectories and alignment with bowtie2.

For one AD and one DB from an interacting protein pair in one yeast colony, only one locus per AD and DB were expected from sequencing result. Instead, multiple loci were found. Multiple reasons were discussed, which could lead to this result: (i) due to shortcomings in sequencing techniques, barcodes may contain errors, which leads to a wrong assignment of a read, (ii) contamination of wells by surrounding wells while PCR reaction, (iii) contamination at sequencing company.

To determine the number of loci per well, to be used for creating candidate interaction combinations from AD and DB sequencing results, all combinations of three hits up to a frequency of 10 % of the aligned reads in the respective AD and DB positions were compiled and tested against literature interactions derived from IntAct and BioGRID to determine up to which threshold an identified locus is included .

For these 314 loci a cherry picking list for the Tecan Evo pipetting robot was built in order to combine all 314 candidate interactors on 96 well plates (node plates). These node plates are used in the next step to distribute the AD and DB clones on mating plates for independent verification.

5.4 Quality assessment

To estimate the quality of the Y2H screen, three measure were determined: completeness, assay sensitivity and sampling sensitivity [16, 12]. The combination of these three measures allows to estimate the overall number of expected interactions in the defined search space.

5.4.1 Completeness

Completeness defines the percentage of interactions between AD and DB clones, which can be tested using the compiled PhyHormORFeome clone collection 5.2 (AD_{PHO} and

DB_{PHO}) compared to the number of possible interactions of the initially defined search space (AD_{SS}, DB_{SS}).

$$Completeness = \frac{AD_{PHO} * DB_{PHO}}{AD_{SS} * DB_{SS}}$$

5.4.2 Sampling sensitivity

Sampling sensitivity describes the percentage of interactions, which have been found in all repeats of the actual experiment compared to all interactions, which could be found with this protein interaction mapping method. The maximum number of identifiable interactions of a protein interaction mapping method can be estimated by fitting a Michaelis-Menten curve through the number of detected interactions per number of primary screens.

5.4.3 Assay sensitivity and specificity

With each protein interaction mapping method only a certain fraction of all true interactions can be detected, due to limitations inherited in each method. With each method also a small fraction of false interactions is detected. The fraction of detectable interactions and false interactions is determined by testing the re-detection of well-known interactions from positive reference set (PRS) (section 5.4.4) and random interactions from a random reference set (RRS) sampled from search space. The assay sensitivity can be evaluated by testing the interactions in the positive and random reference set. Subsequently the number of true positive and false negative interactions from the positive reference and false positive and true negative interactions from the random reference set can be counted. Assay sensitivity is calculated as fraction of the number of true positive interactions (tp) of the number of all positive interactions (p). The assay specificity is the number of true negative interactions by all negative interactions. The standard error of the proportion is calculated by the fraction of the proportion of true interactions (pr) times 1 minus the proportion divided by the number of interactions (n).

$$assay\ sensitivity = \frac{tp}{p}$$

$$assay\ specificity = \frac{tn}{n}$$

$$SE_{proportion} = \sqrt{\frac{pr * (1 - pr)}{n}}$$

5.4.4 Positive reference set

To measure the assay sensitivity of the Y2H mapping pipeline in mapping protein interactions of phytohormone related proteins, a sufficiently large set of high quality binary protein-protein interactions contained in the search space is required.

These protein-protein interactions must have been detected with a binary protein-protein interaction detection method, which means that only two proteins interaction with each other and the interaction is not part of a protein complex consisting of three or more proteins.

To meet the quality requirements, all interactions must have shown with two different methods, where one must be a binary protein-protein interaction detection method, or with the same binary protein-protein interaction detection method in two independent publications.

Candidate interactions for the positive reference set were compiled from IntAct (downloaded august 2014) [146] and BioGRID (Version 3.2.115) [147]. The selected detection methods for binary interactions are listed in tables A.7 and A.8 in the appendix. The Methods are selected to show physical contact between two proteins and not only co-localization. The IntAct dataset contained 17,574 interactions and the BioGRID dataset contained 21,474 interactions, where both interaction partners were from *Arabidopsis thaliana*. After filtering for interactions, where both interaction partner are from *Arabidopsis thaliana*, both datasets were cleaned for protein-RNA interaction, interactions from high-throughput experiments which reported more than 100 interactions, interactions, which are part of a complex (marked as "spoke expansion") and interactions from Arabidopsis Interactome 1 [11]. From the BioGRID dataset all interactions were removed, which were already contained in the IntAct dataset to avoid problems while merging interactions from both datasets. In the IntAct dataset, gene identifier, were mapped to TAIR locus IDs, if necessary. After selecting only protein-protein interactions, where both interaction partners are available in the PhyHormORFeome, 605 interactions from BioGRID and 814 interactions from IntAct were derived. These reduced datasets were used to identify interactions with high confidence. Overall 233 interactions were found, whereby 147 interactions were found in BioGRID and 151 interactions in IntAct with an overlap of 65 interactions. For the final positive reference set 140 interactions out of these 233 interactions were randomly selected and all 247 publications collected, in which they are described. All interactions were manually reviewed by two persons in their respective publications, if the experimental detection satisfies all required quality criteria: (i) result is shown in a figure, (ii) interaction was found in a small scale experiment, and (iii) all required controls of the experiment are shown in the publication to avoid detection of a false positive interactions. The publications were read and discussed by Nora Marin, Antoni Garcia i Molina, Pascal Falter-Braun, Melina Altmann and me. Out of 140 critically reviewed interaction, 108 interactions satisfied all quality criteria and were part of the positive reference set. A test of all clones prior to validation of the interactions reduced the number of interaction to 92.

5.4.5 Random reference set

The specificity of an assay can be detected by testing a negative reference set containing protein pairs which do not interact. This allows to determine the number of false positive interactions in the final set of identified interactions in the search space. As there exists no set of protein pairs which are known to not interact, a random set of protein pairs is used. This random reference set (RRS) contains the same number of protein pairs than the positive reference set and is sampled from all proteins in the search space . The RRS is tested for interactions in the same way than the PRS.

5.5 Network structure

To determine the structure of the selected networks, degree distribution and clustering coefficient distribution was calculated. From the distribution of those two measures it is

possible to determine the underlying structure of the network [7]. Our systematic PhI network was compared against two networks derived from IntAct [146] and BioGRID [147] restricted to our search space SSP_{PHO} including complemented members of IAA and RCAR protein families and only binary interactions. This extended search space SSP_{PHO} was used for all subsequent analyses. As comparison for systematic network maps, clustering coefficient and degree distribution for AI-1_{MAIN} [11] was calculated. Calculations were executed in R [269] using package igraph [149].

5.6 Network visualization

All networks were visualized using Cytoscape [37] and Cytoscape plug-ins. Network layout algorithms used are *Prefuse Force Directed Layout* included in Cytoscape and *yFiles Radial Layout* of yFiles layout algorithm collection [1]. Coloring of nodes using pie charts was done with plug-in *enhancedGraphics* [270]. Symbol names of proteins were derived from Araport version 11 [173].

5.7 Communities

A community is a group of nodes in the network, which are stronger connected to each other than to the rest of the network [25]. For identification of communities different algorithms have been proposed. The algorithms edge betweenness, fast greedy, infomap, label propagation, louvain and walktrap are available in the R package igraph [149] and were applicable and worked with all examined networks.

For all communities, the enrichment in a hormone annotation was determined. For each community, it was tested, if one out of the eight phytohormones is significantly enriched. Therefor, for each hormone the number of proteins annotated with this hormone and number of proteins not annotated with this hormone were counted and tested for significance using fisher's exact test.

Table 5.1: Contingency table to test for enrichment of hormone annotation.

	In community	Not in community	Sum
Is hormone X	A	C	A+C
Is not hormone X	B	D	B+D
Sum	A+B	C+D	A+B+C+D

Best results were shown by the edge betweenness algorithm. The edge betweenness algorithm calculates an edge betweenness score for each edge in a network, whereby the edge betweenness score measures the number of shortest paths crossing an edge. In an iterative process the highest scoring edge is removed and subsequently the edge betweenness score is recalculated for all remaining edges.

All tested community algorithms were compared, in terms of similarity of identified communities. This means, it is tested, if the same nodes are together in the same community. To determine the similarity of community structures, five measures were calculated to compare communities identified by the edge betweenness algorithm with communities identified by above listed algorithms. The measures are *Variation of information* (vi)

[1] `www.yworks.com`

[32], *Normalized Mutual Information* (nmi) [271], *Split-Join Distance* (split join) [272], *Rand Index* (rand) [273] and *adjusted Rand Index* (adjusted rand) [274]. The range of values and interpretation is shown in table 5.2.

Table 5.2: Measures to calculate similarity between community algorithms

Method	Range of values (identical - dissimilar)
Variation of Information [32]	0 - 2logK, where K is number of clusters , (log is to the base e) [32]
Normalized Mutual Information [271]	1 - 0 [271]
Split-Join Distance (also known as van Dongen S) [272]	0 - 2N, where N is number of nodes; higher values mean more different clusterings [272]
Rand Index [273]	1 - 0 [273]
adjusted Rand index [274]	1 - 0 [275]; values > 0.90 excellent recovery, values > 0.80 good recovery, (c) values > 0.65 moderate recovery, < 0.65 poor recovery [276].

5.8 GO enrichment of communities

Communities of PhI were tested for statistically overrepresented Gene Ontology (GO) terms using R package GOstats [277] provided by the R package repository Bioconductor (`https://www.bioconductor.org`). This package uses a hypergeometric test to determine, if a GO term in a subset of nodes is statistically overrepresented compared to all nodes in a given set. The hypergeometric test is implemented in the function hyperGTest. This function was invoked with parameter conditional = TRUE, which means that the structure of the GO graph is used to estimate statistical overrepresentation of a term. As universe (set of gene ids, which define the complete space of genes to be used for testing statistical overrepresentation) the proteins in PhI were defined. The overrepresented GO terms per community are shown in supplemental tables A.12 - A.32.

Overrepresented GO terms were visualized with REVIGO [47], which combines semantically similar terms and removes redundant terms.

5.9 Hormone interconnectedness

To determine the interconnectedness of a hormone signaling pathway, the mean length of shortest paths between proteins annotated with the same hormone was calculated. To determine the interconnectedness between two different hormone signaling pathways, the mean length of shortest paths between proteins belonging to two distinct hormone signaling pathways was calculated. Only shortest paths, which directly connect proteins, with the same and distinct hormone annotation were taken into account, respectively. Shortest paths, which cross a node annotated with the same hormone than the tested protein pair, are excluded. The exclusion of shortest paths, which contain a protein

annotated with the same hormone as one of the two tested proteins has several advantages: two strongly connected clusters, which are not directly connected receive a small mean shortest path, which shows the strong connectivity within the cluster. This strong connectivity within the cluster would be hidden by a higher mean shortest path. A long distance between two strongly connected clusters can be seen by an outlier in the distribution of shortest path lengths for the respective hormone annotation. The number of shortest paths, which are used to calculate the mean length of the shortest path between / within a hormone combination is significantly reduced compared to all possible shortest paths for a hormone combination. Therefor the weight of each path on the mean value is higher. In figure 5.3 the number of shortest paths for the blue nodes is reduced from 108 paths to 20 paths.

```
for i in 1 to number of hormones do
    for j in i to number of hormones do
        find all unique shortest paths between all nodes annotated with hormone[i]
         and hormone[j];
        if shortest path crosses node annotated with hormone[i] or hormone[j]
         then
            skip path;
        else
            calculate length of shortest path;
        end
    end
    Calculate mean length of the shortest path for hormone combination;
end
```

Algorithm 1: Find direct shortest paths and calculate mean short path for hormone combination

5.10 Hormone annotation inference

To infer hormone annotation for proteins an association network was constructed from PhI. Therefore similarity between interaction profiles of all pairs of proteins in PhI was calculated. All measures described in [33] were used to determine similarity (Cosine, CSI, Geometric, Hypergeometric, Jaccard, Simpson, and Pearson Correlation Coefficient) with R. Visual inspection [33] showed that Geometric performed best, whereas PCC, Hypergeometric and CSI showed worse performance in determining profile similarity between proteins. As cutoff for similarity, the N1 value was selected. This means, the respective similarity values at 1 % of the most similar protein pairs was selected.

5.11 Hormone crosstalk

To identify possible points of signal integration between different hormone signaling pathways (crosstalks) on the level of protein-protein interactions, interactions between proteins involved in different hormone signaling pathways were identified.

For the assignment of proteins to a hormone signaling pathway, the respective phytohormone annotations were assigned to proteins. Hormone related annotations of genes were derived from Arabidopsis Hormone Database (AHD) 2.0 [140], which is based on literature published before October 2010. From AHD 2.0, annotations based on genetic evidence (mutant or transgenic study) were used for protein annotation.

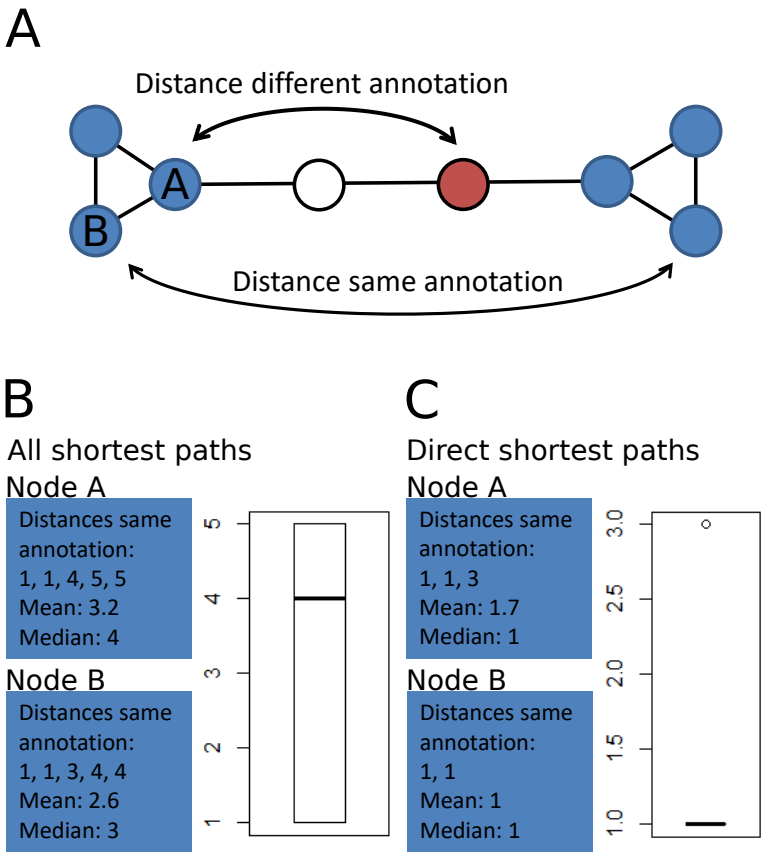


Figure 5.3: Example of the calculation of the mean shortest path length in the network. A) Network consisting of two cluster of blue nodes connected by one red and one white node. B) Lengths, mean and median of all shortest paths from node A and node B to all other blue nodes, respectively. Box plot shows the lengths of all shortest paths between all blue nodes. C) Lengths, mean and median of direct shortest paths from node A and node B to all other blue nodes, without shortest paths, which cross other blue nodes. Box plot shows the lengths of direct shortest paths.

Hormone annotations supported by gene ontology annotations were extracted from gene ontology data provided by The Arabidopsis Information Resource (TAIR) [278]. Gene ontology data in file ATH_GO_GOSLIM.txt were downloaded on 03.08.2018 from TAIR FTP Server. Prior to extraction of hormone annotation, the data set was cleaned for annotations with the evidence code "IEA". These annotations are inferred from electronic annotation, this means that the respective annotation is assigned without curational judgement and are assumed to have a lower confidence than other evidences, although quality has improved in recent years [279]. Gene ontology terms were screened for the occurrence of the following hormone names: *auxin*, *abscisic acid*, *brassinosteroid*, *cytokinin*, *ethylene*, *gibberellin*, *jasmonic acid*, *salicylic acid*, *strigolactone*, and *karrikin*. If one of these hormone names was identified in any of the GO Terms, the respective hormone annotation was assigned to the respective locus.

Subsequently, protein-protein interactions are screened for interacting proteins, which are assigned to different hormone signaling pathways. Protein pairs, which have strictly

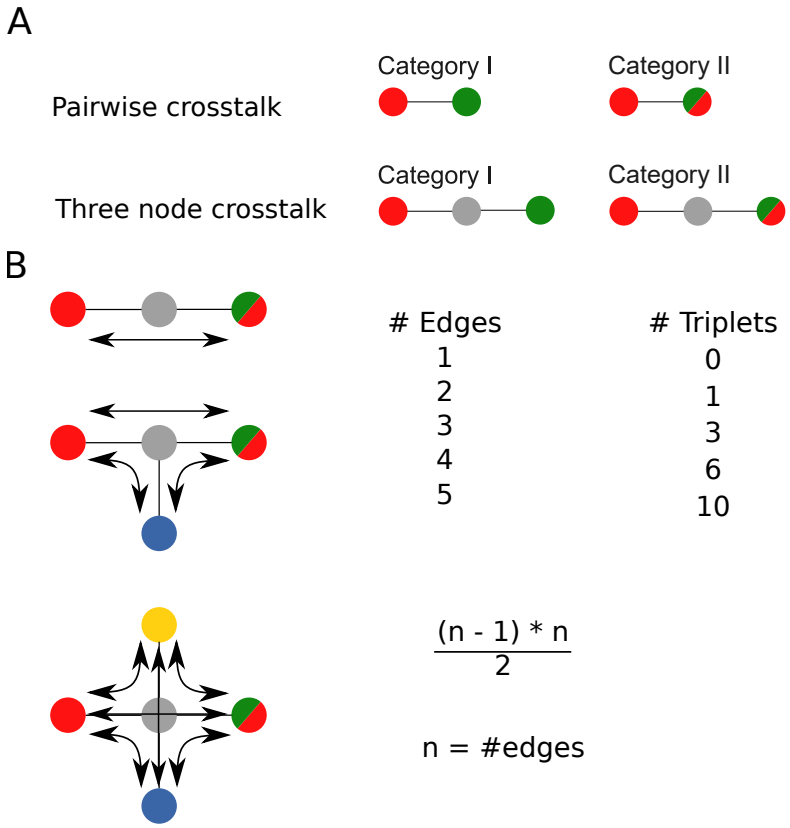


Figure 5.4: A) Crosstalk categories I and II in pairwise crosstalks and three crosstalks. Crosstalks of category I have strictly different hormone annotations. Crosstalks of category II have partially overlapping hormone annotations. B) Sketch to illustrate the number of possible three node crosstalks per connector protein. *Triplet* denotes a unique combination of a connector protein and two hormone annotated proteins.

different annotations are assigned to crosstalk category I; proteins pairs, which have overlapping annotations are assigned to crosstalk category II (fig. 5.4 A).

For identification of proteins involved in crosstalk, so far unknown to be involved in any phytohormone signaling pathway (no hormone annotation at all), networks were screened for proteins with no annotation, which connect two proteins annotated with different phytohormones (three node crosstalk). These three node crosstalks are assigned to the respective crosstalk category I and II (fig. 5.4 A).

Non-annotated proteins, which connect hormone annotated proteins can facilitate a large number of crosstalks due to possible combinations of interacting proteins. The number of crosstalks can be calculated by $\frac{(n-1)*n}{2}$, where n is the number of interactions * hormone annotations. This means, that all hormone annotations of one interacting proteins must be tested for possible crosstalk against hormone annotations of the other interacting proteins (fig. 5.4 B).

5.12 Condition-dependent interactions

RNA-Seq expression data were used as proxy to determine abundance of proteins in the given condition of the RNA-Seq experiment. RNA-Seq experiment data from *Arabidopsis*

thaliana were downloaded from NCBI Sequence Read Archive (SRA)[2]. The experiments had to meet the following criteria in order to guarantee a high quality and comparability of the experiments: (a) RNA-Seq data must be derived by paired-end sequencing, (b) RNA-Seq data must be produced using the same technique / instrument family. Here Illumina HiSeq system was selected. (c) each run must have produced more than 10 million reads to guarantee data for low expressed genes. 1264 RNA-Seq runs of 486 conditions met these requirements, where conditions comprise combinations of wild type plants, mutant plants, tissues, and specific treatments.

Downloaded sra-files were decompressed using fastq-dump and tested for quality using FASTQC. Derived fastq-files were aligned to reference sequence with Kallisto, which was also used to quantify expression of genes as TPM (transcripts per million) values. Replicates of RNA-Seq runs were collapsed by calculating mean values. The resulting expression matrix consists of TPM values for 27.655 genes in 486 conditions.

The expression matrix was used to calculate the interaction score for each interaction in all available treatments by multiplying TPM values of interacting proteins. Where TPM values are used as a proxy to estimate protein levels:

$$InteractionScore[P_A : P_B] = k_p[P_A] * [P_B] = k_T[T_A][T_B]$$

The quantitative information of two transcripts T_A and T_B with with a complex formation constant k_T is used to estimate possibility of complex formation of two proteins P_A and P_B with constant k_P. This is used to model the Interactions Score for protein pair [P_A:P_B] under the assumptions already described in [51]: (i) protein levels are modeled relative to mRNA levels and (ii) involvement of a protein in an interaction does not influence probability to be also involved in another interaction. Interaction scores were afterward transformed condition-wise into z-scores, by

$$z = \frac{x - \mu(col)}{\sigma(col)}$$

where μ is mean value and σ is standard deviation. Interaction scores were calculated for all available interactions from BioGRID, IntAct, PhI and PhI_{out}.

For each condition the following information were used to characterize the experiment: ecotype, wildtype or knock-out with the respective genes, growth stage or age, and treatment.

Calculations for the interaction score matrix were conducted by Gurudutta Panda.

5.13 Identification of potential kinase - substrate interactions

For identification of kinases and classification for their respective kinase family, data from [130] were used. Information about known kinase substrates were downloaded from PhosPhAt[3] database [202, 203, 204]. Phosphatases were identified and classified based on PlantsP[4] [280]

Identification of potential kinase - substrate interactions in protein-protein interactions of the Y2H interaction map was accomplished by identification of potential phosphorylation sites in proteins fitting to interacting kinase. The identification pipeline consists

[2] `https://www.ncbi.nlm.nih.gov/sra/`
[3] `http://phosphat.uni-hohenheim.de/`
[4] `http://plantsp.genomics.purdue.edu`

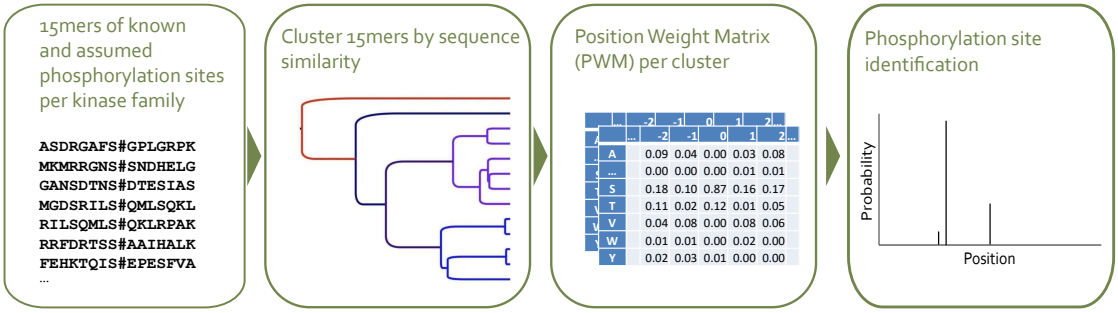


Figure 5.5: Pipeline to determine possible kinase - substrate interactions. For each kinase family a set of 15mers around known and assumed phosphorylation sites were clustered. For each cluster of 15mers a position weight matrix (PWM) was calculated. PWMs were used to identify possible phosphorylation sites in proteins, which interact with kinases from the respective kinase family.

---

of determination of possible phosphorylation motifs per kinase family, identification of phosphorylation sites in proteins interacting with kinase and substrates described in literature, and calculation of significance of potential phosphorylation sites (fig. 5.5).

Determination of possible phosphorylation motifs

Only a low number of phosphorylation motifs of specific kinases or kinase families have been described so far. Therefore possible phosphorylation motifs of kinase families were determined from a set 15 amino acid long peptide sequences around a phosphorylated amino acid in the center (15mers) provided by Waltraud Schulze, Universität Hohenheim, by personal communication. This set is subdivided into subsets of 15mers per kinase family or subfamily, which phosphorylates them or is assumed to phosphorylate them. Each kinase family / subfamily specific subset was used to identify several possible phosphorylation motifs for this kinase family / subfamily. Each kinase family / subfamily consists of several kinases and each subset of 15mers is phosphorylated by unknown number of kinases. To determine the optimal number of phosphorylation motifs, which can be extracted from each subset of 15mers, an elbow plot with up to 100 cluster was generated. From the elbow plot, which shows the percent of variance explained by the number of cluster, the number of cluster for each subset of 15mers was chosen. Additionally a Principal Component Analysis (PCA) was conducted to verify the chosen number of cluster based on the elbow plot, as an elbow plot is not always unambiguous and provides multiple possible numbers of clusters.

Each subset of 15mers were clustered agglomeratively into the respective number of clusters. From the sequences contained in each cluster, a Position Frequency Matrix (PFM) was built and afterwards a Position Probability Matrix (PPM). Where a PPM contains the frequency of each amino acid at the respective position. For visualization of the PPM by a sequence logo, the PPM was converted into in an information content matrix and plotted with the R package ggseqlogo [281]. Where the information content per position is calculated by

$$R_i = log_2(20) - (H_i + e_n)$$

where R_i is the amount of information present at position i, $log_2(20)$ is the maximum uncertainty, H_i is uncertainty at position i, and e_n is a correction factor, which is only need in case of only a few samples [282] and was therefore neglected here.

The uncertainty H_i for position i was calculated by

$$H_i = -\sum_{b=a}^{t} f_{b,i} * log_2 f_{b,i}$$

where b is one of the amino acids, and $f_{b,i}$ is the frequency of amino acid b at position i.

The height h of letter a in column i is given by

$$h_a = f_{b,i} * R_i$$

according to [283, 282].

Identification of potential substrates and their phosphorylation sites

All proteins in the protein interaction network were assigned to their respective kinase family according to kinase classification in [130]. If a protein interacts with a kinase, it was classified for analysis of possible phosphorylation site. Phosphatases were assigned to families according to PlansP database [280]. To identify potential phosphorylation sites in possible substrates, each possible phosphorylation site (amino acids S, T, and Y) was tested for fit to all PWMs (depending on the number of cluster based on elbow plot) of the respective kinase family to which the interacting kinase belongs to. Therefore the product of frequencies of matching amino acids in the PWM was used to determine fit of a PWM at the respective position. The best matching position with the maximum value found with each PWM could show a possible phosphorylation site and was subsequently tested for significance against a background distribution.

All PWMs were also applied to known substrates of the respective kinase family. Known substrates were downloaded from PhosPhAt database on 20.01.2016 [202, 203, 204].

Calculation of significance of potential phosphorylation sites

To test if a possible phosphorylation site and its surrounding amino acids fit better to a PWM than expected by chance, it was compared against background. As background all proteins sequences of primary gene models were used. All sequences were tested against each PWM to generate a background distribution. The best value of each possible substrate was used to calculate an observed p-value against the respective background distribution. Possible substrates with a p-value < 0.05 were treated as significant substrates and are subject for experimental validation.

5.14 Repressor - transcription factor interactions

For identification of transcription factor - repressor proteins interactions in protein-protein interaction networks, a list of repressor proteins was provided by Melina Altmann. The list contains 71 proteins, which are described in literature as repressors in the respective phytohormone pathway. The list is attached in the appendix (see table A.44).

Transcription factors were determined by gene ontology annotation. All proteins were selected, which gene ontology annotation contains "DNA binding" and "transcription factor". Three gene ontology terms are present in *Arabidopsis thaliana* GO annotations, which contain both key terms:

- "transcription factor activity, RNA polymerase II proximal promoter sequence-specific DNA binding"
- "DNA-binding transcription factor activity"
- "DNA-binding transcription activator activity, RNA polymerase II-specific"

In Arabidopsis 1667 proteins are annotated with one of these three GO terms. This list was used to identify transcription factors in the protein-protein interaction network. In the PhI network are contained 80 transcription factors and in the combined network of PhI and PhI_{out} are 140 transcription factors contained. GO annotations were downloaded at 2018-08-03 from TAIR (∗www.arabidopsis.org).

5.15 Natural variation in *Arabidopsis thaliana* ecotypes

Single Nucleotide Polymorphism (SNP) information of 1135 *Arabidopsis thaliana* accessions, published in [58], was downloaded from 1001 Genomes project page [5]. Coding sequences of all representative gene models in TAIR10 [6] were extracted from chromosome sequences (TAIR10_chromosome_files) of *Arabidopsis thaliana Col-0* sequence data downloaded from TAIR [7] according to position information in TAIR10_GFF3_genes.gff file. SNP information of natural accessions was extracted from "1001genomes_snp-short-indel_only_ACGTN.vcf.gz" and integrated into coding sequences.

Selection measures for coding sequences of each primary gene model listed in TAIR10 were calculated using MATLAB and Molecular Biology & Evolution Toolbox (MBE-Toolbox) [284]. The following measure were calculated: Tajima's D, Fu & Li's F* and D*, and Ramos' R2. Loci were selected for further analysis, if the value a gene in one selection measure lies not in the 95 % confidence interval. This means the value is lower than the 2.5^{th} percentile or greater than the 97.5^{th} percentile of the respective measure.

Genes with a Tajima's D value not in the 95 % confidence interval and an additional selection measure not in the 95 % confidence interval were selected for further analysis (deviating genes). These genes were analyzed for network properties and involvement in biological processes. Other network properties were used to interpret and validate Tajima's D results.

Network properties of deviating genes were calculated with R packages igraph [149]. To identify significant properties, mean values of betweenness and clustering coefficient were compared against mean values in 10,000 degree preserved randomized networks. Degree was tested against 10,000 networks with random permuted node ids. Closeness is only defined on connected graphs, therefore this measure was tested on the largest connected subgraph consisting of 218 nodes and 449 interactions. Closeness was compared against randomized networks of the largest connected subgraphs, where no unconnected subgraphs / nodes were allowed.

[5] `http://1001genomes.org/data/GMI-MPI/releases/v3.1/`
[6] from file TAIR10_representative_gene_models
[7] `ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR10_genome_release/`

6 Methods
Host-pathogen networks

6.1 Pathogen convergence analysis

Intraspecies convergence: Significance of intraspecies convergence was determined experimentally based on the experimentally observed number of interacting *A. thaliana* host proteins within Space $8k_{sys}$ for effectors from Hpa, Psy, and Gor provided in table 3.1. For each pathogen effector interactors were sampled randomly from a list of AI-1_{MAIN} proteins (not shown) and from a degree preserved list of AI-1_{MAIN} loci (fig. 3.4 B-D) [11] using the "sample" command in R. The second analysis is more stringent as it increases the probability of repeatedly picking the more connected proteins and therefore leads to a lower number of nodes expected by chance. The distribution obtained from 10,000 samplings were plotted and compared to the experimentally observed value. The experimental P-value was calculated by dividing the number of samplings where the number of common targets is greater or equals the observed number of common targets ($samplings_{MORE}$) by the number of samplings performed ($samplings_{ALL}$). If the observed number of targets is not seen in the simulation, the P-value is set to < 0.001.

$$\text{observed p-value} = \frac{samplings_{MORE}}{samplings_{ALL}}$$

Interspecies convergence statistics: Significance of the convergence of effectors from different pathogens interacting with common host proteins was determined experimentally. The convergence was determined for all possible pathogen combinations based on the numbers of common interaction partners provided in Figure (fig. 3.4 E). For each pathogen the number of host interaction partners was sampled randomly from a unique list of proteins in AI-1_{MAIN} [11] using the "sample" command in R with replacement. The observed number of common proteins for each pathogen combination in each of 10,000 samplings was plotted as a background expectation and compared to the experimentally observed value of common interaction partners provided in Figure (fig. 3.4 F-I). The experimental P-value was calculated by dividing the number of samplings where the number of common targets is greater or equals the observed number of common interactors ($samplings_{MORE}$) by the number of samplings performed ($samplings_{ALL}$). If the observed value of common targets is not seen in the simulation, the P-value is set to < 0.001.

$$\text{observed p-value} = \frac{samplings_{MORE}}{samplings_{ALL}}$$

6.2 Gene ontology enrichment

We used GO enrichment analysis to test, which functional processes are overrepresented i) among effector interacting proteins, and ii) among the top 55 ranking genes. To this

end we performed a GO enrichment analysis using all loci in AI-1_{MAIN} as the background distribution. The analysis is based on GO annotations of TAIR10 (timestamp: 2013-09-03), which we downloaded from the TAIR ftp-server. We removed all annotations with the evidence code "inferred from electronic annotation" (IEA). 17 out of 2661 loci in AI-1MAIN do not have any manually curated GO annotation. For enrichment analysis we used the GOstats package version 2.28.0 [277]. We used the function hyperGTest to perform a hypergeometric test on the GO terms. We used a P-value cut-off of 0.005 and "conditional testing", which means that parent terms are tested without genes, which already have been found to be significant in a children term. Results were plotted with REVIGO [47] using standard parameters.

6.3 Analysis of genetic validation

Fold-change calculation of genetic validation results

For each pathogen / pathogen strain p the mean value x of raw spore / sporangiophores counts for each T-DNA and Col-0 control line were normalized by scaling values between 0 and 1.

$$Normalized(x_{i,p}) = \frac{x_{i,p} - X_{min,p}}{X_{max,p} - X_{min,p}}$$

where $x_{i,p}$ is the raw mean value of the pathogen p and the value of the tested gene i. $X_{min,p}$ is the minimum value of pathogen p and $X_{max,p}$ is the maximum value of pathogen p.

The phenotype of the T-DNA line with respect to the Col-0 control plants was evaluated by calculating the fold change of the mean normalized values of all available batches for each T-DNA line. The average fold change of all batches for a given T-DNA line was converted to a log2 fold change. A log2 fold change of 0 means same pathogen infestation of T-DNA line and control line, a negative log2 fold change means lower infestation (enhanced disease resistance) and a positive log2 fold change means higher infestation (enhanced disease susceptibility).

$$fc_{i,p,k} = \frac{Normalized(x^{M}_{i,p,k})}{Normalized(x^{C}_{i,p,k})}$$

$fc_{i,p,k}$ is the fold change of T-DNA line i inoculated with pathogen p in batch k. $Normalized(x^{M}_{i,p,k})$ is the normalized value of T-DNA line inoculated with pathogen p in batch k and $Normalized(x^{M}_{i,p,k})$ is the normalized value of control line of T-DNA line i inoculated with pathogen p in batch k.

$$Log_2 fc_{i,p} = log_2(\frac{fc_{i,p,k}}{n_i})$$

$Log_2 fc_{i,p}$ is the log_2 fold change of T-DNA line i of pathogen p.$fc_{i,p,k}$ is divided by the number of batches n_i of the respective T-DNA line i.

P-value calculation of genetic validation results

P-values represent the significance of the difference in pathogen infestation between T-DNA lines and Col-0 control plants. Depending on the pathogen the data were collected

in different ways and require individual analysis steps and application of a suitable statistical method for p-value calculation.

Values of *G. orontii* experiments of adult plants and seedlings have been derived from hemocytometer counts and represent spores/g fresh weight. A Gaussian generalized linear model was fitted on the data and used for ANOVA analysis (R package "car") [285]. As we have repeated the inoculation experiments up to four times with a T-DNA line and the control line we treat it as a block experiment, where every T-DNA line and the respective control plants of one batch are treated as one block. The block is treated as second factor in our ANOVA analysis beside the first factor of the knocked out gene. Our data were analyzed as a two-way ANOVA experiment with the factors gene and batch. As correction for multiple testing the Benjamini & Hochberg [286] method was applied.

P. syringae data are represented as cfu / ml and have the same characteristics like *G. orontii* data and have been analyzed the same way as *G. orontii* data.

H. arabidopsidis isolates Emwa1, Emoy2 and Noco2 data sets were collected as counts of sporangiophores per cotyledon. These data do not satisfy the requirement of ANOVA for normality distribution. This requires the use of a non-parametric test. We used the Kruskal-Wallis test to calculate the P-value and corrected the results with Bonferroni multiple testing correction method.

Merging of phenotype for multiple T-DNA lines

Merging phenotypes for multiple T-DNA lines: For the summarized phenotypic analysis of mutants we combined the phenotypic outcome if more than one T-DNA line per gene was available. Therefore we compared the pathogen-specific P values of the phenotypes on the different T-DNA lines representing the same gene. We selected the phenotypic outcome with the lowest (most significant) P value to obtain merged phenotypes for a gene. We found no contradictory phenotypes for any pathogen, i.e. we had no case where we observed an edr phenotype in one mutant line and an eds phenotype in the other mutant line. In 15 cases we observed a statistically significant pathogen-specific phenotype in one allele of a gene and no phenotype for the second allele of this gene. In these instances we selected the outcome of the line with the statistically significant result (i.e. the line showing the altered pathogen infection phenotype).

6.4 Phenotype correlation and density

For infection assay experiments, the number of pathogens targeting a gene was used to group t-DNA line experiments into 3 groups by the respective number of pathogens targeting a host protein. For each pathogen group all carried out experiments were evaluated for the phenotype and the number of experiments with eds, edr and no phenotype were determined. The ratio of edr and eds phenotypes to all experiments per group was calculated to determine density of phenotypes per group.

To determine, if the number of phenotypes observed in t-DNA lines increase with the number pathogens targeting the knocked-out gene, the observed phenotypes per number of pathogens targeting a gene were plotted.

6.5 Natural variation in *Arabidopsis thaliana* ecotypes

Arabidopsis Consensus Sequence Building and AAP evaluation

To evaluate natural variation in Arabidopsis accessions we used the complete genomes of 80 accessions sequenced in the context of the 1001Genomes project and mapped on the Col-0 reference genome. These were collected in eight regions distributed over Europe and Asia, where Arabidopsis naturally occurs and provide a large spatial and phylogenetic distribution adapted in different environments. This dataset was published by Cao et al., 2011 [62] and can be downloaded from 1001genomes.org.

A challenge for the quantitative evaluation of coding variation is the fact that single nucleotide polymorphisms (SNPs) are reported relative to the Columbia (Col-0) reference genome [62]. This results in numerous SNPs being called in all 80 accessions. In contrast, a conservative nomenclature would identify this as a SNP in Col-0 only. We used a majority voting scheme to define the consensus sequence of the Arabidopsis population consisting of the genomic sequences of Col-0 and the dataset MPICao2010 [62], which is available at `1001genomes.org`. In this scheme, the most common base at any position defines the consensus sequence and all other variants that occur in the population are counted as variant SNPs where they occur. The codons in coding regions of the consensus sequence of representative gene models as annotated in TAIR10 were compared against the respective codons in the 81 accessions in the genome matrix. In each accession we examined the codons for synonymous SNPs (sSNPs), non-synonymous SNPs (nsSNPs) and the resulting amino acid. A unique amino acid polymorphism (AAP) is defined qualitatively as a specific amino acid substitution at a given position, independent of the frequency of how often the specific substitution was found in the analyzed population. In other words a SNP leading to a hypothetical G− >A substitution counts as a single unique AAP independent of how often this amino acid replacement occurred; a substitution resulting in a hypothetical G− >T replacement is counted as a second unique AAP. For each protein we counted the unique number of AAPs observed for an individual position. We calculated the sum of unique AAPs per position of a protein, which results in the number of unique AAPs in a protein.

Calculation of Tajima's D and Watterson's Θ

In order to determine Tajima's D and Watterson's Θ values for all genes in AI-1_{MAIN}, we extracted the aligned genomic sequences of all 80 accessions and Col-0 for the respective representative gene models as fasta file from the whole genome alignment file (TAIR10_-genome_matrix_2012_03_13.txt.gz) from the dataset MPICao2010 [62]. This file was used to calculate Tajima's D and Watterson's Θ using the standard settings in the compute program of the analysis software package (version 0.8.4) developed by the Thornton lab [287]. The calculation of Tajima's D and Watterson's Θ was executed by Stefan Henz from Max Planck Institute for Developmental Biology, Tübingen.

To rank genes for their genetic variation within the 81 accessions we calculated a combined rank of Tajima's D and Watterson's Θ (DΘ-ranking). Therefor we sorted all genes in descending order according to their Tajima's D and Watterson's Θ value, respectively, and assigned them a rank depending on their position. Subsequently genes were ranked according to the ascending order of the mean rank of these two estimators.

Statistics of Tajima's D for effector targets and their AI-1_{MAIN} interactors

To determine, if the Tajima's D values of effector-interactors or that of their AI-1_{MAIN} interaction partners deviate significantly from the background distribution we performed a sampling and random rewiring analysis respectively for the four effector groups as indicated in the main text (all targets, targets of two or three pathogens, targets of three pathogens, and targets with phenotype).

To evaluate Tajima's D for effector-interactors we randomly drew without replacement 1,000 samples of the same size as the respective effector group from the unique set of genes encoding proteins in AI-1_{MAIN} as random control. To analyze the significance of Tajima's D of the AI-1_{MAIN} interaction partners of effector-interactors we generated 1,000 randomized degree preserved networks and determined the mean Tajima's D of all interaction partners of the respective effector-interactors in the random networks and compared this to the real mean value obtained from AI-1_{MAIN}. For each of the four sets of effector-interactors and their AI-1_{MAIN} interaction partners we calculated a two-sided P-value for the observed mean compared to the distribution of the 1,000 mean values of the random controls. We counted the number of occurrences greater equal and lower equal than population mean +/- (population mean - sample mean), respectively and divided it by the number of samples.

Statistical evaluation of top Watterson's Θ- and AAP-ranking genes with effector-interactors

DΘ-ranking

We investigated whether proteins encoded by top DΘ-ranking genes preferentially interact with effector-interactors. This analysis was performed for interactions with the four different sets of effector targets: all effector targets, proteins interacting with effectors from two or three pathogens, proteins interacting with effectors of three pathogens and effector targets showing a phenotype in the phenotyping assay.

We performed a degree preserving random rewiring of the AI-1_{MAIN} network by 100,000 times permuting two interaction partners of two randomly selected edges using the rewire function in the igraph R package v0.7.0 [149]. This was repeated to generate 1,000 rewired networks. In each random network we counted the number of effector-interactors interacting with proteins encoded by the cumulative 1 – 70 top DΘ-ranking genes and the analysis was repeated for each class of effector-interactors. The data was used to calculate the experimental P-value for the probability of finding the experimentally observed number of interactions between top DΘ-ranking genes with effector-interactors by chance. We calculated an observed P-value by dividing the number of observations with a value greater equal than the real number of observation by the number of generated rewired networks.

Amino Acid Polymorphism Effector Interactors Evaluation

To determine, if loci having a high number of AAPs interact more often with effector targets than other loci, we evaluated the AAP in the same way than the combined DΘ ranking. We sorted the loci descending by their number of AAPs and determined the cumulative number of interacting effector-interactors. To calculate a P-value we compared the real value against the distribution of number of effector targets from

1,000 rewired networks. P-value calculation and network rewiring was conducted as for evaluation of combined ranking of DΘ.

Bibliography

[1] W. J. Robbins. THE IMPORTANCE OF PLANTS. *Science*, 100(2603):440–443, November 1944. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.100.2603.440, doi:10.1126/science.100.2603.440.

[2] O. Bahcall, A. K. Eggleston, and S. Shadan. Plants. *Nature*, 543:327, March 2017. URL: https://doi.org/10.1038/543327a.

[3] X.-Z. Liang, Y. Wu, R. G. Chambers, D. L. Schmoldt, W. Gao, C. Liu, Y.-A. Liu, C. Sun, and J. A. Kennedy. Determining climate effects on US total agricultural productivity. *Proceedings of the National Academy of Sciences*, 114(12):E2285–E2292, March 2017. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1615922114, doi:10.1073/pnas.1615922114.

[4] M. V. Mickelbart, P. M. Hasegawa, and J. Bailey-Serres. Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. *Nature Reviews Genetics*, 16(4):237–251, April 2015. URL: http://www.nature.com/articles/nrg3901, doi:10.1038/nrg3901.

[5] B. Scheres and W. H. van der Putten. The plant perceptron connects environment to development. *Nature*, 543(7645):337–345, March 2017. URL: http://www.nature.com/articles/nature22010, doi:10.1038/nature22010.

[6] M. Biasiolo, M. Forcato, L. Possamai, F. Ferrari, L. Agnelli, M. Lionetti, K. Todoerti, A. Neri, M. Marchiori, S. Bortoluzzi, and S. Bicciato. Critical analysis of transcriptional and post-transcriptional regulatory networks in multiple myeloma. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, pages 397–408, 2010.

[7] A.-L. Barabási and Z. N. Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, February 2004. URL: http://www.nature.com/articles/nrg1272, doi:10.1038/nrg1272.

[8] P. Braun and A.-C. Gingras. History of protein-protein interactions: from egg-white to complex networks. *Proteomics*, 12(10):1478–1498, May 2012. doi:10.1002/pmic.201100563.

[9] V. S. Rao, K. Srinivas, G. N. Sujini, and G. N. S. Kumar. Protein-protein interaction detection: methods and analysis. *International Journal of Proteomics*, 2014:147648, 2014. doi:10.1155/2014/147648.

[10] B. Suter, S. Kittanakom, and I. Stagljar. Two-hybrid technologies in proteomics research. *Current Opinion in Biotechnology*, 19(4):316–323, August 2008. URL: https://linkinghub.elsevier.com/retrieve/pii/S095816690800075X, doi:10.1016/j.copbio.2008.06.005.

[11] Arabidopsis Interactome Mapping Consortium. Evidence for network evolution in an Arabidopsis interactome map. *Science (New York, N.Y.)*, 333(6042):601–607, July 2011. doi:10.1126/science.1203877.

[12] K. Venkatesan, J.-F. Rual, A. Vazquez, U. Stelzl, I. Lemmens, T. Hirozane-Kishikawa, T. Hao, M. Zenkner, X. Xin, K.-I. Goh, M. A. Yildirim, N. Simonis, K. Heinzmann, F. Gebreab, J. M. Sahalie, S. Cevik, C. Simon, A.-S. d. Smet, E. Dann, A. Smolyar, A. Vinayagam, H. Yu, D. Szeto, H. Borick, A. Dricot, N. Klitgord, R. R. Murray, C. Lin, M. Lalowski, J. Timm, K. Rau, C. Boone, P. Braun, M. E. Cusick, F. P. Roth, D. E. Hill, J. Tavernier, E. E. Wanker, A.-L. Barabási, and M. Vidal. An empirical framework for binary interactome mapping. *Nature Methods*, 6(1):83–90, January 2009. URL: https://www.nature.com/articles/nmeth.1280, doi:10.1038/nmeth.1280.

[13] P. N. Hengen. False positives from the yeast two-hybrid system. *Trends in Biochemical Sciences*, 22(1):33–34, January 1997.

[14] A. Brückner, C. Polge, N. Lentze, D. Auerbach, and U. Schlattner. Yeast Two-Hybrid, a Powerful Tool for Systems Biology. *International Journal of Molecular Sciences*, 10(6):2763–2788, June 2009. URL: http://www.mdpi.com/1422-0067/10/6/2763, doi:10.3390/ijms10062763.

[15] P. Braun, M. Tasan, M. Dreze, M. Barrios-Rodiles, I. Lemmens, H. Yu, J. M. Sahalie, R. R. Murray, L. Roncari, A.-S. de Smet, K. Venkatesan, J.-F. Rual, J. Vandenhaute, M. E. Cusick, T. Pawson, D. E. Hill, J. Tavernier, J. L. Wrana, F. P. Roth, and M. Vidal. An experimentally derived confidence score for binary protein-protein interactions. *Nature Methods*, 6(1):91–97, January 2009. doi:10.1038/nmeth.1281.

[16] P. Braun. Interactome mapping for analysis of complex phenotypes: insights from benchmarking binary interaction assays. *Proteomics*, 12(10):1499–1518, May 2012. doi:10.1002/pmic.201100598.

[17] H. Yu, P. Braun, M. A. Yıldırım, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, N. Simonis, T. Hao, J.-F. Rual, A. Dricot, A. Vazquez, R. R. Murray, C. Simon, L. Tardivo, S. Tam, N. Svrzikapa, C. Fan, A.-S. d. Smet, A. Motyl, M. E. Hudson, J. Park, X. Xin, M. E. Cusick, T. Moore, C. Boone, M. Snyder, F. P. Roth, A.-L. Barabási, J. Tavernier, D. E. Hill, and M. Vidal. High-Quality Binary Protein Interaction Map of the Yeast Interactome Network. *Science*, 322(5898):104–110, October 2008. URL: http://science.sciencemag.org/content/322/5898/104, doi:10.1126/science.1158684.

[18] A.-L. Barabási and M. Pósfai. *Network science*. Cambridge University Press, Cambridge, 2016. URL: http://barabasi.com/networksciencebook/.

[19] T. Yamada and P. Bork. Evolution of biomolecular networks — lessons from metabolic and protein interactions. *Nature Reviews Molecular Cell Biology*, 10(11):791–803, November 2009. URL: http://www.nature.com/articles/nrm2787, doi:10.1038/nrm2787.

[20] L. C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, January 1978. URL: http://linkinghub.elsevier.com/retrieve/pii/0378873378900217, doi:10.1016/0378-8733(78)90021-7.

[21] C. Wang, Y. Liu, S.-S. Li, and G.-Z. Han. Insights into the Origin and Evolution of the Plant Hormone Signaling Machinery. *Plant Physiology*, 167(3):872–886, March 2015. URL: http://www.plantphysiol.org/lookup/doi/10.1104/pp.114.247403, doi:10.1104/pp.114.247403.

[22] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406:378, July 2000. URL: https://doi.org/10.1038/35019019.

[23] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, October 2000. doi:10.1038/35036627.

[24] H. Jeong, S. P. Mason, A. L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–42, May 2001. doi:10.1038/35075138.

[25] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, June 2002. URL: http://www.pnas.org/content/99/12/7821, doi:10.1073/pnas.122653799.

[26] W. W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33:452–473, 1977.

[27] M. Latapy and P. Pons. Computing communities in large networks using random walks. *arXiv:cond-mat/0412368*, December 2004. arXiv: cond-mat/0412368. URL: http://arxiv.org/abs/cond-mat/0412368.

[28] M. Rosvall and C. T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, January 2008. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.0706851105, doi:10.1073/pnas.0706851105.

[29] A. Clauset, M. E. J. Newman, and C. Moore. Finding community structure in very large networks. *Physical Review E*, 70(6), December 2004. URL: https://link.aps.org/doi/10.1103/PhysRevE.70.066111, doi:10.1103/PhysRevE.70.066111.

[30] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, October 2008. URL: http://stacks.iop.org/1742-5468/2008/i=10/a=P10008?key=crossref.46968f6ec61eb8f907a760be1c5ace52, doi:10.1088/1742-5468/2008/10/P10008.

[31] U. N. Raghavan, R. Albert, and S. Kumara. Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, 76(3), September 2007. URL: https://link.aps.org/doi/10.1103/PhysRevE.76.036106, doi:10.1103/PhysRevE.76.036106.

[32] M. Meilă. Comparing Clusterings by the Variation of Information. In B. Schölkopf and M. K. Warmuth, editors, *Learning Theory and Kernel Machines*, pages 173–187. Springer Berlin Heidelberg, 2003.

[33] J. I. Fuxman Bass, A. Diallo, J. Nelson, J. M. Soto, C. L. Myers, and A. J. M. Walhout. Using networks to measure similarity between genes: association index selection. *Nature Methods*, 10(12):1169–1176, December 2013. doi:10.1038/nmeth.2728.

[34] S. Philippi. Data and knowledge integration in the life sciences. *Briefings in Bioinformatics*, 9(6):451–451, July 2008. URL: https://academic.oup.com/bib/article-lookup/doi/10.1093/bib/bbn046, doi:10.1093/bib/bbn046.

[35] D. Gomez-Cabrero, I. Abugessaisa, D. Maier, A. Teschendorff, M. Merkenschlager, A. Gisel, E. Ballestar, E. Bongcam-Rudloff, A. Conesa, and J. Tegnér. Data integration in the era of omics: current and future challenges. *BMC Systems Biology*, 8(Suppl 2):I1, 2014. URL: http://bmcsystbiol.biomedcentral.com/articles/10.1186/1752-0509-8-S2-I1, doi:10.1186/1752-0509-8-S2-I1.

[36] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, 25(1):25–29, May 2000. doi:10.1038/75556.

[37] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–2504, November 2003. doi:10.1101/gr.1239303.

[38] S. Maere, K. Heymans, and M. Kuiper. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics (Oxford, England)*, 21(16):3448–3449, August 2005. doi:10.1093/bioinformatics/bti551.

[39] G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.-H. Fridman, F. Pagès, Z. Trajanoski, and J. Galon. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics (Oxford, England)*, 25(8):1091–1093, April 2009. doi:10.1093/bioinformatics/btp101.

[40] F. Al-Shahrour, R. Díaz-Uriarte, and J. Dopazo. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics (Oxford, England)*, 20(4):578–580, March 2004. doi:10.1093/bioinformatics/btg455.

[41] F. Al-Shahrour, P. Minguez, J. Tárraga, I. Medina, E. Alloza, D. Montaner, and J. Dopazo. FatiGO +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. *Nucleic Acids Research*, 35(Web Server issue):W91–96, July 2007. doi:10.1093/nar/gkm260.

[42] G. Dennis, B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane, and R. A. Lempicki. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology*, 4(5):P3, 2003.

[43] T. Beissbarth and T. P. Speed. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics (Oxford, England)*, 20(9):1464–1465, June 2004. doi:10.1093/bioinformatics/bth088.

[44] E. Eden, R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, 10(1), December 2009. URL: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-10-48, doi:10.1186/1471-2105-10-48.

[45] M. Morgan, S. Falcon, and R. Gentleman. *GSEABase: Gene set enrichment data structures and methods*. 2018.

[46] A. Alexa and J. Rahnenfuhrer. *topGO: Enrichment Analysis for Gene Ontology*. 2018.

[47] F. Supek, M. Bošnjak, N. Škunca, and T. Šmuc. REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS One*, 6(7):e21800, 2011. doi:10.1371/journal.pone.0021800.

[48] Z. Wang, M. Gerstein, and M. Snyder. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews. Genetics*, 10(1):57–63, January 2009. doi:10.1038/nrg2484.

[49] S. Pepke, B. Wold, and A. Mortazavi. Computation for ChIP-seq and RNA-seq studies. *Nature Methods*, 6(11):S22–S32, November 2009. URL: http://www.nature.com/articles/nmeth.1371, doi:10.1038/nmeth.1371.

[50] C. Soneson and M. Delorenzi. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, 14(1), December 2013. URL: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-91, doi:10.1186/1471-2105-14-91.

[51] A. Celaj, U. Schlecht, J. D. Smith, W. Xu, S. Suresh, M. Miranda, A. M. Aparicio, M. Proctor, R. W. Davis, F. P. Roth, and R. P. St.Onge. Quantitative analysis of protein interaction network dynamics in yeast. *Molecular Systems Biology*, 13(7):934, July 2017. URL: http://msb.embopress.org/lookup/doi/10.15252/msb.20177532, doi:10.15252/msb.20177532.

[52] M. Bamshad and S. P. Wooding. Signatures of natural selection in the human genome. *Nature Reviews Genetics*, 4(2):99–110, February 2003. URL: http://www.nature.com/articles/nrg999, doi:10.1038/nrg999.

[53] R. Nielsen. Molecular Signatures of Natural Selection. *Annual Review of Genetics*, 39(1):197–218, December 2005. URL: http://www.annualreviews.org/doi/10.1146/annurev.genet.39.073003.112420, doi:10.1146/annurev.genet.39.073003.112420.

[54] J. J. Vitti, S. R. Grossman, and P. C. Sabeti. Detecting Natural Selection in Genomic Data. *Annual Review of Genetics*, 47(1):97–120, November 2013. URL: http://www.annualreviews.org/doi/10.1146/annurev-genet-111212-133526, doi:10.1146/annurev-genet-111212-133526.

[55] P. Luisi, D. Alvarez-Ponce, M. Pybus, M. A. Fares, J. Bertranpetit, and H. Laayouni. Recent Positive Selection Has Acted on Genes Encoding Proteins with More Interactions within the Whole Human Interactome. *Genome Biology and Evolution*, 7(4):1141–1154, April 2015. URL: https://academic.oup.com/gbe/article-lookup/doi/10.1093/gbe/evv055, doi:10.1093/gbe/evv055.

[56] L. Excoffier and G. Heckel. Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics*, 7(10):745–758, October 2006. URL: http://www.nature.com/articles/nrg1904, doi:10.1038/nrg1904.

[57] E. Paradis, T. Gosselin, J. Goudet, T. Jombart, and K. Schliep. Linking genomics and population genetics with R. *Molecular Ecology Resources*, 17(1):54–66, January 2017. URL: http://doi.wiley.com/10.1111/1755-0998.12577, doi:10.1111/1755-0998.12577.

[58] C. Alonso-Blanco, J. Andrade, C. Becker, F. Bemm, J. Bergelson, K. M. Borgwardt, J. Cao, E. Chae, T. M. Dezwaan, W. Ding, J. R. Ecker, M. Expósito-Alonso, A. Farlow, J. Fitz, X. Gan, D. G. Grimm, A. M. Hancock, S. R. Henz, S. Holm, M. Horton, M. Jarsulic, R. A. Kerstetter, A. Korte, P. Korte, C. Lanz, C.-R. Lee, D. Meng, T. P. Michael, R. Mott, N. W. Muliyati, T. Nägele, M. Nagler, V. Nizhynska, M. Nordborg, P. Y. Novikova, F. X. Picó, A. Platzer, F. A. Rabanal, A. Rodriguez, B. A. Rowan, P. A. Salomé, K. J. Schmid, R. J. Schmitz, m. Seren, F. G. Sperone, M. Sudkamp, H. Svardal, M. M. Tanzer, D. Todd, S. L. Volchenboum, C. Wang, G. Wang, X. Wang, W. Weckwerth, D. Weigel, and X. Zhou. 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell*, 166(2):481–491, July 2016. URL: https://linkinghub.elsevier.com/retrieve/pii/S0092867416306675, doi:10.1016/j.cell.2016.05.063.

[59] M. Koornneef, C. Alonso-Blanco, and D. Vreugdenhil. NATURALLY OCCURRING GENETIC VARIATION IN *ARABIDOPSIS THALIANA*. *Annual Review of Plant Biology*, 55(1):141–172, June 2004. URL: http://www.annualreviews.org/doi/10.1146/annurev.arplant.55.031903.141605, doi:10.1146/annurev.arplant.55.031903.141605.

[60] A. Fournier-Level, A. Korte, M. D. Cooper, M. Nordborg, J. Schmitt, and A. M. Wilczek. A Map of Local Adaptation in Arabidopsis thaliana. *Science*, 334(6052):86–89, October 2011. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1209271, doi:10.1126/science.1209271.

[61] A. M. Hancock, B. Brachi, N. Faure, M. W. Horton, L. B. Jarymowycz, F. G. Sperone, C. Toomajian, F. Roux, and J. Bergelson. Adaptation to Climate Across the Arabidopsis thaliana Genome. *Science*, 334(6052):83–86, October 2011. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1209244, doi:10.1126/science.1209244.

[62] J. Cao, K. Schneeberger, S. Ossowski, T. Günther, S. Bender, J. Fitz, D. Koenig, C. Lanz, O. Stegle, C. Lippert, X. Wang, F. Ott, J. Müller, C. Alonso-Blanco, K. Borgwardt, K. J. Schmid, and D. Weigel. Whole-genome sequencing of multiple Arabidopsis thaliana populations. *Nature Genetics*, 43(10):956–963, August 2011. doi:10.1038/ng.911.

[63] Q. Long, F. A. Rabanal, D. Meng, C. D. Huber, A. Farlow, A. Platzer, Q. Zhang, B. J. Vilhjálmsson, A. Korte, V. Nizhynska, V. Voronin, P. Korte, L. Sedman, T. Mandáková, M. A. Lysak, m. Seren, I. Hellmann, and M. Nordborg. Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. *Nature Genetics*, 45(8):884–890, August 2013. URL: http://www.nature.com/articles/ng.2678, doi:10.1038/ng.2678.

[64] E. Sasaki, P. Zhang, S. Atwell, D. Meng, and M. Nordborg. "Missing" G x E Variation Controls Flowering Time in Arabidopsis thaliana. *PLOS Genetics*, 11(10):e1005597, October 2015. URL: https://dx.plos.org/10.1371/journal.pgen.1005597, doi:10.1371/journal.pgen.1005597.

[65] C. Alonso-Blanco, M. G. Aarts, L. Bentsink, J. J. Keurentjes, M. Reymond, D. Vreugdenhil, and M. Koornneef. What Has Natural Variation Taught Us about Plant Development, Physiology, and Adaptation? *THE PLANT CELL ONLINE*, 21(7):1877–1896, July 2009. URL: http://www.plantcell.org/cgi/doi/10.1105/tpc.109.068114, doi:10.1105/tpc.109.068114.

[66] A. Nemri, S. Atwell, A. M. Tarone, Y. S. Huang, K. Zhao, D. J. Studholme, M. Nordborg, and J. D. G. Jones. Genome-wide survey of Arabidopsis natural variation in downy mildew resistance using combined association and linkage mapping. *Proceedings of the National Academy of Sciences*, 107(22):10302–10307, June 2010. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.0913160107, doi:10.1073/pnas.0913160107.

[67] E. Puerma and M. Aguadé. Polymorphism at genes involved in salt tolerance in *Arabidopsis thaliana* (Brassicaceae). *American Journal of Botany*, 100(2):384–390, February 2013. URL: http://doi.wiley.com/10.3732/ajb.1200332, doi:10.3732/ajb.1200332.

[68] G. Bauchet, S. Munos, C. Sauvage, J. Bonnet, L. Grivet, and M. Causse. Genes involved in floral meristem in tomato exhibit drastically reduced genetic diversity and signature of selection. *BMC Plant Biology*, 14(1), December 2014. URL: http://bmcplantbiol.biomedcentral.com/articles/10.1186/s12870-014-0279-2, doi:10.1186/s12870-014-0279-2.

[69] A. Haudry, A. E. Platts, E. Vello, D. R. Hoen, M. Leclercq, R. J. Williamson, E. Forczek, Z. Joly-Lopez, J. G. Steffen, K. M. Hazzouri, K. Dewar, J. R. Stinchcombe, D. J. Schoen, X. Wang, J. Schmutz, C. D. Town, P. P. Edger, J. C. Pires, K. S. Schumaker, D. E. Jarvis, T. Mandáková, M. A. Lysak, E. van den Bergh, M. E. Schranz, P. M. Harrison, A. M. Moses, T. E. Bureau, S. I. Wright, and M. Blanchette. An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nature Genetics*, 45(8):891–898, August 2013. URL: http://www.nature.com/articles/ng.2684, doi:10.1038/ng.2684.

[70] H. B. Fraser, D. P. Wall, and A. E. Hirsh. A simple dependence between protein evolution rate and the number of protein-protein interactions. *BMC evolutionary biology*, 3:11, May 2003. doi:10.1186/1471-2148-3-11.

[71] F. Leal Valentim, F. Neven, P. Boyen, and A. D. J. van Dijk. Interactome-Wide Prediction of Protein-Protein Binding Sites Reveals Effects of Protein Sequence Variation in Arabidopsis thaliana. *PLoS ONE*, 7(10):e47022, October 2012. URL: https://dx.plos.org/10.1371/journal.pone.0047022, doi:10.1371/journal.pone.0047022.

[72] P. M. Kim, J. O. Korbel, and M. B. Gerstein. Positive selection at the protein network periphery: Evaluation in terms of structural constraints and cellular context. *Proceedings of the National Academy of Sciences*, 104(51):20274–20279, December 2007. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.0710183104, doi:10.1073/pnas.0710183104.

[73] A.-L. Barabási, N. Gulbahce, and J. Loscalzo. Network medicine: a network-based approach to human disease. *Nature Reviews. Genetics*, 12(1):56–68, January 2011. doi:10.1038/nrg2918.

[74] L. Garcia-Alonso, J. Jimenez-Almazan, J. Carbonell-Caballero, A. Vela-Boza, J. Santoyo-Lopez, G. Antinolo, and J. Dopazo. The role of the interactome in the maintenance of deleterious variability in human populations. *Molecular Systems Biology*, 10(9):752–752, September 2014. URL: http://msb.embopress.org/cgi/doi/10.15252/msb.20145222, doi:10.15252/msb.20145222.

[75] A. Santner, L. I. A. Calderon-Villalobos, and M. Estelle. Plant hormones are versatile chemical regulators of plant growth. *Nature Chemical Biology*, 5(5):301–307, May 2009. doi:10.1038/nchembio.165.

[76] P. L. Rodriguez, G. Benning, and E. Grill. ABI2, a second protein phosphatase 2c involved in abscisic acid signal transduction in *Arabidopsis*. *FEBS Letters*, 421(3):185–190, January 1998. URL: http://doi.wiley.com/10.1016/S0014-5793%2897%2901558-5, doi:10.1016/S0014-5793(97)01558-5.

[77] K. Melcher, L.-M. Ng, X. E. Zhou, F.-F. Soon, Y. Xu, K. M. Suino-Powell, S.-Y. Park, J. J. Weiner, H. Fujii, V. Chinnusamy, A. Kovach, J. Li, Y. Wang, J. Li, F. C. Peterson, D. R. Jensen, E.-L. Yong, B. F. Volkman, S. R. Cutler, J.-K. Zhu, and H. E. Xu. A gate–latch–lock mechanism for hormone signalling by abscisic acid receptors. *Nature*, 462(7273):602–608, December 2009. URL: http://www.nature.com/articles/nature08613, doi:10.1038/nature08613.

[78] S.-Y. Park, P. Fung, N. Nishimura, D. R. Jensen, H. Fujii, Y. Zhao, S. Lumba, J. Santiago, A. Rodrigues, T.-f. F. Chow, S. E. Alfred, D. Bonetta, R. Finkelstein, N. J. Provart, D. Desveaux, P. L. Rodriguez, P. McCourt, J.-K. Zhu, J. I. Schroeder, B. F. Volkman, and S. R. Cutler. Abscisic Acid Inhibits Type 2c Protein Phosphatases via the PYR/PYL Family of START Proteins. *Science*, April 2009. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1173041, doi:10.1126/science.1173041.

[79] H. Fujii, P. E. Verslues, and J.-K. Zhu. Identification of Two Protein Kinases Required for Abscisic Acid Regulation of Seed Germination, Root Growth, and

Gene Expression in Arabidopsis. *THE PLANT CELL ONLINE*, 19(2):485–494, February 2007. URL: http://www.plantcell.org/cgi/doi/10.1105/tpc.106.048538, doi:10.1105/tpc.106.048538.

[80] J. Li, J. Wen, K. A. Lease, J. T. Doke, F. E. Tax, and J. C. Walker. BAK1, an Arabidopsis LRR Receptor-like Protein Kinase, Interacts with BRI1 and Modulates Brassinosteroid Signaling. *Cell*, 110(2):213–222, July 2002. URL: https://linkinghub.elsevier.com/retrieve/pii/S0092867402008127, doi:10.1016/S0092-8674(02)00812-7.

[81] S. Mora-Garcia. Nuclear protein phosphatases with Kelch-repeat domains modulate the response to brassinosteroids in Arabidopsis. *Genes & Development*, 18(4):448–460, February 2004. URL: http://www.genesdev.org/cgi/doi/10.1101/gad.1174204, doi:10.1101/gad.1174204.

[82] G. Vert and J. Chory. Downstream nuclear events in brassinosteroid signalling. *Nature*, 441(7089):96–100, May 2006. URL: http://www.nature.com/articles/nature04681, doi:10.1038/nature04681.

[83] T. Inoue, M. Higuchi, Y. Hashimoto, M. Seki, M. Kobayashi, T. Kato, S. Tabata, K. Shinozaki, and T. Kakimoto. Identification of CRE1 as a cytokinin receptor from Arabidopsis. *Nature*, 409(6823):1060–1063, February 2001. URL: http://www.nature.com/articles/35059117, doi:10.1038/35059117.

[84] T. Suzuki, K. Miwa, K. Ishikawa, H. Yamada, H. Aiba, and T. Mizuno. The Arabidopsis Sensor His-kinase, AHK4, Can Respond to Cytokinins. *Plant and Cell Physiology*, 42(2):107–113, February 2001. URL: http://academic.oup.com/pcp/article/42/2/107/1930040/The-Arabidopsis-Sensor-Hiskinase-AHK4-Can-Respond, doi:10.1093/pcp/pce037.

[85] I. Hwang and J. Sheen. Two-component circuitry in Arabidopsis cytokinin signal transduction. *Nature*, 413(6854):383–389, September 2001. URL: http://www.nature.com/articles/35096500, doi:10.1038/35096500.

[86] C. E. Hutchison, J. Li, C. Argueso, M. Gonzalez, E. Lee, M. W. Lewis, B. B. Maxwell, T. D. Perdue, G. E. Schaller, J. M. Alonso, J. R. Ecker, and J. J. Kieber. The Arabidopsis Histidine Phosphotransfer Proteins Are Redundant Positive Regulators of Cytokinin Signaling. *THE PLANT CELL ONLINE*, 18(11):3073–3087, November 2006. URL: http://www.plantcell.org/cgi/doi/10.1105/tpc.106.045674, doi:10.1105/tpc.106.045674.

[87] H. Sakai. ARR1, a Transcription Factor for Genes Immediately Responsive to Cytokinins. *Science*, 294(5546):1519–1521, November 2001. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1065201, doi:10.1126/science.1065201.

[88] G. E. Schaller and A. B. Bleecker. Ethylene-binding sites generated in yeast expressing the Arabidopsis ETR1 gene. *Science (New York, N.Y.)*, 270(5243):1809–1811, December 1995.

[89] H. Qiao, Z. Shen, S.-s. C. Huang, R. J. Schmitz, M. A. Urich, S. P. Briggs, and J. R. Ecker. Processing and Subcellular Trafficking of ER-Tethered EIN2 Control Response to Ethylene Gas. *Science*, 338(6105):390–393, October 2012. URL: `http://www.sciencemag.org/cgi/doi/10.1126/science.1225974`, `doi:10.1126/science.1225974`.

[90] H. Qiao, K. N. Chang, J. Yazaki, and J. R. Ecker. Interplay between ethylene, ETP1/ETP2 F-box proteins, and degradation of EIN2 triggers ethylene responses in Arabidopsis. *Genes & Development*, 23(4):512–521, February 2009. URL: `http://genesdev.cshlp.org/cgi/doi/10.1101/gad.1765709`, `doi:10.1101/gad.1765709`.

[91] H. Guo and J. R. Ecker. Plant responses to ethylene gas are mediated by SCF(EBF1/EBF2)-dependent proteolysis of EIN3 transcription factor. *Cell*, 115(6):667–677, December 2003.

[92] T. Potuschak, E. Lechner, Y. Parmentier, S. Yanagisawa, S. Grava, C. Koncz, and P. Genschik. EIN3-dependent regulation of plant ethylene hormone signaling by two arabidopsis F box proteins: EBF1 and EBF2. *Cell*, 115(6):679–689, December 2003.

[93] Z. Q. Fu, S. Yan, A. Saleh, W. Wang, J. Ruble, N. Oka, R. Mohan, S. H. Spoel, Y. Tada, N. Zheng, and X. Dong. NPR3 and NPR4 are receptors for the immune signal salicylic acid in plants. *Nature*, 486(7402):228–232, June 2012. URL: `http://www.nature.com/articles/nature11162`, `doi:10.1038/nature11162`.

[94] B. Ülker and I. E. Somssich. WRKY transcription factors: from DNA binding towards biological function. *Current Opinion in Plant Biology*, 7(5):491–498, October 2004. URL: `https://linkinghub.elsevier.com/retrieve/pii/S1369526604001049`, `doi:10.1016/j.pbi.2004.07.012`.

[95] K. Murase, Y. Hirano, T.-p. Sun, and T. Hakoshima. Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. *Nature*, 456(7221):459–463, November 2008. URL: `http://www.nature.com/articles/nature07519`, `doi:10.1038/nature07519`.

[96] A. Shimada, M. Ueguchi-Tanaka, T. Nakatsu, M. Nakajima, Y. Naoe, H. Ohmiya, H. Kato, and M. Matsuoka. Structural basis for gibberellin recognition by its receptor GID1. *Nature*, 456(7221):520–523, November 2008. URL: `http://www.nature.com/articles/nature07546`, `doi:10.1038/nature07546`.

[97] A. Santner and M. Estelle. Recent advances and emerging trends in plant hormone signalling. *Nature*, 459(7250):1071–1078, June 2009. URL: `http://www.nature.com/articles/nature08122`, `doi:10.1038/nature08122`.

[98] A. Larrieu and T. Vernoux. Comparison of plant hormone signalling systems. *Essays In Biochemistry*, 58(0):165–181, September 2015. URL: `http://essays.biochemistry.org/cgi/doi/10.1042/bse0580165`, `doi:10.1042/bse0580165`.

[99] G. R. Flematti, K. W. Dixon, and S. M. Smith. What are karrikins and how were they 'discovered' by plants? *BMC Biology*, 13(1), December 2015. URL: `http:`

//bmcbiol.biomedcentral.com/articles/10.1186/s12915-015-0219-0, doi:10.1186/s12915-015-0219-0.

[100] S. M. Smith and J. Li. Signalling and responses to strigolactones and karrikins. *Current Opinion in Plant Biology*, 21:23–29, October 2014. URL: https://linkinghub.elsevier.com/retrieve/pii/S1369526614000880, doi:10.1016/j.pbi.2014.06.003.

[101] F. Zhou, Q. Lin, L. Zhu, Y. Ren, K. Zhou, N. Shabek, F. Wu, H. Mao, W. Dong, L. Gan, W. Ma, H. Gao, J. Chen, C. Yang, D. Wang, J. Tan, X. Zhang, X. Guo, J. Wang, L. Jiang, X. Liu, W. Chen, J. Chu, C. Yan, K. Ueno, S. Ito, T. Asami, Z. Cheng, J. Wang, C. Lei, H. Zhai, C. Wu, H. Wang, N. Zheng, and J. Wan. D14–SCFD3-dependent degradation of D53 regulates strigolactone signalling. *Nature*, 504(7480):406–410, December 2013. URL: http://www.nature.com/articles/nature12878, doi:10.1038/nature12878.

[102] A. de Saint Germain, S. Bonhomme, F.-D. Boyer, and C. Rameau. Novel insights into strigolactone distribution and signalling. *Current Opinion in Plant Biology*, 16(5):583–589, October 2013. URL: https://linkinghub.elsevier.com/retrieve/pii/S1369526613000903, doi:10.1016/j.pbi.2013.06.007.

[103] N. Dharmasiri, S. Dharmasiri, and M. Estelle. The F-box protein TIR1 is an auxin receptor. *Nature*, 435(7041):441–445, May 2005. URL: http://www.nature.com/articles/nature03543, doi:10.1038/nature03543.

[104] J. A. Long. TOPLESS Regulates Apical Embryonic Fate in Arabidopsis. *Science*, 312(5779):1520–1523, June 2006. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1123841, doi:10.1126/science.1123841.

[105] F. dos Santos Maraschin, J. Memelink, and R. Offringa. Auxin-induced, SCF ^{tir1} -mediated poly-ubiquitination marks AUX/IAA proteins for degradation. *The Plant Journal*, 59(1):100–109, July 2009. URL: http://doi.wiley.com/10.1111/j.1365-313X.2009.03854.x, doi:10.1111/j.1365-313X.2009.03854.x.

[106] L. B. Sheard, X. Tan, H. Mao, J. Withers, G. Ben-Nissan, T. R. Hinds, Y. Kobayashi, F.-F. Hsu, M. Sharon, J. Browse, S. Y. He, J. Rizo, G. A. Howe, and N. Zheng. Jasmonate perception by inositol-phosphate-potentiated COI1–JAZ co-receptor. *Nature*, 468(7322):400–405, November 2010. URL: http://www.nature.com/articles/nature09430, doi:10.1038/nature09430.

[107] L. Pauwels, G. F. Barbero, J. Geerinck, S. Tilleman, W. Grunewald, A. C. Pérez, J. M. Chico, R. V. Bossche, J. Sewell, E. Gil, G. García-Casado, E. Witters, D. Inzé, J. A. Long, G. De Jaeger, R. Solano, and A. Goossens. NINJA connects the co-repressor TOPLESS to jasmonate signalling. *Nature*, 464(7289):788–791, April 2010. URL: http://www.nature.com/articles/nature08854, doi:10.1038/nature08854.

[108] V. Verma, P. Ravindran, and P. P. Kumar. Plant hormone-mediated regulation of stress responses. *BMC plant biology*, 16:86, April 2016. doi:10.1186/s12870-016-0771-y.

[109] S. H. Spoel. NPR1 Modulates Cross-Talk between Salicylate- and Jasmonate-Dependent Defense Pathways through a Novel Function in the Cytosol. *THE PLANT CELL ONLINE*, 15(3):760–770, March 2003. URL: http://www.plantcell.org/cgi/doi/10.1105/tpc.009159, doi:10.1105/tpc.009159.

[110] J. Li. The WRKY70 Transcription Factor: A Node of Convergence for Jasmonate-Mediated and Salicylate-Mediated Signals in Plant Defense. *THE PLANT CELL ONLINE*, 16(2):319–331, February 2004. URL: http://www.plantcell.org/cgi/doi/10.1105/tpc.016980, doi:10.1105/tpc.016980.

[111] D. Wang, K. Pajerowska-Mukhtar, A. H. Culler, and X. Dong. Salicylic Acid Inhibits Pathogen Growth in Plants through Repression of the Auxin Signaling Pathway. *Current Biology*, 17(20):1784–1790, October 2007. URL: https://linkinghub.elsevier.com/retrieve/pii/S0960982207019847, doi:10.1016/j.cub.2007.09.025.

[112] Z. Zhu, F. An, Y. Feng, P. Li, L. Xue, M. A, Z. Jiang, J.-M. Kim, T. K. To, W. Li, X. Zhang, Q. Yu, Z. Dong, W.-Q. Chen, M. Seki, J.-M. Zhou, and H. Guo. Derepression of ethylene-stabilized transcription factors (EIN3/EIL1) mediates jasmonate and ethylene signaling synergy in Arabidopsis. *Proceedings of the National Academy of Sciences*, 108(30):12539–12544, July 2011. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1103959108, doi:10.1073/pnas.1103959108.

[113] O. Lorenzo. ETHYLENE RESPONSE FACTOR1 Integrates Signals from Ethylene and Jasmonate Pathways in Plant Defense. *THE PLANT CELL ONLINE*, 15(1):165–178, January 2003. URL: http://www.plantcell.org/cgi/doi/10.1105/tpc.007468, doi:10.1105/tpc.007468.

[114] A. N. Stepanova. A Link between Ethylene and Auxin Uncovered by the Characterization of Two Root-Specific Ethylene-Insensitive Mutants in Arabidopsis. *THE PLANT CELL ONLINE*, 17(8):2230–2242, August 2005. URL: http://www.plantcell.org/cgi/doi/10.1105/tpc.105.033365, doi:10.1105/tpc.105.033365.

[115] A. N. Stepanova, J. Robertson-Hoyt, J. Yun, L. M. Benavente, D.-Y. Xie, K. Doležal, A. Schlereth, G. Jürgens, and J. M. Alonso. TAA1-Mediated Auxin Biosynthesis Is Essential for Hormone Crosstalk and Plant Development. *Cell*, 133(1):177–191, April 2008. URL: https://linkinghub.elsevier.com/retrieve/pii/S0092867408002122, doi:10.1016/j.cell.2008.01.047.

[116] Y. Tao, J.-L. Ferrer, K. Ljung, F. Pojer, F. Hong, J. A. Long, L. Li, J. E. Moreno, M. E. Bowman, L. J. Ivans, Y. Cheng, J. Lim, Y. Zhao, C. L. Ballaré, G. Sandberg, J. P. Noel, and J. Chory. Rapid Synthesis of Auxin via a New Tryptophan-Dependent Pathway Is Required for Shade Avoidance in Plants. *Cell*, 133(1):164–176, April 2008. URL: https://linkinghub.elsevier.com/retrieve/pii/S0092867408002146, doi:10.1016/j.cell.2008.01.049.

[117] A. Tsuchisaka. Unique and Overlapping Expression Patterns among the Arabidopsis 1-Amino-Cyclopropane-1-Carboxylate Synthase Gene Family Members. *PLANT PHYSIOLOGY*, 136(2):2982–3000, October 2004. URL: http://www.plantphysiol.org/cgi/doi/10.1104/pp.104.049999, doi:10.1104/pp.104.049999.

[118] P. Nagpal. Auxin response factors ARF6 and ARF8 promote jasmonic acid production and flower maturation. *Development*, 132(18):4107–4118, August 2005. URL: http://dev.biologists.org/cgi/doi/10.1242/dev.01955, doi:10.1242/dev.01955.

[119] L. Laplaze, E. Benkova, I. Casimiro, L. Maes, S. Vanneste, R. Swarup, D. Weijers, V. Calvo, B. Parizot, M. B. Herrera-Rodriguez, R. Offringa, N. Graham, P. Doumas, J. Friml, D. Bogusz, T. Beeckman, and M. Bennett. Cytokinins Act Directly on Lateral Root Founder Cells to Inhibit Root Initiation. *THE PLANT CELL ONLINE*, 19(12):3889–3900, December 2007. URL: http://www.plantcell.org/cgi/doi/10.1105/tpc.107.055863, doi:10.1105/tpc.107.055863.

[120] I. Blilou, J. Xu, M. Wildwater, V. Willemsen, I. Paponov, J. Friml, R. Heidstra, M. Aida, K. Palme, and B. Scheres. The PIN auxin efflux facilitator network controls growth and patterning in Arabidopsis roots. *Nature*, 433(7021):39–44, January 2005. URL: http://www.nature.com/articles/nature03184, doi:10.1038/nature03184.

[121] G. Vert, C. L. Walcher, J. Chory, and J. L. Nemhauser. Integration of auxin and brassinosteroid pathways by Auxin Response Factor 2. *Proceedings of the National Academy of Sciences of the United States of America*, 105(28):9829–9834, July 2008. doi:10.1073/pnas.0803996105.

[122] S. J. Unterholzner, W. Rozhon, M. Papacek, J. Ciomas, T. Lange, K. G. Kugler, K. F. Mayer, T. Sieberer, and B. Poppenberger. Brassinosteroids Are Master Regulators of Gibberellin Biosynthesis in Arabidopsis. *The Plant Cell*, 27(8):2261–2272, August 2015. URL: http://www.plantcell.org/lookup/doi/10.1105/tpc.15.00433, doi:10.1105/tpc.15.00433.

[123] Y. Jaillais and J. Chory. Unraveling the paradoxes of plant hormone signaling integration. *Nature Structural & Molecular Biology*, 17(6):642–645, June 2010. doi:10.1038/nsmb0610-642.

[124] Y. F. Huang, C. T. Chen, and C. H. Kao. Salicylic acid inhibits the biosynthesis of ethylene in detached rice leaves. *Plant Growth Regulation*, 12(1-2):79–82, January 1993. URL: http://link.springer.com/10.1007/BF00144586, doi:10.1007/BF00144586.

[125] C.-J. Jiang, M. Shimono, S. Sugano, M. Kojima, K. Yazawa, R. Yoshida, H. Inoue, N. Hayashi, H. Sakakibara, and H. Takatsuji. Abscisic Acid Interacts Antagonistically with Salicylic Acid Signaling Pathway in Rice– *Magnaporthe grisea* Interaction. *Molecular Plant-Microbe Interactions*, 23(6):791–798, June 2010. URL: http://apsjournals.apsnet.org/doi/10.1094/MPMI-23-6-0791, doi:10.1094/MPMI-23-6-0791.

[126] Y.-X. Yang, G. Ahammed, C. Wu, S.-y. Fan, and Y.-H. Zhou. Crosstalk among Jasmonate, Salicylate and Ethylene Signaling Pathways in Plant Disease and Immune Responses. *Current Protein & Peptide Science*, 16(5):450–461, May 2015. URL: http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1389-2037&volume=16&issue=5&spage=450, doi:10.2174/1389203716666150330141638.

[127] Q.-F. Li, C. Wang, L. Jiang, S. Li, S. S. M. Sun, and J.-X. He. An interaction between BZR1 and DELLAs mediates direct signaling crosstalk between brassinosteroids and gibberellins in Arabidopsis. *Science Signaling*, 5(244):ra72, October 2012. doi:10.1126/scisignal.2002908.

[128] J. M. Stone and J. C. Walker. Plant protein kinase families and signal transduction. *Plant Physiology*, 108(2):451–457, June 1995.

[129] T. Hunter. Why nature chose phosphate to modify proteins. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1602):2513–2516, September 2012. URL: http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2012.0013, doi:10.1098/rstb.2012.0013.

[130] M. Zulawski, G. Schulze, R. Braginets, S. Hartmann, and W. X. Schulze. The Arabidopsis Kinome: phylogeny and evolutionary insights into functional diversification. *BMC genomics*, 15:548, 2014. doi:10.1186/1471-2164-15-548.

[131] S. Reiland, G. Messerli, K. Baerenfaller, B. Gerrits, A. Endler, J. Grossmann, W. Gruissem, and S. Baginsky. Large-Scale Arabidopsis Phosphoproteome Profiling Reveals Novel Chloroplast Kinase Substrates and Phosphorylation Networks. *PLANT PHYSIOLOGY*, 150(2):889–903, June 2009. URL: http://www.plantphysiol.org/cgi/doi/10.1104/pp.109.138677, doi:10.1104/pp.109.138677.

[132] R. Linding, L. J. Jensen, G. J. Ostheimer, M. A. van Vugt, C. Jørgensen, I. M. Miron, F. Diella, K. Colwill, L. Taylor, K. Elder, P. Metalnikov, V. Nguyen, A. Pasculescu, J. Jin, J. G. Park, L. D. Samson, J. R. Woodgett, R. Russell, P. Bork, M. B. Yaffe, and T. Pawson. Systematic Discovery of In Vivo Phosphorylation Networks. *Cell*, 129(7):1415–1426, June 2007. URL: https://linkinghub.elsevier.com/retrieve/pii/S0092867407007271, doi:10.1016/j.cell.2007.05.052.

[133] J. A. Ubersax and J. E. Ferrell. Mechanisms of specificity in protein phosphorylation. *Nature Reviews. Molecular Cell Biology*, 8(7):530–541, July 2007. doi:10.1038/nrm2203.

[134] F. Huang, M. Kemel Zago, L. Abas, A. van Marion, C. S. Galván-Ampudia, and R. Offringa. Phosphorylation of Conserved PIN Motifs Directs *Arabidopsis* PIN1 Polarity and Auxin Transport. *The Plant Cell*, 22(4):1129–1142, April 2010. URL: http://www.plantcell.org/lookup/doi/10.1105/tpc.109.072678, doi:10.1105/tpc.109.072678.

[135] P. N. Dodds and J. P. Rathjen. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nature Reviews. Genetics*, 11(8):539–548, August 2010. doi:10.1038/nrg2812.

[136] J. L. Dangl, D. M. Horvath, and B. J. Staskawicz. Pivoting the plant immune system from dissection to deployment. *Science (New York, N.Y.)*, 341(6147):746–751, August 2013. doi:10.1126/science.1236011.

[137] G. Wang, U. Ellendorff, B. Kemp, J. W. Mansfield, A. Forsyth, K. Mitchell, K. Bastas, C.-M. Liu, A. Woods-Tor, C. Zipfel, P. J. de Wit, J. D. Jones, M. Tor, and B. P.

Thomma. A Genome-Wide Functional Investigation into the Roles of Receptor-Like Proteins in Arabidopsis. *PLANT PHYSIOLOGY*, 147(2):503–517, April 2008. URL: http://www.plantphysiol.org/cgi/doi/10.1104/pp.108.119487, doi:10.1104/pp.108.119487.

[138] J. D. G. Jones and J. L. Dangl. The plant immune system. *Nature*, 444(7117):323–329, November 2006. doi:10.1038/nature05286.

[139] M. S. Mukhtar, A.-R. Carvunis, M. Dreze, P. Epple, 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, European Union Effectoromics Consortium, J. Vandenhaute, F. P. Roth, D. E. Hill, J. R. Ecker, M. Vidal, J. Beynon, P. Braun, and J. L. Dangl. Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science (New York, N.Y.)*, 333(6042):596–601, July 2011. doi:10.1126/science.1203659.

[140] Z. Jiang, X. Liu, Z. Peng, Y. Wan, Y. Ji, W. He, W. Wan, J. Luo, and H. Guo. AHD2.0: an update version of Arabidopsis Hormone Database for plant systematic studies. *Nucleic Acids Research*, 39(Database issue):D1123–1129, January 2011. doi:10.1093/nar/gkq1066.

[141] M. Schmid, T. S. Davison, S. R. Henz, U. J. Pape, M. Demar, M. Vingron, B. Schölkopf, D. Weigel, and J. U. Lohmann. A gene expression map of Arabidopsis thaliana development. *Nature Genetics*, 37(5):501–506, May 2005. doi:10.1038/ng1543.

[142] R. Weßling, P. Epple, S. Altmann, Y. He, L. Yang, S. R. Henz, N. McDonald, K. Wiley, K. C. Bader, C. Gläßer, M. S. Mukhtar, S. Haigis, L. Ghamsari, A. E. Stephens, J. R. Ecker, M. Vidal, J. D. G. Jones, K. F. X. Mayer, E. Ver Loren van Themaat, D. Weigel, P. Schulze-Lefert, J. L. Dangl, R. Panstruga, and P. Braun. Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell Host & Microbe*, 16(3):364–375, September 2014. doi:10.1016/j.chom.2014.08.004.

[143] J.-F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D. S. Goldberg, L. V. Zhang, S. L. Wong, G. Franklin, S. Li, J. S. Albala, J. Lim, C. Fraughton, E. Llamosas, S. Cevik, C. Bex, P. Lamesch, R. S. Sikorski, J. Vandenhaute, H. Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M. E. Cusick, D. E. Hill, F. P. Roth, and M. Vidal. Towards a proteome-scale map of the human protein–protein interaction network. *Nature*, 437(7062):1173–1178, October 2005. URL: https://www.nature.com/articles/nature04209, doi:10.1038/nature04209.

[144] N. Simonis, J.-F. Rual, A.-R. Carvunis, M. Tasan, I. Lemmens, T. Hirozane-Kishikawa, T. Hao, J. M. Sahalie, K. Venkatesan, F. Gebreab, S. Cevik, N. Klitgord, C. Fan, P. Braun, N. Li, N. Ayivi-Guedehoussou, E. Dann, N. Bertin, D. Szeto, A. Dricot, M. A. Yildirim, C. Lin, A.-S. de Smet, H.-L. Kao, C. Simon, A. Smolyar, J. S. Ahn, M. Tewari, M. Boxem, S. Milstein, H. Yu, M. Dreze,

J. Vandenhaute, K. C. Gunsalus, M. E. Cusick, D. E. Hill, J. Tavernier, F. P. Roth, and M. Vidal. Empirically controlled mapping of the Caenorhabditis elegans protein-protein interactome network. *Nature Methods*, 6(1):47–54, January 2009.

[145] M. Dreze, D. Monachello, C. Lurin, M. E. Cusick, D. E. Hill, M. Vidal, and P. Braun. Chapter 12 - High-Quality Binary Interactome Mapping. In *Methods in Enzymology*, volume 470 of *Guide to Yeast Genetics: Functional Genomics, Proteomics, and Other Systems Analysis*, pages 281–315. Academic Press, January 2010. URL: `http://www.sciencedirect.com/science/article/pii/S0076687910700124`, `doi:10.1016/S0076-6879(10)70012-4`.

[146] S. Orchard, M. Ammari, B. Aranda, L. Breuza, L. Briganti, F. Broackes-Carter, N. H. Campbell, G. Chavali, C. Chen, N. del Toro, M. Duesbury, M. Dumousseau, E. Galeota, U. Hinz, M. Iannuccelli, S. Jagannathan, R. Jimenez, J. Khadake, A. Lagreid, L. Licata, R. C. Lovering, B. Meldal, A. N. Melidoni, M. Milagros, D. Peluso, L. Perfetto, P. Porras, A. Raghunath, S. Ricard-Blum, B. Roechert, A. Stutz, M. Tognolli, K. van Roey, G. Cesareni, and H. Hermjakob. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Research*, 42(D1):D358–D363, January 2014. URL: `https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt1115`, `doi:10.1093/nar/gkt1115`.

[147] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers. BioGRID: a general repository for interaction datasets. *Nucleic Acids Research*, 34(Database issue):D535–539, January 2006. `doi:10.1093/nar/gkj109`.

[148] K. Mitra, A.-R. Carvunis, S. K. Ramesh, and T. Ideker. Integrative approaches for finding modular structure in biological networks. *Nature Reviews Genetics*, 14(10):719–732, October 2013. URL: `http://www.nature.com/articles/nrg3552`, `doi:10.1038/nrg3552`.

[149] G. Csárdi and T. Nepusz. The igraph software package for complex network research. *InterJournal Complex Systems*, 2006. URL: `http://wbldb.lievers.net/10011687.html`.

[150] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2), February 2004. URL: `https://link.aps.org/doi/10.1103/PhysRevE.69.026113`, `doi:10.1103/PhysRevE.69.026113`.

[151] S. Redner. Teasing out the missing links: Networks. *Nature*, 453(7191):47–48, May 2008. URL: `http://www.nature.com/articles/453047a`, `doi:10.1038/453047a`.

[152] Q. Ye, W. Zhu, L. Li, S. Zhang, Y. Yin, H. Ma, and X. Wang. Brassinosteroids control male fertility by regulating the expression of key genes involved in Arabidopsis anther and pollen development. *Proceedings of the National Academy of Sciences of the United States of America*, 107(13):6100–6105, March 2010. `doi:10.1073/pnas.0912333107`.

[153] A. Wasilewska, F. Vlad, C. Sirichandra, Y. Redko, F. Jammes, C. Valon, N. Frei dit Frey, and J. Leung. An update on abscisic acid signaling in plants and more.. *Molecular Plant*, 1(2):198–217, March 2008. doi:10.1093/mp/ssm022.

[154] J. Leung and J. Giraudat. ABSCISIC ACID SIGNAL TRANSDUCTION. *Annual Review of Plant Physiology and Plant Molecular Biology*, 49:199–222, June 1998. doi:10.1146/annurev.arplant.49.1.199.

[155] B. Péret, B. De Rybel, I. Casimiro, E. Benková, R. Swarup, L. Laplaze, T. Beeckman, and M. J. Bennett. Arabidopsis lateral root development: an emerging story. *Trends in Plant Science*, 14(7):399–408, July 2009. doi:10.1016/j.tplants.2009.05.002.

[156] C. Wasternack and B. Hause. Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. *Annals of Botany*, 111(6):1021–1058, June 2013. doi:10.1093/aob/mct067.

[157] Y. Wu, D. Zhang, J. Y. Chu, P. Boyle, Y. Wang, I. D. Brindle, V. De Luca, and C. Després. The Arabidopsis NPR1 protein is a receptor for the plant defense hormone salicylic acid. *Cell Reports*, 1(6):639–647, June 2012. doi:10.1016/j.celrep.2012.05.008.

[158] A. C. Vlot, D. A. Dempsey, and D. F. Klessig. Salicylic Acid, a multifaceted hormone to combat disease. *Annual Review of Phytopathology*, 47:177–206, 2009. doi:10.1146/annurev.phyto.050908.135202.

[159] I. Hwang, J. Sheen, and B. Müller. Cytokinin Signaling Networks. *Annual Review of Plant Biology*, 63(1):353–380, June 2012. URL: http://www.annualreviews.org/doi/10.1146/annurev-arplant-042811-105503, doi:10.1146/annurev-arplant-042811-105503.

[160] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray. From molecular to modular cell biology. *Nature*, 402(S6761):C47–C52, December 1999. URL: http://www.nature.com/articles/35011540, doi:10.1038/35011540.

[161] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabasi. The human disease network. *Proceedings of the National Academy of Sciences*, 104(21):8685–8690, May 2007. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.0701361104, doi:10.1073/pnas.0701361104.

[162] J. Menche, A. Sharma, M. Kitsak, S. D. Ghiassian, M. Vidal, J. Loscalzo, and A.-L. Barabasi. Uncovering disease-disease relationships through the incomplete interactome. *Science*, 347(6224):1257601–1257601, February 2015. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1257601, doi:10.1126/science.1257601.

[163] M. Costanzo, A. Baryshnikova, J. Bellay, Y. Kim, E. D. Spear, C. S. Sevier, H. Ding, J. L. Koh, K. Toufighi, S. Mostafavi, J. Prinz, R. P. St. Onge, B. VanderSluis, T. Makhnevych, F. J. Vizeacoumar, S. Alizadeh, S. Bahr, R. L. Brost, Y. Chen, M. Cokol, R. Deshpande, Z. Li, Z.-Y. Lin, W. Liang, M. Marback, J. Paw, B.-J. San Luis, E. Shuteriqi, A. H. Y. Tong, N. van Dyk, I. M. Wallace,

J. A. Whitney, M. T. Weirauch, G. Zhong, H. Zhu, W. A. Houry, M. Brudno, S. Ragibizadeh, B. Papp, C. Pal, F. P. Roth, G. Giaever, C. Nislow, O. G. Troyanskaya, H. Bussey, G. D. Bader, A.-C. Gingras, Q. D. Morris, P. M. Kim, C. A. Kaiser, C. L. Myers, B. J. Andrews, and C. Boone. The Genetic Landscape of a Cell. *Science*, 327(5964):425–431, January 2010. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1180823, doi:10.1126/science.1180823.

[164] R. Deshpande, B. VanderSluis, and C. L. Myers. Comparison of Profile Similarity Measures for Genetic Interaction Networks. *PLoS ONE*, 8(7):e68664, July 2013. URL: https://dx.plos.org/10.1371/journal.pone.0068664, doi:10.1371/journal.pone.0068664.

[165] X. Hou, L. Y. C. Lee, K. Xia, Y. Yan, and H. Yu. DELLAs Modulate Jasmonate Signaling via Competitive Binding to JAZs. *Developmental Cell*, 19(6):884–894, December 2010. URL: https://linkinghub.elsevier.com/retrieve/pii/S1534580710004776, doi:10.1016/j.devcel.2010.10.024.

[166] A. Wager and J. Browse. Social Network: JAZ Protein Interactions Expand Our Knowledge of Jasmonate Signaling. *Frontiers in Plant Science*, 3, 2012. URL: http://journal.frontiersin.org/article/10.3389/fpls.2012.00041/abstract, doi:10.3389/fpls.2012.00041.

[167] S. Song, H. Huang, H. Gao, J. Wang, D. Wu, X. Liu, S. Yang, Q. Zhai, C. Li, T. Qi, and D. Xie. Interaction between MYC2 and ETHYLENE INSENSITIVE3 modulates antagonism between jasmonate and ethylene signaling in Arabidopsis. *The Plant Cell*, 26(1):263–279, January 2014. doi:10.1105/tpc.113.120394.

[168] M. A.-U. Nahar, T. Ishida, D. R. Smyth, M. Tasaka, and M. Aida. Interactions of CUP-SHAPED COTYLEDON and SPATULA Genes Control Carpel Margin Development in Arabidopsis thaliana. *Plant and Cell Physiology*, 53(6):1134–1143, June 2012. URL: https://academic.oup.com/pcp/article-lookup/doi/10.1093/pcp/pcs057, doi:10.1093/pcp/pcs057.

[169] F. E. Vaistij, Y. Gan, S. Penfield, A. D. Gilday, A. Dave, Z. He, E.-M. Josse, G. Choi, K. J. Halliday, and I. A. Graham. Differential control of seed primary dormancy in Arabidopsis ecotypes by the transcription factor SPATULA. *Proceedings of the National Academy of Sciences*, 110(26):10866–10871, June 2013. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1301647110, doi:10.1073/pnas.1301647110.

[170] M. Groszmann, T. Paicu, and D. R. Smyth. Functional domains of SPATULA, a bHLH transcription factor involved in carpel and fruit development in Arabidopsis. *The Plant Journal: For Cell and Molecular Biology*, 55(1):40–52, July 2008. doi:10.1111/j.1365-313X.2008.03469.x.

[171] T. Girin, T. Paicu, P. Stephenson, S. Fuentes, E. Körner, M. O'Brien, K. Sorefan, T. A. Wood, V. Balanzá, C. Ferrándiz, D. R. Smyth, and L. Østergaard. INDEHISCENT and SPATULA interact to specify carpel and valve margin tissue and thus promote seed dispersal in Arabidopsis. *The Plant Cell*, 23(10):3641–3653, October 2011. doi:10.1105/tpc.111.090944.

[172] K. Sidaway-Lee, E.-M. Josse, A. Brown, Y. Gan, K. J. Halliday, I. A. Graham, and S. Penfield. SPATULA links daytime temperature and plant growth rate. *Current biology: CB*, 20(16):1493–1497, August 2010. doi:10.1016/j.cub.2010.07.028.

[173] C.-Y. Cheng, V. Krishnakumar, A. P. Chan, F. Thibaud-Nissen, S. Schobel, and C. D. Town. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *The Plant Journal*, 89(4):789–804, February 2017. URL: http://doi.wiley.com/10.1111/tpj.13415, doi:10.1111/tpj.13415.

[174] I. Hwang, H.-C. Chen, and J. Sheen. Two-component signal transduction pathways in Arabidopsis. *Plant Physiology*, 129(2):500–515, June 2002. doi:10.1104/pp.005504.

[175] N. Marín-de la Rosa, B. Sotillo, P. Miskolczi, D. J. Gibbs, J. Vicente, P. Carbonero, L. Oñate-Sánchez, M. J. Holdsworth, R. Bhalerao, D. Alabadí, and M. A. Blázquez. Large-scale identification of gibberellin-related transcription factors defines group VII ETHYLENE RESPONSE FACTORS as functional DELLA partners. *Plant Physiology*, 166(2):1022–1032, October 2014. doi:10.1104/pp.114.244723.

[176] E.-M. Josse, Y. Gan, J. Bou-Torrent, K. L. Stewart, A. D. Gilday, C. E. Jeffree, F. E. Vaistij, J. F. Martínez-García, F. Nagy, I. A. Graham, and K. J. Halliday. A DELLA in disguise: SPATULA restrains the growth of the developing Arabidopsis seedling. *The Plant Cell*, 23(4):1337–1351, April 2011. doi:10.1105/tpc.110.082594.

[177] A. Boccaccini, S. Santopolo, D. Capauto, R. Lorrai, E. Minutello, G. Serino, P. Costantino, and P. Vittorioso. The DOF protein DAG1 and the DELLA protein GAI cooperate in negatively regulating the AtGA3ox1 gene. *Molecular Plant*, 7(9):1486–1489, September 2014. doi:10.1093/mp/ssu046.

[178] J. W. Chandler, B. Jacobs, M. Cole, P. Comelli, and W. Werr. DORNRÖSCHEN-LIKE expression marks Arabidopsis floral organ founder cells and precedes auxin response maxima. *Plant Molecular Biology*, 76(1-2):171–185, May 2011. doi:10.1007/s11103-011-9779-8.

[179] A. Nag, Y. Yang, and T. Jack. DORNROSCHEN-LIKE, an AP2 gene, is necessary for stamen emergence in Arabidopsis. *Plant Molecular Biology*, 65(3):219–232, October 2007. doi:10.1007/s11103-007-9210-7.

[180] L. Yant, J. Mathieu, T. T. Dinh, F. Ott, C. Lanz, H. Wollmann, X. Chen, and M. Schmid. Orchestration of the Floral Transition and Floral Development in Arabidopsis by the Bifunctional Transcription Factor APETALA2. *THE PLANT CELL ONLINE*, 22(7):2156–2170, July 2010. URL: http://www.plantcell.org/cgi/doi/10.1105/tpc.110.075606, doi:10.1105/tpc.110.075606.

[181] E. Balsemão-Pires, L. R. Andrade, and G. Sachetto-Martins. Functional study of TCP23 in Arabidopsis thaliana during plant development. *Plant Physiology and Biochemistry*, 67:120–125, June 2013. URL: https://linkinghub.elsevier.com/retrieve/pii/S0981942813000971, doi:10.1016/j.plaphy.2013.03.009.

[182] S. Bandyopadhyay, M. Mehta, D. Kuo, M.-K. Sung, R. Chuang, E. J. Jaehnig, B. Bodenmiller, K. Licon, W. Copeland, M. Shales, D. Fiedler, J. Dutkowski,

A. Guenole, H. van Attikum, K. M. Shokat, R. D. Kolodner, W.-K. Huh, R. Aebersold, M.-C. Keogh, N. J. Krogan, and T. Ideker. Rewiring of Genetic Networks in Response to DNA Damage. *Science*, 330(6009):1385–1389, December 2010. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1195618, doi:10.1126/science.1195618.

[183] L. Ou-Yang, D.-Q. Dai, X.-L. Li, M. Wu, X.-F. Zhang, and P. Yang. Detecting temporal protein complexes from dynamic protein-protein interaction networks. *BMC Bioinformatics*, 15(1):335, 2014. URL: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-15-335, doi:10.1186/1471-2105-15-335.

[184] X.-H. Li, P. L. Chavali, and M. M. Babu. Capturing dynamic protein interactions. *Science*, 359(6380):1105–1106, March 2018. URL: http://www.sciencemag.org/lookup/doi/10.1126/science.aat0576, doi:10.1126/science.aat0576.

[185] J.-D. J. Han, N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupuy, A. J. M. Walhout, M. E. Cusick, F. P. Roth, and M. Vidal. Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature*, 430(6995):88–93, July 2004. URL: https://doi.org/10.1038/nature02555, doi:10.1038/nature02555.

[186] G. Toledo-Ortiz, E. Huq, and P. H. Quail. The Arabidopsis basic/helix-loop-helix transcription factor family. *The Plant Cell*, 15(8):1749–1770, August 2003.

[187] P. C. Bailey, C. Martin, G. Toledo-Ortiz, P. H. Quail, E. Huq, M. A. Heim, M. Jakoby, M. Werber, and B. Weisshaar. Update on the basic helix-loop-helix transcription factor gene family in Arabidopsis thaliana. *The Plant Cell*, 15(11):2497–2502, November 2003. doi:10.1105/tpc.151140.

[188] M. A. Heim, M. Jakoby, M. Werber, C. Martin, B. Weisshaar, and P. C. Bailey. The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Molecular Biology and Evolution*, 20(5):735–747, May 2003. doi:10.1093/molbev/msg088.

[189] D. M. Friedrichsen, J. Nemhauser, T. Muramitsu, J. N. Maloof, J. Alonso, J. R. Ecker, M. Furuya, and J. Chory. Three redundant brassinosteroid early response genes encode putative bHLH transcription factors required for normal growth. *Genetics*, 162(3):1445–1456, November 2002.

[190] N. Cifuentes-Esquivel, J. Bou-Torrent, A. Galstyan, M. Gallemí, G. Sessa, M. Salla Martret, I. Roig-Villanova, I. Ruberti, and J. F. Martínez-García. The bHLH proteins BEE and BIM positively modulate the shade avoidance syndrome in Arabidopsis seedlings. *The Plant Journal: For Cell and Molecular Biology*, 75(6):989–1002, September 2013. doi:10.1111/tpj.12264.

[191] P. Brodersen, M. Petersen, H. Bjørn Nielsen, S. Zhu, M.-A. Newman, K. M. Shokat, S. Rietz, J. Parker, and J. Mundy. Arabidopsis MAP kinase 4 regulates salicylic acid- and jasmonic acid/ethylene-dependent responses via EDS1 and PAD4. *The Plant Journal: For Cell and Molecular Biology*, 47(4):532–546, August 2006. doi:10.1111/j.1365-313X.2006.02806.x.

[192] R. Catala, J. Ouyang, I. A. Abreu, Y. Hu, H. Seo, X. Zhang, and N.-H. Chua. The Arabidopsis E3 SUMO ligase SIZ1 regulates plant growth and drought responses. *The Plant Cell*, 19(9):2952–2966, September 2007. doi:10.1105/tpc.106.049981.

[193] B.-h. Lee, D. A. Henderson, and J.-K. Zhu. The Arabidopsis cold-responsive transcriptome and its regulation by ICE1. *The Plant Cell*, 17(11):3155–3175, November 2005. doi:10.1105/tpc.105.035568.

[194] B. C. W. Crawford and M. F. Yanofsky. HALF FILLED promotes reproductive tract development and fertilization efficiency in Arabidopsis thaliana. *Development (Cambridge, England)*, 138(14):2999–3009, July 2011. doi:10.1242/dev.067793.

[195] B. Poppenberger, W. Rozhon, M. Khan, S. Husar, G. Adam, C. Luschnig, S. Fujioka, and T. Sieberer. CESTA, a positive regulator of brassinosteroid biosynthesis. *The EMBO journal*, 30(6):1149–1161, March 2011. doi:10.1038/emboj.2011.35.

[196] D. Huang, W. Wu, S. R. Abrams, and A. J. Cutler. The relationship of drought-related gene expression in Arabidopsis thaliana to hormonal and environmental factors. *Journal of Experimental Botany*, 59(11):2991–3007, 2008. doi:10.1093/jxb/ern155.

[197] F. G. Malinovsky, M. Batoux, B. Schwessinger, J. H. Youn, L. Stransfeld, J. Win, S.-K. Kim, and C. Zipfel. Antagonistic regulation of growth and immunity by the Arabidopsis basic helix-loop-helix transcription factor homolog of brassinosteroid enhanced expression2 interacting with increased leaf inclination1 binding bHLH1. *Plant Physiology*, 164(3):1443–1455, March 2014. doi:10.1104/pp.113.234625.

[198] E. Oh, J.-Y. Zhu, M.-Y. Bai, R. A. Arenhart, Y. Sun, and Z.-Y. Wang. Cell elongation is regulated through a central circuit of interacting transcription factors in the Arabidopsis hypocotyl. *eLife*, 3, May 2014. doi:10.7554/eLife.03031.

[199] B. T. Seet, I. Dikic, M.-M. Zhou, and T. Pawson. Reading protein modifications with interaction domains. *Nature Reviews Molecular Cell Biology*, 7(7):473–483, July 2006. URL: http://www.nature.com/doifinder/10.1038/nrm1960, doi:10.1038/nrm1960.

[200] G. Manning, D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam. The protein kinase complement of the human genome. *Science (New York, N.Y.)*, 298(5600):1912–1934, 2002. doi:10.1126/science.1075762.

[201] M. D. Lehti-Shiu and S.-H. Shiu. Diversity, classification and function of the plant protein kinase superfamily. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 367(1602):2619–2639, September 2012. doi:10.1098/rstb.2012.0003.

[202] J. L. Heazlewood, P. Durek, J. Hummel, J. Selbig, W. Weckwerth, D. Walther, and W. X. Schulze. PhosPhAt: a database of phosphorylation sites in Arabidopsis thaliana and a plant-specific phosphorylation site predictor. *Nucleic Acids Research*, 36(Database issue):D1015–1021, January 2008. doi:10.1093/nar/gkm812.

[203] P. Durek, R. Schmidt, J. L. Heazlewood, A. Jones, D. MacLean, A. Nagel, B. Kersten, and W. X. Schulze. PhosPhAt: the Arabidopsis thaliana phosphorylation site database. An update. *Nucleic Acids Research*, 38(Database issue):D828–834, January 2010. doi:10.1093/nar/gkp810.

[204] M. Zulawski, R. Braginets, and W. X. Schulze. PhosPhAt goes kinases–searchable protein kinase target information in the plant phosphorylation site database PhosPhAt. *Nucleic Acids Research*, 41(Database issue):D1176–1184, January 2013. doi:10.1093/nar/gks1081.

[205] T. Viaene, C. F. Delwiche, S. A. Rensing, and J. Friml. Origin and evolution of PIN auxin transporters in the green lineage. *Trends in Plant Science*, 18(1):5–10, January 2013. doi:10.1016/j.tplants.2012.08.009.

[206] M. Zourelidou, B. Absmanner, B. Weller, I. C. Barbosa, B. C. Willige, A. Fastner, V. Streit, S. A. Port, J. Colcombet, S. de la Fuente van Bentem, H. Hirt, B. Kuster, W. X. Schulze, U. Z. Hammes, and C. Schwechheimer. Auxin efflux by PIN-FORMED proteins is activated by two different protein kinases, D6 PROTEIN KINASE and PINOID. *eLife*, 3, June 2014. URL: https://elifesciences.org/articles/02860, doi:10.7554/eLife.02860.

[207] P. Dhonukshe, F. Huang, C. S. Galvan-Ampudia, A. P. Mahonen, J. Kleine-Vehn, J. Xu, A. Quint, K. Prasad, J. Friml, B. Scheres, and R. Offringa. Plasma membrane-bound AGC3 kinases phosphorylate PIN auxin carriers at TPRXS(N/S) motifs to direct apical PIN recycling. *Development*, 137(19):3245–3255, October 2010. URL: http://dev.biologists.org/cgi/doi/10.1242/dev.052456, doi:10.1242/dev.052456.

[208] A. V. Garcia, M. Al-Yousif, and H. Hirt. Role of AGC kinases in plant growth and stress responses. *Cellular and molecular life sciences: CMLS*, 69(19):3259–3267, October 2012. doi:10.1007/s00018-012-1093-3.

[209] S. D. Clouse. Brassinosteroid Signal Transduction: From Receptor Kinase Activation to Transcriptional Networks Regulating Plant Development. *The Plant Cell*, 23(4):1219–1230, April 2011. URL: http://www.plantcell.org/lookup/doi/10.1105/tpc.111.084475, doi:10.1105/tpc.111.084475.

[210] J.-Y. Zhu, J. Sae-Seaw, and Z.-Y. Wang. Brassinosteroid signalling. *Development*, 140(8):1615–1620, April 2013. URL: http://dev.biologists.org/cgi/doi/10.1242/dev.060590, doi:10.1242/dev.060590.

[211] F. Tajima. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3):585–595, November 1989.

[212] Y. X. Fu and W. H. Li. Statistical tests of neutrality of mutations. *Genetics*, 133(3):693–709, March 1993.

[213] S. E. Ramos-Onsins and J. Rozas. Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, 19(12):2092–2100, December 2002. doi:10.1093/oxfordjournals.molbev.a004034.

[214] M. Pybus, G. M. Dall'Olio, P. Luisi, M. Uzkudun, A. Carreño-Torres, P. Pavlidis, H. Laayouni, J. Bertranpetit, and J. Engelken. 1000 Genomes Selection Browser 1.0: a genome browser dedicated to signatures of natural selection in modern humans. *Nucleic Acids Research*, 42(D1):D903–D909, January 2014. URL: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt1188, doi:10.1093/nar/gkt1188.

[215] B. E. Campitelli and J. R. Stinchcombe. Population Dynamics and Evolutionary History of the Weedy Vine *Ipomoea hederacea* in North America. *G3: Genes|Genomes|Genetics*, 4(8):1407–1416, August 2014. URL: http://g3journal.org/lookup/doi/10.1534/g3.114.011700, doi:10.1534/g3.114.011700.

[216] A. Ramirez-Soriano, S. E. Ramos-Onsins, J. Rozas, F. Calafell, and A. Navarro. Statistical Power Analysis of Neutrality Tests Under Demographic Expansions, Contractions and Bottlenecks With Recombination. *Genetics*, 179(1):555–567, May 2008. URL: http://www.genetics.org/cgi/doi/10.1534/genetics.107.083006, doi:10.1534/genetics.107.083006.

[217] M. T. Webster and N. G. Smith. Fixation biases affecting human SNPs. *Trends in Genetics*, 20(3):122–126, March 2004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0168952504000204, doi:10.1016/j.tig.2004.01.005.

[218] J. Rozas. DNA Sequence Polymorphism Analysis Using DnaSP. In D. Posada, editor, *Bioinformatics for DNA Sequence Analysis*, volume 537, pages 337–350. Humana Press, Totowa, NJ, 2009. URL: http://link.springer.com/10.1007/978-1-59745-251-9_17, doi:10.1007/978-1-59745-251-9_17.

[219] G. Bedada, A. Westerbergh, E. Nevo, A. Korol, and K. J. Schmid. DNA sequence variation of wild barley Hordeum spontaneum (L.) across environmental gradients in Israel. *Heredity*, 112:646, March 2014. URL: https://doi.org/10.1038/hdy.2014.2.

[220] S. Wuchty, A.-L. Barabási, and M. T. Ferdig. Stable evolutionary signal in a yeast protein interaction network. *BMC evolutionary biology*, 6:8, January 2006. doi:10.1186/1471-2148-6-8.

[221] L. B. Barreiro and L. Quintana-Murci. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nature Reviews. Genetics*, 11(1):17–30, January 2010. doi:10.1038/nrg2698.

[222] Y. Wang, J. Li, S. Hou, X. Wang, Y. Li, D. Ren, S. Chen, X. Tang, and J.-M. Zhou. A Pseudomonas syringae ADP-ribosyltransferase inhibits Arabidopsis mitogen-activated protein kinase kinases. *The Plant Cell*, 22(6):2033–2044, June 2010. doi:10.1105/tpc.110.075697.

[223] N. Leonhardt, J. M. Kwak, N. Robert, D. Waner, G. Leonhardt, and J. I. Schroeder. Microarray expression analyses of Arabidopsis guard cells and isolation of a recessive abscisic acid hypersensitive protein phosphatase 2c mutant. *The Plant Cell*, 16(3):596–615, March 2004. doi:10.1105/tpc.019000.

[224] S. C. Lee, C. W. Lim, W. Lan, K. He, and S. Luan. ABA signaling in guard cells entails a dynamic protein-protein interaction relay from the PYL-RCAR family receptors to ion channels. *Molecular Plant*, 6(2):528–538, March 2013. doi:10.1093/mp/sss078.

[225] X. Saavedra, A. Modrego, D. Rodríguez, M. P. González-García, L. Sanz, G. Nicolás, and O. Lorenzo. The nuclear interactor PYL8/RCAR3 of Fagus sylvatica FsPP2c1 is a positive regulator of abscisic acid signaling in seeds and stress. *Plant Physiology*, 152(1):133–150, January 2010. doi:10.1104/pp.109.146381.

[226] R. Gupta, Y. Huang, J. Kieber, and S. Luan. Identification of a dual-specificity protein phosphatase that inactivates a MAP kinase from Arabidopsis. *The Plant Journal: For Cell and Molecular Biology*, 16(5):581–589, December 1998.

[227] R. Liu, Y. Liu, N. Ye, G. Zhu, M. Chen, L. Jia, Y. Xia, L. Shi, W. Jia, and J. Zhang. AtDsPTP1 acts as a negative regulator in osmotic stress signalling during Arabidopsis seed germination and seedling establishment. *Journal of Experimental Botany*, 66(5):1339–1353, March 2015. doi:10.1093/jxb/eru484.

[228] K. Ishida, T. Yamashino, A. Yokoyama, and T. Mizuno. Three Type-B Response Regulators, ARR1, ARR10 and ARR12, Play Essential but Redundant Roles in Cytokinin Signal Transduction Throughout the Life Cycle of Arabidopsis thaliana. *Plant and Cell Physiology*, 49(1):47–57, January 2008. URL: https://academic.oup.com/pcp/article-lookup/doi/10.1093/pcp/pcm165, doi:10.1093/pcp/pcm165.

[229] R. Thilmony, W. Underwood, and S. Y. He. Genome-wide transcriptional analysis of the *Arabidopsis thaliana* interaction with the plant pathogen *Pseudomonas syringae* pv. *tomato* DC3000 and the human pathogen *Escherichia coli* O157:H7. *The Plant Journal*, 46(1):34–53, April 2006. URL: http://doi.wiley.com/10.1111/j.1365-313X.2006.02725.x, doi:10.1111/j.1365-313X.2006.02725.x.

[230] P. Sukumar, K. S. Edwards, A. Rahman, A. Delong, and G. K. Muday. PINOID kinase regulates root gravitropism through modulation of PIN2-dependent basipetal auxin transport in Arabidopsis. *Plant Physiology*, 150(2):722–735, June 2009. doi:10.1104/pp.108.131607.

[231] J. J. Blakeslee, H.-W. Zhou, J. T. Heath, K. R. Skottke, J. A. R. Barrios, S.-Y. Liu, and A. DeLong. Specificity of RCN1-mediated protein phosphatase 2a regulation in meristem organization and stress response in roots. *Plant Physiology*, 146(2):539–553, February 2008. doi:10.1104/pp.107.112995.

[232] M. Singh, A. Gupta, and A. Laxmi. Glucose control of root growth direction in Arabidopsis thaliana. *Journal of Experimental Botany*, 65(12):2981–2993, July 2014. doi:10.1093/jxb/eru146.

[233] E. Demarsy, I. Schepens, K. Okajima, M. Hersch, S. Bergmann, J. Christie, K.-I. Shimazaki, S. Tokutomi, and C. Fankhauser. Phytochrome Kinase Substrate 4 is phosphorylated by the phototropin 1 photoreceptor. *The EMBO journal*, 31(16):3457–3467, August 2012. doi:10.1038/emboj.2012.186.

[234] G. Wu, X. Wang, X. Li, Y. Kamiya, M. S. Otegui, and J. Chory. Methylation of a phosphatase specifies dephosphorylation and degradation of activated brassinosteroid receptors. *Science Signaling*, 4(172):ra29, May 2011. doi:10.1126/scisignal.2001258.

[235] N. Saito, S. Munemasa, Y. Nakamura, Y. Shimoishi, I. C. Mori, and Y. Murata. Roles of RCN1, regulatory A subunit of protein phosphatase 2a, in methyl jasmonate signaling and signal crosstalk between methyl jasmonate and abscisic acid. *Plant & Cell Physiology*, 49(9):1396–1401, September 2008. doi:10.1093/pcp/pcn106.

[236] T.-S. Tseng and W. R. Briggs. The Arabidopsis rcn1-1 mutation impairs dephosphorylation of Phot2, resulting in enhanced blue light responses. *The Plant Cell*, 22(2):392–402, February 2010. doi:10.1105/tpc.109.066423.

[237] L. A. Rymarquis, F. F. Souret, and P. J. Green. Evidence that XRN4, an Arabidopsis homolog of exoribonuclease XRN1, preferentially impacts transcripts with certain sequences or in particular functional categories. *RNA (New York, N.Y.)*, 17(3):501–511, March 2011. doi:10.1261/rna.2467911.

[238] F. Vogel, D. Hofius, K. E. Paulus, I. Jungkunz, and U. Sonnewald. The second face of a known player: Arabidopsis silencing suppressor AtXRN4 acts organ-specifically. *The New Phytologist*, 189(2):484–493, January 2011. doi:10.1111/j.1469-8137.2010.03482.x.

[239] A. H. Nguyen, A. Matsui, M. Tanaka, K. Mizunashi, K. Nakaminami, M. Hayashi, K. Iida, T. Toyoda, D. V. Nguyen, and M. Seki. Loss of Arabidopsis 5'-3' Exoribonuclease AtXRN4 Function Enhances Heat Stress Tolerance of Plants Subjected to Severe Heat Stress. *Plant & Cell Physiology*, 56(9):1762–1772, September 2015. doi:10.1093/pcp/pcv096.

[240] S. Xia, Z. Zhu, L. Hao, J.-G. Chen, L. Xiao, Y. Zhang, and X. Li. Negative regulation of systemic acquired resistance by replication factor C subunit3 in Arabidopsis. *Plant Physiology*, 150(4):2009–2017, August 2009. doi:10.1104/pp.109.138321.

[241] S. Xia, L. Xiao, P. Gannon, and X. Li. RFC3 regulates cell proliferation and pathogen resistance in Arabidopsis. *Plant Signaling & Behavior*, 5(2):168–170, February 2010.

[242] S. Hirsch and G. E. D. Oldroyd. GRAS-domain transcription factors that regulate plant development. *Plant Signaling & Behavior*, 4(8):698–700, August 2009.

[243] P. Torres-Galea, B. Hirtreiter, and C. Bolle. Two GRAS proteins, SCARECROW-LIKE21 and PHYTOCHROME A SIGNAL TRANSDUCTION1, function cooperatively in phytochrome A signal transduction. *Plant Physiology*, 161(1):291–304, January 2013. doi:10.1104/pp.112.206607.

[244] G. Huang, R. Dong, R. Allen, E. L. Davis, T. J. Baum, and R. S. Hussey. A root-knot nematode secretory peptide functions as a ligand for a plant transcription factor. *Molecular plant-microbe interactions: MPMI*, 19(5):463–470, May 2006. doi:10.1094/MPMI-19-0463.

[245] X. Xu, C. Chen, B. Fan, and Z. Chen. Physical and functional interactions between pathogen-induced Arabidopsis WRKY18, WRKY40, and WRKY60 transcription factors. *The Plant Cell*, 18(5):1310–1326, May 2006. doi:10.1105/tpc.105.037523.

[246] H. Chen, Z. Lai, J. Shi, Y. Xiao, Z. Chen, and X. Xu. Roles of arabidopsis WRKY18, WRKY40 and WRKY60 transcription factors in plant responses to abscisic acid and abiotic stress. *BMC plant biology*, 10:281, December 2010. doi:10.1186/1471-2229-10-281.

[247] Z.-Q. Liu, L. Yan, Z. Wu, C. Mei, K. Lu, Y.-T. Yu, S. Liang, X.-F. Zhang, X.-F. Wang, and D.-P. Zhang. Cooperation of three WRKY-domain transcription factors WRKY18, WRKY40, and WRKY60 in repressing two ABA-responsive genes ABI4 and ABI5 in Arabidopsis. *Journal of Experimental Botany*, 63(18):6371–6392, November 2012. doi:10.1093/jxb/ers293.

[248] R. Kalladan, J. R. Lasky, T. Z. Chang, S. Sharma, T. E. Juenger, and P. E. Verslues. Natural variation identifies genes affecting drought-induced abscisic acid accumulation in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, 114(43):11536–11541, October 2017. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1705884114, doi:10.1073/pnas.1705884114.

[249] N. H. Davila Olivas, E. Frago, M. P. M. Thoen, K. J. Kloth, F. F. M. Becker, J. J. A. van Loon, G. Gort, J. J. B. Keurentjes, J. van Heerwaarden, and M. Dicke. Natural variation in life history strategy of Arabidopsis thaliana determines stress responses to drought and insects of different feeding guilds. *Molecular Ecology*, 26(11):2959–2977, June 2017. doi:10.1111/mec.14100.

[250] M. T. Nishimura and J. L. Dangl. Arabidopsis and the plant immune system. *The Plant Journal: For Cell and Molecular Biology*, 61(6):1053–1066, March 2010. doi:10.1111/j.1365-313X.2010.04131.x.

[251] J. Fu and S. Wang. Insights into auxin signaling in plant-pathogen interactions. *Frontiers in Plant Science*, 2:74, 2011. doi:10.3389/fpls.2011.00074.

[252] H. Liu, B. Zhang, T. Wu, Y. Ding, X. Ding, and Z. Chu. Copper Ion Elicits Defense Response in Arabidopsis thaliana by Activating Salicylate- and Ethylene-Dependent Signaling Pathways. *Molecular Plant*, 8(10):1550–1553, October 2015. doi:10.1016/j.molp.2015.07.008.

[253] C.-C. Chen, W.-F. Chien, N.-C. Lin, and K.-C. Yeh. Alternative functions of Arabidopsis Yellow Stripe-Like3: from metal translocation to pathogen defense. *PloS One*, 9(5):e98008, 2014. doi:10.1371/journal.pone.0098008.

[254] J. M. Alonso and J. R. Ecker. Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in Arabidopsis. *Nature Reviews Genetics*, 7:524, June 2006. URL: https://doi.org/10.1038/nrg1893.

[255] P. D. Spanu, J. C. Abbott, J. Amselem, T. A. Burgis, D. M. Soanes, K. Stüber, E. Ver Loren van Themaat, J. K. M. Brown, S. A. Butcher, S. J. Gurr, M.-H. Lebrun, C. J. Ridout, P. Schulze-Lefert, N. J. Talbot, N. Ahmadinejad, C. Ametz, G. R. Barton, M. Benjdia, P. Bidzinski, L. V. Bindschedler, M. Both, M. T.

Brewer, L. Cadle-Davidson, M. M. Cadle-Davidson, J. Collemare, R. Cramer, O. Frenkel, D. Godfrey, J. Harriman, C. Hoede, B. C. King, S. Klages, J. Kleemann, D. Knoll, P. S. Koti, J. Kreplak, F. J. López-Ruiz, X. Lu, T. Maekawa, S. Mahanil, C. Micali, M. G. Milgroom, G. Montana, S. Noir, R. J. O'Connell, S. Oberhaensli, F. Parlange, C. Pedersen, H. Quesneville, R. Reinhardt, M. Rott, S. Sacristán, S. M. Schmidt, M. Schön, P. Skamnioti, H. Sommer, A. Stephens, H. Takahara, H. Thordal-Christensen, M. Vigouroux, R. Wessling, T. Wicker, and R. Panstruga. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science (New York, N.Y.)*, 330(6010):1543–1546, December 2010. doi:10.1126/science.1194573.

[256] C. Micali, K. Göllner, M. Humphry, C. Consonni, and R. Panstruga. The Powdery Mildew Disease of Arabidopsis: A Paradigm for the Interaction between Plants and Biotrophic Fungi. *The Arabidopsis Book*, 6:e0115, 2008. doi:10.1199/tab.0115.

[257] R. Weßling, S. M. Schmidt, C. O. Micali, F. Knaust, R. Reinhardt, U. Neumann, E. Ver Loren van Themaat, and R. Panstruga. Transcriptome analysis of enriched Golovinomyces orontii haustoria by deep 454 pyrosequencing. *Fungal genetics and biology: FG & B*, 49(6):470–482, June 2012. doi:10.1016/j.fgb.2012.04.001.

[258] E. A. van der Biezen, C. T. Freddie, K. Kahn, J. E. Parker, and J. D. G. Jones. Arabidopsis RPP4 is a member of the RPP5 multigene family of TIR-NB-LRR genes and confers downy mildew resistance through multiple signalling components. *The Plant Journal: For Cell and Molecular Biology*, 29(4):439–451, February 2002.

[259] G. A. Watterson. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, 7(2):256–276, April 1975.

[260] Q. Chao, M. Rothenberg, R. Solano, G. Roman, W. Terzaghi, and J. R. Ecker. Activation of the ethylene gas response pathway in Arabidopsis by the nuclear protein ETHYLENE-INSENSITIVE3 and related proteins. *Cell*, 89(7):1133–1144, June 1997.

[261] M. Naseem, M. Kaltdorf, and T. Dandekar. The nexus between growth and defence signalling: auxin and cytokinin modulate plant immune response pathways. *Journal of Experimental Botany*, 66(16):4885–4896, August 2015. URL: https://academic.oup.com/jxb/article-lookup/doi/10.1093/jxb/erv297, doi:10.1093/jxb/erv297.

[262] L. Yang, B. Li, X.-y. Zheng, J. Li, M. Yang, X. Dong, G. He, C. An, and X. W. Deng. Salicylic acid biosynthesis is enhanced and contributes to increased biotrophic pathogen resistance in Arabidopsis hybrids. *Nature Communications*, 6:7309, June 2015. doi:10.1038/ncomms8309.

[263] T. A. E. Rahman, M. E. Oirdi, R. Gonzalez-Lamothe, and K. Bouarab. Necrotrophic pathogens use the salicylic acid signaling pathway to promote disease development in tomato. *Molecular plant-microbe interactions: MPMI*, 25(12):1584–1593, December 2012. doi:10.1094/MPMI-07-12-0187-R.

[264] R. F. Doolittle. Convergent evolution: the need to be explicit. *Trends in Biochemical Sciences*, 19(1):15–18, January 1994. URL: http://linkinghub.elsevier.com/retrieve/pii/0968000494901678, doi:10.1016/0968-0004(94)90167-8.

[265] S. H. Kim, G. H. Son, S. Bhattacharjee, H. J. Kim, J. C. Nam, P. D. T. Nguyen, J. C. Hong, and W. Gassmann. The Arabidopsis immune adaptor SRFR1 interacts with TCP transcription factors that redundantly contribute to effector-triggered immunity. *The Plant Journal: For Cell and Molecular Biology*, 78(6):978–989, June 2014. doi:10.1111/tpj.12527.

[266] E. Pang, Y. Hao, Y. Sun, and K. Lin. Differential variation patterns between hubs and bottlenecks in human protein-protein interaction networks. *BMC Evolutionary Biology*, 16(1), December 2016. URL: http://bmcevolbiol.biomedcentral.com/articles/10.1186/s12862-016-0840-8, doi:10.1186/s12862-016-0840-8.

[267] M. Altmann, S. Altmann, C. Falter, and P. Falter-Braun. High-Quality Yeast-2-Hybrid Interaction Network Mapping. *Current Protocols in Plant Biology*, 3(3):e20067, September 2018. doi:10.1002/cppb.20067.

[268] B. Langmead and S. L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359, April 2012. URL: https://www.nature.com/articles/nmeth.1923, doi:10.1038/nmeth.1923.

[269] R. C. Team. R: A Language and Environment for Statistical Computing, 2018. URL: https://www.R-project.org.

[270] J. H. Morris, A. Kuchinsky, T. E. Ferrin, and A. R. Pico. enhancedGraphics: a Cytoscape app for enhanced node graphics. *F1000Research*, 3:147, 2014. doi:10.12688/f1000research.4460.1.

[271] L. Danon, A. Díaz-Guilera, J. Duch, and A. Arenas. Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09):P09008–P09008, September 2005. URL: http://stacks.iop.org/1742-5468/2005/i=09/a=P09008?key=crossref.5420f964e99dd130e25dd14c3f1af547, doi:10.1088/1742-5468/2005/09/P09008.

[272] S. v. Dongen. RAPPORT Performance Criteria for Graph Clustering and Markov Cluster Experiments. 2000.

[273] W. M. Rand. Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association*, 66(336):846–850, 1971. URL: http://www.jstor.org/stable/2284239, doi:10.2307/2284239.

[274] L. Hubert and P. Arabie. Comparing partitions. *Journal of Classification*, 2(1):193–218, December 1985. URL: https://doi.org/10.1007/BF01908075, doi:10.1007/BF01908075.

[275] S. Wagner and D. Wagner. *Comparing Clusterings- An Overview*. 2007.

[276] D. Steinley. Properties of the Hubert-Arabie adjusted Rand index. *Psychological Methods*, 9(3):386–396, September 2004. doi:10.1037/1082-989X.9.3.386.

[277] S. Falcon and R. Gentleman. Using GOstats to test gene lists for GO term association. *Bioinformatics (Oxford, England)*, 23(2):257–258, January 2007. doi:10.1093/bioinformatics/btl567.

[278] P. Lamesch, T. Z. Berardini, D. Li, D. Swarbreck, C. Wilks, R. Sasidharan, R. Muller, K. Dreher, D. L. Alexander, M. Garcia-Hernandez, A. S. Karthikeyan, C. H. Lee, W. D. Nelson, L. Ploetz, S. Singh, A. Wensel, and E. Huala. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research*, 40(D1):D1202–D1210, January 2012. URL: `https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkr1090`, `doi:10.1093/nar/gkr1090`.

[279] N. Skunca, A. Altenhoff, and C. Dessimoz. Quality of computationally inferred gene ontology annotations. *PLoS computational biology*, 8(5):e1002533, May 2012. `doi:10.1371/journal.pcbi.1002533`.

[280] M. Gribskov, F. Fana, J. Harper, D. A. Hope, A. C. Harmon, D. W. Smith, F. E. Tax, and G. Zhang. PlantsP: a functional genomics database for plant phosphorylation. *Nucleic Acids Research*, 29(1):111–113, January 2001.

[281] O. Wagih. *ggseqlogo: A 'ggplot2' Extension for Drawing Publication-Ready Sequence Logos*. 2017. URL: `https://CRAN.R-project.org/package=ggseqlogo`.

[282] T. D. Schneider and R. M. Stephens. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research*, 18(20):6097–6100, October 1990.

[283] T. D. Schneider, G. D. Stormo, L. Gold, and A. Ehrenfeucht. Information content of binding sites on nucleotide sequences. *Journal of Molecular Biology*, 188(3):415–431, April 1986.

[284] J. J. Cai, D. K. Smith, X. Xia, and K.-Y. Yuen. MBEToolbox 2.0: an enhanced version of a MATLAB toolbox for molecular biology and evolution. *Evolutionary Bioinformatics Online*, 2:179–182, February 2007.

[285] J. Fox, S. Weisberg, and J. Fox. *An R companion to applied regression*. SAGE Publications, Thousand Oaks, Calif, 2nd ed edition, 2011. OCLC: ocn648922089.

[286] Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995. URL: `http://www.jstor.org/stable/2346101`.

[287] K. Thornton. Libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics (Oxford, England)*, 19(17):2325–2327, November 2003.

7 Publications

- Ravikumar R, Kalbfuß N, Gendre D, Steiner A, Altmann M, *Altmann S*, Rybak K, Edelmann H, Stephan F, Lampe M, Facher E, Wanner G, Falter-Braun P, Bhalerao RP, Assaad FF. Independent yet overlapping pathways ensure the robustness and responsiveness of trans-Golgi network functions in Arabidopsis. *Development.* 2018 Nov 7;145(21). pii: dev169201. doi:10.1242/dev.169201.

- Altmann M, *Altmann S*, Falter C, Falter-Braun P. High-Quality Yeast-2-Hybrid Interaction Network Mapping. *Curr Protoc Plant Biol.* 2018 Sep;3(3):e20067. doi:10.1002/cppb.20067. Epub 2018 Jun 26.

- Yazaki J, Galli M, Kim AY, Nito K, Aleman F, Chang KN, Carvunis AR, Quan R, Nguyen H, Song L, Alvarez JM, Huang SS, Chen H, Ramachandran N, *Altmann S*, Gutiérrez RA, Hill DE, Schroeder JI, Chory J, LaBaer J, Vidal M, Braun P, Ecker JR. Mapping transcription factor interactome networks using HaloTag protein arrays. *Proc Natl Acad Sci U S A*. 2016 Jul 19;113(29):E4238-47. doi: 10.1073/pnas.1603229113. Epub 2016 Jun 29.

- Petri T, *Altmann S*, Geistlinger L, Zimmer R, Küffner R. Addressing false discoveries in network inference. *Bioinformatics.* 2015 Sep 01; 31(17):2836-43. doi: 10.1093/bioinformatics/btv215. Epub 2015 Apr 24.

- Weßling R, Epple P, *Altmann S*, He Y, Yang L, Henz SR, McDonald N, Wiley K, Bader KC, Gläßer C, Mukhtar MS, Haigis S, Ghamsari L, Stephens AE, Ecker JR, Vidal M, Jones JD, Mayer KF, Ver Loren van Themaat E, Weigel D, Schulze-Lefert P, Dangl JL, Panstruga R, Braun P. Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell Host Microbe.* 2014 Sep 10;16(3):364-75. doi:10.1016/j.chom.2014.08.004.

8 Contributions

8.1 Phytohormone signal transduction project

Melina Altmann (Institute of Network Biology, Helmholtz Zentrum München, München) performed ORFeome preparation, Y2H screens, and plant experiments, which are analyzed in this theses.

Dr. Gurudutta Panda compiled RNA-Seq data and calculated the edge score for protein-protein interactions.

Dr. Pascal Falter-Braun (Institute of Network Biology, Helmholtz Zentrum München, München) compiled the search space of proteins involved in phytohormone signaling.

8.2 Host-pathogen protein interaction project

Ralf Weßling (Department of Plant Microbe Interactions, Max Planck Institute for Plant Breeding Research, Cologne) identified and cloned OECs mapped OEC - Arabidopsis interactions, on which the analyses of this theses are based.

Dr. Stefan Henz (Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen) calculated Tajima's D and Watterson's Θ values.

Dr. Emil Ver Loren van Themaat (Department of Plant Microbe Interactions, Max Planck Institute for Plant Breeding Research, Cologne) did statistical analysis of genetic validation of results plant treatment with *Golovinomyces orontii*.

Christine Gläßer, Klaus F.X. Mayer, Jeffery L. Dangle, and Pascal Falter-Braun supported statistical analysis of convergence.

Petra Epple, Ralf Weßling, Emil Ver Loren van Themaat, and Pascal Falter-Braun supported phenotype statistics.

Kai C. Bader, Klaus F.X. Mayer, Detlef Weigel, Jeffery L. Dangle, and Pascal Falter-Braun supported analysis of natural variation.

A full list of contributions can be found in [142].

9 Acknowledgments

Ich möchte mich bei meinem Betreuer Pascal Falter-Braun bedanken für die Betreuung bei diesen Projekten und dass er mich für die Biologie hinter den Daten sensibilisiert hat.

Mein Dank gilt auch Claus Schwechheimer für Möglichkeit an seinem Lehrstuhl zu promovieren und die kritische Betrachtung meiner Analysen.

Chris-Carolin Schön möchte ich für die Bereitschaft danken, den Vorsitz des Prüfungskomitees übernommen zu haben.

Ich danke Thomas Rattei, der sich bereiterklärt hat meine Arbeit zu begutachten und ein Teil des Prüfungskomitees zu sein.

Meinen Eltern danke ich für die Unterstützung und das Verständnis für meine Arbeit.

Meiner Frau Melina gilt mein größter Dank. Ohne dich hätte ich das nicht geschafft.

A Appendix

A.1 Enriched gene families in AHD 2.0

Table A.1: Table of enriched gene families in AHD 2.0. Compiled by Pascal Falter-Braun.

Gene Family	Source Family	Hormones	Examples in AHD2	Total in Genome	Total in ADH2	Enrichment	P-Value	Add to HormORFeome	Available in AtORFeome2	ORFs to clone
BZR TF Family	TAIR 10	BR	BES1, BZR1	6	6	15,0	0,0001	0	6	0
Histidine Kinase	TAIR 10	CK, ET	AHK1-4, EIN1, 4	16	10	9,4	0,0001	6	4	12
ARF TF Family	TAIR 10	IAA	ARF1, 3-8	23	12	7,8	0,0001	11	6	17
Response Regulator	TAIR 10	CK	ARR3-6	32	10	4,7	0,0001	22	12	20
WRKY TFFamily	TAIR 10	ABA, JA, SA	WRKY2, 18, 70	75	15	3,0	0,0001	60	40	35
AP2-EREBP TF Family	TAIR 10	ABA, BR, CK, GA, JA, ET	AIL5, ERF1, CBF2	146	29	3,0	0,0001	117	87	59
MYB TF Family	TAIR 10	ABA, BR, GA, IAA, JA, ET	AS1, MYB21, MYB24	140	25	2,7	0,0001	115	59	81
AUX-IAA	DATF	ABA, IAA	AXR2, SLR, SHY2	36	13	3,8	0,0001	23	36	0
Whirly TF Family	TAIR 10	SA	WHY1, 3	3	2	10,0	0,0003	1	3	0
ABI3VP1 TF Family	TAIR 10	ABA, IAA	ABI3, ARF16, FUS3	11	5	6,8	0,0004	6	2	9
MAP Kinase (MAPK) Family	TAIR 10	ABA, ET, JA, SA	MPK3,4,6	24	7	4,4	0,0008	17	22	2
ARR-B TF Family	TAIR 10	CK	ARR1,5,10,12	15	4	4,0	0,002	11	3	12
GRAS TF Family	TAIR 10	GA, SA	GAI, RGA, RGL1,2	33	7	3,2	0,006	26	20	13
MLO proteins	TAIR 10	IAA	MLO2, 11, 14	14	4	4,3	0,01	10	4	10
TGA3-like	TAIR 10	SA	TGA1,3,4	4	3	11,3	0,011	1	3	1
Protein tyrosine phosphatase (PTP)	TAIR 10	ABA, ET, SA	PHS1, MKP1, IBR5	29	5	2,6	0,011	24	16	13
EIL - Family	Agris	ABA, ET	EIL1, EIN3	6	3	5,0	0,016	3	2	4
ABC transporters (PDR subfamily)	TAIR 10	ABA, IAA	MDR1, CER5, PEN3	16	3	2,8	0,019	13	0	16
CBL-interacting serione-threonine Protein Kinases (AtCIPKs)	TAIR 10	ABA, IAA, SA	CIPK3, 6, 14	25	3	1,8	0,022	22	18	7
MAP Kinase Kinase (MAPKK) Family	TAIR 10	ABA, ET, IAA, JA	MEK1, MKK3,9, SIS1	11	3	4,1	0,033	8	11	0
RAV TF Family	TAIR 10	IAA	NGA1, 2, 4	11	3	4,1	0,033	8	4	7
bZIP	DATF/Agris	ABA, SA	ABI5, TGA1-6	107	17	1,6	0,046	90	72	35
ZIM	DATF	IAA, JA	JAZ1, 3	27	6	2,2	0,046	21	25	2
Non redundant TOTAL								462	241	221

A.2 Expression values for loci from mRNA

Table A.2: Expression values for 217 loci to be cloned from cDNA. For each locus, three tissues with the highest expression and the respective expression values are shown.

Locus	1. tissue	expr 1st	2. tissue	expr 2nd	3. tissue	expr 3rd
At3g13380	seed	7.54	internode	7.364	node	7.259
At1g09570	petiole	8.896	hypocotyl	8.892	root	8.649
At2g47430	pollen.sac	4.231	rosette.leaf	4	petiole	3.988
At4g18130	seed	10.415	inflorescence, shoot.apex	9.143	flower	9.065
At5g35840	cauline.leaf	7.547	leaf	7.411	flower	7.31
At2g21770	seed	7.178	node	6.871	pollen.sac	6.51
At3g44730	node	8.51	internode	8.247	hypocotyl	7.858
At1g19220	internode	7.781	node	7.595	cauline.leaf	7.389
At2g42010	pollen.sac	9.542	silique	8.044	internode	7.97
At2g25540	silique	11.492	seed	8.458	pollen.sac	6.884
At2g31660	hypocotyl	8.167	leaf, shoot.apex	8.056	petiole	8.024
At1g08420	pollen.sac	8.298	internode	6.87	leaf	6.784
At3g22980	seed	8.141	pollen.sac	7.637	inflorescence, shoot.apex	7.539
At5g44790	pollen.sac	11.719	internode	10.921	root	10.747
At1g63440	pollen.sac	8.995	root	8.258	seed	7.403
At4g15530	pollen.sac	12.119	seed	11.433	cauline.leaf	11.306
At3g51770	pollen.sac	10.791	cauline.leaf	7.551	internode	7.551
At4g00240	pollen.sac	7.211	root	5.181	cotyledon	5.11
At5g10720	root	5.906	pollen.sac	4.381	seed	4.3
At4g32180	hypocotyl	8.819	node	8.769	root	8.753
At3g45140	cotyledon	12.431	rosette.leaf	12.41	flower	12.179
At4g03080	cotyledon	9.064	cauline.leaf	9.048	rosette.leaf	9.034
At4g11830	internode	5.768	node	5.356	hypocotyl	5.309
At3g10550	pollen.sac	7.939	seed	7.361	internode	6.975
At5g04540	pollen.sac	9.894	seed	9.343	root	7.956
At5g04040	seed	9.687	pollen.sac	9.549	shoot	8.441
At5g16780	pollen.sac	8.797	inflorescence, shoot.apex	7.876	seed	7.616
At1g16540	seed	6.341	cauline.leaf	6.099	leaf	5.936
At1g52570	pollen.sac	10.213	flower	6.482	cauline.leaf	6.312
At4g02780	pollen.sac	5.311	leaf, shoot.apex	5.024	seed	5.004
At1g03445	pollen.sac	8.543	cotyledon	6.832	rosette.leaf	6.557
At5g10560	inflorescence, shoot.apex	9.003	leaf, shoot.apex	8.664	cotyledon	8.57
At2g30470	seed	7.626	silique	7.089	inflorescence, shoot.apex	6.453

Table A.2 continued from previous page

Locus	1. tissue	expr 1st	2. tissue	expr 2nd	3. tissue	expr 3rd
At5g60450	internode	9.458	node	9.203	inflorescence, shoot.apex	8.722
At1g79460	silique	7.195	inflorescence, shoot.apex	6.921	node	6.553
At3g55270	leaf	7.657	cauline.leaf	7.46	root	7.395
At4g32010	internode	9.415	node	8.351	inflorescence, shoot.apex	7.508
At3g04580	hypocotyl	7.532	node	7.342	internode	7.213
At4g16280	leaf	7.177	pollen.sac	7.134	internode	7.117
At1g26830	pollen.sac	8.924	internode	8.717	seed	8.528
At1g69670	leaf	6.55	internode	6.487	seed	6.487
At5g36210	leaf, shoot.apex	8.947	inflorescence, shoot.apex	8.787	petiole	8.642
At3g06860	seed	10.674	flower	9.756	leaf	9.62
At1g23540	pollen.sac	9.605	seed	4.609	cotyledon	4.522
At3g24650	seed	10.155	silique	7.257	pollen.sac	5.5
At2g37650	internode	6.946	cauline.leaf	6.575	leaf	6.067
At4g21550	inflorescence, shoot.apex	7.121	pollen.sac	6.853	cauline.leaf	6.732
At5g52310	internode	10.526	leaf	10.23	petiole	10.223
At5g19330	inflorescence, shoot.apex	8.969	shoot	8.921	root	8.92
At2g29060	seed	6.785	internode	6.381	cauline.leaf	6.281
At3g09100	inflorescence, shoot.apex	8.733	leaf	8.381	cauline.leaf	8.309
At4g23850	cauline.leaf	11.374	cotyledon	11.145	leaf	11.075
At1g59750	inflorescence, shoot.apex	9.006	seed	8.448	hypocotyl	8.395
At4g18010	shoot	9.867	cotyledon	9.635	rosette.leaf	8.783
At1g02205	flower	11.62	internode	11.491	node	11.023
At1g02190	flower	9.137	inflorescence, shoot.apex	9.085	pollen.sac	5.614
At5g28210	pollen.sac	6.132	cauline.leaf	4.606	internode	4.561
At5g05730	hypocotyl	10.238	shoot	9.672	petiole	9.648
At5g25350	seed	11.304	root	9.9	leaf	9.888
At2g46530	silique	7.203	inflorescence, shoot.apex	6.993	cauline.leaf	6.888
At1g49190	pollen.sac	6.052	cotyledon	4.593	seed	4.587
At5g07210	silique	6.77	seed	6.161	pollen.sac	4.818
At1g23080	shoot	8.363	hypocotyl	8.352	petiole	8.186
At3g11980	pollen.sac	4.649	internode	4.542	node	4.235
At2g37700	pollen.sac	6.412	root	5.256	seed	5.006
At1g34410	pollen.sac	5.328	cotyledon	4.941	seed	4.935
At3g14440	leaf	6.501	cauline.leaf	6.138	cotyledon	5.697

Table A.2 continued from previous page

Locus	1. tissue	expr 1st	2. tissue	expr 2nd	3. tissue	expr 3rd
At1g63650	inflorescence, shoot.apex	7.688	leaf, shoot.apex	6.678	silique	6.103
At1g34310	rosette.leaf	10.966	cotyledon	10.943	petiole	10.655
At3g01510	cotyledon	8.61	petiole	8.571	rosette.leaf	8.51
At1g77850	silique	7.577	seed	7.551	pollen.sac	7.187
At3g46600	cotyledon	9.193	shoot	9.051	cauline.leaf	8.881
At2g39200	root	7.974	shoot	7.315	pollen.sac	6.407
At5g13320	cauline.leaf	7.845	leaf	7.619	flower	6.945
At2g41510	hypocotyl	7.209	pollen.sac	6.346	cauline.leaf	5.449
At1g11310	cauline.leaf	10.16	leaf	9.929	shoot	9.761
At2g16910	pollen.sac	5.326	seed	4.289	flower	4.064
At1g77110	inflorescence, shoot.apex	8.61	node	7.615	leaf, shoot.apex	7.592
At5g26920	cotyledon	9.277	leaf	9.006	cauline.leaf	8.811
At5g57390	seed	10.421	silique	7.788	inflorescence, shoot.apex	6.813
At4g12400	root	6.604	flower	6.119	seed	6.005
At1g26700	pollen.sac	5.877	seed	4.862	cotyledon	4.82
At1g34170	pollen.sac	4.924	cotyledon	4.745	rosette.leaf	4.638
At2g22330	hypocotyl	11.023	root	10.051	petiole	9.481
At2g17430	root	5.409	internode	5.311	leaf	5.253
At3g52430	cauline.leaf	6.981	cotyledon	6.591	rosette.leaf	6.411
At4g22070	pollen.sac	6.959	root	6.626	shoot	4.621
At5g14930	cauline.leaf	6.354	cotyledon	5.982	leaf	5.975
At3g56700	node	6.807	internode	5.28	pollen.sac	5.251
At1g79360	pollen.sac	9.798	seed	8.326	flower	6.851
At4g36380	seed	8.636	root	7.356	hypocotyl	7.198
At5g55020	pollen.sac	6.793	seed	4.837	rosette.leaf	4.495
At2g32410	inflorescence, shoot.apex	6.738	leaf, shoot.apex	6.232	flower	6.195
At5g60100	cauline.leaf	8.107	rosette.leaf	7.866	node	7.854
At3g03540	seed	8.2	hypocotyl	5.218	root	5.046
At1g29230	pollen.sac	6.791	flower	6.596	seed	6.351
At5g16260	inflorescence, shoot.apex	9.183	seed	8.633	leaf, shoot.apex	8.629
At2g26870	silique	7.261	pollen.sac	7.211	inflorescence, shoot.apex	6.904
At5g25900	seed	10.484	silique	10.308	hypocotyl	9.373
At3g45290	cauline.leaf	8.807	leaf	7.824	internode	7.718
At2g34180	root	5.411	pollen.sac	5.002	hypocotyl	4.474
At1g32130	internode	9.465	cauline.leaf	9.115	root	9.099
At2g33670	pollen.sac	9.231	seed	4.884	cotyledon	4.674
At4g09570	petiole	8.288	rosette.leaf	7.991	root	7.895
At1g07340	pollen.sac	9.305	seed	4.414	flower	4.359
At2g44110	root	6.904	pollen.sac	5.872	seed	4.679

Table A.2 continued from previous page

Locus	1. tissue	expr 1st	2. tissue	expr 2nd	3. tissue	expr 3rd
At3g44560	pollen.sac	4.6	seed	3.791	root	3.712
At3g44550	hypocotyl	7.403	root	7.378	seed	6.314
At4g33790	silique	9.686	node	9.532	internode	9.296
At4g15900	inflorescence, shoot.apex	9.23	leaf, shoot.apex	9.134	hypocotyl	8.432
At5g45810	pollen.sac	10.449	root	5.703	leaf	5.681
At1g17060	seed	9.877	silique	8.071	root	7.513
At3g23770	pollen.sac	5.809	cotyledon	4.777	shoot	4.604
At4g26770	pollen.sac	8.199	seed	5.722	root	5.299
At5g65800	pollen.sac	4.422	root	3.824	seed	3.762
At3g49700	pollen.sac	6.013	cotyledon	5.223	leaf, shoot.apex	5.091
At3g19270	node	7.221	leaf, shoot.apex	6.407	inflorescence, shoot.apex	5.856
At1g42560	pollen.sac	8.381	cotyledon	5.454	shoot	5.437
At2g43840	pollen.sac	6.466	root	6.133	inflorescence, shoot.apex	5.857
At1g67820	seed	8.681	silique	7.697	leaf, shoot.apex	7.135
At3g50280	flower	7.856	cauline.leaf	7.044	hypocotyl	7.024
At1g53940	pollen.sac	5.466	root	5.17	seed	4.628
At4g18470	inflorescence, shoot.apex	5.99	pollen.sac	5.184	leaf, shoot.apex	4.995
At5g01990	seed	9.202	root	8.413	silique	8.041
At5g58850	pollen.sac	5.899	seed	4.528	petiole	4.267
At4g14580	internode	8.13	root	7.795	node	7.034
At3g01080	pollen.sac	6.334	cauline.leaf	5.77	leaf	5.568
At1g62430	leaf	9.508	node	9.021	cotyledon	8.976
At4g05140	pollen.sac	6.029	cotyledon	5.717	leaf	5.569
At1g61630	pollen.sac	6.877	seed	4.877	cotyledon	4.699
At1g72570	seed	4.459	pollen.sac	3.713	cauline.leaf	3.314
At1g52920	pollen.sac	7.962	seed	7.634	cauline.leaf	5.897
At5g22420	pollen.sac	5.877	seed	4.76	flower	4.718
At4g26930	pollen.sac	10.054	cotyledon	4.823	seed	4.738
At1g19640	pollen.sac	11.402	silique	10.253	flower	8.376
At2g31470	pollen.sac	5.351	flower	3.957	seed	3.933
At5g55250	silique	9.109	inflorescence, shoot.apex	7.222	seed	6.544
At1g44090	pollen.sac	6.167	seed	5.324	rosette.leaf	4.83
At3g11480	flower	8.416	cauline.leaf	5.607	pollen.sac	4.658
At1g60980	pollen.sac	5.676	seed	4.503	cotyledon	4.385
At4g26420	pollen.sac	5.879	rosette.leaf	5.211	shoot	5.016
At5g15100	pollen.sac	9.088	flower	4.713	inflorescence, shoot.apex	4.708

Table A.2 continued from previous page

Locus	1. tissue	expr 1st	2. tissue	expr 2nd	3. tissue	expr 3rd
At2g39880	inflorescence, shoot.apex	7.492	leaf, shoot.apex	7.48	pollen.sac	6.97
At1g28300	silique	6.018	pollen.sac	5.741	seed	4.841
At3g63010	root	8.384	cauline.leaf	7.631	leaf	7.488
At1g01030	pollen.sac	5.978	flower	5.751	cotyledon	5.478
At2g44810	seed	6.496	pollen.sac	5.907	cotyledon	5.67
At1g80330	silique	6.059	seed	6.021	pollen.sac	5.587
At5g16530	cauline.leaf	5.781	pollen.sac	5.254	hypocotyl	5.22
At4g25500	internode	9.179	inflorescence, shoot.apex	8.875	silique	8.677
At5g56330	pollen.sac	5.017	cotyledon	4.613	seed	4.46
At4g21690	pollen.sac	6.835	silique	5.332	seed	5.033
At5g52830	root	6.81	pollen.sac	6.737	cotyledon	6.458
At5g18930	internode	7.315	hypocotyl	7.189	node	6.861
At2g40750	rosette.leaf	8.513	cotyledon	8.01	cauline.leaf	7.899
At4g21200	root	3.988	inflorescence, shoot.apex	3.911	cotyledon	3.865
At4g00760	cauline.leaf	7.453	pollen.sac	7.325	hypocotyl	7.266
At1g50960	pollen.sac	5.204	hypocotyl	4.983	leaf	4.414
At2g14920	internode	4.997	hypocotyl	4.483	node	4.253
At1g69560	pollen.sac	5.753	inflorescence, shoot.apex	4.994	seed	4.786
At3g19160	pollen.sac	6.548	rosette.leaf	5.809	cotyledon	5.782
At5g28650	inflorescence, shoot.apex	7.461	node	7.374	pollen.sac	6.756
At5g23000	pollen.sac	6.035	root	4.804	inflorescence, shoot.apex	4.757
At1g09400	pollen.sac	5.763	seed	4.831	leaf	4.737
At1g07745	pollen.sac	6.989	cotyledon	5.769	inflorescence, shoot.apex	5.701
At1g49660	root	9.637	hypocotyl	9.445	leaf	9.022
At4g24650	pollen.sac	3.875	seed	3.861	leaf	3.828
At2g44910	inflorescence, shoot.apex	6.925	seed	5.195	pollen.sac	5.181
At4g24470	inflorescence, shoot.apex	9.552	cauline.leaf	9.162	seed	9.066
At1g48500	root	6.426	seed	5.984	pollen.sac	5.974
At1g35140	cotyledon	10.702	shoot	10.316	rosette.leaf	10.043
At4g37780	pollen.sac	4.823	hypocotyl	4.079	root	4.03
At3g29020	pollen.sac	4.632	flower	3.873	rosette.leaf	3.845
At3g17010	inflorescence, shoot.apex	8.089	pollen.sac	5.641	cotyledon	5.197
At1g56160	pollen.sac	5.454	seed	4.379	internode	4.016
At5g49240	pollen.sac	4.869	seed	4.462	cotyledon	4.412
At2g04450	rosette.leaf	7.62	cotyledon	7.205	cauline.leaf	7.204

Table A.2 continued from previous page

Locus	1. tissue	expr 1st	2. tissue	expr 2nd	3. tissue	expr 3rd
At5g06250	inflorescence, shoot.apex	6.241	pollen.sac	5.743	node	5.692
At1g29860	pollen.sac	6.748	cauline.leaf	5.679	leaf	5.613
At1g16490	internode	7.811	node	6.841	silique	5.836
At5g40430	pollen.sac	3.65	seed	3.605	silique	3.59
At3g29030	flower	9.542	petiole	9.419	rosette.leaf	9.413
At1g80590	pollen.sac	5.783	seed	5.036	leaf	4.668
At2g25230	pollen.sac	3.858	petiole	3.394	internode	3.392
At3g23610	pollen.sac	10.048	internode	7.704	cauline.leaf	7.616
At2g25820	pollen.sac	6.937	petiole	5.682	rosette.leaf	5.28
At2g41310	root	6.728	hypocotyl	6.59	node	6.11
At4g19850	node	4.666	hypocotyl	4.2	internode	4.188
At1g69935	cauline.leaf	8.791	cotyledon	8.687	shoot	8.558
At5g39700	pollen.sac	4.31	internode	3.93	leaf	3.752
At4g28395	pollen.sac	7.195	seed	5.839	cauline.leaf	5.729
At3g56380	pollen.sac	3.599	cauline.leaf	3.37	inflorescence, shoot.apex	3.357
At5g26140	pollen.sac	4.612	cotyledon	3.778	seed	3.722
At3g04280	silique	8.347	seed	7.684	pollen.sac	5.073
At1g29440	silique	6.439	internode	5.869	rosette.leaf	5.701
At3g53250	leaf	4.694	shoot	4.283	cotyledon	3.573
At2g26020	cauline.leaf	10.071	petiole	9.002	rosette.leaf	8.492
At1g08260	inflorescence, shoot.apex	5.34	leaf, shoot.apex	4.693	hypocotyl	4.308
At2g27120	pollen.sac	5.59	flower	3.777	silique	3.767
At5g02310	seed	8.418	cauline.leaf	8.325	internode	8.315
At4g12020	cauline.leaf	7.22	seed	6.703	pollen.sac	6.569
At3g43300	internode	9.087	pollen.sac	9.007	root	8.977
At1g04120	seed	10.564	node	9.257	silique	8.885
At1g59870	cauline.leaf	12.035	shoot	11.839	cotyledon	11.598
At1g13980	internode	9.062	pollen.sac	8.952	inflorescence, shoot.apex	8.908
At5g45050	internode	7.185	cauline.leaf	7.173	node	6.96
At4g39850	seed	8.814	cauline.leaf	8.809	leaf	8.671
At5g04240	seed	8.712	inflorescence, shoot.apex	8.008	silique	8.003
At2g27150	leaf	7.518	cauline.leaf	7.277	seed	6.565
At5g58160	seed	7.347	pollen.sac	7.161	cotyledon	5.052
At5g13680	inflorescence, shoot.apex	8.193	leaf, shoot.apex	7.679	seed	7.35
At2g47000	root	10.138	hypocotyl	7.356	cauline.leaf	7.053
At5g45510	root	10.256	cotyledon	9.458	cauline.leaf	9.274
At2g17820	hypocotyl	8.637	node	8.349	cauline.leaf	7.8
At4g39400	inflorescence, shoot.apex	9.848	shoot	9.848	cauline.leaf	9.827

Table A.2 continued from previous page

Locus	1. tissue	expr 1st	2. tissue	expr 2nd	3. tissue	expr 3rd
At1g49040	pollen.sac	8.655	inflorescence, shoot.apex	8.623	node	8.505
At2g18790	inflorescence, shoot.apex	9.138	root	8.681	seed	8.678

A.3 Search space PHO

Table A.3: Loci in search space PHO of this project including 25 additional loci.

Locus	Locus	Locus	Locus	Locus	Locus
AT1G01030	AT1G56160	AT2G30980	AT3G24650	AT4G25480	AT5G22570
AT1G01060	AT1G56650	AT2G31180	AT3G25730	AT4G25490	AT5G23000
AT1G01090	AT1G57560	AT2G31190	AT3G25890	AT4G25500	AT5G23720
AT1G01140	AT1G59580	AT2G31230	AT3G26090	AT4G25560	AT5G24110
AT1G01250	AT1G59750	AT2G31470	AT3G26790	AT4G26070	AT5G24470
AT1G01360	AT1G59870	AT2G31660	AT3G26810	AT4G26080	AT5G24520
AT1G01560	AT1G59900	AT2G32410	AT3G27140	AT4G26110	AT5G25110
AT1G02120	AT1G59940	AT2G32460	AT3G27310	AT4G26150	AT5G25190
AT1G02190	AT1G60980	AT2G32960	AT3G27320	AT4G26420	AT5G25350
AT1G02205	AT1G61110	AT2G33150	AT3G27670	AT4G26440	AT5G25370
AT1G02400	AT1G61560	AT2G33670	AT3G27810	AT4G26640	AT5G25390
AT1G02630	AT1G61630	AT2G33710	AT3G27920	AT4G26770	AT5G25620
AT1G03445	AT1G62300	AT2G33790	AT3G28470	AT4G26840	AT5G25810
AT1G03800	AT1G62430	AT2G33830	AT3G28860	AT4G26930	AT5G25890
AT1G04120	AT1G62740	AT2G33860	AT3G28910	AT4G27260	AT5G25900
AT1G04240	AT1G62830	AT2G34180	AT3G29020	AT4G27330	AT5G26140
AT1G04250	AT1G63030	AT2G34600	AT3G29030	AT4G27440	AT5G26170
AT1G04310	AT1G63100	AT2G34650	AT3G29350	AT4G27450	AT5G26660
AT1G04370	AT1G63160	AT2G34830	AT3G30180	AT4G27630	AT5G26780
AT1G04550	AT1G63440	AT2G34900	AT3G30210	AT4G27780	AT5G26920
AT1G04710	AT1G63650	AT2G35320	AT3G42960	AT4G27920	AT5G27320
AT1G05000	AT1G63910	AT2G35635	AT3G43300	AT4G27950	AT5G27740
AT1G05180	AT1G64000	AT2G35680	AT3G43440	AT4G28110	AT5G28210
AT1G06160	AT1G64280	AT2G35700	AT3G43700	AT4G28140	AT5G28650
AT1G06180	AT1G64380	AT2G36080	AT3G44540	AT4G28395	AT5G35410
AT1G06390	AT1G64400	AT2G36270	AT3G44550	AT4G28910	AT5G35550
AT1G06400	AT1G64520	AT2G36450	AT3G44560	AT4G29010	AT5G35750
AT1G07230	AT1G64990	AT2G36800	AT3G44620	AT4G29740	AT5G35840
AT1G07340	AT1G65620	AT2G36890	AT3G44730	AT4G29800	AT5G36210
AT1G07420	AT1G66230	AT2G37260	AT3G45140	AT4G29810	AT5G37020
AT1G07430	AT1G66340	AT2G37340	AT3G45290	AT4G30080	AT5G38860
AT1G07520	AT1G66350	AT2G37550	AT3G45640	AT4G30160	AT5G38970
AT1G07530	AT1G66370	AT2G37630	AT3G45780	AT4G30480	AT5G39400
AT1G07745	AT1G66380	AT2G37650	AT3G46130	AT4G30610	AT5G39700
AT1G07880	AT1G66390	AT2G37700	AT3G46600	AT4G30935	AT5G40280
AT1G08260	AT1G66550	AT2G37940	AT3G46930	AT4G30960	AT5G40330
AT1G08420	AT1G66560	AT2G38050	AT3G47600	AT4G31060	AT5G40350
AT1G08810	AT1G66600	AT2G38120	AT3G48040	AT4G31500	AT5G40360
AT1G08920	AT1G67080	AT2G38310	AT3G48090	AT4G31550	AT5G40430
AT1G09400	AT1G67260	AT2G38340	AT3G48100	AT4G31780	AT5G40440
AT1G09540	AT1G67560	AT2G38470	AT3G48360	AT4G31800	AT5G40990
AT1G09570	AT1G67710	AT2G38490	AT3G48430	AT4G31920	AT5G41570
AT1G09700	AT1G67820	AT2G39200	AT3G48610	AT4G32010	AT5G41920
AT1G09950	AT1G68150	AT2G39220	AT3G48690	AT4G32180	AT5G42650
AT1G10210	AT1G68210	AT2G39250	AT3G48700	AT4G32280	AT5G42750
AT1G10470	AT1G68320	AT2G39550	AT3G48920	AT4G32285	AT5G43290
AT1G10940	AT1G68550	AT2G39760	AT3G49530	AT4G32410	AT5G43410
AT1G11000	AT1G68620	AT2G39880	AT3G49690	AT4G32540	AT5G43590

Table A.3 continued from previous page

Locus	Locus	Locus	Locus	Locus	Locus
AT1G11310	AT1G68840	AT2G39940	AT3G49700	AT4G32570	AT5G43830
AT1G11680	AT1G69010	AT2G40180	AT3G49950	AT4G32730	AT5G43940
AT1G12270	AT1G69270	AT2G40220	AT3G50060	AT4G32800	AT5G44030
AT1G12610	AT1G69310	AT2G40330	AT3G50070	AT4G33430	AT5G44200
AT1G12630	AT1G69560	AT2G40340	AT3G50110	AT4G33450	AT5G44210
AT1G12820	AT1G69670	AT2G40350	AT3G50260	AT4G33520	AT5G44420
AT1G12890	AT1G69810	AT2G40670	AT3G50280	AT4G33790	AT5G44430
AT1G12980	AT1G69935	AT2G40740	AT3G50650	AT4G33950	AT5G44790
AT1G13260	AT1G70330	AT2G40750	AT3G50660	AT4G34410	AT5G45050
AT1G13960	AT1G70560	AT2G40940	AT3G50750	AT4G34990	AT5G45340
AT1G13980	AT1G70700	AT2G41310	AT3G51060	AT4G35230	AT5G45510
AT1G14280	AT1G70940	AT2G41510	AT3G51590	AT4G35790	AT5G45550
AT1G14350	AT1G71130	AT2G41710	AT3G51770	AT4G36380	AT5G45710
AT1G14410	AT1G71260	AT2G42010	AT3G52180	AT4G36450	AT5G45810
AT1G14920	AT1G71450	AT2G42430	AT3G52430	AT4G36540	AT5G45820
AT1G15100	AT1G71520	AT2G42680	AT3G52930	AT4G36710	AT5G46350
AT1G15360	AT1G71830	AT2G42880	AT3G53200	AT4G36780	AT5G46570
AT1G15550	AT1G71860	AT2G43710	AT3G53250	AT4G36830	AT5G46790
AT1G16060	AT1G71960	AT2G43790	AT3G53480	AT4G36900	AT5G47100
AT1G16370	AT1G72360	AT2G43820	AT3G53710	AT4G36920	AT5G47220
AT1G16390	AT1G72450	AT2G43840	AT3G54220	AT4G37050	AT5G47230
AT1G16490	AT1G72520	AT2G44050	AT3G54320	AT4G37060	AT5G48150
AT1G16540	AT1G72570	AT2G44110	AT3G54720	AT4G37070	AT5G48170
AT1G17060	AT1G72770	AT2G44745	AT3G54950	AT4G37260	AT5G48870
AT1G17380	AT1G73030	AT2G44810	AT3G54990	AT4G37390	AT5G48880
AT1G17420	AT1G73220	AT2G44840	AT3G55270	AT4G37580	AT5G49240
AT1G17440	AT1G73410	AT2G44910	AT3G55530	AT4G37650	AT5G49330
AT1G17550	AT1G73500	AT2G44940	AT3G55730	AT4G37750	AT5G49520
AT1G17730	AT1G73590	AT2G44950	AT3G56380	AT4G37780	AT5G49620
AT1G17950	AT1G73670	AT2G45150	AT3G56400	AT4G37870	AT5G49720
AT1G17990	AT1G73740	AT2G45160	AT3G56700	AT4G38130	AT5G49980
AT1G18080	AT1G73830	AT2G45420	AT3G57040	AT4G38620	AT5G50080
AT1G18150	AT1G74080	AT2G45820	AT3G57140	AT4G38630	AT5G51190
AT1G18350	AT1G74430	AT2G46070	AT3G57600	AT4G38830	AT5G51550
AT1G18400	AT1G74650	AT2G46130	AT3G58680	AT4G38850	AT5G51760
AT1G18570	AT1G74710	AT2G46310	AT3G58710	AT4G39030	AT5G51810
AT1G18710	AT1G74890	AT2G46370	AT3G59790	AT4G39350	AT5G51990
AT1G18860	AT1G74910	AT2G46400	AT3G60460	AT4G39400	AT5G52020
AT1G18970	AT1G74930	AT2G46530	AT3G60490	AT4G39410	AT5G52040
AT1G19050	AT1G74950	AT2G46790	AT3G60620	AT4G39780	AT5G52240
AT1G19180	AT1G75000	AT2G46830	AT3G60630	AT4G39850	AT5G52260
AT1G19190	AT1G75080	AT2G46870	AT3G61250	AT4G39950	AT5G52300
AT1G19210	AT1G75490	AT2G47000	AT3G61630	AT5G01240	AT5G52310
AT1G19220	AT1G75830	AT2G47190	AT3G61830	AT5G01290	AT5G52510
AT1G19350	AT1G76090	AT2G47240	AT3G61850	AT5G01540	AT5G52600
AT1G19610	AT1G76420	AT2G47260	AT3G61860	AT5G01550	AT5G52830
AT1G19640	AT1G76680	AT2G47430	AT3G61970	AT5G01600	AT5G53160
AT1G19850	AT1G76690	AT2G47460	AT3G62340	AT5G01810	AT5G53280
AT1G20330	AT1G77110	AT2G47520	AT3G62610	AT5G01820	AT5G53290
AT1G20780	AT1G77120	AT2G47770	AT3G62670	AT5G01900	AT5G53470
AT1G21450	AT1G77200	AT3G01040	AT3G62980	AT5G01990	AT5G53760

Table A.3 continued from previous page

Locus	Locus	Locus	Locus	Locus	Locus
AT1G21690	AT1G77470	AT3G01080	AT3G63010	AT5G02310	AT5G53950
AT1G21910	AT1G77640	AT3G01140	AT3G63200	AT5G02320	AT5G54190
AT1G21970	AT1G77690	AT3G01510	AT3G63210	AT5G02810	AT5G54230
AT1G22070	AT1G77760	AT3G01530	AT4G00120	AT5G03280	AT5G54510
AT1G22190	AT1G77850	AT3G01970	AT4G00150	AT5G03310	AT5G55020
AT1G22640	AT1G77920	AT3G02410	AT4G00240	AT5G03455	AT5G55160
AT1G22770	AT1G78080	AT3G02610	AT4G00540	AT5G03690	AT5G55170
AT1G22810	AT1G78440	AT3G02620	AT4G00710	AT5G03730	AT5G55250
AT1G22985	AT1G78590	AT3G02630	AT4G00760	AT5G04040	AT5G56010
AT1G23080	AT1G78700	AT3G02800	AT4G01026	AT5G04190	AT5G56110
AT1G23540	AT1G79040	AT3G02940	AT4G01250	AT5G04240	AT5G56270
AT1G23860	AT1G79180	AT3G03450	AT4G01370	AT5G04430	AT5G56300
AT1G24180	AT1G79360	AT3G03520	AT4G01500	AT5G04540	AT5G56330
AT1G24260	AT1G79410	AT3G03530	AT4G01680	AT5G04870	AT5G56580
AT1G24590	AT1G79460	AT3G03540	AT4G01720	AT5G05170	AT5G56610
AT1G25220	AT1G79530	AT3G03740	AT4G02570	AT5G05410	AT5G56860
AT1G25340	AT1G79700	AT3G03850	AT4G02600	AT5G05440	AT5G56970
AT1G25470	AT1G80330	AT3G04240	AT4G02780	AT5G05580	AT5G57050
AT1G25490	AT1G80340	AT3G04280	AT4G03080	AT5G05690	AT5G57090
AT1G25560	AT1G80580	AT3G04580	AT4G03190	AT5G05700	AT5G57390
AT1G26120	AT1G80590	AT3G04670	AT4G03560	AT5G05730	AT5G57620
AT1G26700	AT1G80840	AT3G05120	AT4G03960	AT5G06100	AT5G57630
AT1G26780	AT2G01420	AT3G05420	AT4G04450	AT5G06250	AT5G57685
AT1G26830	AT2G01450	AT3G06110	AT4G04500	AT5G06950	AT5G57740
AT1G27320	AT2G01570	AT3G06230	AT4G05100	AT5G06960	AT5G57800
AT1G28160	AT2G01650	AT3G06460	AT4G05110	AT5G07070	AT5G58080
AT1G28300	AT2G01760	AT3G06470	AT4G05120	AT5G07100	AT5G58140
AT1G28330	AT2G01830	AT3G06490	AT4G05130	AT5G07210	AT5G58160
AT1G28360	AT2G02560	AT3G06860	AT4G05140	AT5G07310	AT5G58220
AT1G28370	AT2G02740	AT3G07390	AT4G08150	AT5G07340	AT5G58230
AT1G29230	AT2G02820	AT3G08500	AT4G08250	AT5G07580	AT5G58350
AT1G29280	AT2G02950	AT3G08550	AT4G08260	AT5G07690	AT5G58380
AT1G29440	AT2G03340	AT3G09100	AT4G08920	AT5G07700	AT5G58850
AT1G29450	AT2G03760	AT3G09230	AT4G08950	AT5G08130	AT5G58950
AT1G29460	AT2G04240	AT3G09370	AT4G09460	AT5G09870	AT5G59220
AT1G29500	AT2G04350	AT3G09990	AT4G09570	AT5G10030	AT5G59450
AT1G29510	AT2G04450	AT3G10550	AT4G11030	AT5G10280	AT5G59780
AT1G29860	AT2G04550	AT3G10940	AT4G11070	AT5G10510	AT5G60100
AT1G30135	AT2G04880	AT3G11020	AT4G11140	AT5G10560	AT5G60120
AT1G30270	AT2G04890	AT3G11170	AT4G11260	AT5G10720	AT5G60410
AT1G30330	AT2G06050	AT3G11240	AT4G11330	AT5G10930	AT5G60450
AT1G30650	AT2G11810	AT3G11410	AT4G11830	AT5G11050	AT5G60890
AT1G31340	AT2G13540	AT3G11440	AT4G11840	AT5G11190	AT5G61380
AT1G31812	AT2G14920	AT3G11480	AT4G11850	AT5G11270	AT5G61420
AT1G31880	AT2G16070	AT3G11540	AT4G12020	AT5G11320	AT5G61430
AT1G32130	AT2G16720	AT3G11580	AT4G12110	AT5G11510	AT5G61590
AT1G32320	AT2G16910	AT3G11980	AT4G12350	AT5G11590	AT5G61600
AT1G32640	AT2G16940	AT3G12120	AT4G12400	AT5G12210	AT5G61790
AT1G33270	AT2G17230	AT3G12250	AT4G12470	AT5G12870	AT5G61890
AT1G33760	AT2G17290	AT3G12390	AT4G13260	AT5G13080	AT5G62000
AT1G34120	AT2G17430	AT3G12490	AT4G13480	AT5G13170	AT5G62470

Table A.3 continued from previous page

Locus	Locus	Locus	Locus	Locus	Locus
AT1G34170	AT2G17480	AT3G12720	AT4G13520	AT5G13220	AT5G62920
AT1G34310	AT2G17820	AT3G12820	AT4G13620	AT5G13320	AT5G63980
AT1G34390	AT2G18170	AT3G13224	AT4G14400	AT5G13330	AT5G64740
AT1G34410	AT2G18470	AT3G13380	AT4G14550	AT5G13630	AT5G64750
AT1G34670	AT2G18550	AT3G13540	AT4G14560	AT5G13680	AT5G64810
AT1G35140	AT2G18790	AT3G13730	AT4G14580	AT5G13910	AT5G64813
AT1G35240	AT2G19500	AT3G13840	AT4G14713	AT5G13930	AT5G65130
AT1G35515	AT2G20180	AT3G13890	AT4G14720	AT5G14310	AT5G65210
AT1G35520	AT2G20350	AT3G13920	AT4G15530	AT5G14340	AT5G65230
AT1G35540	AT2G20610	AT3G14230	AT4G15560	AT5G14750	AT5G65510
AT1G35670	AT2G20880	AT3G14440	AT4G15900	AT5G14930	AT5G65790
AT1G36060	AT2G20950	AT3G14720	AT4G16110	AT5G15100	AT5G65800
AT1G37130	AT2G21050	AT3G15150	AT4G16250	AT5G15130	AT5G65970
AT1G42560	AT2G21770	AT3G15210	AT4G16280	AT5G15310	AT5G66400
AT1G43160	AT2G21900	AT3G15356	AT4G16420	AT5G15860	AT5G66770
AT1G43800	AT2G22090	AT3G15540	AT4G16750	AT5G15970	AT5G67000
AT1G43950	AT2G22200	AT3G15730	AT4G17230	AT5G16080	AT5G67030
AT1G44090	AT2G22330	AT3G16280	AT4G17490	AT5G16230	AT5G67190
AT1G44830	AT2G23290	AT3G16420	AT4G17500	AT5G16240	AT5G67300
AT1G46768	AT2G23320	AT3G16570	AT4G17720	AT5G16260	AT5G67500
AT1G48000	AT2G23340	AT3G16770	AT4G17780	AT5G16480	ATMG01090
AT1G48260	AT2G23430	AT3G16857	AT4G17870	AT5G16530	AT5G62320
AT1G48500	AT2G24570	AT3G17010	AT4G18010	AT5G16600	AT1G03790
AT1G48630	AT2G25000	AT3G17510	AT4G18020	AT5G16770	AT3G50500
AT1G49040	AT2G25090	AT3G17860	AT4G18130	AT5G16780	AT4G36930
AT1G49120	AT2G25180	AT3G18040	AT4G18170	AT5G17420	AT5G45830
AT1G49190	AT2G25230	AT3G18130	AT4G18450	AT5G17430	AT1G35560
AT1G49430	AT2G25490	AT3G18240	AT4G18470	AT5G17490	AT5G66880
AT1G49640	AT2G25540	AT3G18910	AT4G18700	AT5G17690	AT1G04400
AT1G49660	AT2G25820	AT3G18980	AT4G18710	AT5G17800	AT5G55910
AT1G50200	AT2G26010	AT3G18990	AT4G18770	AT5G18010	AT3G32130
AT1G50420	AT2G26020	AT3G19160	AT4G18780	AT5G18020	AT2G42620
AT1G50600	AT2G26070	AT3G19270	AT4G18880	AT5G18030	AT1G15050
AT1G50640	AT2G26290	AT3G19420	AT4G18890	AT5G18050	AT1G15580
AT1G50680	AT2G26560	AT3G19770	AT4G19230	AT5G18060	AT1G80390
AT1G50960	AT2G26710	AT3G19820	AT4G19850	AT5G18080	AT4G28640
AT1G51120	AT2G26870	AT3G20310	AT4G21200	AT5G18450	AT3G16500
AT1G51190	AT2G26950	AT3G20550	AT4G21440	AT5G18560	AT2G01200
AT1G51500	AT2G26960	AT3G20660	AT4G21550	AT5G18930	AT3G17600
AT1G51600	AT2G26980	AT3G20770	AT4G21670	AT5G19000	AT3G04730
AT1G51660	AT2G27050	AT3G20840	AT4G21690	AT5G19010	AT3G62100
AT1G51965	AT2G27070	AT3G21175	AT4G22070	AT5G19140	AT4G29080
AT1G52340	AT2G27120	AT3G21220	AT4G22340	AT5G19180	AT5G43700
AT1G52570	AT2G27150	AT3G22850	AT4G22680	AT5G19330	AT1G04100
AT1G52920	AT2G27210	AT3G22980	AT4G22750	AT5G19770	AT2G33310
AT1G53170	AT2G28305	AT3G23000	AT4G23550	AT5G19780	AT1G73000
AT1G53510	AT2G28350	AT3G23050	AT4G23650	AT5G19790	AT2G26040
AT1G53910	AT2G28550	AT3G23140	AT4G23750	AT5G20270	AT5G45870
AT1G53940	AT2G29060	AT3G23150	AT4G23810	AT5G20410	AT4G23980
AT1G54270	AT2G29090	AT3G23220	AT4G23850	AT5G20570	AT2G24765
AT1G54350	AT2G29380	AT3G23230	AT4G24210	AT5G20730	AT1G51950

Table A.3 continued from previous page

Locus	Locus	Locus	Locus	Locus	Locus
AT1G54490	AT2G29390	AT3G23240	AT4G24230	AT5G20900	AT3G23030
AT1G54990	AT2G29980	AT3G23250	AT4G24240	AT5G20910	AT2G22670
AT1G55010	AT2G30020	AT3G23610	AT4G24250	AT5G21010	AT2G46990
AT1G55180	AT2G30250	AT3G23770	AT4G24400	AT5G22010	AT4G18620
AT1G55580	AT2G30360	AT3G24280	AT4G24470	AT5G22110	AT5G45860
AT1G55600	AT2G30470	AT3G24310	AT4G24650	AT5G22260	AT4G37790
AT1G55610	AT2G30590	AT3G24500	AT4G25420	AT5G22420	
AT1G56070	AT2G30860	AT3G24520	AT4G25470	AT5G22500	

A.4 PhI interactions

Table A.4: PhI interactions

IntA	IntB	IntA	IntB	IntA	IntB
AT1G01030	AT3G17600	AT1G35560	AT3G24520	AT1G31880	AT2G25090
AT1G01140	AT2G23290	AT1G35560	AT3G29350	AT1G31880	AT5G08130
AT1G01360	AT4G26080	AT1G35560	AT1G64520	AT1G31880	AT2G42880
AT1G04100	AT1G15050	AT1G35560	AT3G57040	AT1G06400	AT4G17720
AT1G04100	AT2G33310	AT1G35560	AT5G62000	AT2G26980	AT5G25190
AT1G04100	AT3G04730	AT1G35560	AT2G46130	AT1G23860	AT3G61860
AT1G04100	AT3G16500	AT1G15550	AT1G35560	AT1G14920	AT2G42880
AT1G04100	AT3G23030	AT1G21690	AT1G35560	AT5G53160	AT5G59220
AT1G04100	AT4G14560	AT4G24400	AT5G47100	AT1G18350	AT2G34650
AT1G04240	AT1G04250	AT1G04250	AT4G14560	AT4G32570	AT5G41920
AT1G04250	AT3G15540	AT4G37260	AT5G58950	AT1G51660	AT4G32570
AT1G04550	AT1G04550	AT1G53170	AT4G37260	AT4G14713	AT4G32570
AT1G04550	AT2G33310	AT2G29380	AT4G37260	AT4G14720	AT4G32570
AT1G04550	AT3G04730	AT4G37260	AT5G17690	AT1G18350	AT4G32570
AT1G04550	AT3G15540	AT1G01140	AT4G37260	AT1G10210	AT5G40440
AT1G07430	AT4G01026	AT2G01760	AT4G37260	AT4G27920	AT5G57050
AT1G07430	AT4G18620	AT2G30590	AT4G37260	AT1G69010	AT5G38860
AT1G10940	AT2G36270	AT1G19220	AT3G23050	AT4G39780	AT5G21010
AT1G15050	AT3G23050	AT4G32010	AT5G55170	AT4G00120	AT4G36930
AT1G15550	AT5G60120	AT1G51600	AT4G24470	AT2G30020	AT2G43790
AT1G17380	AT3G29350	AT3G21175	AT4G24470	AT5G01810	AT5G47100
AT1G17380	AT4G28910	AT3G23610	AT3G45640	AT2G40750	AT4G26110
AT1G18400	AT5G41920	AT2G29380	AT5G45830	AT1G25490	AT5G51760
AT1G19050	AT3G29350	AT1G35560	AT4G11070	AT1G25490	AT5G45810
AT1G22770	AT1G51950	AT1G35560	AT4G29800	AT4G29810	AT5G58950
AT1G22770	AT2G23290	AT1G35560	AT1G50600	AT1G51660	AT2G43790
AT1G22770	AT2G37630	AT1G18400	AT1G35560	AT2G43790	AT4G29810
AT1G22770	AT2G44950	AT1G35560	AT1G53170	AT2G04550	AT5G47100
AT1G22770	AT3G16500	AT1G35560	AT5G13080	AT3G21175	AT5G11270
AT1G22770	AT4G08150	AT1G35560	AT5G43290	AT2G02560	AT3G21175
AT1G22770	AT4G32570	AT1G14280	AT1G35560	AT1G19180	AT3G29350
AT1G22770	AT4G37260	AT1G35560	AT2G39250	AT3G52930	AT5G03690
AT1G22770	AT5G60120	AT1G35560	AT2G01760	AT1G64280	AT5G65210
AT1G25470	AT5G08130	AT1G35560	AT2G30590	AT2G02950	AT2G39760
AT1G30270	AT5G47100	AT1G35560	AT3G45640	AT2G39760	AT3G24520
AT1G32640	AT3G29350	AT1G35560	AT4G36540	AT1G25490	AT3G48090
AT1G37130	AT1G37130	AT1G28360	AT1G35560	AT1G21690	AT1G25490
AT1G37130	AT5G18930	AT1G35560	AT5G55170	AT1G15550	AT5G25890
AT1G37130	AT5G60120	AT1G35560	AT3G01970	AT1G51600	AT3G21175
AT1G50600	AT2G20350	AT1G17550	AT4G27920	AT1G69010	AT1G69010
AT1G51660	AT4G08150	AT1G04250	AT3G61830	AT5G43830	AT5G43830
AT1G51950	AT3G24520	AT1G04250	AT1G04550	AT1G04240	AT4G14560
AT1G51950	AT3G45640	AT2G25090	AT4G37260	AT1G04240	AT3G15540
AT1G51950	AT3G61830	AT1G51660	AT4G37260	AT3G15540	AT3G23050
AT1G52340	AT1G52340	AT2G42880	AT4G37260	AT2G39760	AT4G39780
AT1G53170	AT5G19000	AT1G10210	AT4G37260	AT4G39780	AT5G19000
AT1G53510	AT2G01570	AT3G14720	AT4G37260	AT1G19180	AT1G30135
AT1G53910	AT2G38490	AT4G32010	AT5G41920	AT1G50600	AT5G67000

Table A.4 continued from previous page

IntA	IntB	IntA	IntB	IntA	IntB
AT1G54490	AT5G60120	AT3G61830	AT4G32010	AT3G42960	AT3G42960
AT1G64280	AT5G06950	AT1G18350	AT4G32010	AT4G00120	AT5G55170
AT1G64520	AT4G08150	AT1G07340	AT1G10210	AT1G01140	AT5G47100
AT1G65620	AT2G37630	AT1G07340	AT1G18350	AT3G21175	AT3G21175
AT1G73410	AT2G20350	AT3G15150	AT3G23610	AT3G01970	AT3G21175
AT1G73410	AT4G08150	AT1G35560	AT3G15150	AT5G58220	AT5G58220
AT1G75080	AT2G01570	AT1G35560	AT2G26980	AT3G20550	AT5G58220
AT2G01570	AT2G30590	AT1G35560	AT2G42880	AT2G26560	AT2G26560
AT2G01570	AT2G42880	AT1G35560	AT4G14713	AT4G26110	AT4G26110
AT2G01570	AT3G14720	AT1G35560	AT1G80340	AT1G19180	AT1G19180
AT2G01570	AT4G16110	AT1G10210	AT1G35560	AT3G45640	AT3G63210
AT2G01570	AT5G41920	AT1G18350	AT1G35560	AT1G53170	AT2G39760
AT2G01760	AT2G23290	AT1G35560	AT3G14720	AT2G39760	AT2G39760
AT2G01760	AT2G46990	AT2G30020	AT4G01370	AT1G25490	AT4G29800
AT2G01760	AT3G16500	AT2G44050	AT5G01810	AT1G25490	AT5G64813
AT2G01760	AT3G17600	AT1G73410	AT4G30080	AT1G18400	AT1G25490
AT2G01760	AT3G62100	AT2G25490	AT4G36930	AT1G25490	AT1G53170
AT2G01760	AT4G00120	AT1G74950	AT4G28910	AT1G25490	AT4G36540
AT2G02560	AT4G08150	AT1G50960	AT3G54220	AT1G04240	AT5G25890
AT2G23290	AT2G25090	AT1G31880	AT2G44745	AT3G23050	AT5G25890
AT2G23290	AT5G53160	AT1G31880	AT5G51760	AT4G14560	AT5G25890
AT2G23290	AT5G58950	AT1G31880	AT5G55910	AT1G54990	AT2G38120
AT2G23290	AT5G67300	AT1G31880	AT1G69560	AT1G69010	AT5G08130
AT2G25490	AT2G44950	AT1G66560	AT5G22570	AT1G04240	AT1G04550
AT2G26980	AT5G47100	AT1G35560	AT5G05440	AT2G29380	AT4G27920
AT2G28350	AT3G17600	AT1G73410	AT5G25190	AT3G15540	AT3G61830
AT2G29380	AT4G01026	AT1G64280	AT3G12250	AT2G04890	AT5G67000
AT2G29380	AT5G45860	AT3G29350	AT4G16110	AT2G43790	AT5G56580
AT2G30590	AT2G46990	AT1G74910	AT3G03540	AT3G23050	AT3G61830
AT2G30590	AT3G62100	AT1G14920	AT4G36930	AT2G25090	AT5G47100
AT2G33310	AT3G15540	AT4G32570	AT5G51760	AT3G17510	AT5G47100
AT2G33310	AT3G23050	AT3G03450	AT5G54190	AT2G26560	AT5G41920
AT2G38490	AT3G17600	AT4G01370	AT4G29810	AT1G19180	AT5G41920
AT2G38490	AT4G28640	AT5G01600	AT5G01600	AT1G10210	AT3G63210
AT2G39220	AT2G44950	AT5G67300	AT5G67300	AT1G25490	AT2G25090
AT2G40330	AT4G26080	AT5G53160	AT5G67300	AT1G25490	AT2G37630
AT2G44950	AT2G44950	AT1G72450	AT4G28910	AT1G25490	AT5G08130
AT2G44950	AT3G24520	AT2G44050	AT2G44050	AT1G25490	AT2G26980
AT2G44950	AT3G48090	AT2G38490	AT2G44050	AT3G61830	AT5G25890
AT2G44950	AT4G08150	AT2G24570	AT2G44050	AT1G04550	AT5G25890
AT2G44950	AT5G51760	AT1G68550	AT2G38490	AT2G30980	AT4G18890
AT3G04240	AT3G28910	AT1G07430	AT5G53160	AT1G24590	AT1G73410
AT3G04240	AT5G60120	AT1G01360	AT1G07430	AT1G24590	AT4G36930
AT3G04730	AT3G15540	AT2G25490	AT4G02570	AT1G22070	AT3G12250
AT3G15540	AT3G17600	AT4G28910	AT5G13220	AT2G37630	AT5G51760
AT3G15540	AT3G23030	AT2G25490	AT4G08150	AT1G75000	AT5G08130
AT3G15540	AT4G14560	AT1G21690	AT1G77470	AT5G08130	AT5G38860
AT3G16500	AT3G23050	AT1G31880	AT1G51660	AT5G05730	AT5G08130
AT3G16500	AT3G61830	AT1G31880	AT1G54490	AT4G33790	AT5G08130
AT3G16500	AT4G14560	AT1G31880	AT5G16080	AT1G73410	AT5G41920
AT3G16500	AT5G17690	AT1G31880	AT1G37130	AT1G69560	AT5G48870

Table A.4 continued from previous page

IntA	IntB	IntA	IntB	IntA	IntB
AT3G17600	AT4G30080	AT1G31880	AT2G02950	AT1G13980	AT3G11410
AT3G23030	AT3G23050	AT1G15550	AT1G31880	AT3G11410	AT5G45830
AT3G62100	AT4G11070	AT1G31880	AT5G61380	AT5G05730	AT5G20900
AT3G62100	AT4G24240	AT1G31880	AT2G32960	AT4G32010	AT5G17490
AT3G62100	AT5G61380	AT1G31880	AT2G24570	AT5G05410	AT5G45830
AT4G08150	AT4G08150	AT5G22570	AT5G62000	AT1G15550	AT1G24590
AT4G08150	AT5G60120	AT1G64280	AT5G06960	AT1G22070	AT1G64280
AT4G08150	AT5G64810	AT1G14920	AT4G24210	AT1G04550	AT3G23050
AT4G11070	AT5G60120	AT3G29350	AT4G31920	AT1G04550	AT4G14560
AT4G14560	AT4G14560	AT1G07430	AT2G25000	AT1G04550	AT2G01760
AT4G14713	AT5G17690	AT2G25000	AT3G24520	AT2G01760	AT2G22090
AT4G15560	AT4G15560	AT2G25000	AT3G29350	AT4G33520	AT5G17490
AT4G16420	AT5G51760	AT2G25000	AT2G25000	AT1G23860	AT2G01760
AT4G26080	AT4G27920	AT2G25490	AT4G32570	AT5G07310	AT5G48150
AT4G26080	AT5G05440	AT4G08150	AT4G32570	AT1G18350	AT1G24590
AT4G26080	AT5G53160	AT1G37130	AT4G32570	AT2G04890	AT5G07310
AT4G28910	AT4G32570	AT3G29350	AT4G32570	AT2G37630	AT2G37630
AT5G04870	AT5G17690	AT1G15550	AT4G32570	AT1G37130	AT2G37630
AT5G06950	AT5G06960	AT2G40330	AT5G57050	AT2G37630	AT2G42880
AT5G10930	AT5G47100	AT5G05440	AT5G57050	AT5G08130	AT5G08130
AT5G17690	AT5G17690	AT5G53160	AT5G57050	AT3G11410	AT4G27920
AT5G35410	AT5G47100	AT1G01360	AT5G57050	AT3G61830	AT3G61830
AT5G45820	AT5G47100	AT1G53510	AT3G03450	AT1G04550	AT3G61830
AT5G47100	AT5G58380	AT3G03450	AT5G64813	AT2G22090	AT2G22090
AT5G60120	AT5G60120	AT1G51660	AT3G45640	AT4G27920	AT5G59220
AT1G04250	AT1G04250	AT4G01370	AT4G26070	AT2G04890	AT4G14720
AT1G73410	AT4G37260	AT2G45820	AT3G44620	AT4G14713	AT4G14720
AT1G78590	AT1G78590	AT2G01760	AT2G44050	AT1G23860	AT1G23860
AT4G32010	AT5G51760	AT2G30590	AT2G44050	AT2G38490	AT3G25890
AT4G24470	AT4G24470	AT4G27450	AT5G43830	AT4G08150	AT4G32010
AT3G23610	AT5G45810	AT1G19180	AT4G28910	AT1G15550	AT4G32010
AT4G36930	AT4G36930	AT1G70700	AT4G28910	AT2G01760	AT4G36930
AT5G45830	AT5G51760	AT1G31880	AT1G53510	AT2G44050	AT5G13930
AT5G45830	AT5G45830	AT1G31880	AT5G60120	AT1G07430	AT4G27920
AT1G35560	AT5G51760	AT1G31880	AT1G53170	AT4G14713	AT4G28910
AT1G35560	AT1G73410	AT1G31880	AT2G46870	AT1G35540	AT3G61830
AT1G35560	AT4G00240	AT1G22640	AT1G31880	AT2G37630	AT3G24520
AT1G35560	AT5G45810	AT1G31880	AT1G69810	AT1G15550	AT2G37630
AT1G35560	AT4G36930	AT1G01140	AT1G31880	AT2G04550	AT3G11410
AT1G35560	AT1G35560	AT1G14280	AT1G31880	AT2G40330	AT3G11410
AT1G77200	AT4G08150	AT1G31880	AT2G01760	AT3G11410	AT5G05440
AT1G17550	AT2G40330	AT4G11070	AT5G22570	AT2G38310	AT3G11410
AT1G17550	AT5G05440	AT1G80840	AT5G22570	AT3G11410	AT5G53160
AT1G17550	AT5G53160	AT4G18880	AT5G64813	AT1G01360	AT3G11410
AT1G01360	AT1G17550	AT1G14920	AT2G44950	AT4G28910	AT5G20900
AT4G37260	AT5G67300	AT1G14920	AT1G53510	AT2G22090	AT5G61380
AT2G25490	AT4G37260	AT1G14920	AT1G70700	AT2G33830	AT3G03530
AT3G29350	AT4G37260	AT2G29380	AT5G53160	AT4G08150	AT5G17490
AT4G37260	AT5G53160	AT1G32640	AT1G70700	AT5G17490	AT5G54190
AT2G04550	AT3G62340	AT1G32640	AT3G20550	AT1G75080	AT5G17490
AT2G04550	AT3G23610	AT1G17380	AT1G32640	AT1G07430	AT4G14720

Table A.4 continued from previous page

IntA	IntB	IntA	IntB	IntA	IntB
AT3G23610	AT4G29810	AT1G69810	AT2G25000	AT1G23860	AT4G08150
AT3G23610	AT3G29350	AT2G01760	AT2G25000	AT2G18170	AT4G29810
AT1G21690	AT3G23610	AT2G01760	AT4G32570	AT1G24590	AT4G11070
AT1G15550	AT4G36930	AT1G01360	AT2G29380	AT1G24590	AT5G60120
AT1G07430	AT5G45830	AT2G42880	AT3G03450	AT1G22070	AT5G06950
AT1G35560	AT2G25490	AT4G01370	AT5G56580	AT5G08130	AT5G60120
AT1G35560	AT4G08150	AT4G14720	AT4G28910	AT1G32640	AT5G01820
AT1G35560	AT2G04550	AT1G63160	AT1G77470		
AT1G35560	AT4G29810	AT1G31880	AT5G04190		

A.5 PhI_{out} interactions

Table A.5: PhI_{out} Interactions

IntA	IntB	IntA	IntB	IntA	IntB
AT1G01360	AT4G26080	AT3G08710	AT5G18930	AT4G32570	AT5G62600
AT1G07430	AT4G01026	AT1G13690	AT3G23610	AT1G68160	AT4G32570
AT1G17380	AT4G28910	AT1G64000	AT2G22880	AT4G32570	AT5G60690
AT1G30270	AT5G47100	AT4G32010	AT5G05790	AT1G01360	AT1G21600
AT1G37130	AT1G37130	AT4G32010	AT4G33925	AT1G51600	AT5G18580
AT1G52340	AT1G52340	AT2G41680	AT4G32010	AT1G51600	AT5G04820
AT2G23290	AT5G67300	AT3G03450	AT3G13740	AT1G51600	AT3G01990
AT2G26980	AT5G47100	AT3G48150	AT4G32010	AT1G51600	AT4G28880
AT4G08150	AT4G08150	AT1G56170	AT4G24470	AT1G51600	AT4G03420
AT4G26080	AT5G05440	AT4G24470	AT5G63470	AT1G51600	AT2G47450
AT4G28910	AT4G32570	AT4G28860	AT5G67300	AT1G01260	AT1G51600
AT5G10930	AT5G47100	AT1G29260	AT4G32010	AT3G09370	AT3G18210
AT5G35410	AT5G47100	AT1G01225	AT4G32010	AT1G51600	AT3G58380
AT5G45820	AT5G47100	AT3G17090	AT3G53250	AT1G21690	AT3G29090
AT5G47100	AT5G58380	AT1G64000	AT5G08480	AT1G51600	AT1G54380
AT1G17550	AT2G40330	AT1G64000	AT4G37710	AT1G69010	AT3G18960
AT1G17550	AT5G05440	AT1G76080	AT5G67300	AT1G01360	AT1G54830
AT1G17550	AT5G53160	AT1G29260	AT2G33150	AT2G25010	AT4G13260
AT4G37260	AT5G67300	AT3G48510	AT4G32010	AT4G11070	AT5G11010
AT4G37260	AT5G53160	AT1G60950	AT4G32010	AT3G15450	AT5G43830
AT3G23610	AT4G29810	AT4G32010	AT5G51940	AT4G14710	AT5G43830
AT4G24400	AT5G47100	AT1G80940	AT4G32010	AT3G29370	AT4G36540
AT4G37260	AT5G58950	AT3G18430	AT4G32010	AT4G14716	AT5G43830
AT1G01140	AT4G37260	AT1G21600	AT4G32010	AT2G29380	AT3G02140
AT2G01760	AT4G37260	AT4G32010	AT4G36030	AT1G04240	AT4G02150
AT3G21175	AT4G24470	AT3G56090	AT5G01600	AT1G04240	AT3G23030
AT1G51660	AT4G37260	AT3G03270	AT5G49620	AT1G04240	AT3G04730
AT2G42880	AT4G37260	AT2G29540	AT4G32010	AT3G17090	AT4G38850
AT4G32010	AT5G41920	AT4G32010	AT5G64150	AT4G38850	AT5G02760
AT1G18350	AT4G32010	AT1G68160	AT4G32010	AT3G21175	AT4G09060
AT3G03450	AT5G54190	AT1G80370	AT4G32010	AT1G51090	AT4G00120
AT4G01370	AT4G29810	AT1G05730	AT4G32010	AT1G02690	AT4G00120
AT5G01600	AT5G01600	AT3G51030	AT5G18930	AT3G21175	AT4G37470
AT5G67300	AT5G67300	AT3G22810	AT5G01810	AT2G24090	AT4G26110
AT1G72450	AT4G28910	AT1G15200	AT4G08150	AT1G77770	AT4G00120
AT2G44050	AT2G44050	AT2G44050	AT3G13672	AT2G32840	AT4G00120
AT2G24570	AT2G44050	AT1G51090	AT2G44050	AT2G38270	AT4G00120
AT1G68550	AT2G38490	AT4G08150	AT5G59730	AT4G00120	AT5G27690
AT1G07430	AT5G53160	AT4G08150	AT5G50680	AT3G21175	AT5G04820
AT4G28910	AT5G13220	AT4G01370	AT5G14600	AT4G00120	AT5G61230
AT1G21690	AT1G77470	AT2G44050	AT5G01820	AT1G04240	AT1G51950
AT1G31880	AT1G51660	AT2G44050	AT4G18650	AT4G00120	AT5G52010
AT1G31880	AT5G16080	AT2G25490	AT2G46900	AT1G04240	AT1G09500
AT1G15550	AT1G31880	AT1G50170	AT3G60630	AT1G09500	AT4G14560
AT1G31880	AT2G32960	AT1G72030	AT3G60630	AT2G18730	AT2G43790
AT1G31880	AT2G24570	AT4G08150	AT5G11980	AT1G12860	AT2G43790
AT1G64280	AT5G06960	AT2G33610	AT4G08150	AT3G21175	AT5G24660
AT2G25490	AT4G32570	AT3G56270	AT4G08150	AT1G30880	AT4G26110

Table A.5 continued from previous page

IntA	IntB	IntA	IntB	IntA	IntB
AT1G51660	AT3G45640	AT1G62520	AT4G08150	AT3G29770	AT4G26110
AT4G01370	AT4G26070	AT2G32840	AT4G08150	AT1G09700	AT5G41070
AT4G27450	AT5G43830	AT1G05860	AT4G08150	AT1G09700	AT3G58380
AT1G19180	AT4G28910	AT2G01620	AT4G08150	AT3G18140	AT4G14560
AT1G70700	AT4G28910	AT3G07090	AT4G38630	AT3G54130	AT5G05690
AT1G31880	AT2G46870	AT2G45820	AT4G02150	AT1G18710	AT3G54130
AT4G11070	AT5G22570	AT2G44050	AT5G62520	AT3G50330	AT4G00120
AT1G80840	AT5G22570	AT2G44050	AT3G50670	AT1G07210	AT4G00120
AT2G01760	AT4G32570	AT1G50420	AT4G22220	AT1G04100	AT1G04240
AT4G14720	AT4G28910	AT3G60630	AT4G28880	AT1G04240	AT5G52547
AT1G31880	AT5G08130	AT2G44050	AT3G05640	AT3G23030	AT4G14560
AT1G31880	AT2G42880	AT2G44050	AT3G60360	AT3G56910	AT4G26110
AT1G23860	AT3G61860	AT1G50420	AT3G52155	AT1G08580	AT4G26110
AT5G53160	AT5G59220	AT1G20780	AT5G03050	AT2G43060	AT4G26110
AT4G32570	AT5G41920	AT1G17650	AT1G50420	AT3G06590	AT4G26110
AT4G29810	AT5G58950	AT1G31240	AT1G50420	AT3G62870	AT4G26110
AT1G51660	AT2G43790	AT2G44050	AT5G22890	AT2G43790	AT5G07260
AT3G52930	AT5G03690	AT1G50420	AT5G52210	AT1G09700	AT1G50670
AT2G39760	AT3G24520	AT2G45820	AT5G23750	AT1G09700	AT3G22490
AT1G15550	AT5G25890	AT2G45820	AT3G57870	AT1G09700	AT4G19200
AT1G69010	AT1G69010	AT1G16705	AT2G45820	AT4G33950	AT5G39360
AT1G19180	AT1G30135	AT2G20610	AT5G37478	AT2G26560	AT4G35620
AT1G54990	AT2G38120	AT1G63090	AT2G25490	AT2G39760	AT4G22720
AT1G69010	AT5G08130	AT4G01370	AT5G07260	AT1G12120	AT2G39760
AT2G25090	AT5G47100	AT2G25490	AT2G29540	AT2G39760	AT4G14160
AT2G26560	AT5G41920	AT3G60630	AT5G25280	AT1G25490	AT3G62550
AT1G10210	AT3G63210	AT1G68160	AT3G60630	AT1G25490	AT5G17710
AT1G15550	AT1G24590	AT1G31880	AT5G16400	AT1G25490	AT3G02140
AT1G22070	AT1G64280	AT1G31880	AT1G66160	AT1G05410	AT1G25490
AT2G01760	AT2G22090	AT1G31880	AT3G02140	AT1G25490	AT2G35900
AT2G04890	AT5G07310	AT1G12120	AT1G31880	AT1G25490	AT1G51520
AT3G61830	AT3G61830	AT1G31880	AT4G26610	AT5G25890	AT5G59730
AT2G22090	AT2G22090	AT1G31880	AT1G63480	AT5G18580	AT5G25890
AT2G38490	AT3G25890	AT1G31880	AT4G30860	AT4G14160	AT5G19000
AT4G08150	AT4G32010	AT1G31880	AT4G02770	AT1G25490	AT3G14080
AT2G40330	AT3G11410	AT1G31880	AT5G23130	AT1G25490	AT3G56270
AT3G11410	AT5G05440	AT4G17720	AT5G03520	AT1G11430	AT1G25490
AT2G38310	AT3G11410	AT4G17720	AT5G47200	AT4G17615	AT5G10930
AT4G28910	AT5G20900	AT3G46060	AT4G17720	AT2G38270	AT5G25890
AT1G15200	AT4G37260	AT4G17720	AT4G18800	AT4G17615	AT5G35410
AT1G02690	AT1G04250	AT1G55190	AT4G17720	AT1G28360	AT1G77770
AT1G19010	AT4G37260	AT1G31880	AT5G22310	AT5G14070	AT5G65210
AT4G37260	AT5G16400	AT1G31880	AT3G11400	AT2G39760	AT5G28770
AT3G13672	AT4G37260	AT1G31880	AT5G03740	AT1G51950	AT5G25890
AT4G37260	AT5G59730	AT1G31880	AT5G50180	AT1G25490	AT1G53000
AT1G66160	AT4G37260	AT1G31880	AT1G49850	AT5G13180	AT5G19000
AT3G02140	AT4G37260	AT1G31880	AT5G65683	AT2G30540	AT5G65210
AT4G37260	AT5G39340	AT1G05710	AT1G31880	AT1G25490	AT5G57950
AT4G17615	AT4G24400	AT1G31880	AT3G15660	AT1G28360	AT5G24660
AT2G17670	AT4G37260	AT1G31880	AT5G59490	AT1G17380	AT3G02090
AT4G37260	AT5G42900	AT1G31880	AT5G41070	AT4G15670	AT5G65210

Table A.5 continued from previous page

IntA	IntB	IntA	IntB	IntA	IntB
AT2G44840	AT3G05000	AT1G31880	AT5G22890	AT1G28360	AT2G44740
AT1G04250	AT1G51950	AT1G31880	AT3G54170	AT3G44610	AT4G17490
AT1G04250	AT4G16143	AT1G01210	AT1G31880	AT3G02000	AT5G65210
AT1G04250	AT4G02150	AT1G31880	AT2G27840	AT1G76080	AT5G19000
AT4G16143	AT4G37260	AT1G31880	AT5G05190	AT1G04710	AT1G29260
AT4G37260	AT5G37055	AT1G31880	AT3G27580	AT3G04300	AT5G10930
AT4G37260	AT5G61010	AT1G31880	AT5G19340	AT4G17500	AT5G41650
AT4G31300	AT4G37260	AT1G31880	AT3G18295	AT1G52340	AT4G33460
AT4G37260	AT5G62520	AT1G31880	AT1G51100	AT1G80780	AT4G37580
AT4G37260	AT5G14070	AT1G03330	AT1G31880	AT1G66160	AT5G08130
AT4G02150	AT4G37260	AT1G31880	AT4G22745	AT3G57290	AT5G08130
AT4G37260	AT5G61230	AT1G06460	AT1G31880	AT1G22070	AT5G27560
AT4G25670	AT4G37260	AT1G31880	AT3G49810	AT4G11140	AT5G04820
AT4G37260	AT5G45680	AT1G03900	AT1G31880	AT1G24590	AT1G68590
AT3G52120	AT4G37260	AT1G31880	AT5G18110	AT1G24590	AT3G56270
AT4G18040	AT4G37260	AT1G31880	AT1G76080	AT1G24590	AT4G18630
AT3G50670	AT4G37260	AT1G31880	AT4G31050	AT1G24590	AT2G38300
AT2G44840	AT4G16143	AT1G31880	AT1G68130	AT3G59810	AT5G48870
AT1G04250	AT1G09500	AT1G31880	AT5G11460	AT3G13720	AT3G53710
AT3G05640	AT4G37260	AT1G31880	AT4G37710	AT5G17300	AT5G24520
AT4G37260	AT5G49210	AT1G31880	AT4G03250	AT2G41710	AT2G45640
AT2G33430	AT4G37260	AT1G09810	AT1G31880	AT1G24590	AT5G14070
AT3G60360	AT4G37260	AT1G31880	AT1G54200	AT1G22070	AT3G27560
AT3G19120	AT4G37260	AT1G31880	AT2G35010	AT1G24590	AT2G42750
AT2G42260	AT4G37260	AT1G31880	AT3G48510	AT1G24590	AT2G24860
AT4G37260	AT5G49690	AT1G23220	AT1G31880	AT1G24590	AT5G49210
AT3G18210	AT4G37260	AT1G10760	AT1G31880	AT1G24590	AT2G33430
AT2G30360	AT4G37260	AT1G14685	AT1G31880	AT5G08130	AT5G67220
AT4G37260	AT5G14170	AT1G31880	AT4G14020	AT1G01260	AT3G43440
AT3G06430	AT4G37260	AT1G31880	AT4G39540	AT2G41710	AT3G56900
AT3G01690	AT4G37260	AT1G31880	AT4G21660	AT1G24590	AT5G24660
AT1G51580	AT4G37260	AT1G31880	AT3G16560	AT4G00180	AT5G08130
AT2G32650	AT4G37260	AT1G31880	AT1G76070	AT1G17980	AT5G08130
AT4G11070	AT4G37260	AT1G31880	AT5G58960	AT1G22070	AT5G06960
AT3G18140	AT4G37260	AT1G20140	AT2G04550	AT1G22070	AT2G30540
AT4G18372	AT4G37260	AT4G17615	AT5G58380	AT1G22070	AT3G18210
AT4G23910	AT4G37260	AT3G50670	AT3G61860	AT1G06180	AT2G04630
AT3G55150	AT4G37260	AT3G12720	AT5G57860	AT3G23060	AT5G08130
AT4G09650	AT4G37260	AT1G74910	AT2G39770	AT1G22070	AT4G12100
AT3G21150	AT4G37260	AT3G23240	AT4G17680	AT5G23710	AT5G57685
AT4G37260	AT5G22890	AT1G15570	AT3G12720	AT1G54380	AT5G48870
AT1G13690	AT4G39410	AT3G61860	AT4G32660	AT5G08130	AT5G18260
AT4G15770	AT4G25470	AT4G31800	AT5G42980	AT2G22880	AT3G01970
AT1G17550	AT4G01026	AT2G43370	AT3G61860	AT2G30120	AT5G08130
AT1G07210	AT4G37260	AT1G53910	AT2G16030	AT1G22070	AT4G15670
AT3G12830	AT4G37260	AT1G53910	AT4G21450	AT1G06460	AT4G11140
AT4G15770	AT4G37260	AT1G64280	AT5G64150	AT1G22070	AT1G28480
AT3G17668	AT4G37260	AT1G53910	AT2G25880	AT1G11810	AT2G33860
AT4G37260	AT5G48335	AT1G07350	AT3G61860	AT1G67340	AT2G32460
AT4G37260	AT5G62770	AT4G09060	AT4G32570	AT4G02485	AT5G04190
AT1G51100	AT4G37260	AT5G04430	AT5G46190	AT5G05360	AT5G41920

Table A.5 continued from previous page

IntA	IntB	IntA	IntB	IntA	IntB
AT3G50910	AT5G44420	AT2G25000	AT5G52650	AT1G22070	AT3G02000
AT1G06460	AT4G37260	AT4G32570	AT5G38410	AT1G10760	AT1G71260
AT4G37260	AT5G05790	AT4G32570	AT5G64960	AT1G76680	AT3G63260
AT3G49810	AT4G37260	AT4G32570	AT5G59730	AT5G24520	AT5G43650
AT3G57720	AT4G37260	AT3G57290	AT4G32570	AT1G06180	AT5G51940
AT2G39100	AT4G37260	AT4G22720	AT4G32570	AT3G28710	AT5G14750
AT3G13200	AT4G37260	AT4G32570	AT4G37470	AT2G03710	AT5G41920
AT4G02770	AT4G37260	AT1G05410	AT4G32570	AT1G22070	AT5G11010
AT1G67090	AT4G37260	AT3G54850	AT4G32570	AT2G32600	AT4G11140
AT1G02140	AT4G37260	AT3G06790	AT4G16420	AT2G33860	AT5G41650
AT3G11850	AT3G30210	AT2G32840	AT4G16420	AT3G11410	AT5G16080
AT2G43370	AT4G37260	AT3G05420	AT4G28880	AT2G30980	AT5G65300
AT1G77710	AT4G37260	AT4G16420	AT5G49210	AT1G02690	AT1G04550
AT4G20300	AT4G37260	AT3G21215	AT5G04430	AT1G04550	AT5G18580
AT4G37260	AT5G20920	AT1G72360	AT2G38490	AT2G35900	AT5G59220
AT4G28860	AT4G37260	AT1G68450	AT5G41570	AT3G07090	AT3G11410
AT4G18700	AT4G37260	AT2G32960	AT2G32960	AT3G11410	AT3G56270
AT1G56170	AT2G46790	AT4G21560	AT4G32570	AT3G11410	AT4G37240
AT2G46790	AT5G63470	AT1G73060	AT4G32570	AT1G04550	AT2G36145
AT2G25250	AT4G37260	AT3G59910	AT4G32570	AT5G07690	AT5G62520
AT1G09810	AT4G37260	AT1G18660	AT4G32570	AT3G11410	AT4G25670
AT1G54200	AT4G37260	AT3G06430	AT4G32570	AT1G04550	AT1G51950
AT3G02460	AT4G37260	AT1G51580	AT4G32570	AT1G04550	AT5G62520
AT2G37630	AT4G37260	AT2G04630	AT4G32570	AT2G33430	AT3G11410
AT1G76890	AT4G37260	AT4G12100	AT4G32570	AT1G35210	AT2G30980
AT3G48510	AT4G37260	AT4G17680	AT4G32570	AT2G22090	AT5G06770
AT4G18830	AT4G37260	AT1G01630	AT4G32570	AT1G48450	AT2G22090
AT4G37260	AT5G25510	AT4G32570	AT5G10710	AT3G19820	AT3G54130
AT1G10650	AT4G37260	AT3G05420	AT5G27630	AT1G03457	AT2G22090
AT1G21600	AT4G37260	AT3G18140	AT4G32570	AT1G04100	AT3G61830
AT3G57730	AT4G37260	AT4G32570	AT5G43820	AT1G04550	AT3G17600
AT2G32180	AT4G37260	AT3G55150	AT4G32570	AT1G04550	AT3G23030
AT2G27020	AT4G37260	AT4G09650	AT4G32570	AT1G04100	AT1G04550
AT1G10650	AT2G44840	AT3G49060	AT4G32570	AT2G22090	AT2G38610
AT1G03130	AT4G37260	AT3G48680	AT4G32570	AT1G54080	AT2G22090
AT2G25880	AT4G37260	AT2G22880	AT5G41570	AT2G22090	AT3G19130
AT1G72670	AT4G37260	AT4G15770	AT4G32570	AT1G27650	AT2G16940
AT4G37260	AT5G25280	AT1G77950	AT4G32570	AT3G22490	AT5G59220
AT4G37260	AT5G27720	AT3G16980	AT4G32570	AT2G17190	AT2G35635
AT4G09060	AT4G32010	AT4G32570	AT5G50230	AT3G61830	AT5G13930
AT3G62550	AT4G32010	AT4G32570	AT5G48470	AT5G17490	AT5G63670
AT4G32010	AT5G59730	AT2G44740	AT4G32570	AT1G03860	AT5G17490
AT1G79650	AT4G32010	AT4G32570	AT5G05790	AT3G07670	AT5G17490
AT3G02140	AT4G32010	AT1G11810	AT4G32570	AT2G37560	AT5G17490
AT4G22720	AT4G32010	AT1G28480	AT4G32570	AT1G02690	AT5G05410
AT4G32010	AT5G39340	AT3G57720	AT4G32570	AT1G76420	AT4G25740
AT4G05450	AT4G32010	AT1G67090	AT4G32570	AT1G76420	AT5G52650
AT2G25490	AT4G32010	AT4G32570	AT5G18110	AT3G02140	AT4G14720
AT2G35940	AT4G32010	AT3G48150	AT4G32570	AT3G06790	AT5G17490
AT4G00780	AT4G24470	AT1G56590	AT4G32570	AT4G18650	AT5G17490
AT4G24470	AT5G18580	AT3G05420	AT4G28860	AT1G11430	AT5G17490

Table A.5 continued from previous page

IntA	IntB	IntA	IntB	IntA	IntB
AT3G03450	AT5G59450	AT2G46130	AT5G08480	AT5G15802	AT5G17490
AT2G31955	AT3G03450	AT2G46130	AT4G37710	AT4G14720	AT4G32295
AT3G03450	AT5G23130	AT5G08480	AT5G41570	AT3G46630	AT5G17490
AT1G02690	AT5G01600	AT3G02000	AT4G32570	AT5G14070	AT5G17490
AT3G18240	AT5G18580	AT1G29260	AT4G32570	AT4G14720	AT4G16143
AT3G23610	AT5G04820	AT1G01225	AT4G32570	AT3G29350	AT4G14720
AT3G14080	AT3G53250	AT1G80720	AT4G32570	AT1G56170	AT4G14713
AT3G03450	AT5G25580	AT3G07300	AT4G32570	AT4G14713	AT5G63470
AT1G68590	AT5G01600	AT4G02485	AT4G32570	AT1G56170	AT4G14720
AT5G12190	AT5G67300	AT1G16705	AT4G32570	AT4G14720	AT5G63470
AT3G06790	AT5G67300	AT3G47910	AT4G32570	AT3G17090	AT5G18020
AT1G19220	AT1G51950	AT3G02460	AT4G32570	AT5G02760	AT5G18020
AT4G16143	AT4G24470	AT2G34590	AT4G32570	AT1G66370	AT5G18580
AT3G03450	AT5G28770	AT4G32570	AT5G47830	AT1G66370	AT2G01620
AT4G16143	AT5G01600	AT4G32570	AT5G08290	AT1G01560	AT1G08780
AT4G02150	AT5G01600	AT3G12210	AT4G16420	AT1G23860	AT5G54580
AT2G04550	AT3G58680	AT5G41570	AT5G67520	AT1G23860	AT4G32660
AT5G18930	AT5G39950	AT3G50800	AT5G40440	AT1G66370	AT3G58650
AT1G61150	AT3G23610	AT4G32570	AT5G64180	AT1G31880	AT4G18830
AT5G57950	AT5G60890	AT1G56450	AT4G32570	AT1G31880	AT5G42050
AT3G03450	AT5G65683	AT4G32570	AT5G07260	AT1G31880	AT1G56450
AT4G00180	AT5G67300	AT1G10650	AT4G32570	AT1G10650	AT1G31880
AT1G05510	AT4G18010	AT1G21600	AT4G32570	AT3G49950	AT4G32570
AT3G48680	AT4G24470	AT3G28710	AT4G32570	AT3G24730	AT4G32570
AT4G20130	AT4G24470	AT3G51130	AT4G32570	AT1G12840	AT4G32570
AT3G03450	AT3G48680	AT2G23760	AT4G32570	AT2G05170	AT4G32570
AT1G04100	AT1G19220	AT4G18060	AT4G32570	AT2G39760	AT5G51100
AT1G19220	AT3G04730	AT1G50710	AT4G32570	AT2G39760	AT5G25340
AT4G32010	AT5G50230	AT2G28000	AT4G32570	AT1G24590	AT5G48335
AT4G32010	AT5G42970	AT4G32570	AT5G22480	AT1G24590	AT5G05790
AT4G32010	AT5G55620	AT2G21630	AT4G32570	AT1G23860	AT3G50670
AT4G32010	AT5G48470	AT3G15650	AT4G32570		

A.6 Network representations

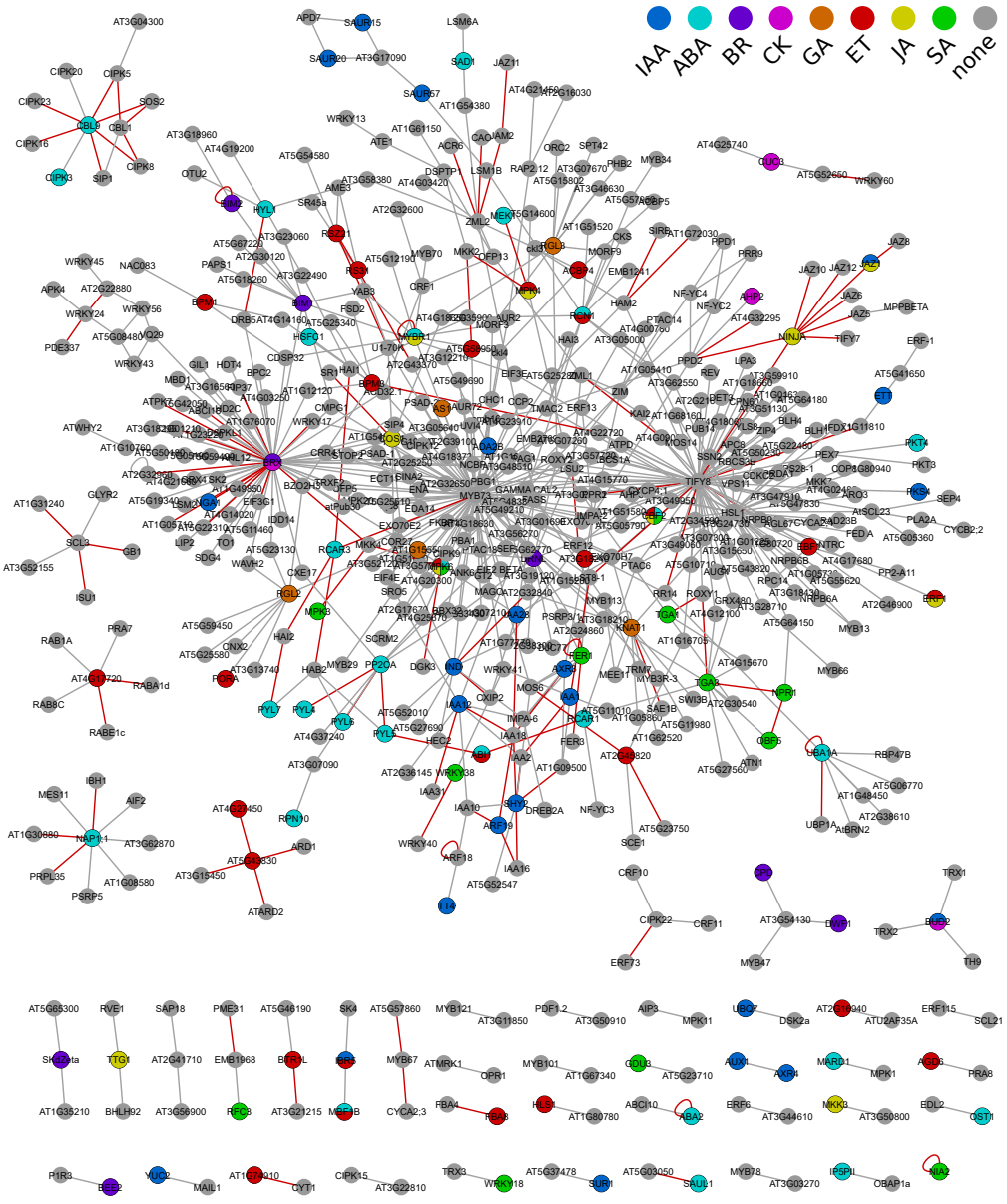


Figure A.1: Network representation of PhI_{out} screen. Proteins are annotated with phytohormone annotations from AHD 2.0 restricted to annotations derived from genetic evidence. Red edges show recovered interactions curated in IntAct or BioGRID database.

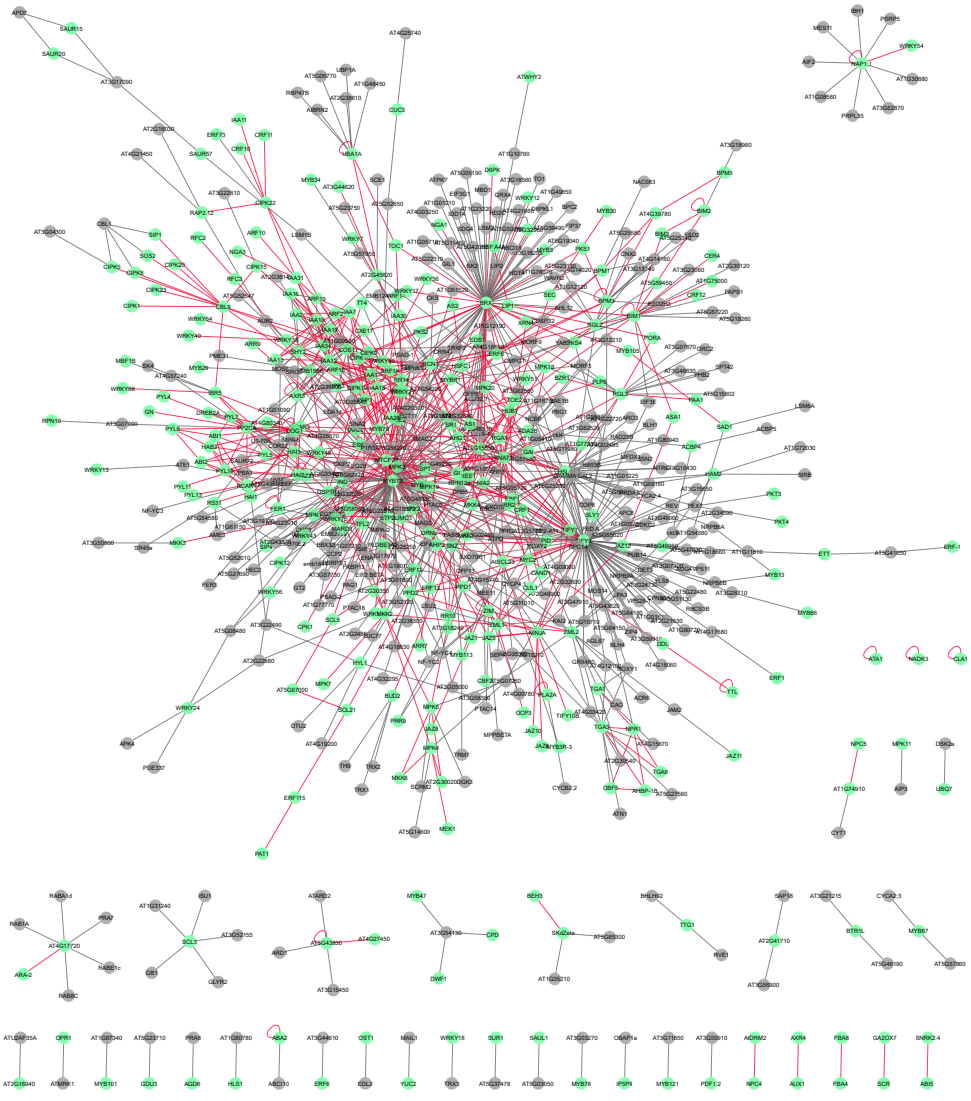


Figure A.2: $\mathrm{PhI}_{\mathrm{ext}}$ combined, proteins in search space $\mathrm{SSP}_{\mathrm{PHO}}$ in green, interactions from PhI in red

A.7 IntAct binary methods

Table A.6: Methods contained in IntAct protein-protein interactions filtered for *Arabidopsis thaliana* and interaction partners in search space SSP_{PHO}. Classification for binary protein-protein interaction detection method in column "Binary".

Experimental System	Interaction type	Binary
psi-mi:"MI:0096"(pull down)	psi-mi:"MI:0915"(physical association)	No
psi-mi:"MI:0018"(two hybrid)	psi-mi:"MI:0915"(physical association)	Yes
psi-mi:"MI:0007"(anti tag coimmunoprecipitation)	psi-mi:"MI:0915"(physical association)	No
psi-mi:"MI:0416"(fluorescence microscopy)	psi-mi:"MI:0403"(colocalization)	No
psi-mi:"MI:0045"(experimental interaction detection)	psi-mi:"MI:0217"(phosphorylation reaction)	Yes
psi-mi:"MI:0006"(anti bait coip)	psi-mi:"MI:0915"(physical association)	No
psi-mi:"MI:0071"(molecular sieving)	psi-mi:"MI:0915"(physical association)	Yes
psi-mi:"MI:0047"(far western blotting)	psi-mi:"MI:0407"(direct interaction)	Yes
psi-mi:"MI:0809"(bimolecular fluorescence complementation)	psi-mi:"MI:0915"(physical association)	No
psi-mi:"MI:0423"(in-gel kinase assay)	psi-mi:"MI:0217"(phosphorylation reaction)	Yes
psi-mi:"MI:0424"(protein kinase assay)	psi-mi:"MI:0217"(phosphorylation reaction)	Yes
psi-mi:"MI:0030"(cross-linking study)	psi-mi:"MI:0915"(physical association)	No
psi-mi:"MI:0112"(ubiquitin reconstruction)	psi-mi:"MI:0915"(physical association)	Yes
psi-mi:"MI:0055"(fluorescent resonance energy transfer)	psi-mi:"MI:0407"(direct interaction)	Yes
psi-mi:"MI:0663"(confocal microscopy)	psi-mi:"MI:0403"(colocalization)	No
psi-mi:"MI:0676"(tandem affinity purification)	psi-mi:"MI:0915"(physical association)	No
psi-mi:"MI:0807"(comigration in gel electrophoresis)	psi-mi:"MI:0915"(physical association)	No
psi-mi:"MI:0841"(phosphotransferase assay)	psi-mi:"MI:0844"(phosphotransfer reaction)	No
psi-mi:"MI:0004"(affinity chromatography technology)	psi-mi:"MI:0407"(direct interaction)	No
psi-mi:"MI:0276"(blue native page)	psi-mi:"MI:0915"(physical association)	Yes
psi-mi:"MI:0434"(phosphatase assay)	psi-mi:"MI:0203"(dephosphorylation reaction)	No
psi-mi:"MI:0090"(protein complementation assay)	psi-mi:"MI:0915"(physical association)	No
psi-mi:"MI:0065"(isothermal titration calorimetry)	psi-mi:"MI:0407"(direct interaction)	Yes
psi-mi:"MI:0013"(biophysical)	psi-mi:"MI:0407"(direct interaction)	No
psi-mi:"MI:0017"(classical fluorescence spectroscopy)	psi-mi:"MI:0407"(direct interaction)	No
psi-mi:"MI:0397"(two hybrid array)	psi-mi:"MI:0915"(physical association)	Yes
psi-mi:"MI:0114"(x-ray crystallography)	psi-mi:"MI:0407"(direct interaction)	Yes
psi-mi:"MI:0077"(nuclear magnetic resonance)	psi-mi:"MI:0407"(direct interaction)	Yes
psi-mi:"MI:1204"(split firefly luciferase complementation)	psi-mi:"MI:0915"(physical association)	No
psi-mi:"MI:1037"(Split renilla luciferase complementation)	psi-mi:"MI:0915"(physical association)	No

A.8 BioGRID binary methods

Table A.7: Methods contained in BioGRID protein-protein interactions filtered for *Arabidopsis thaliana* and interaction partners in search space SSP_{PHO}. Classification for binary protein-protein interaction detection method in column "Binary".

Experimental System	Binary
Two-hybrid	Yes
Affinity Capture-Western	No
Reconstituted Complex	Yes
Phenotypic Enhancement	No
Phenotypic Suppression	No
Dosage Growth Defect	No
Synthetic Rescue	No
Synthetic Growth Defect	No
FRET	Yes
Biochemical Activity	Yes
PCA	No
Co-crystal Structure	Yes
Affinity Capture-MS	No
Co-purification	No
Co-fractionation	Yes
Far Western	Yes
Co-localization	No

A.9 PhI communities

Table A.8: Community number - AGI assignment.

Locus	Community	Locus	Community	Locus	Community
AT1G01030	1	AT4G02570	6	AT1G28360	16
AT1G53910	1	AT2G34650	6	AT3G15150	16
AT2G28350	1	AT1G17380	7	AT5G45810	16
AT2G38490	1	AT1G19050	7	AT4G00240	16
AT3G17600	1	AT1G32640	7	AT3G57040	16
AT1G68550	1	AT4G28910	7	AT2G46130	16
AT4G28640	1	AT3G29350	7	AT4G29800	16
AT4G30080	1	AT1G74950	7	AT5G13080	16
AT3G25890	1	AT1G72450	7	AT5G43290	16
AT1G01140	2	AT1G19180	7	AT2G39250	16
AT2G01570	2	AT1G70700	7	AT4G36540	16
AT2G23290	2	AT5G58220	7	AT1G80340	16
AT4G37260	2	AT3G20550	7	AT1G75080	17
AT2G25090	2	AT5G05730	7	AT4G33520	17
AT2G42880	2	AT4G16110	7	AT5G17490	17
AT3G14720	2	AT5G13220	7	AT5G54190	17
AT5G67300	2	AT4G31920	7	AT2G01760	18
AT5G58950	2	AT1G30135	7	AT2G30590	18
AT1G01360	3	AT5G20900	7	AT3G62100	18
AT1G07430	3	AT5G01820	7	AT2G44050	18
AT2G29380	3	AT1G25470	8	AT2G25000	18
AT2G40330	3	AT1G69010	8	AT1G23860	18
AT4G26080	3	AT1G75000	8	AT2G46990	18
AT5G45830	3	AT5G08130	8	AT4G24240	18
AT1G17550	3	AT4G33790	8	AT3G61860	18
AT5G53160	3	AT5G38860	8	AT5G13930	18
AT5G05440	3	AT1G30270	9	AT2G02560	19
AT4G27920	3	AT2G26980	9	AT4G24470	19
AT1G13980	3	AT5G10930	9	AT1G51600	19
AT3G11410	3	AT5G35410	9	AT3G21175	19
AT5G05410	3	AT5G45820	9	AT3G01970	19
AT2G38310	3	AT5G47100	9	AT5G11270	19
AT4G01026	3	AT2G04550	9	AT3G04240	20
AT4G18620	3	AT4G24400	9	AT3G28910	20
AT5G45860	3	AT5G01810	9	AT4G14713	21
AT5G57050	3	AT3G17510	9	AT5G04870	21
AT5G59220	3	AT5G58380	9	AT5G17690	21
AT1G04100	4	AT3G62340	9	AT4G15560	22
AT1G04240	4	AT5G25190	9	AT1G78590	23
AT1G04250	4	AT1G50600	10	AT1G21690	24
AT1G04550	4	AT4G14720	10	AT1G63160	24
AT1G15050	4	AT2G04890	10	AT1G77470	24
AT2G33310	4	AT5G07310	10	AT1G10210	25
AT3G04730	4	AT2G20350	10	AT1G07340	25
AT3G15540	4	AT5G67000	10	AT3G45640	25
AT3G16500	4	AT5G48150	10	AT5G40440	25
AT3G23030	4	AT1G52340	11	AT3G63210	25
AT4G14560	4	AT1G53170	12	AT2G30020	26

Table A.8 continued from previous page

Locus	Community	Locus	Community	Locus	Community
AT1G19220	4	AT4G39780	12	AT4G01370	26
AT3G61830	4	AT2G39760	12	AT4G29810	26
AT3G23050	4	AT5G19000	12	AT2G43790	26
AT1G35540	4	AT5G21010	12	AT2G18170	26
AT5G25890	4	AT1G53510	13	AT4G26070	26
AT1G10940	5	AT4G36930	13	AT5G56580	26
AT2G36270	5	AT1G14920	13	AT1G50960	27
AT1G15550	6	AT4G00120	13	AT3G54220	27
AT1G18400	6	AT5G55170	13	AT1G66560	28
AT1G22770	6	AT4G24210	13	AT5G22570	28
AT1G37130	6	AT1G54490	14	AT1G80840	28
AT1G51660	6	AT1G14280	14	AT5G62000	28
AT1G51950	6	AT1G31880	14	AT1G74910	29
AT1G65620	6	AT2G24570	14	AT3G03540	29
AT1G73410	6	AT1G22640	14	AT3G03450	30
AT2G25490	6	AT1G69810	14	AT4G18880	30
AT2G39220	6	AT2G02950	14	AT5G64813	30
AT2G44950	6	AT1G69560	14	AT5G01600	31
AT4G08150	6	AT2G22090	14	AT2G45820	32
AT4G11070	6	AT5G61380	14	AT3G44620	32
AT4G16420	6	AT2G44745	14	AT4G27450	33
AT5G60120	6	AT5G55910	14	AT5G43830	33
AT4G32010	6	AT5G16080	14	AT1G06400	34
AT1G77200	6	AT2G32960	14	AT4G17720	34
AT1G18350	6	AT2G46870	14	AT2G40750	35
AT4G32570	6	AT5G04190	14	AT4G26110	35
AT1G25490	6	AT5G48870	14	AT3G52930	36
AT2G26560	6	AT1G64280	15	AT5G03690	36
AT1G24590	6	AT5G06950	15	AT3G42960	37
AT2G37630	6	AT1G22070	15	AT1G54990	38
AT5G41920	6	AT5G06960	15	AT2G38120	38
AT5G18930	6	AT3G12250	15	AT2G30980	39
AT3G24520	6	AT5G65210	15	AT4G18890	39
AT3G48090	6	AT1G64520	16	AT2G33830	40
AT5G51760	6	AT3G23610	16	AT3G03530	40
AT5G64810	6	AT1G35560	16		

A.10 Community algorithm comparison results

A.10.1 PhI

Table A.9: Similarity of communities in PhI network, identified by different algorithms. eb: edge betweenness

	vi	nmi	split.join	rand	adjusted.rand
eb-Walktrap	0.962254	0.851224	88	0.946805	0.522346
eb-Infomap	0.838272	0.874249	85	0.970614	0.647311
eb-Fast Greedy	1.184477	0.803939	117	0.936287	0.489007
eb-Label Prop	1.261409	0.804855	125	0.945976	0.466887
eb-Louvain	1.313307	0.781107	133	0.933227	0.445494

A.10.2 IntAct

Table A.10: Similarity of communities in IntAct network, identified by different algorithms. eb: edge betweenness

	vi	nmi	split.join	rand	adjusted.rand
eb-Walktrap	0.371117	0.943235	45	0.983284	0.834574
eb-Infomap	0.358938	0.947611	53	0.987374	0.852977
eb-Fast Greedy	0.318589	0.950003	49	0.981332	0.824515
eb-Label Prop	0.703635	0.897817	96	0.975742	0.743509
eb-Louvain	0.307253	0.951888	51	0.984053	0.847503

A.10.3 BioGRID

Table A.11: Similarity of communities in BioGRID network, identified by different algorithms. eb: edge betweenness

	vi	nmi	split.join	rand	adjusted.rand
eb-Walktrap	0.414431	0.931881	58	0.978804	0.847405
eb-Infomap	0.741482	0.887839	114	0.963989	0.676777
eb-Fast Greedy	0.162772	0.972366	26	0.990848	0.935915
eb-Label Prop	0.683993	0.893981	93	0.973034	0.783851
eb-Louvain	0.279807	0.951918	39	0.982216	0.879586

A.11 Enriched communities in BioGRID

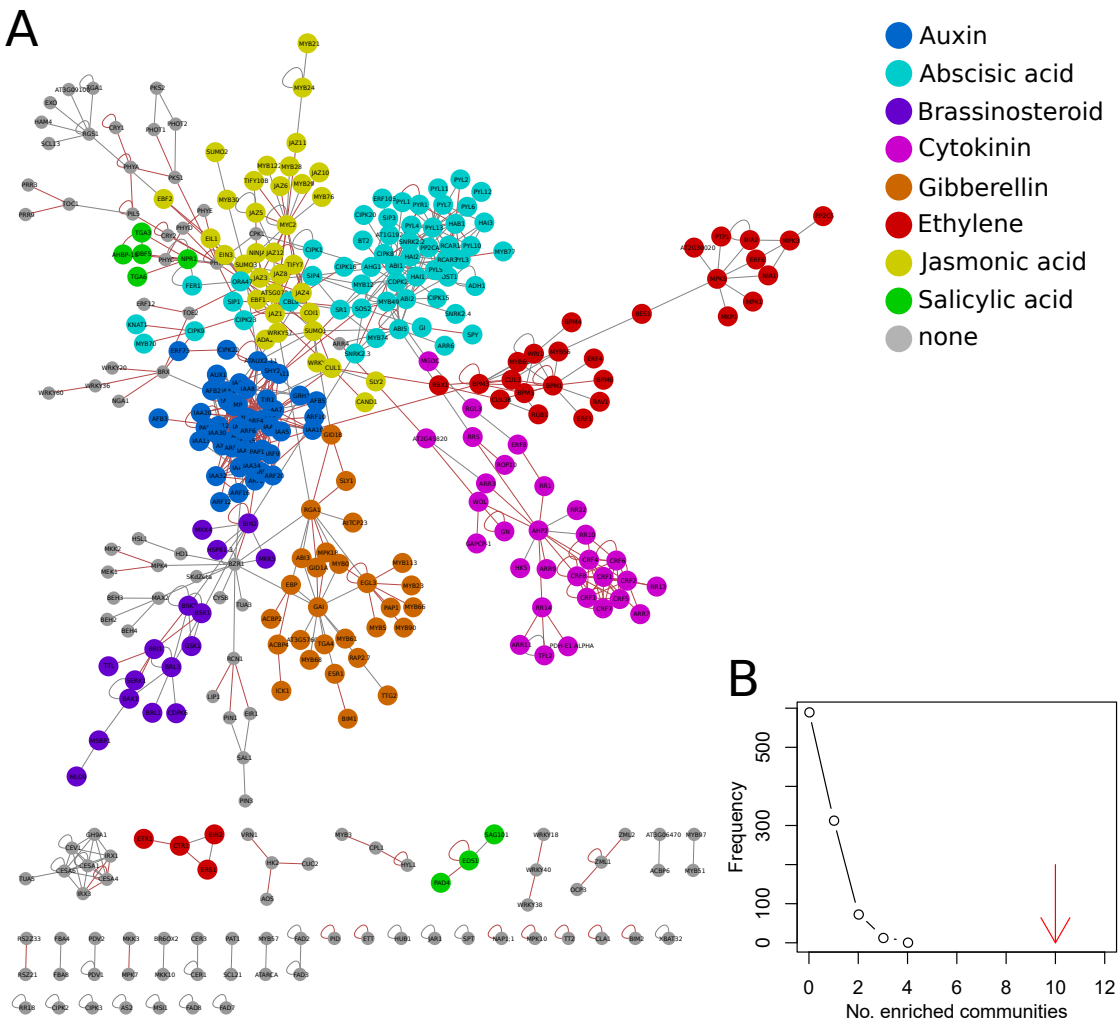


Figure A.3: A) BioGRID network restricted to search space SSP_{PHO} and binary interactions colored by hormone enriched communities. B) Number of enriched communities in BioGRID using AHD genetic annotations compared against 1000 randomized networks.

A.12 Enriched communities in AI-1_{MAIN}

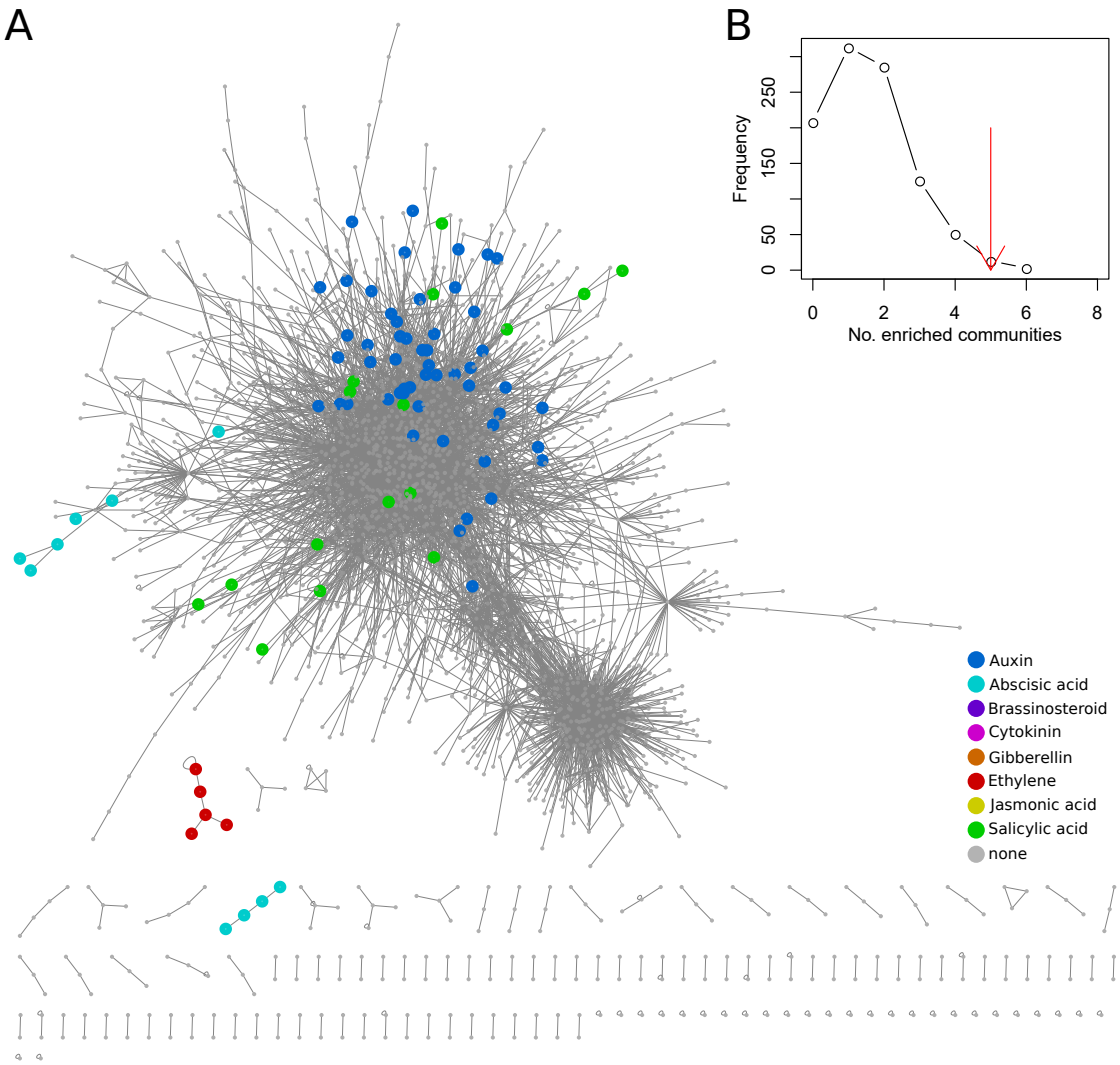


Figure A.4: A) AI-1_{MAIN} network colored by hormone enriched communities. B) Number of enriched communities in AI-1_{MAIN} using AHD genetic annotations compared against 1000 randomized networks.

A.13 GO enrichment in communities

Table A.12: Overrepresented GO terms in Community 1

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0048829	0.003080278	74.33333333	0.103448276	2	3	root cap development
GO:0048442	0.034482759	Inf	0.034482759	1	1	sepal development
GO:0051301	0.049046007	7.962962963	0.379310345	2	11	cell division

Table A.13: Overrepresented GO terms in Community 2

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0032879	0.026327747	12.11111111	0.275862069	2	8	regulation of localization
GO:0006812	0.033257798	10.33333333	0.310344828	2	9	cation transport
GO:0055075	0.034482759	Inf	0.034482759	1	1	potassium ion homeostasis
GO:0051365	0.034482759	Inf	0.034482759	1	1	cellular response to potassium ion starvation
GO:0043266	0.034482759	Inf	0.034482759	1	1	regulation of potassium ion transport
GO:0006814	0.034482759	Inf	0.034482759	1	1	sodium ion transport
GO:0042176	0.034482759	Inf	0.034482759	1	1	regulation of protein catabolic process

Table A.14: Overrepresented GO terms in Community 3

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:1905957	1.56811E-06	58.09090909	0.586206897	6	8	regulation of cellular response to alcohol
GO:1901420	6.61787E-06	89.16666667	0.439655172	5	6	negative regulation of response to alcohol
GO:0009788	6.61787E-06	89.16666667	0.439655172	5	6	negative regulation of abscisic acid-activated signaling pathway
GO:0071396	0.000130801	7.878787879	3.150862069	10	43	cellular response to lipid
GO:0009737	0.000553126	6.25	3.663793103	10	50	response to abscisic acid
GO:0010648	0.000665085	12.38095238	0.879310345	5	12	negative regulation of cell communication
GO:0023057	0.000665085	12.38095238	0.879310345	5	12	negative regulation of signaling
GO:0009966	0.001723073	7.272727273	1.538793103	6	21	regulation of signal transduction
GO:0010119	0.003013416	22.82142857	0.36637931	3	5	regulation of stomatal movement
GO:0009738	0.004246734	7.330827068	1.274336283	5	24	abscisic acid-activated signaling pathway
GO:0071229	0.004966124	4.259259259	4.689655172	10	64	cellular response to acid chemical
GO:0050790	0.005750373	15.14285714	0.439655172	3	6	regulation of catalytic activity
GO:0009414	0.008996379	5.555555556	1.465517241	5	20	response to water deprivation
GO:0009408	0.014657579	9	0.586206897	3	8	response to heat
GO:0097306	0.014702849	6.19047619	1.080357143	4	22	cellular response to alcohol
GO:0010029	0.01646158	5.706293706	1.099137931	4	15	regulation of seed germination
GO:0042538	0.026101933	4.781065089	1.245689655	4	17	hyperosmotic salinity response
GO:0009644	0.027864988	14.2	0.293103448	2	4	response to high light intensity
GO:0032870	0.029386612	2.959183673	5.862068966	10	80	cellular response to hormone stimulus
GO:0050793	0.035638939	3.237536657	2.711206897	6	37	regulation of developmental process
GO:2000033	0.044432702	9.422222222	0.36637931	2	5	regulation of seed dormancy process
GO:0048580	0.045112494	3.315972222	2.125	5	29	regulation of post-embryonic development

Table A.17: Overrepresented GO terms in Community 7

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0009695	9.85468E-09	84.8	0.775862069	8	10	jasmonic acid biosynthetic process
GO:0009611	1.76168E-08	41.8	1.086206897	9	14	response to wounding
GO:0007165	7.3506E-07	Inf	3.938461538	12	64	signal transduction
GO:0032870	1.08029E-06	18.75	6.206896552	16	80	cellular response to hormone stimulus
GO:0009867	5.88161E-06	11.36263736	2.870689655	11	37	jasmonic acid mediated signaling pathway

Table A.17 continued from previous page

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0009753	1.44035E-05	10.02463054	3.103448276	11	40	response to jasmonic acid
GO:0043207	4.17848E-05	8.619047619	3.413793103	11	44	response to external biotic stimulus
GO:0009812	6.37947E-05	17.33333333	0.931034483	6	12	flavonoid metabolic process
GO:0009620	6.76559E-05	9.19047619	2.327586207	9	30	response to fungus
GO:0042445	6.76559E-05	9.19047619	2.327586207	9	30	hormone metabolic process
GO:0051704	8.56743E-05	7.76984127	3.646551724	11	47	multi-organism process
GO:0010033	0.000225089	15.77477477	9.931034483	17	128	response to organic substance
GO:0070887	0.000257482	51.63636364	0.474683544	4	15	cellular response to chemical stimulus
GO:0032787	0.000272775	7.230769231	2.715517241	9	35	monocarboxylic acid metabolic process
GO:0046394	0.000272775	7.230769231	2.715517241	9	35	carboxylic acid biosynthetic process
GO:0009719	0.000486296	8.95049505	9.077586207	16	117	response to endogenous stimulus
GO:0044283	0.000556739	6.379310345	2.948275862	9	38	small molecule biosynthetic process
GO:0031537	0.000782731	20.0952381	0.543103448	4	7	regulation of anthocyanin metabolic process
GO:0065008	0.001240749	5.274390244	3.956896552	10	51	regulation of biological quality
GO:0040008	0.001488232	15	0.620689655	4	8	regulation of growth
GO:0009736	0.001488232	15	0.620689655	4	8	cytokinin-activated signaling pathway
GO:0071229	0.00199394	4.773584906	4.965517241	11	64	cellular response to acid chemical
GO:0048831	0.00404963	7.097902098	1.24137931	5	16	regulation of shoot system development
GO:1901701	0.004983704	5.789189189	2.613861386	7	44	cellular response to oxygen-containing compound
GO:0042538	0.005456456	6.474358974	1.318965517	5	17	hyperosmotic salinity response
GO:1900674	0.005709807	Inf	0.155172414	2	2	olefin biosynthetic process
GO:0009693	0.005709807	Inf	0.155172414	2	2	ethylene biosynthetic process
GO:0080113	0.005709807	Inf	0.155172414	2	2	regulation of seed growth
GO:0043449	0.005709807	Inf	0.155172414	2	2	cellular alkene metabolic process
GO:0009308	0.005709807	Inf	0.155172414	2	2	amine metabolic process
GO:0006568	0.005709807	Inf	0.155172414	2	2	tryptophan metabolic process
GO:0006576	0.005709807	Inf	0.155172414	2	2	cellular biogenic amine metabolic process
GO:0042440	0.00602549	8.448979592	0.853448276	4	11	pigment metabolic process
GO:0043436	0.006303	4.744444444	2.585152838	7	37	oxoacid metabolic process
GO:0031347	0.006533199	5.131578947	1.939655172	6	25	regulation of defense response
GO:0009735	0.00859006	7.357142857	0.931034483	4	12	response to cytokinin
GO:0045087	0.009081816	4.227272727	2.715517241	7	35	innate immune response
GO:0009414	0.011754771	5.102564103	1.551724138	5	20	response to water deprivation
GO:0002376	0.012632325	3.903030303	2.870689655	7	37	immune system process
GO:0050789	0.012981402	Inf	10.375	14	166	regulation of biological process
GO:0009963	0.016335014	26.625	0.232758621	2	3	positive regulation of flavonoid biosynthetic process
GO:0016144	0.016335014	26.625	0.232758621	2	3	S-glycoside biosynthetic process
GO:1901657	0.016335014	26.625	0.232758621	2	3	glycosyl compound metabolic process
GO:0019761	0.016335014	26.625	0.232758621	2	3	glucosinolate biosynthetic process
GO:0019757	0.016335014	26.625	0.232758621	2	3	glycosinolate metabolic process
GO:0010082	0.016335014	26.625	0.232758621	2	3	regulation of root meristem growth
GO:0097306	0.017108942	3.958333333	2.327586207	6	30	cellular response to alcohol
GO:0009723	0.017108942	3.958333333	2.327586207	6	30	response to ethylene
GO:0009738	0.017108942	3.958333333	2.327586207	6	30	abscisic acid-activated signaling pathway
GO:0006612	0.018044265	4.457013575	1.706896552	5	22	protein targeting to membrane
GO:0010363	0.018044265	4.457013575	1.706896552	5	22	regulation of plant-type hypersensitive response
GO:0072657	0.018044265	4.457013575	1.706896552	5	22	protein localization to membrane
GO:0010941	0.018044265	4.457013575	1.706896552	5	22	regulation of cell death
GO:0034050	0.018044265	4.457013575	1.706896552	5	22	host programmed cell death induced by symbiont
GO:0090351	0.018044265	4.457013575	1.706896552	5	22	seedling development
GO:0010029	0.02032862	5.272727273	1.163793103	4	15	regulation of seed germination
GO:0045184	0.021915735	4.188034188	1.784482759	5	23	establishment of protein localization
GO:0050776	0.021915735	4.188034188	1.784482759	5	23	regulation of immune response
GO:0006886	0.021915735	4.188034188	1.784482759	5	23	intracellular protein transport
GO:0070727	0.026310746	3.947368421	1.862068966	5	24	cellular macromolecule localization
GO:0015833	0.026310746	3.947368421	1.862068966	5	24	peptide transport
GO:0071396	0.02954021	3.146464646	3.336206897	7	43	cellular response to lipid
GO:0035266	0.031159266	13.25	0.310344828	2	4	meristem growth

Table A.17 continued from previous page

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0042435	0.031159266	13.25	0.310344828	2	4	indole-containing compound biosynthetic process
GO:0051649	0.031255133	3.730769231	1.939655172	5	25	establishment of localization in cell
GO:1901576	0.031964645	3.592592593	9.043478261	13	130	organic substance biosynthetic process
GO:0044249	0.033543811	3.128318584	9.853448276	14	127	cellular biosynthetic process
GO:0035556	0.047142246	2.951612903	2.870689655	6	37	intracellular signal transduction
GO:0048509	0.04953787	8.791666667	0.387931034	2	5	regulation of meristem development
GO:0010073	0.04953787	8.791666667	0.387931034	2	5	meristem maintenance
GO:0006520	0.04953787	8.791666667	0.387931034	2	5	cellular amino acid metabolic process
GO:0071705	0.049595391	3.193979933	2.172413793	5	28	nitrogen compound transport

Table A.30: Overrepresented GO terms in Community 26

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0050832	9.06345E-05	30.58823529	0.663793103	5	22	defense response to fungus
GO:0010374	0.000327488	55.5	0.181034483	3	6	stomatal complex development
GO:0042542	0.000565537	41.4375	0.211206897	3	7	response to hydrogen peroxide
GO:0048481	0.000565537	41.4375	0.211206897	3	7	plant ovule development
GO:0043069	0.000782731	20.0952381	0.543103448	4	18	negative regulation of programmed cell death
GO:0045727	0.000783699	Inf	0.060344828	2	2	positive regulation of translation
GO:0048440	0.000892881	33	0.24137931	3	8	carpel development
GO:0031348	0.000980535	18.66666667	0.573275862	4	19	negative regulation of defense response
GO:0006091	0.001321546	27.375	0.271551724	3	9	generation of precursor metabolites and energy
GO:0042742	0.001480715	16.31372549	0.63362069	4	21	defense response to bacterium
GO:0000165	0.001480715	16.31372549	0.63362069	4	21	MAPK cascade
GO:0006612	0.001789622	15.33333333	0.663793103	4	22	protein targeting to membrane
GO:0010363	0.001789622	15.33333333	0.663793103	4	22	regulation of plant-type hypersensitive response
GO:0072657	0.001789622	15.33333333	0.663793103	4	22	protein localization to membrane
GO:0032268	0.001789622	15.33333333	0.663793103	4	22	regulation of cellular protein metabolic process
GO:0010941	0.001789622	15.33333333	0.663793103	4	22	regulation of cell death
GO:0034050	0.001789622	15.33333333	0.663793103	4	22	host programmed cell death induced by symbiont
GO:0045184	0.002142193	14.45614035	0.693965517	4	23	establishment of protein localization
GO:0050776	0.002142193	14.45614035	0.693965517	4	23	regulation of immune response
GO:0006886	0.002142193	14.45614035	0.693965517	4	23	intracellular protein transport
GO:0043043	0.002317023	89.6	0.090517241	2	3	peptide biosynthetic process
GO:0051247	0.002317023	89.6	0.090517241	2	3	positive regulation of protein metabolic process
GO:0070727	0.002541831	13.66666667	0.724137931	4	24	cellular macromolecule localization
GO:0015833	0.002541831	13.66666667	0.724137931	4	24	peptide transport
GO:0009862	0.002541831	13.66666667	0.724137931	4	24	systemic acquired resistance. salicylic acid mediated signaling pathway
GO:0080134	0.002991972	12.95238095	0.754310345	4	25	regulation of response to stress
GO:0051649	0.002991972	12.95238095	0.754310345	4	25	establishment of localization in cell
GO:0009409	0.002991972	12.95238095	0.754310345	4	25	response to cold
GO:0051707	0.003122506	11.92307692	1.327586207	5	44	response to other organism
GO:0035304	0.003324497	18	0.362068966	3	12	regulation of protein dephosphorylation
GO:0006986	0.004263885	16.125	0.392241379	3	13	response to unfolded protein
GO:0030968	0.004263885	16.125	0.392241379	3	13	endoplasmic reticulum unfolded protein response
GO:0035967	0.004263885	16.125	0.392241379	3	13	cellular response to topologically incorrect protein
GO:0007030	0.004566792	44.6	0.120689655	2	4	Golgi organization
GO:1901292	0.004566792	44.6	0.120689655	2	4	nucleoside phosphate catabolic process
GO:0072522	0.004566792	44.6	0.120689655	2	4	purine-containing compound biosynthetic process
GO:0006096	0.004566792	44.6	0.120689655	2	4	glycolytic process
GO:0046031	0.004566792	44.6	0.120689655	2	4	ADP metabolic process
GO:0046034	0.004566792	44.6	0.120689655	2	4	ATP metabolic process
GO:0006163	0.004566792	44.6	0.120689655	2	4	purine nucleotide metabolic process

Table A.30 continued from previous page

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0006165	0.004566792	44.6	0.120689655	2	4	nucleoside diphosphate phosphorylation
GO:0009199	0.004566792	44.6	0.120689655	2	4	ribonucleoside triphosphate metabolic process
GO:0009161	0.004566792	44.6	0.120689655	2	4	ribonucleoside monophosphate metabolic process
GO:0009168	0.004566792	44.6	0.120689655	2	4	purine ribonucleoside monophosphate biosynthetic process
GO:0009152	0.004566792	44.6	0.120689655	2	4	purine ribonucleotide biosynthetic process
GO:0009185	0.004566792	44.6	0.120689655	2	4	ribonucleoside diphosphate metabolic process
GO:0009124	0.004566792	44.6	0.120689655	2	4	nucleoside monophosphate biosynthetic process
GO:0009126	0.004566792	44.6	0.120689655	2	4	purine nucleoside monophosphate metabolic process
GO:0009142	0.004566792	44.6	0.120689655	2	4	nucleoside triphosphate biosynthetic process
GO:0009144	0.004566792	44.6	0.120689655	2	4	purine nucleoside triphosphate metabolic process
GO:0009135	0.004566792	44.6	0.120689655	2	4	purine nucleoside diphosphate metabolic process
GO:0006833	0.004566792	44.6	0.120689655	2	4	water transport
GO:0009206	0.004566792	44.6	0.120689655	2	4	purine ribonucleoside triphosphate biosynthetic process
GO:0071705	0.004680164	11.16666667	0.844827586	4	28	nitrogen compound transport
GO:0009595	0.005353793	14.59090909	0.422413793	3	14	detection of biotic stimulus
GO:0009697	0.005353793	14.59090909	0.422413793	3	14	salicylic acid biosynthetic process
GO:0043900	0.005353793	14.59090909	0.422413793	3	14	regulation of multi-organism process
GO:0006979	0.005353793	14.59090909	0.422413793	3	14	response to oxidative stress
GO:0018958	0.005353793	14.59090909	0.422413793	3	14	phenol-containing compound metabolic process
GO:0048438	0.005353793	14.59090909	0.422413793	3	14	floral whorl development
GO:0006468	0.006122034	10.20512821	0.905172414	4	30	protein phosphorylation
GO:0097306	0.006122034	10.20512821	0.905172414	4	30	cellular response to alcohol
GO:0009738	0.006122034	10.20512821	0.905172414	4	30	abscisic acid-activated signaling pathway
GO:0042537	0.006602021	13.3125	0.452586207	3	15	benzene-containing compound metabolic process
GO:0019220	0.006602021	13.3125	0.452586207	3	15	regulation of phosphate metabolic process
GO:1901617	0.006602021	13.3125	0.452586207	3	15	organic hydroxy compound biosynthetic process
GO:0019363	0.0075007	29.6	0.150862069	2	5	pyridine nucleotide biosynthetic process
GO:0006090	0.0075007	29.6	0.150862069	2	5	pyruvate metabolic process
GO:0046390	0.0075007	29.6	0.150862069	2	5	ribose phosphate biosynthetic process
GO:0009259	0.0075007	29.6	0.150862069	2	5	ribonucleotide metabolic process
GO:0071446	0.007849569	9.380952381	0.965517241	4	32	cellular response to salicylic acid stimulus
GO:0010310	0.008015736	12.23076923	0.482758621	3	16	regulation of hydrogen peroxide metabolic process
GO:0016311	0.00960149	11.30357143	0.512931034	3	17	dephosphorylation
GO:1901135	0.00960149	11.30357143	0.512931034	3	17	carbohydrate derivative metabolic process
GO:0005975	0.00960149	11.30357143	0.512931034	3	17	carbohydrate metabolic process
GO:0046686	0.011087306	22.1	0.181034483	2	6	response to cadmium ion
GO:0072524	0.011087306	22.1	0.181034483	2	6	pyridine-containing compound metabolic process
GO:0046496	0.011087306	22.1	0.181034483	2	6	nicotinamide nucleotide metabolic process
GO:0010200	0.013312374	9.796875	0.573275862	3	19	response to chitin
GO:0016999	0.013312374	9.796875	0.573275862	3	19	antibiotic metabolic process
GO:0071395	0.013597564	7.757575758	1.11637931	4	37	cellular response to jasmonic acid stimulus
GO:0034248	0.015296014	17.6	0.211206897	2	7	regulation of cellular amide metabolic process

Table A.30 continued from previous page

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0044550	0.015296014	17.6	0.211206897	2	7	secondary metabolite biosynthetic process
GO:0006733	0.015296014	17.6	0.211206897	2	7	oxidoreduction coenzyme metabolic process
GO:0009605	0.01834455	7.033898305	1.931034483	5	64	response to external stimulus
GO:0055086	0.020097058	14.6	0.24137931	2	8	nucleobase-containing small molecule metabolic process
GO:0044272	0.020097058	14.6	0.24137931	2	8	sulfur compound biosynthetic process
GO:0072593	0.02029964	8.131578947	0.663793103	3	22	reactive oxygen species metabolic process
GO:0009814	0.022555258	8.375	0.701298701	3	27	defense response. incompatible interaction
GO:0071396	0.023653483	6.358974359	1.297413793	4	43	cellular response to lipid
GO:0051188	0.02546149	12.45714286	0.271551724	2	9	cofactor biosynthetic process
GO:0048523	0.025706581	6.166666667	1.327586207	4	44	negative regulation of cellular process
GO:0023052	0.027968791	8.210526316	3.047413793	6	101	signaling
GO:1901698	0.029079703	6.920454545	0.754310345	3	25	response to nitrogen compound
GO:0009864	0.030172414	Inf	0.030172414	1	1	induced systemic resistance. jasmonic acid mediated signaling pathway
GO:0009868	0.030172414	Inf	0.030172414	1	1	jasmonic acid and ethylene-dependent systemic resistance. jasmonic acid mediated signaling pathway
GO:0002684	0.030172414	Inf	0.030172414	1	1	positive regulation of immune system process
GO:0042539	0.030172414	Inf	0.030172414	1	1	hypotonic salinity response
GO:0000023	0.030172414	Inf	0.030172414	1	1	maltose metabolic process
GO:0048232	0.030172414	Inf	0.030172414	1	1	male gamete generation
GO:0019252	0.030172414	Inf	0.030172414	1	1	starch biosynthetic process
GO:0007112	0.030172414	Inf	0.030172414	1	1	male meiosis cytokinesis
GO:0045089	0.030172414	Inf	0.030172414	1	1	positive regulation of innate immune response
GO:0002253	0.030172414	Inf	0.030172414	1	1	activation of immune response
GO:0009861	0.030172414	Inf	0.030172414	1	1	jasmonic acid and ethylene-dependent systemic resistance
GO:0007154	0.034674158	7.636363636	3.168103448	6	105	cell communication
GO:1901361	0.037768715	9.6	0.331896552	2	11	organic cyclic compound catabolic process
GO:0051301	0.037768715	9.6	0.331896552	2	11	cell division
GO:0019439	0.037768715	9.6	0.331896552	2	11	aromatic compound catabolic process
GO:0010608	0.037768715	9.6	0.331896552	2	11	posttranscriptional regulation of gene expression
GO:0044270	0.037768715	9.6	0.331896552	2	11	cellular nitrogen compound catabolic process
GO:0046700	0.037768715	9.6	0.331896552	2	11	heterocycle catabolic process
GO:0014070	0.040526615	5.188405797	1.50862069	4	50	response to organic cyclic compound
GO:0009737	0.040526615	5.188405797	1.50862069	4	50	response to abscisic acid
GO:0009888	0.043678603	5.740384615	0.875	3	29	tissue development
GO:0019637	0.044657559	8.6	0.362068966	2	12	organophosphate metabolic process
GO:0022402	0.044657559	8.6	0.362068966	2	12	cell cycle process
GO:0051716	0.047006734	6.857142857	3.349137931	6	111	cellular response to stimulus
GO:0042445	0.047851555	5.5	0.905172414	3	30	hormone metabolic process
GO:0071495	0.049571094	5	2.413793103	5	80	cellular response to endogenous stimulus
GO:0036211	0.049644248	4.789115646	1.599137931	4	53	protein modification process
GO:0002376	0.049975823	5.818181818	0.935064935	3	36	immune system process

Table A.31: Overrepresented GO terms in Community 28

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0016999	0.001787852	39.75	0.327586207	3	19	antibiotic metabolic process
GO:0016045	0.002200326	75	0.086206897	2	5	detection of bacterium
GO:0098581	0.002200326	75	0.086206897	2	5	detection of external biotic stimulus
GO:0072593	0.002811913	33	0.379310345	3	22	reactive oxygen species metabolic process
GO:0051186	0.004155609	28.09090909	0.431034483	3	25	cofactor metabolic process

Table A.31 continued from previous page

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:2001141	0.017215787	Inf	1.465517241	4	85	regulation of RNA biosynthetic process
GO:0006355	0.017215787	Inf	1.465517241	4	85	regulation of transcription. DNA-templated
GO:0002697	0.017241379	Inf	0.017241379	1	1	regulation of immune effector process
GO:0050691	0.017241379	Inf	0.017241379	1	1	regulation of defense response to virus by host
GO:0009911	0.017241379	Inf	0.017241379	1	1	positive regulation of flower development
GO:0008285	0.017241379	Inf	0.017241379	1	1	negative regulation of cell proliferation
GO:0010047	0.017241379	Inf	0.017241379	1	1	fruit dehiscence
GO:0097659	0.018055581	Inf	1.482758621	4	86	nucleic acid-templated transcription
GO:0009697	0.018984236	18	0.24137931	2	14	salicylic acid biosynthetic process
GO:0043900	0.018984236	18	0.24137931	2	14	regulation of multi-organism process
GO:0018958	0.018984236	18	0.24137931	2	14	phenol-containing compound metabolic process
GO:0019219	0.019826955	Inf	1.517241379	4	88	regulation of nucleobase-containing compound metabolic process
GO:0034654	0.020759988	Inf	1.534482759	4	89	nucleobase-containing compound biosynthetic process
GO:0042537	0.021773948	16.53846154	0.25862069	2	15	benzene-containing compound metabolic process
GO:1901617	0.021773948	16.53846154	0.25862069	2	15	organic hydroxy compound biosynthetic process
GO:0051606	0.021773948	16.53846154	0.25862069	2	15	detection of stimulus
GO:0016070	0.022724446	Inf	1.568965517	4	91	RNA metabolic process
GO:0009889	0.023757375	Inf	1.586206897	4	92	regulation of biosynthetic process
GO:0010310	0.024735378	15.28571429	0.275862069	2	16	regulation of hydrogen peroxide metabolic process
GO:0010467	0.025928461	Inf	1.620689655	4	94	gene expression
GO:0034976	0.027864988	14.2	0.293103448	2	17	response to endoplasmic reticulum stress
GO:0034645	0.028245051	Inf	1.655172414	4	96	cellular macromolecule biosynthetic process
GO:0060255	0.028245051	Inf	1.655172414	4	96	regulation of macromolecule metabolic process
GO:0043069	0.031159266	13.25	0.310344828	2	18	negative regulation of programmed cell death
GO:0002831	0.034258845	75.66666667	0.034482759	1	2	regulation of response to biotic stimulus
GO:0042744	0.034258845	75.66666667	0.034482759	1	2	hydrogen peroxide catabolic process
GO:0010227	0.034258845	75.66666667	0.034482759	1	2	floral organ abscission
GO:0010200	0.034614726	12.41176471	0.327586207	2	19	response to chitin
GO:0031348	0.034614726	12.41176471	0.327586207	2	19	negative regulation of defense response
GO:0048519	0.036124894	10.95918367	0.896551724	3	52	negative regulation of biological process
GO:0046483	0.040644029	Inf	1.810344828	4	105	heterocycle metabolic process
GO:0042742	0.041995378	11	0.362068966	2	21	defense response to bacterium
GO:0000165	0.041995378	11	0.362068966	2	21	MAPK cascade
GO:0034641	0.042237912	Inf	1.827586207	4	106	cellular nitrogen compound metabolic process
GO:0006612	0.045913724	10.4	0.379310345	2	22	protein targeting to membrane
GO:0010363	0.045913724	10.4	0.379310345	2	22	regulation of plant-type hypersensitive response
GO:0050832	0.045913724	10.4	0.379310345	2	22	defense response to fungus
GO:0072657	0.045913724	10.4	0.379310345	2	22	protein localization to membrane
GO:0010941	0.045913724	10.4	0.379310345	2	22	regulation of cell death
GO:0034050	0.045913724	10.4	0.379310345	2	22	host programmed cell death induced by symbiont
GO:0045184	0.049979564	9.857142857	0.396551724	2	23	establishment of protein localization
GO:0050776	0.049979564	9.857142857	0.396551724	2	23	regulation of immune response
GO:0006886	0.049979564	9.857142857	0.396551724	2	23	intracellular protein transport

Table A.15: Overrepresented GO terms in Community 4

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0009733	1.49365E-10	87.26666667	2.844827586	14	44	response to auxin
GO:0010583	0.000160897	39.09090909	0.387931034	4	6	response to cyclopentenone
GO:0009719	0.000383231	15.49514563	7.564655172	14	117	response to endogenous stimulus
GO:0009741	0.001200165	15.41818182	0.581896552	4	9	response to brassinosteroid
GO:0010033	0.001327203	12.64912281	8.275862069	14	128	response to organic substance
GO:0048527	0.002684138	21.1	0.343612335	3	6	lateral root development
GO:0090696	0.006187332	6.28125	1.357758621	5	21	post-embryonic plant organ development
GO:0048364	0.008478052	7.461538462	0.931506849	4	17	root development
GO:0010102	0.034927199	10.97435897	0.323275862	2	5	lateral root morphogenesis
GO:0010015	0.042373477	5.175	0.840517241	3	13	root morphogenesis

Table A.16: Overrepresented GO terms in Community 6

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0048825	0.000386637	20.9375	0.875	5	7	cotyledon development
GO:0009880	0.002521133	16.08	0.75	4	6	embryonic pattern specification
GO:0010014	0.005367128	10.66666667	0.875	4	7	meristem initiation
GO:0009955	0.006508794	23.30769231	0.5	3	4	adaxial/abaxial pattern specification
GO:0009944	0.006508794	23.30769231	0.5	3	4	polarity specification of adaxial/abaxial axis
GO:0048532	0.009180204	5.833333333	1.5	5	12	anatomical structure arrangement
GO:0044248	0.012114524	3.485714286	3.5	8	28	cellular catabolic process
GO:0042752	0.014890266	11.59615385	0.625	3	5	regulation of circadian rhythm
GO:0045595	0.014890266	11.59615385	0.625	3	5	regulation of cell differentiation
GO:0007062	0.014890266	11.59615385	0.625	3	5	sister chromatid cohesion
GO:0010564	0.014890266	11.59615385	0.625	3	5	regulation of cell cycle process
GO:0010072	0.014890266	11.59615385	0.625	3	5	primary shoot apical meristem specification
GO:0009798	0.014890266	11.59615385	0.625	3	5	axis specification
GO:0098813	0.014890266	11.59615385	0.625	3	5	nuclear chromosome segregation
GO:0009954	0.015151515	Inf	0.25	2	2	proximal/distal pattern formation
GO:1903409	0.015151515	Inf	0.25	2	2	reactive oxygen species biosynthetic process
GO:0009057	0.019277875	4.490740741	1.75	5	14	macromolecule catabolic process
GO:0048367	0.021940408	2.550802139	7	12	56	shoot system development
GO:0009314	0.023205109	2.8125	4.625	9	37	response to radiation
GO:0009640	0.024468053	5.253333333	1.25	4	10	photomorphogenesis
GO:0048869	0.024932357	3.52173913	2.5	6	20	cellular developmental process
GO:0006511	0.0272594	7.692307692	0.75	3	6	ubiquitin-dependent protein catabolic process
GO:0043632	0.0272594	7.692307692	0.75	3	6	modification-dependent macromolecule catabolic process
GO:0065008	0.028288314	2.490277778	6.375	11	51	regulation of biological quality
GO:0042742	0.031798744	3.269565217	2.625	6	21	defense response to bacterium
GO:0051301	0.035092628	4.48	1.375	4	11	cell division
GO:0034645	0.035672109	2.223628692	12	17	96	cellular macromolecule biosynthetic process
GO:0050832	0.039828512	3.048913043	2.75	6	22	defense response to fungus
GO:0048443	0.041897233	14.96296296	0.375	2	3	stamen development
GO:0009799	0.041897233	14.96296296	0.375	2	3	specification of symmetry
GO:0048608	0.043384505	2.273901809	6.75	11	54	reproductive structure development
GO:0009637	0.043678603	5.740384615	0.875	3	7	response to blue light
GO:0051603	0.043678603	5.740384615	0.875	3	7	proteolysis involved in cellular protein catabolic process
GO:0048507	0.045112494	3.315972222	2.125	5	17	meristem development
GO:0016070	0.048003514	2.100512821	11.375	16	91	RNA metabolic process

Table A.18: Overrepresented GO terms in Community 8

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0048229	0.01461015	18.33333333	0.206896552	2	8	gametophyte development
GO:0009556	0.025862069	Inf	0.025862069	1	1	microsporogenesis
GO:0034293	0.025862069	Inf	0.025862069	1	1	sexual sporulation
GO:1901568	0.025862069	Inf	0.025862069	1	1	fatty acid derivative metabolic process
GO:0010025	0.025862069	Inf	0.025862069	1	1	wax biosynthetic process

Table A.19: Overrepresented GO terms in Community 9

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0019722	0.000761599	41.0625	0.237068966	3	5	calcium-mediated signaling
GO:0010107	0.002052545	Inf	0.094827586	2	2	potassium ion import
GO:0071804	0.002052545	Inf	0.094827586	2	2	cellular potassium ion transport
GO:0010118	0.010703045	9.984375	0.521551724	3	11	stomatal movement
GO:1901700	0.011611123	6.079787234	4.88362069	9	103	response to oxygen-containing compound
GO:0009737	0.014540918	4.827272727	2.370689655	6	50	response to abscisic acid
GO:0035556	0.017752151	4.921875	1.754310345	5	37	intracellular signal transduction
GO:0098662	0.027699704	12.05555556	0.284482759	2	6	inorganic cation transmembrane transport
GO:0034220	0.027699704	12.05555556	0.284482759	2	6	ion transmembrane transport
GO:0030001	0.027699704	12.05555556	0.284482759	2	6	metal ion transport
GO:0015672	0.037768715	9.6	0.331896552	2	7	monovalent inorganic cation transport
GO:0001101	0.03875984	7.655172414	0.879120879	3	32	response to acid chemical

Table A.20: Overrepresented GO terms in Community 12

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0031396	0.012931034	Inf	0.012931034	1	1	regulation of protein ubiquitination

Table A.21: Overrepresented GO terms in Community 13

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0010325	0.005405051	37.16666667	0.129310345	2	5	raffinose family oligosaccharide biosynthetic process
GO:0010187	0.005405051	37.16666667	0.129310345	2	5	negative regulation of seed germination
GO:0009937	0.005405051	37.16666667	0.129310345	2	5	regulation of gibberellic acid mediated signaling pathway
GO:0009311	0.008013072	27.75	0.155172414	2	6	oligosaccharide metabolic process
GO:0090351	0.012366775	10.89473684	0.568965517	3	22	seedling development
GO:0051093	0.01461015	18.33333333	0.206896552	2	8	negative regulation of developmental process
GO:0051241	0.01461015	18.33333333	0.206896552	2	8	negative regulation of multicellular organismal process
GO:0048444	0.022933059	13.625	0.25862069	2	10	floral organ morphogenesis
GO:0006997	0.025862069	Inf	0.025862069	1	1	nucleus organization
GO:0010197	0.025862069	Inf	0.025862069	1	1	polar nucleus fusion
GO:0010114	0.025862069	Inf	0.025862069	1	1	response to red light
GO:0010233	0.025862069	Inf	0.025862069	1	1	phloem transport
GO:0048284	0.025862069	Inf	0.025862069	1	1	organelle fusion
GO:0006808	0.025862069	Inf	0.025862069	1	1	regulation of nitrogen utilization
GO:0010476	0.027699704	12.05555556	0.284482759	2	11	gibberellin mediated signaling pathway
GO:0016051	0.032848113	10.8	0.310344828	2	12	carbohydrate biosynthetic process
GO:0009739	0.032848113	10.8	0.310344828	2	12	response to gibberellin
GO:0010154	0.035835837	6.793103448	0.827586207	3	32	fruit development

Table A.22: Overrepresented GO terms in Community 14

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0009638	0.000851103	54	0.25862069	3	4	phototropism
GO:0071482	0.003918495	Inf	0.129310345	2	2	cellular response to light stimulus
GO:0010017	0.003918495	Inf	0.129310345	2	2	red or far-red light signaling pathway
GO:0034765	0.011312526	33.23076923	0.193965517	2	3	regulation of ion transmembrane transport
GO:0010155	0.011312526	33.23076923	0.193965517	2	3	regulation of proton transport
GO:0051049	0.034927199	10.97435897	0.323275862	2	5	regulation of transport
GO:0071214	0.034927199	10.97435897	0.323275862	2	5	cellular response to abiotic stimulus
GO:0006820	0.034927199	10.97435897	0.323275862	2	5	anion transport
GO:0046777	0.034927199	10.97435897	0.323275862	2	5	protein autophosphorylation
GO:0009892	0.035217404	4.278074866	1.357758621	4	21	negative regulation of metabolic process

Table A.23: Overrepresented GO terms in Community 15

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0009410	4.92843E-12	Inf	0.155172414	6	6	response to xenobiotic stimulus
GO:0006986	8.45718E-09	Inf	0.336206897	6	13	response to unfolded protein
GO:0030968	8.45718E-09	Inf	0.336206897	6	13	endoplasmic reticulum unfolded protein response
GO:0035967	8.45718E-09	Inf	0.336206897	6	13	cellular response to topologically incorrect protein
GO:0009862	6.63347E-07	Inf	0.620689655	6	24	systemic acquired resistance. salicylic acid mediated signaling pathway
GO:0009814	1.85674E-06	Inf	0.724137931	6	28	defense response. incompatible interaction
GO:0071446	4.4661E-06	Inf	0.827586207	6	32	cellular response to salicylic acid stimulus
GO:0009617	6.47893E-06	Inf	0.071090047	3	5	response to bacterium
GO:0006955	7.99963E-06	Inf	0.905172414	6	35	immune response
GO:0046677	1.36058E-05	Inf	0.982758621	6	38	response to antibiotic
GO:0042493	2.21601E-05	Inf	1.060344828	6	41	response to drug
GO:0006952	2.58535E-05	Inf	1.086206897	6	42	defense response
GO:0051704	5.29194E-05	Inf	1.215517241	6	47	multi-organism process
GO:0014070	7.83162E-05	Inf	1.293103448	6	50	response to organic cyclic compound
GO:0043207	0.000763462	Inf	0.68722467	4	39	response to external biotic stimulus
GO:0071495	0.001480994	Inf	2.068965517	6	80	cellular response to endogenous stimulus
GO:0001101	0.001725748	Inf	2.120689655	6	82	response to acid chemical
GO:0070887	0.002863945	Inf	2.301724138	6	89	cellular response to chemical stimulus
GO:0016045	0.005405051	37.16666667	0.129310345	2	5	detection of bacterium
GO:0098581	0.005405051	37.16666667	0.129310345	2	5	detection of external biotic stimulus
GO:0023052	0.006245995	Inf	2.612068966	6	101	signaling
GO:1901700	0.007046867	Inf	2.663793103	6	103	response to oxygen-containing compound
GO:0007154	0.007931538	Inf	2.715517241	6	105	cell communication
GO:0042742	0.010792047	11.55555556	0.543103448	3	21	defense response to bacterium
GO:0008219	0.01407423	10.3	0.594827586	3	23	cell death
GO:0009725	0.01541413	Inf	3.025862069	6	117	response to hormone
GO:1901698	0.017899247	9.272727273	0.646551724	3	25	response to nitrogen compound
GO:0048522	0.020022148	8.826086957	0.672413793	3	26	positive regulation of cellular process
GO:0045132	0.025862069	Inf	0.025862069	1	1	meiotic chromosome segregation
GO:0010112	0.025862069	Inf	0.025862069	1	1	regulation of systemic acquired resistance
GO:0010260	0.025862069	Inf	0.025862069	1	1	animal organ senescence
GO:0015706	0.025862069	Inf	0.025862069	1	1	nitrate transport
GO:0010043	0.025862069	Inf	0.025862069	1	1	response to zinc ion
GO:0045893	0.027699704	12.05555556	0.284482759	2	11	positive regulation of transcription. DNA-templated
GO:1902680	0.027699704	12.05555556	0.284482759	2	11	positive regulation of RNA biosynthetic process
GO:0045935	0.027699704	12.05555556	0.284482759	2	11	positive regulation of nucleobase-containing compound metabolic process
GO:0035304	0.032848113	10.8	0.310344828	2	12	regulation of protein dephosphorylation
GO:0009697	0.044227185	8.916666667	0.362068966	2	14	salicylic acid biosynthetic process
GO:0018958	0.044227185	8.916666667	0.362068966	2	14	phenol-containing compound metabolic process

Table A.24: Overrepresented GO terms in Community 16

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0044260	0.002628398	11.67567568	6.892241379	12	123	cellular macromolecule metabolic process
GO:0042023	0.002910882	Inf	0.112068966	2	2	DNA endoreduplication
GO:0006261	0.002910882	Inf	0.112068966	2	2	DNA-dependent DNA replication
GO:0044249	0.003805715	10.85217391	7.11637931	12	127	cellular biosynthetic process
GO:0009889	0.005913099	5.569105691	5.155172414	10	92	regulation of biosynthetic process
GO:0090304	0.007838744	5.254901961	5.323275862	10	95	nucleic acid metabolic process
GO:0006635	0.008454214	39.63636364	0.168103448	2	3	fatty acid beta-oxidation
GO:0034440	0.008454214	39.63636364	0.168103448	2	3	lipid oxidation
GO:0060255	0.008587785	5.15503876	5.379310345	10	96	regulation of macromolecule metabolic process
GO:0009059	0.010267245	4.962121212	5.49137931	10	98	macromolecule biosynthetic process
GO:0006725	0.017034227	4.432624113	5.827586207	10	104	cellular aromatic compound metabolic process
GO:0044255	0.01747335	5.638888889	1.120689655	4	20	cellular lipid metabolic process
GO:0046483	0.01845823	4.350877193	5.88362069	10	105	heterocycle metabolic process
GO:0019219	0.019239593	3.987341772	4.931034483	9	88	regulation of nucleobase-containing compound metabolic process
GO:2000112	0.019239593	3.987341772	4.931034483	9	88	regulation of cellular macromolecule biosynthetic process
GO:0034641	0.019978942	4.270833333	5.939655172	10	106	cellular nitrogen compound metabolic process
GO:1901360	0.021601184	4.192439863	5.995689655	10	107	organic cyclic compound metabolic process
GO:0009735	0.02266502	7	0.672413793	3	12	response to cytokinin
GO:0072329	0.02641473	13.09090909	0.280172414	2	5	monocarboxylic acid catabolic process
GO:0048509	0.02641473	13.09090909	0.280172414	2	5	regulation of meristem development
GO:0010467	0.031062969	3.547058824	5.267241379	9	94	gene expression
GO:0009826	0.038362441	9.772727273	0.336206897	2	6	unidimensional cell growth
GO:0048588	0.038362441	9.772727273	0.336206897	2	6	developmental cell growth
GO:0048638	0.038362441	9.772727273	0.336206897	2	6	regulation of developmental growth
GO:0071695	0.044124701	3.97979798	1.456896552	4	26	anatomical structure maturation
GO:0080090	0.047229865	10.66666667	0.388888889	2	14	regulation of primary metabolic process
GO:0032502	0.047946779	3.164835165	5.603448276	9	100	developmental process

Table A.25: Overrepresented GO terms in Community 17

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0035434	0.017241379	Inf	0.017241379	1	1	copper ion transmembrane transport
GO:0000041	0.034258845	75.66666667	0.034482759	1	2	transition metal ion transport
GO:0009767	0.034258845	75.66666667	0.034482759	1	2	photosynthetic electron transport chain

Table A.26: Overrepresented GO terms in Community 18

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0030422	0.009614189	27.5	0.172413793	2	4	production of siRNA involved in RNA interference
GO:0017148	0.009614189	27.5	0.172413793	2	4	negative regulation of translation
GO:0031050	0.009614189	27.5	0.172413793	2	4	dsRNA fragmentation
GO:1901699	0.009614189	27.5	0.172413793	2	4	cellular response to nitrogen compound
GO:0035196	0.009614189	27.5	0.172413793	2	4	production of miRNAs involved in gene silencing by miRNA
GO:0043331	0.009614189	27.5	0.172413793	2	4	response to dsRNA
GO:0009629	0.010413142	10.14285714	0.517241379	3	12	response to gravity
GO:0000398	0.015651899	18.25	0.215517241	2	5	mRNA splicing. via spliceosome
GO:0000375	0.015651899	18.25	0.215517241	2	5	RNA splicing. via transesterification reactions
GO:0043687	0.015651899	18.25	0.215517241	2	5	post-translational protein modification
GO:0042221	0.029644324	7.112903226	5.732758621	9	133	response to chemical
GO:0034248	0.031361163	10.85	0.301724138	2	7	regulation of cellular amide metabolic process
GO:0035194	0.031361163	10.85	0.301724138	2	7	posttranscriptional gene silencing by RNA
GO:0051248	0.031361163	10.85	0.301724138	2	7	negative regulation of protein metabolic process
GO:0009642	0.04084434	9	0.344827586	2	8	response to light intensity
GO:0016458	0.04084434	9	0.344827586	2	8	gene silencing
GO:0040029	0.04084434	9	0.344827586	2	8	regulation of gene expression. epigenetic
GO:0042726	0.043103448	Inf	0.043103448	1	1	flavin-containing compound metabolic process
GO:0010262	0.043103448	Inf	0.043103448	1	1	somatic embryogenesis
GO:0010588	0.043103448	Inf	0.043103448	1	1	cotyledon vascular tissue pattern formation
GO:0048598	0.043103448	Inf	0.043103448	1	1	embryonic morphogenesis
GO:0006767	0.043103448	Inf	0.043103448	1	1	water-soluble vitamin metabolic process
GO:0009110	0.043103448	Inf	0.043103448	1	1	vitamin biosynthetic process
GO:0006857	0.043103448	Inf	0.043103448	1	1	oligopeptide transport
GO:0009231	0.043103448	Inf	0.043103448	1	1	riboflavin biosynthetic process
GO:0009715	0.043103448	Inf	0.043103448	1	1	chalcone biosynthetic process

Table A.27: Overrepresented GO terms in Community 21

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0045857	0.012931034	Inf	0.012931034	1	1	negative regulation of molecular function. epigenetic
GO:0009910	0.012931034	Inf	0.012931034	1	1	negative regulation of flower development
GO:0018022	0.025750112	114	0.025862069	1	2	peptidyl-lysine methylation
GO:0051567	0.025750112	114	0.025862069	1	2	histone H3-K9 methylation
GO:0010048	0.025750112	114	0.025862069	1	2	vernalization response
GO:0009825	0.038457719	56.75	0.038793103	1	3	multidimensional cell growth
GO:0008213	0.038457719	56.75	0.038793103	1	3	protein alkylation
GO:0031048	0.038457719	56.75	0.038793103	1	3	chromatin silencing by small RNA
GO:0016571	0.038457719	56.75	0.038793103	1	3	histone methylation
GO:0006346	0.038457719	56.75	0.038793103	1	3	methylation-dependent chromatin silencing

Table A.28: Overrepresented GO terms in Community 24

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0048449	0.00037319	Inf	0.043103448	2	5	floral organ formation
GO:0006325	0.001343484	Inf	0.077586207	2	9	chromatin organization
GO:0009909	0.001679355	Inf	0.086206897	2	10	regulation of flower development
GO:0048646	0.002910882	Inf	0.112068966	2	13	anatomical structure formation involved in morphogenesis
GO:0090697	0.003396029	Inf	0.120689655	2	14	post-embryonic plant organ morphogenesis
GO:0016570	0.004366812	Inf	0.004366812	1	1	histone modification
GO:0009560	0.00862069	Inf	0.00862069	1	1	embryo sac egg cell differentiation
GO:0001510	0.00862069	Inf	0.00862069	1	1	RNA methylation
GO:0072528	0.00862069	Inf	0.00862069	1	1	pyrimidine-containing compound biosynthetic process
GO:0006312	0.00862069	Inf	0.00862069	1	1	mitotic recombination
GO:0006220	0.00862069	Inf	0.00862069	1	1	pyrimidine nucleotide metabolic process
GO:0009220	0.00862069	Inf	0.00862069	1	1	pyrimidine ribonucleotide biosynthetic process
GO:0006996	0.010300045	Inf	0.206896552	2	24	organelle organization
GO:0048437	0.010300045	Inf	0.206896552	2	24	floral organ development
GO:0009553	0.01720406	229	0.017241379	1	2	embryo sac development
GO:0007276	0.01720406	229	0.017241379	1	2	gamete generation
GO:0018022	0.01720406	229	0.017241379	1	2	peptidyl-lysine methylation
GO:0051567	0.01720406	229	0.017241379	1	2	histone H3-K9 methylation
GO:2000026	0.017353336	Inf	0.267241379	2	31	regulation of multicellular organismal development
GO:0071840	0.020935961	Inf	0.293103448	2	34	cellular component organization or biogenesis
GO:0043414	0.021645022	Inf	0.021645022	1	5	macromolecule methylation
GO:0090567	0.024854456	Inf	0.318965517	2	37	reproductive shoot system development
GO:0008213	0.025750112	114	0.025862069	1	3	protein alkylation
GO:0031048	0.025750112	114	0.025862069	1	3	chromatin silencing by small RNA
GO:0016571	0.025750112	114	0.025862069	1	3	histone methylation
GO:0044703	0.025750112	114	0.025862069	1	3	multi-organism reproductive process
GO:0006346	0.025750112	114	0.025862069	1	3	methylation-dependent chromatin silencing
GO:0000398	0.042730258	56.5	0.043103448	1	5	mRNA splicing. via spliceosome
GO:0000375	0.042730258	56.5	0.043103448	1	5	RNA splicing. via transesterification reactions
GO:0046390	0.042730258	56.5	0.043103448	1	5	ribose phosphate biosynthetic process
GO:0009259	0.042730258	56.5	0.043103448	1	5	ribonucleotide metabolic process
GO:0018193	0.042730258	56.5	0.043103448	1	5	peptidyl-amino acid modification

Table A.29: Overrepresented GO terms in Community 25

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0010374	0.005405051	37.16666667	0.129310345	2	6	stomatal complex development
GO:0000165	0.005733487	17.41666667	0.452586207	3	21	MAPK cascade
GO:0048481	0.0075007	29.6	0.150862069	2	7	plant ovule development
GO:0048440	0.009913025	24.55555556	0.172413793	2	8	carpel development
GO:0009814	0.013445185	12.12	0.603448276	3	28	defense response. incompatible interaction
GO:0016310	0.0197828	10.24137931	0.689655172	3	32	phosphorylation
GO:1902533	0.021551724	Inf	0.021551724	1	1	positive regulation of intracellular signal transduction
GO:0071470	0.021551724	Inf	0.021551724	1	1	cellular response to osmotic stress
GO:0045937	0.021551724	Inf	0.021551724	1	1	positive regulation of phosphate metabolic process
GO:0032147	0.021551724	Inf	0.021551724	1	1	activation of protein kinase activity
GO:0043085	0.021551724	Inf	0.021551724	1	1	positive regulation of catalytic activity
GO:0033674	0.021551724	Inf	0.021551724	1	1	positive regulation of kinase activity
GO:0072663	0.021551724	Inf	0.021551724	1	1	establishment of protein localization to peroxisome
GO:0001934	0.021551724	Inf	0.021551724	1	1	positive regulation of protein phosphorylation
GO:0000169	0.021551724	Inf	0.021551724	1	1	activation of MAPK activity involved in osmosensory signaling pathway
GO:0016558	0.021551724	Inf	0.021551724	1	1	protein import into peroxisome matrix
GO:0008643	0.021551724	Inf	0.021551724	1	1	carbohydrate transport
GO:0071806	0.021551724	Inf	0.021551724	1	1	protein transmembrane transport
GO:0043406	0.021551724	Inf	0.021551724	1	1	positive regulation of MAP kinase activity
GO:0015749	0.021551724	Inf	0.021551724	1	1	monosaccharide transmembrane transport
GO:0043574	0.021551724	Inf	0.021551724	1	1	peroxisomal transport
GO:0017038	0.021551724	Inf	0.021551724	1	1	protein import
GO:0035304	0.022551296	14.46666667	0.25862069	2	12	regulation of protein dephosphorylation
GO:0006955	0.02554437	9.140625	0.754310345	3	35	immune response
GO:0006979	0.030542782	11.94444444	0.301724138	2	14	response to oxidative stress
GO:0048438	0.030542782	11.94444444	0.301724138	2	14	floral whorl development
GO:0051174	0.034927199	10.97435897	0.323275862	2	15	regulation of phosphorus metabolic process
GO:0010310	0.039559845	10.14285714	0.344827586	2	16	regulation of hydrogen peroxide metabolic process
GO:0006952	0.042557522	7.230769231	0.905172414	3	42	defense response
GO:0033365	0.042730258	56.5	0.043103448	1	2	protein localization to organelle
GO:0010120	0.042730258	56.5	0.043103448	1	2	camalexin biosynthetic process
GO:0052315	0.042730258	56.5	0.043103448	1	2	phytoalexin biosynthetic process
GO:0007031	0.042730258	56.5	0.043103448	1	2	peroxisome organization
GO:0010584	0.042730258	56.5	0.043103448	1	2	pollen exine formation
GO:0080136	0.042730258	56.5	0.043103448	1	2	priming of cellular response to stress
GO:0043062	0.042730258	56.5	0.043103448	1	2	extracellular structure organization
GO:0010927	0.042730258	56.5	0.043103448	1	2	cellular component assembly involved in morphogenesis
GO:0085029	0.042730258	56.5	0.043103448	1	2	extracellular matrix assembly
GO:0046217	0.042730258	56.5	0.043103448	1	2	indole phytoalexin metabolic process
GO:0016311	0.044432702	9.422222222	0.36637931	2	17	dephosphorylation
GO:0009607	0.048369346	6.804878049	0.948275862	3	44	response to biotic stimulus
GO:0051707	0.048369346	6.804878049	0.948275862	3	44	response to other organism
GO:0043069	0.04953787	8.791666667	0.387931034	2	18	negative regulation of programmed cell death

Table A.32: Overrepresented GO terms in Community 30

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0032922	0.012931034	Inf	0.012931034	1	1	circadian regulation of gene expression
GO:0009938	0.038457719	56.75	0.038793103	1	3	negative regulation of gibberellic acid mediated signaling pathway
GO:0009628	0.046381355	Inf	1.086206897	3	84	response to abiotic stimulus

A.14 Association network clustering

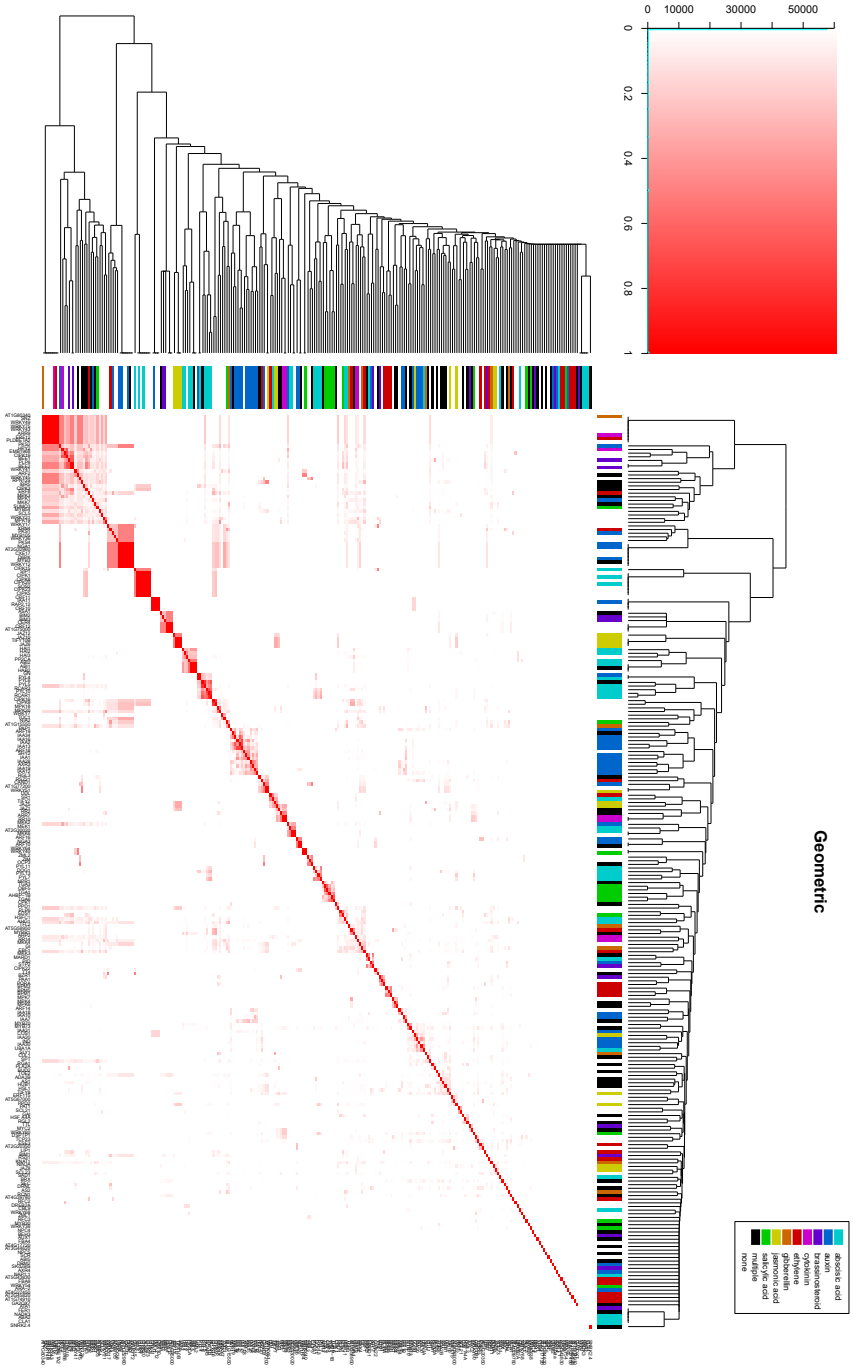


Figure A.5: Clustering of association values of PhI interactions. Hormone annotations from AHD 2.0 and GO combined.

A.15 PhI_{ext} direct crosstalks

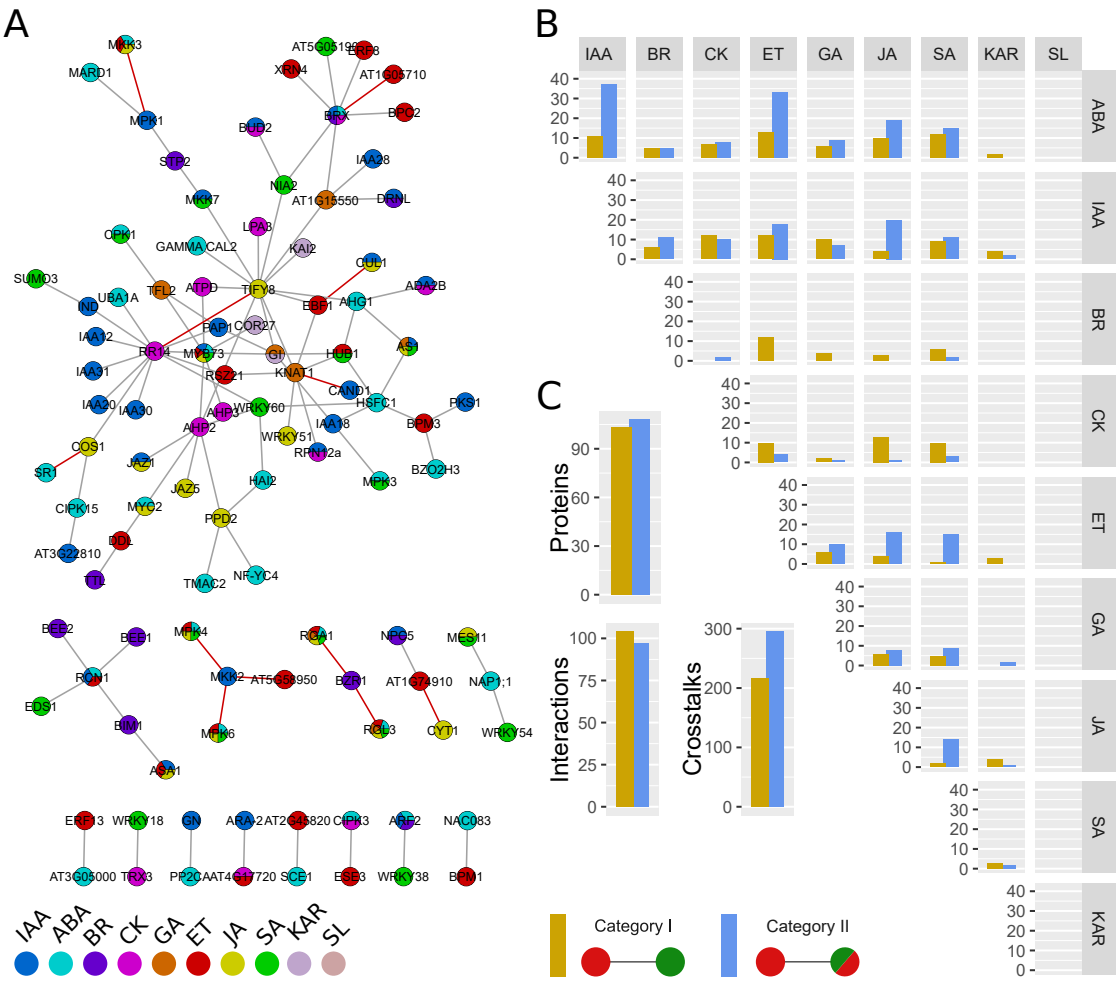


Figure A.6: A) PhI_{ext} pairwise crosstalks category I. B) PhI_{ext} number of pairwise crosstalks category I (orange) and category II (blue). C) Network properties

Table A.37: List of all pairwise crosstalks in PhI_{ext}

Interactor A	Interactor B	Annotation A	Annotation B	Crosstalk
AT1G01360	AT4G26080	abscisic acid	abscisic acid, ethylene	Cat II
AT1G10940	AT2G36270	abscisic acid, auxin	abscisic acid, gibberellin	Cat II
AT1G15050	AT3G23050	auxin	abscisic acid, auxin, jasmonic acid	Cat II
AT1G17380	AT3G29350	jasmonic acid	cytokinin	Cat I
AT1G22770	AT1G51950	gibberellin, karrikin	auxin	Cat I
AT1G22770	AT2G37630	gibberellin, karrikin	auxin, gibberellin, jasmonic acid, salicylic acid	Cat II
AT1G22770	AT2G44950	gibberellin, karrikin	ethylene, salicylic acid	Cat I
AT1G22770	AT3G16500	gibberellin, karrikin	auxin	Cat I
AT1G22770	AT4G08150	gibberellin, karrikin	gibberellin	Cat II
AT1G22770	AT4G32570	gibberellin, karrikin	jasmonic acid	Cat I
AT1G22770	AT4G37260	gibberellin, karrikin	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	Cat I
AT1G32640	AT3G29350	abscisic acid, jasmonic acid	cytokinin	Cat I
AT1G37130	AT5G18930	salicylic acid	auxin, cytokinin	Cat I
AT1G51950	AT3G24520	auxin	abscisic acid	Cat I
AT1G51950	AT3G45640	auxin	abscisic acid, salicylic acid	Cat I
AT1G64280	AT5G06950	jasmonic acid, salicylic acid	salicylic acid	Cat II
AT1G64520	AT4G08150	auxin, cytokinin	gibberellin	Cat I
AT1G65620	AT2G37630	gibberellin	auxin, gibberellin, jasmonic acid, salicylic acid	Cat II
AT1G75080	AT2G01570	brassinosteroid	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	Cat I

Table A.37 continued from previous page

Interactor A	Interactor B	Annotation A	Annotation B	Crosstalk
AT2G01570	AT4G16110	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	cytokinin, ethylene	Cat II
AT2G01760	AT2G46990	cytokinin	auxin	Cat I
AT2G01760	AT3G16500	cytokinin	auxin	Cat I
AT2G01760	AT3G17600	cytokinin	auxin	Cat I
AT2G01760	AT3G62100	cytokinin	auxin	Cat I
AT2G01760	AT4G00120	cytokinin	auxin	Cat I
AT2G02560	AT4G08150	auxin	gibberellin	Cat I
AT2G25490	AT2G44950	ethylene	ethylene, salicylic acid	Cat II
AT2G26980	AT5G47100	abscisic acid, cytokinin	abscisic acid	Cat II
AT2G28350	AT3G17600	abscisic acid, auxin	auxin	Cat II
AT2G33310	AT3G23050	auxin	abscisic acid, auxin, jasmonic acid	Cat II
AT2G40330	AT4G26080	abscisic acid, jasmonic acid	abscisic acid, ethylene	Cat II
AT2G44950	AT3G24520	ethylene, salicylic acid	abscisic acid	Cat I
AT2G44950	AT3G48090	ethylene, salicylic acid	salicylic acid	Cat II
AT2G44950	AT4G08150	ethylene, salicylic acid	gibberellin	Cat I
AT2G44950	AT5G51760	ethylene, salicylic acid	abscisic acid	Cat I
AT3G16500	AT3G23050	auxin	abscisic acid, auxin, jasmonic acid	Cat II
AT3G16500	AT5G17690	auxin	gibberellin	Cat I
AT3G23030	AT3G23050	auxin	abscisic acid, auxin, jasmonic acid	Cat II
AT4G08150	AT5G64810	gibberellin	jasmonic acid	Cat I
AT4G16420	AT5G51760	auxin, cytokinin	abscisic acid	Cat I
AT4G26080	AT4G27920	abscisic acid, ethylene	abscisic acid	Cat II
AT4G26080	AT5G05440	abscisic acid, ethylene	abscisic acid	Cat II
AT4G26080	AT5G53160	abscisic acid, ethylene	abscisic acid	Cat II
AT5G04870	AT5G17690	abscisic acid, salicylic acid	gibberellin	Cat I
AT4G37260	AT5G67300	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, auxin, ethylene, gibberellin, jasmonic acid, salicylic acid	Cat II
AT2G25490	AT4G37260	ethylene	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	Cat II
AT3G29350	AT4G37260	cytokinin	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	Cat I
AT4G37260	AT5G53160	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid	Cat II
AT4G37260	AT5G58950	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	ethylene	Cat II
AT1G53170	AT4G37260	ethylene	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	Cat II
AT4G37260	AT5G17690	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	gibberellin	Cat I
AT2G01760	AT4G37260	cytokinin	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	Cat I
AT1G19220	AT3G23050	auxin, ethylene	abscisic acid, auxin, jasmonic acid	Cat II
AT1G10210	AT4G37260	auxin	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	Cat II
AT1G07340	AT1G10210	brassinosteroid	auxin	Cat I
AT1G07340	AT1G18350	brassinosteroid	auxin, salicylic acid	Cat I
AT2G30020	AT4G01370	abscisic acid	abscisic acid, ethylene, jasmonic acid, salicylic acid	Cat II
AT2G44050	AT5G01810	jasmonic acid	abscisic acid	Cat I
AT1G31880	AT5G51760	abscisic acid, auxin, brassinosteroid, cytokinin	abscisic acid	Cat II
AT1G31880	AT5G55910	abscisic acid, auxin, brassinosteroid, cytokinin	auxin	Cat II
AT1G64280	AT3G12250	jasmonic acid, salicylic acid	salicylic acid	Cat II
AT3G29350	AT4G16110	cytokinin	cytokinin, ethylene	Cat II
AT1G74910	AT3G03540	ethylene	auxin, brassinosteroid	Cat I
AT4G32570	AT5G51760	jasmonic acid	abscisic acid	Cat I
AT3G03450	AT5G54190	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	ethylene	Cat II
AT4G01370	AT4G29810	abscisic acid, ethylene, jasmonic acid, salicylic acid	auxin	Cat I
AT5G53160	AT5G67300	abscisic acid	abscisic acid, auxin, ethylene, gibberellin, jasmonic acid, salicylic acid	Cat II
AT2G25490	AT4G02570	ethylene	auxin, jasmonic acid	Cat I
AT2G25490	AT4G08150	ethylene	gibberellin	Cat I
AT1G31880	AT1G54490	abscisic acid, auxin, brassinosteroid, cytokinin	ethylene	Cat I
AT1G31880	AT1G37130	abscisic acid, auxin, brassinosteroid, cytokinin	salicylic acid	Cat I
AT1G31880	AT2G02950	abscisic acid, auxin, brassinosteroid, cytokinin	auxin	Cat II
AT1G15550	AT1G31880	gibberellin	abscisic acid, auxin, brassinosteroid, cytokinin	Cat I
AT5G22570	AT5G62000	salicylic acid	abscisic acid, auxin, brassinosteroid	Cat I
AT1G64280	AT5G06960	jasmonic acid, salicylic acid	salicylic acid	Cat II
AT1G14920	AT4G24210	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	gibberellin	Cat II
AT1G07430	AT2G25000	abscisic acid	salicylic acid	Cat I
AT2G25000	AT3G24520	salicylic acid	abscisic acid	Cat I
AT2G25000	AT3G29350	salicylic acid	cytokinin	Cat I
AT2G25490	AT4G32570	ethylene	jasmonic acid	Cat I
AT4G08150	AT4G32570	gibberellin	jasmonic acid	Cat I
AT1G37130	AT4G32570	salicylic acid	jasmonic acid	Cat I
AT3G29350	AT4G32570	cytokinin	jasmonic acid	Cat I
AT1G15550	AT4G32570	gibberellin	jasmonic acid	Cat I
AT2G40330	AT5G57050	abscisic acid, jasmonic acid	abscisic acid	Cat II
AT3G03450	AT5G64813	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	ethylene	Cat II

Table A.37 continued from previous page

Interactor A	Interactor B	Annotation A	Annotation B	Crosstalk
AT4G01370	AT4G26070	abscisic acid, ethylene, jasmonic acid, salicylic acid	abscisic acid	Cat II
AT2G01760	AT2G44050	cytokinin	jasmonic acid	Cat I
AT1G19180	AT4G28910	auxin, jasmonic acid	jasmonic acid	Cat II
AT1G31880	AT1G53170	abscisic acid, auxin, brassinosteroid, cytokinin	ethylene	Cat I
AT1G31880	AT2G46870	abscisic acid, auxin, brassinosteroid, cytokinin	auxin	Cat II
AT1G22640	AT1G31880	abscisic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin	Cat II
AT1G14280	AT1G31880	auxin	abscisic acid, auxin, brassinosteroid, cytokinin	Cat II
AT1G31880	AT2G01760	abscisic acid, auxin, brassinosteroid, cytokinin	cytokinin	Cat II
AT1G14920	AT2G44950	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	ethylene, salicylic acid	Cat II
AT1G14920	AT1G70700	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	jasmonic acid	Cat II
AT1G32640	AT1G70700	abscisic acid, jasmonic acid	jasmonic acid	Cat II
AT1G32640	AT3G20550	abscisic acid, jasmonic acid	ethylene	Cat I
AT1G17380	AT1G32640	jasmonic acid	abscisic acid, jasmonic acid	Cat II
AT2G01760	AT2G25000	cytokinin	salicylic acid	Cat I
AT2G01760	AT4G32570	cytokinin	jasmonic acid	Cat I
AT1G31880	AT5G04190	abscisic acid, auxin, brassinosteroid, cytokinin	auxin	Cat II
AT1G31880	AT5G08130	abscisic acid, auxin, brassinosteroid, cytokinin	brassinosteroid	Cat II
AT1G06400	AT4G17720	auxin	cytokinin, ethylene	Cat I
AT2G26980	AT5G25190	abscisic acid, cytokinin	ethylene	Cat I
AT1G18350	AT2G34650	auxin, salicylic acid	auxin, karrikin	Cat II
AT1G18350	AT4G32570	auxin, salicylic acid	jasmonic acid	Cat I
AT1G10210	AT5G40440	auxin	abscisic acid, ethylene, jasmonic acid	Cat I
AT2G30020	AT2G43790	abscisic acid	abscisic acid, ethylene, jasmonic acid, salicylic acid	Cat II
AT2G40750	AT4G26110	salicylic acid	abscisic acid	Cat I
AT1G25490	AT5G51760	abscisic acid, auxin, ethylene	abscisic acid	Cat II
AT4G29810	AT5G58950	auxin	ethylene	Cat I
AT2G43790	AT4G29810	abscisic acid, ethylene, jasmonic acid, salicylic acid	auxin	Cat I
AT2G04550	AT5G47100	abscisic acid, auxin, ethylene	abscisic acid	Cat II
AT1G19180	AT3G29350	auxin, jasmonic acid	cytokinin	Cat I
AT1G64280	AT5G65210	jasmonic acid, salicylic acid	salicylic acid	Cat II
AT2G02950	AT2G39760	auxin	ethylene	Cat I
AT2G39760	AT3G24520	ethylene	abscisic acid	Cat I
AT1G25490	AT3G48090	abscisic acid, auxin, ethylene	salicylic acid	Cat I
AT1G15550	AT5G25890	gibberellin	auxin	Cat I
AT3G15540	AT3G23050	auxin	abscisic acid, auxin, jasmonic acid	Cat II
AT1G19180	AT1G30135	auxin, jasmonic acid	jasmonic acid	Cat II
AT4G00120	AT5G55170	auxin	salicylic acid	Cat I
AT3G20550	AT5G58220	ethylene	brassinosteroid	Cat I
AT3G45640	AT3G63210	abscisic acid, salicylic acid	abscisic acid	Cat II
AT1G25490	AT5G64813	abscisic acid, auxin, ethylene	ethylene	Cat II
AT1G18400	AT1G25490	brassinosteroid	abscisic acid, auxin, ethylene	Cat I
AT1G25490	AT1G53170	abscisic acid, auxin, ethylene	ethylene	Cat II
AT1G25490	AT4G36540	abscisic acid, auxin, ethylene	brassinosteroid	Cat I
AT3G23050	AT5G25890	abscisic acid, auxin, jasmonic acid	auxin	Cat II
AT1G54990	AT2G38120	auxin	auxin, ethylene	Cat II
AT1G10210	AT3G63210	auxin	abscisic acid	Cat I
AT1G25490	AT2G37630	abscisic acid, auxin, ethylene	auxin, gibberellin, jasmonic acid, salicylic acid	Cat II
AT1G25490	AT5G08130	abscisic acid, auxin, ethylene	brassinosteroid	Cat I
AT1G25490	AT2G26980	abscisic acid, auxin, ethylene	abscisic acid, cytokinin	Cat II
AT2G37630	AT5G51760	auxin, gibberellin, jasmonic acid, salicylic acid	abscisic acid	Cat I
AT5G05730	AT5G08130	auxin, ethylene, jasmonic acid	brassinosteroid	Cat I
AT1G13980	AT3G11410	auxin	abscisic acid	Cat I
AT5G05730	AT5G20900	auxin, ethylene, jasmonic acid	jasmonic acid	Cat II
AT1G15550	AT1G24590	gibberellin	auxin, brassinosteroid	Cat I
AT1G22070	AT1G64280	salicylic acid	jasmonic acid, salicylic acid	Cat II
AT1G04550	AT3G23050	auxin	abscisic acid, auxin, jasmonic acid	Cat II
AT1G04550	AT2G01760	auxin	cytokinin	Cat I
AT2G01760	AT2G22090	cytokinin	abscisic acid	Cat I
AT1G23860	AT2G01760	ethylene	cytokinin	Cat I
AT1G18350	AT1G24590	auxin, salicylic acid	auxin, brassinosteroid	Cat II
AT1G37130	AT2G37630	salicylic acid	auxin, gibberellin, jasmonic acid, salicylic acid	Cat II
AT2G44050	AT5G13930	jasmonic acid	auxin, jasmonic acid	Cat II
AT2G37630	AT3G24520	auxin, gibberellin, jasmonic acid, salicylic acid	abscisic acid	Cat I
AT1G15550	AT2G37630	gibberellin	auxin, gibberellin, jasmonic acid, salicylic acid	Cat II
AT2G04550	AT3G11410	abscisic acid, auxin, ethylene	abscisic acid	Cat II
AT2G40330	AT3G11410	abscisic acid, jasmonic acid	abscisic acid	Cat II
AT2G33830	AT3G03530	auxin	auxin, brassinosteroid	Cat II
AT4G08150	AT5G17490	gibberellin	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	Cat II
AT5G17490	AT5G54190	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	ethylene	Cat II
AT1G75080	AT5G17490	brassinosteroid	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	Cat I

Table A.37 continued from previous page

Interactor A	Interactor B	Annotation A	Annotation B	Crosstalk
AT1G07430	AT4G14720	abscisic acid	jasmonic acid	Cat I
AT1G23860	AT4G08150	ethylene	gibberellin	Cat I
AT1G32640	AT5G01820	abscisic acid, jasmonic acid	abscisic acid	Cat II
AT3G02140	AT4G37260	abscisic acid	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	Cat II
AT4G37260	AT5G39340	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	cytokinin	Cat I
AT4G37260	AT5G42900	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	karrikin	Cat I
AT2G44840	AT3G05000	ethylene	abscisic acid	Cat I
AT4G09650	AT4G37260	cytokinin	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	Cat I
AT3G12830	AT4G37260	auxin	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	Cat II
AT2G37630	AT4G37260	auxin, gibberellin, jasmonic acid, salicylic acid	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	Cat II
AT1G19220	AT1G51950	auxin, ethylene	auxin	Cat II
AT3G03450	AT5G28770	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	abscisic acid	Cat II
AT2G04550	AT3G58680	abscisic acid, auxin, ethylene	abscisic acid, ethylene	Cat II
AT3G03450	AT3G48680	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	abscisic acid	Cat II
AT1G04100	AT1G19220	auxin	auxin, ethylene	Cat II
AT1G19220	AT3G04730	auxin, ethylene	auxin	Cat II
AT3G56090	AT5G01600	cytokinin	cytokinin, salicylic acid	Cat II
AT3G22810	AT5G01810	auxin	abscisic acid	Cat I
AT2G44050	AT5G01820	jasmonic acid	abscisic acid	Cat I
AT2G45820	AT3G57870	ethylene	abscisic acid	Cat I
AT1G31880	AT3G02140	abscisic acid, auxin, brassinosteroid, cytokinin	abscisic acid	Cat II
AT1G31880	AT4G26610	abscisic acid, auxin, brassinosteroid, cytokinin	auxin	Cat II
AT3G46060	AT4G17720	ethylene	cytokinin, ethylene	Cat II
AT1G31880	AT5G03740	abscisic acid, auxin, brassinosteroid, cytokinin	abscisic acid	Cat II
AT1G05710	AT1G31880	ethylene	abscisic acid, auxin, brassinosteroid, cytokinin	Cat I
AT1G31880	AT5G05190	abscisic acid, auxin, brassinosteroid, cytokinin	salicylic acid	Cat I
AT1G31880	AT3G27580	abscisic acid, auxin, brassinosteroid, cytokinin	auxin	Cat II
AT1G31880	AT1G68130	abscisic acid, auxin, brassinosteroid, cytokinin	auxin	Cat II
AT1G14685	AT1G31880	ethylene	abscisic acid, auxin, brassinosteroid, cytokinin	Cat I
AT1G74910	AT2G39770	ethylene	jasmonic acid	Cat I
AT4G31800	AT5G42980	salicylic acid	cytokinin	Cat I
AT4G32570	AT4G37470	jasmonic acid	karrikin	Cat I
AT1G73060	AT4G32570	cytokinin	jasmonic acid	Cat I
AT4G09650	AT4G32570	cytokinin	jasmonic acid	Cat I
AT3G48680	AT4G32570	abscisic acid	jasmonic acid	Cat I
AT1G28480	AT4G32570	jasmonic acid, salicylic acid	jasmonic acid	Cat II
AT3G29770	AT4G26110	jasmonic acid, salicylic acid	abscisic acid	Cat I
AT1G25490	AT3G02140	abscisic acid, auxin, ethylene	abscisic acid	Cat II
AT2G39760	AT5G28770	ethylene	abscisic acid	Cat I
AT5G13180	AT5G19000	abscisic acid	ethylene	Cat I
AT5G17300	AT5G24520	auxin	auxin, ethylene, jasmonic acid	Cat II
AT1G22070	AT1G28480	salicylic acid	jasmonic acid, salicylic acid	Cat II
AT3G02140	AT4G14720	abscisic acid	jasmonic acid	Cat I
AT3G29350	AT4G14720	cytokinin	jasmonic acid	Cat I
AT4G14720	AT5G63470	jasmonic acid	abscisic acid	Cat I

Table A.33: Number of pairwise crosstalks in PhI, Category I

	ABA	IAA	BR	CK	ET	GA	JA	SA	KAR	SL
ABA	28	10	5	5	7	6	5	10	1	0
IAA	0	31	6	10	10	10	4	8	3	0
BR	0	0	8	0	10	4	3	5	0	0
CK	0	0	0	2	6	2	8	6	0	0
ET	0	0	0	0	10	6	3	1	2	0
GA	0	0	0	0	0	2	6	5	0	0
JA	0	0	0	0	0	0	10	2	2	0
SA	0	0	0	0	0	0	0	6	2	0
KAR	0	0	0	0	0	0	0	0	0	0
SL	0	0	0	0	0	0	0	0	0	0

Table A.34: Number of pairwise crosstalks in PhI, Category II

	ABA	IAA	BR	CK	ET	GA	JA	SA	KAR	SL
ABA	23	27	3	6	27	6	15	11	0	0
IAA	0	21	8	7	11	6	16	8	2	0
BR	0	0	1	2	0	0	0	2	0	0
CK	0	0	0	2	3	1	1	2	0	0
ET	0	0	0	0	12	9	15	14	0	0
GA	0	0	0	0	0	6	7	8	2	0
JA	0	0	0	0	0	0	8	10	1	0
SA	0	0	0	0	0	0	0	9	2	0
KAR	0	0	0	0	0	0	0	0	0	0
SL	0	0	0	0	0	0	0	0	0	0

Table A.35: Number of pairwise crosstalks in $\mathrm{PhI}_{\mathrm{ext}}$, Category I

	ABA	IAA	BR	CK	ET	GA	JA	SA	KAR	SL
ABA	30	11	5	7	13	6	10	12	2	0
IAA	0	42	6	12	12	10	4	9	4	0
BR	0	0	8	0	12	4	3	6	0	0
CK	0	0	0	2	10	2	13	10	0	0
ET	0	0	0	0	10	6	4	1	3	0
GA	0	0	0	0	0	2	6	5	0	0
JA	0	0	0	0	0	0	10	2	4	0
SA	0	0	0	0	0	0	0	7	3	0
KAR	0	0	0	0	0	0	0	0	0	0
SL	0	0	0	0	0	0	0	0	0	0

Table A.36: Number of pairwise crosstalks in PhI$_{ext}$, Category II

	ABA	IAA	BR	CK	ET	GA	JA	SA	KAR	SL
ABA	30	37	5	8	33	9	19	15	0	0
IAA	0	30	11	10	18	7	20	11	2	0
BR	0	0	1	2	0	0	0	2	0	0
CK	0	0	0	3	4	1	1	3	0	0
ET	0	0	0	0	14	10	16	15	0	0
GA	0	0	0	0	0	6	8	9	2	0
JA	0	0	0	0	0	0	10	14	1	0
SA	0	0	0	0	0	0	0	11	2	0
KAR	0	0	0	0	0	0	0	0	0	0
SL	0	0	0	0	0	0	0	0	0	0

A.16 PhI$_{ext}$ three node crosstalks

Table A.38: List of all three node crosstalks in PhI$_{ext}$

Interactor A	Interactor B	Interactor C	Type	Annotation A	Annotation C
AT5G47100	AT1G01140	AT4G37260	Cat II	abscisic acid	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT5G47100	AT1G01140	AT1G31880	Cat II	abscisic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G37260	AT1G01140	AT1G31880	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G08150	AT1G51660	AT4G37260	Cat I	gibberellin	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT4G08150	AT1G51660	AT1G31880	Cat I	gibberellin	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G08150	AT1G51660	AT2G43790	Cat I	gibberellin	abscisic acid, ethylene, jasmonic acid, salicylic acid
AT4G08150	AT1G51660	AT3G45640	Cat I	gibberellin	abscisic acid, salicylic acid
AT4G37260	AT1G51660	AT1G31880	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G37260	AT1G51660	AT4G32570	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT1G51660	AT2G43790	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, ethylene, jasmonic acid, salicylic acid
AT4G37260	AT1G51660	AT3G45640	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, salicylic acid
AT1G31880	AT1G51660	AT4G32570	Cat I	abscisic acid, auxin, brassinosteroid, cytokinin	jasmonic acid
AT1G31880	AT1G51660	AT2G43790	Cat II	abscisic acid, auxin, brassinosteroid, cytokinin	abscisic acid, ethylene, jasmonic acid, salicylic acid
AT1G31880	AT1G51660	AT3G45640	Cat II	abscisic acid, auxin, brassinosteroid, cytokinin	abscisic acid, salicylic acid
AT4G32570	AT1G51660	AT2G43790	Cat II	jasmonic acid	abscisic acid, ethylene, jasmonic acid, salicylic acid
AT4G32570	AT1G51660	AT3G45640	Cat I	jasmonic acid	abscisic acid, salicylic acid
AT2G43790	AT1G51660	AT3G45640	Cat II	abscisic acid, ethylene, jasmonic acid, salicylic acid	abscisic acid, salicylic acid
AT2G01570	AT1G53510	AT1G31880	Cat II	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT1G31880	AT1G53510	AT1G14920	Cat II	abscisic acid, auxin, brassinosteroid, cytokinin	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT1G31880	AT1G53510	AT3G03450	Cat II	abscisic acid, auxin, brassinosteroid, cytokinin	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT4G08150	AT1G73410	AT4G37260	Cat I	gibberellin	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT4G08150	AT1G73410	AT1G24590	Cat I	gibberellin	auxin, brassinosteroid
AT4G08150	AT1G73410	AT4G30080	Cat I	gibberellin	auxin
AT4G08150	AT1G73410	AT5G25190	Cat I	gibberellin	ethylene
AT4G37260	AT1G73410	AT1G24590	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	auxin, brassinosteroid
AT4G37260	AT1G73410	AT4G30080	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	auxin
AT4G37260	AT1G73410	AT5G25190	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	ethylene
AT1G24590	AT1G73410	AT4G30080	Cat II	auxin, brassinosteroid	auxin
AT1G24590	AT1G73410	AT5G25190	Cat I	auxin, brassinosteroid	ethylene

Table A.38 continued from previous page

Interactor A	Interactor B	Interactor C	Type	Annotation A	Annotation C
AT4G30080	AT1G73410	AT5G25190	Cat I	auxin	ethylene
AT1G22770	AT2G23290	AT2G01760	Cat I	gibberellin, karrikin	cytokinin
AT1G22770	AT2G23290	AT5G67300	Cat II	gibberellin, karrikin	abscisic acid, auxin, ethylene, gibberellin, jasmonic acid, salicylic acid
AT1G22770	AT2G23290	AT5G53160	Cat I	gibberellin, karrikin	abscisic acid
AT1G22770	AT2G23290	AT5G58950	Cat I	gibberellin, karrikin	ethylene
AT2G01760	AT2G23290	AT5G67300	Cat I	cytokinin	abscisic acid, auxin, ethylene, gibberellin, jasmonic acid, salicylic acid
AT2G01760	AT2G23290	AT5G53160	Cat I	cytokinin	abscisic acid
AT2G01760	AT2G23290	AT5G58950	Cat I	cytokinin	ethylene
AT5G67300	AT2G23290	AT5G58950	Cat II	abscisic acid, auxin, ethylene, gibberellin, jasmonic acid, salicylic acid	ethylene
AT5G53160	AT2G23290	AT5G58950	Cat I	abscisic acid	ethylene
AT1G01360	AT2G29380	AT4G37260	Cat II	abscisic acid	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT5G45830	AT2G29380	AT4G37260	Cat II	abscisic acid	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT4G37260	AT2G29380	AT4G27920	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid
AT4G37260	AT2G29380	AT4G01026	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid
AT4G37260	AT2G29380	AT5G45860	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid
AT2G01570	AT2G30590	AT3G62100	Cat I	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	auxin
AT2G01570	AT2G30590	AT4G37260	Cat II	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT2G01570	AT2G30590	AT2G44050	Cat II	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	jasmonic acid
AT2G01570	AT2G30590	AT2G46990	Cat I	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	auxin
AT3G62100	AT2G30590	AT4G37260	Cat II	auxin	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT3G62100	AT2G30590	AT2G44050	Cat I	auxin	jasmonic acid
AT4G37260	AT2G30590	AT2G44050	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT2G30590	AT2G46990	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	auxin
AT2G44050	AT2G30590	AT2G46990	Cat I	jasmonic acid	auxin
AT3G17600	AT2G38490	AT2G44050	Cat I	auxin	jasmonic acid
AT3G17600	AT2G38490	AT1G72360	Cat I	auxin	ethylene
AT2G44050	AT2G38490	AT1G72360	Cat I	jasmonic acid	ethylene
AT2G44050	AT2G38490	AT4G28640	Cat I	jasmonic acid	auxin
AT1G72360	AT2G38490	AT4G28640	Cat I	ethylene	auxin
AT3G62100	AT4G11070	AT4G37260	Cat II	auxin	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT3G62100	AT4G11070	AT5G22570	Cat I	auxin	salicylic acid
AT3G62100	AT4G11070	AT1G24590	Cat II	auxin	auxin, brassinosteroid
AT4G37260	AT4G11070	AT5G22570	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	salicylic acid
AT4G37260	AT4G11070	AT1G24590	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	auxin, brassinosteroid
AT5G22570	AT4G11070	AT1G24590	Cat I	salicylic acid	auxin, brassinosteroid
AT4G28910	AT4G14713	AT5G17690	Cat I	jasmonic acid	gibberellin
AT4G28910	AT4G14713	AT5G63470	Cat I	jasmonic acid	abscisic acid
AT5G17690	AT4G14713	AT4G32570	Cat I	gibberellin	jasmonic acid
AT5G17690	AT4G14713	AT4G14720	Cat I	gibberellin	jasmonic acid
AT5G17690	AT4G14713	AT5G63470	Cat I	gibberellin	abscisic acid
AT4G32570	AT4G14713	AT5G63470	Cat I	jasmonic acid	abscisic acid
AT1G15550	AT5G60120	AT1G22770	Cat II	gibberellin	gibberellin, karrikin
AT1G15550	AT5G60120	AT1G37130	Cat I	gibberellin	salicylic acid
AT1G15550	AT5G60120	AT1G54490	Cat I	gibberellin	ethylene
AT1G15550	AT5G60120	AT5G08130	Cat I	gibberellin	brassinosteroid
AT1G22770	AT5G60120	AT1G37130	Cat I	gibberellin, karrikin	salicylic acid
AT1G22770	AT5G60120	AT1G54490	Cat I	gibberellin, karrikin	ethylene
AT1G22770	AT5G60120	AT1G31880	Cat I	gibberellin, karrikin	abscisic acid, auxin, brassinosteroid, cytokinin
AT1G22770	AT5G60120	AT1G24590	Cat I	gibberellin, karrikin	auxin, brassinosteroid
AT1G22770	AT5G60120	AT5G08130	Cat I	gibberellin, karrikin	brassinosteroid
AT1G37130	AT5G60120	AT1G54490	Cat I	salicylic acid	ethylene
AT1G37130	AT5G60120	AT4G08150	Cat I	salicylic acid	gibberellin
AT1G37130	AT5G60120	AT1G24590	Cat I	salicylic acid	auxin, brassinosteroid
AT1G37130	AT5G60120	AT5G08130	Cat I	salicylic acid	brassinosteroid
AT1G54490	AT5G60120	AT4G08150	Cat I	ethylene	gibberellin
AT1G54490	AT5G60120	AT1G24590	Cat I	ethylene	auxin, brassinosteroid
AT1G54490	AT5G60120	AT5G08130	Cat I	ethylene	brassinosteroid
AT4G08150	AT5G60120	AT1G31880	Cat I	gibberellin	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G08150	AT5G60120	AT1G24590	Cat I	gibberellin	auxin, brassinosteroid
AT4G08150	AT5G60120	AT5G08130	Cat I	gibberellin	brassinosteroid
AT1G31880	AT5G60120	AT1G24590	Cat II	abscisic acid, auxin, brassinosteroid, cytokinin	auxin, brassinosteroid
AT1G24590	AT5G60120	AT5G08130	Cat II	auxin, brassinosteroid	brassinosteroid

Table A.38 continued from previous page

Interactor A	Interactor B	Interactor C	Type	Annotation A	Annotation C
AT1G15550	AT4G32010	AT2G25490	Cat I	gibberellin	ethylene
AT1G15550	AT4G32010	AT1G18350	Cat I	gibberellin	auxin, salicylic acid
AT1G15550	AT4G32010	AT5G17490	Cat II	gibberellin	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT1G15550	AT4G32010	AT3G02140	Cat I	gibberellin	abscisic acid
AT1G15550	AT4G32010	AT2G35940	Cat I	gibberellin	abscisic acid
AT1G15550	AT4G32010	AT5G51760	Cat I	gibberellin	abscisic acid
AT1G15550	AT4G32010	AT5G55170	Cat I	gibberellin	salicylic acid
AT1G15550	AT4G32010	AT5G39340	Cat I	gibberellin	cytokinin
AT1G15550	AT4G32010	AT5G55620	Cat I	gibberellin	ethylene
AT2G25490	AT4G32010	AT1G18350	Cat I	ethylene	auxin, salicylic acid
AT2G25490	AT4G32010	AT5G17490	Cat II	ethylene	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT2G25490	AT4G32010	AT3G02140	Cat I	ethylene	abscisic acid
AT2G25490	AT4G32010	AT2G35940	Cat I	ethylene	abscisic acid
AT2G25490	AT4G32010	AT5G51760	Cat I	ethylene	abscisic acid
AT2G25490	AT4G32010	AT5G55170	Cat I	ethylene	salicylic acid
AT2G25490	AT4G32010	AT5G39340	Cat I	ethylene	cytokinin
AT4G08150	AT4G32010	AT1G18350	Cat I	gibberellin	auxin, salicylic acid
AT4G08150	AT4G32010	AT3G02140	Cat I	gibberellin	abscisic acid
AT4G08150	AT4G32010	AT2G35940	Cat I	gibberellin	abscisic acid
AT4G08150	AT4G32010	AT5G51760	Cat I	gibberellin	abscisic acid
AT4G08150	AT4G32010	AT5G55170	Cat I	gibberellin	salicylic acid
AT4G08150	AT4G32010	AT5G39340	Cat I	gibberellin	cytokinin
AT4G08150	AT4G32010	AT5G55620	Cat I	gibberellin	ethylene
AT1G18350	AT4G32010	AT5G17490	Cat II	auxin, salicylic acid	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT1G18350	AT4G32010	AT3G02140	Cat I	auxin, salicylic acid	abscisic acid
AT1G18350	AT4G32010	AT2G35940	Cat I	auxin, salicylic acid	abscisic acid
AT1G18350	AT4G32010	AT5G51760	Cat I	auxin, salicylic acid	abscisic acid
AT1G18350	AT4G32010	AT5G55170	Cat II	auxin, salicylic acid	salicylic acid
AT1G18350	AT4G32010	AT5G39340	Cat I	auxin, salicylic acid	cytokinin
AT1G18350	AT4G32010	AT5G55620	Cat I	auxin, salicylic acid	ethylene
AT5G17490	AT4G32010	AT3G02140	Cat II	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	abscisic acid
AT5G17490	AT4G32010	AT2G35940	Cat II	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	abscisic acid
AT5G17490	AT4G32010	AT5G51760	Cat II	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	abscisic acid
AT5G17490	AT4G32010	AT5G55170	Cat II	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	salicylic acid
AT5G17490	AT4G32010	AT5G39340	Cat I	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	cytokinin
AT5G17490	AT4G32010	AT5G55620	Cat II	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	ethylene
AT3G02140	AT4G32010	AT5G55170	Cat I	abscisic acid	salicylic acid
AT3G02140	AT4G32010	AT5G39340	Cat I	abscisic acid	cytokinin
AT3G02140	AT4G32010	AT5G55620	Cat I	abscisic acid	ethylene
AT2G35940	AT4G32010	AT5G55170	Cat I	abscisic acid	salicylic acid
AT2G35940	AT4G32010	AT5G39340	Cat I	abscisic acid	cytokinin
AT2G35940	AT4G32010	AT5G55620	Cat I	abscisic acid	ethylene
AT5G51760	AT4G32010	AT5G55170	Cat I	abscisic acid	salicylic acid
AT5G51760	AT4G32010	AT5G39340	Cat I	abscisic acid	cytokinin
AT5G51760	AT4G32010	AT5G55620	Cat I	abscisic acid	ethylene
AT5G55170	AT4G32010	AT5G39340	Cat I	salicylic acid	cytokinin
AT5G55170	AT4G32010	AT5G55620	Cat I	salicylic acid	ethylene
AT5G39340	AT4G32010	AT5G55620	Cat I	cytokinin	ethylene
AT3G29350	AT3G23610	AT2G04550	Cat I	cytokinin	abscisic acid, auxin, ethylene
AT3G29350	AT3G23610	AT4G29810	Cat I	cytokinin	auxin
AT3G29350	AT3G23610	AT3G45640	Cat I	cytokinin	abscisic acid, salicylic acid
AT2G04550	AT3G23610	AT3G15150	Cat I	abscisic acid, auxin, ethylene	cytokinin
AT2G04550	AT3G23610	AT4G29810	Cat II	abscisic acid, auxin, ethylene	auxin
AT2G04550	AT3G23610	AT3G45640	Cat II	abscisic acid, auxin, ethylene	abscisic acid, salicylic acid
AT3G15150	AT3G23610	AT4G29810	Cat I	cytokinin	auxin
AT3G15150	AT3G23610	AT3G45640	Cat I	cytokinin	abscisic acid, salicylic acid
AT4G29810	AT3G23610	AT3G45640	Cat I	auxin	abscisic acid, salicylic acid
AT1G15550	AT4G36930	AT2G01760	Cat I	gibberellin	cytokinin
AT1G15550	AT4G36930	AT2G25490	Cat I	gibberellin	ethylene
AT1G15550	AT4G36930	AT1G14920	Cat II	gibberellin	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT1G15550	AT4G36930	AT4G00120	Cat I	gibberellin	auxin
AT2G01760	AT4G36930	AT2G25490	Cat I	cytokinin	ethylene
AT2G01760	AT4G36930	AT1G14920	Cat I	cytokinin	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT2G01760	AT4G36930	AT1G24590	Cat I	cytokinin	auxin, brassinosteroid
AT2G25490	AT4G36930	AT1G14920	Cat II	ethylene	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT2G25490	AT4G36930	AT4G00120	Cat I	ethylene	auxin

Table A.38 continued from previous page

Interactor A	Interactor B	Interactor C	Type	Annotation A	Annotation C
AT2G25490	AT4G36930	AT1G24590	Cat I	ethylene	auxin, brassinosteroid
AT1G14920	AT4G36930	AT4G00120	Cat I	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	auxin
AT1G14920	AT4G36930	AT1G24590	Cat I	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	auxin, brassinosteroid
AT4G00120	AT4G36930	AT1G24590	Cat II	auxin	auxin, brassinosteroid
AT1G15550	AT1G35560	AT1G18400	Cat I	gibberellin	brassinosteroid
AT1G15550	AT1G35560	AT1G53170	Cat I	gibberellin	ethylene
AT1G15550	AT1G35560	AT1G64520	Cat I	gibberellin	auxin, cytokinin
AT1G15550	AT1G35560	AT2G01760	Cat I	gibberellin	cytokinin
AT1G15550	AT1G35560	AT2G25490	Cat I	gibberellin	ethylene
AT1G15550	AT1G35560	AT2G26980	Cat I	gibberellin	abscisic acid, cytokinin
AT1G15550	AT1G35560	AT3G29350	Cat I	gibberellin	cytokinin
AT1G15550	AT1G35560	AT2G04550	Cat I	gibberellin	abscisic acid, auxin, ethylene
AT1G15550	AT1G35560	AT1G14280	Cat I	gibberellin	auxin
AT1G15550	AT1G35560	AT1G28360	Cat I	gibberellin	ethylene
AT1G15550	AT1G35560	AT1G10210	Cat I	gibberellin	auxin
AT1G15550	AT1G35560	AT1G18350	Cat I	gibberellin	auxin, salicylic acid
AT1G15550	AT1G35560	AT3G15150	Cat I	gibberellin	cytokinin
AT1G15550	AT1G35560	AT5G05440	Cat I	gibberellin	abscisic acid
AT1G15550	AT1G35560	AT4G29810	Cat I	gibberellin	auxin
AT1G15550	AT1G35560	AT3G45640	Cat I	gibberellin	abscisic acid, salicylic acid
AT1G15550	AT1G35560	AT3G24520	Cat I	gibberellin	abscisic acid
AT1G15550	AT1G35560	AT5G51760	Cat I	gibberellin	abscisic acid
AT1G15550	AT1G35560	AT3G57040	Cat I	gibberellin	cytokinin
AT1G15550	AT1G35560	AT5G62000	Cat I	gibberellin	abscisic acid, auxin, brassinosteroid
AT1G15550	AT1G35560	AT5G55170	Cat I	gibberellin	salicylic acid
AT1G15550	AT1G35560	AT4G36540	Cat I	gibberellin	brassinosteroid
AT1G18400	AT1G35560	AT1G53170	Cat I	brassinosteroid	ethylene
AT1G18400	AT1G35560	AT1G64520	Cat I	brassinosteroid	auxin, cytokinin
AT1G18400	AT1G35560	AT2G01760	Cat I	brassinosteroid	cytokinin
AT1G18400	AT1G35560	AT2G25490	Cat I	brassinosteroid	ethylene
AT1G18400	AT1G35560	AT2G26980	Cat I	brassinosteroid	abscisic acid, cytokinin
AT1G18400	AT1G35560	AT4G08150	Cat I	brassinosteroid	gibberellin
AT1G18400	AT1G35560	AT3G29350	Cat I	brassinosteroid	cytokinin
AT1G18400	AT1G35560	AT2G04550	Cat I	brassinosteroid	abscisic acid, auxin, ethylene
AT1G18400	AT1G35560	AT1G14280	Cat I	brassinosteroid	auxin
AT1G18400	AT1G35560	AT1G28360	Cat I	brassinosteroid	ethylene
AT1G18400	AT1G35560	AT1G10210	Cat I	brassinosteroid	auxin
AT1G18400	AT1G35560	AT1G18350	Cat I	brassinosteroid	auxin, salicylic acid
AT1G18400	AT1G35560	AT3G15150	Cat I	brassinosteroid	cytokinin
AT1G18400	AT1G35560	AT5G05440	Cat I	brassinosteroid	abscisic acid
AT1G18400	AT1G35560	AT4G29810	Cat I	brassinosteroid	auxin
AT1G18400	AT1G35560	AT3G45640	Cat I	brassinosteroid	abscisic acid, salicylic acid
AT1G18400	AT1G35560	AT3G24520	Cat I	brassinosteroid	abscisic acid
AT1G18400	AT1G35560	AT5G51760	Cat I	brassinosteroid	abscisic acid
AT1G18400	AT1G35560	AT3G57040	Cat I	brassinosteroid	cytokinin
AT1G18400	AT1G35560	AT5G62000	Cat II	brassinosteroid	abscisic acid, auxin, brassinosteroid
AT1G18400	AT1G35560	AT5G55170	Cat I	brassinosteroid	salicylic acid
AT1G18400	AT1G35560	AT1G80340	Cat I	brassinosteroid	gibberellin
AT1G53170	AT1G35560	AT1G64520	Cat I	ethylene	auxin, cytokinin
AT1G53170	AT1G35560	AT2G01760	Cat I	ethylene	cytokinin
AT1G53170	AT1G35560	AT2G26980	Cat I	ethylene	abscisic acid, cytokinin
AT1G53170	AT1G35560	AT4G08150	Cat I	ethylene	gibberellin
AT1G53170	AT1G35560	AT3G29350	Cat I	ethylene	cytokinin
AT1G53170	AT1G35560	AT2G04550	Cat II	ethylene	abscisic acid, auxin, ethylene
AT1G53170	AT1G35560	AT1G14280	Cat I	ethylene	auxin
AT1G53170	AT1G35560	AT1G10210	Cat I	ethylene	auxin
AT1G53170	AT1G35560	AT1G18350	Cat I	ethylene	auxin, salicylic acid
AT1G53170	AT1G35560	AT3G15150	Cat I	ethylene	cytokinin
AT1G53170	AT1G35560	AT5G05440	Cat I	ethylene	abscisic acid
AT1G53170	AT1G35560	AT4G29810	Cat I	ethylene	auxin
AT1G53170	AT1G35560	AT3G45640	Cat I	ethylene	abscisic acid, salicylic acid
AT1G53170	AT1G35560	AT3G24520	Cat I	ethylene	abscisic acid
AT1G53170	AT1G35560	AT5G51760	Cat I	ethylene	abscisic acid
AT1G53170	AT1G35560	AT3G57040	Cat I	ethylene	cytokinin
AT1G53170	AT1G35560	AT5G62000	Cat I	ethylene	abscisic acid, auxin, brassinosteroid
AT1G53170	AT1G35560	AT5G55170	Cat I	ethylene	salicylic acid
AT1G53170	AT1G35560	AT4G36540	Cat I	ethylene	brassinosteroid
AT1G53170	AT1G35560	AT1G80340	Cat I	ethylene	gibberellin
AT1G64520	AT1G35560	AT2G01760	Cat II	auxin, cytokinin	cytokinin
AT1G64520	AT1G35560	AT2G25490	Cat I	auxin, cytokinin	ethylene
AT1G64520	AT1G35560	AT2G26980	Cat II	auxin, cytokinin	abscisic acid, cytokinin
AT1G64520	AT1G35560	AT3G29350	Cat II	auxin, cytokinin	cytokinin
AT1G64520	AT1G35560	AT2G04550	Cat II	auxin, cytokinin	abscisic acid, auxin, ethylene
AT1G64520	AT1G35560	AT1G14280	Cat II	auxin, cytokinin	auxin
AT1G64520	AT1G35560	AT1G28360	Cat I	auxin, cytokinin	ethylene
AT1G64520	AT1G35560	AT1G10210	Cat II	auxin, cytokinin	auxin
AT1G64520	AT1G35560	AT1G18350	Cat II	auxin, cytokinin	auxin, salicylic acid
AT1G64520	AT1G35560	AT3G15150	Cat II	auxin, cytokinin	cytokinin
AT1G64520	AT1G35560	AT5G05440	Cat I	auxin, cytokinin	abscisic acid
AT1G64520	AT1G35560	AT4G29810	Cat II	auxin, cytokinin	auxin
AT1G64520	AT1G35560	AT3G45640	Cat I	auxin, cytokinin	abscisic acid, salicylic acid
AT1G64520	AT1G35560	AT3G24520	Cat I	auxin, cytokinin	abscisic acid
AT1G64520	AT1G35560	AT5G51760	Cat I	auxin, cytokinin	abscisic acid

Table A.38 continued from previous page

Interactor A	Interactor B	Interactor C	Type	Annotation A	Annotation C
AT1G64520	AT1G35560	AT3G57040	Cat II	auxin, cytokinin	cytokinin
AT1G64520	AT1G35560	AT5G62000	Cat II	auxin, cytokinin	abscisic acid, auxin, brassinosteroid
AT1G64520	AT1G35560	AT5G55170	Cat I	auxin, cytokinin	salicylic acid
AT1G64520	AT1G35560	AT4G36540	Cat I	auxin, cytokinin	brassinosteroid
AT1G64520	AT1G35560	AT1G80340	Cat I	auxin, cytokinin	gibberellin
AT2G01760	AT1G35560	AT2G25490	Cat I	cytokinin	ethylene
AT2G01760	AT1G35560	AT2G26980	Cat II	cytokinin	abscisic acid, cytokinin
AT2G01760	AT1G35560	AT4G08150	Cat I	cytokinin	gibberellin
AT2G01760	AT1G35560	AT2G04550	Cat I	cytokinin	abscisic acid, auxin, ethylene
AT2G01760	AT1G35560	AT1G14280	Cat I	cytokinin	auxin
AT2G01760	AT1G35560	AT1G28360	Cat I	cytokinin	ethylene
AT2G01760	AT1G35560	AT1G10210	Cat I	cytokinin	auxin
AT2G01760	AT1G35560	AT1G18350	Cat I	cytokinin	auxin, salicylic acid
AT2G01760	AT1G35560	AT5G05440	Cat I	cytokinin	abscisic acid
AT2G01760	AT1G35560	AT4G29810	Cat I	cytokinin	auxin
AT2G01760	AT1G35560	AT3G45640	Cat I	cytokinin	abscisic acid, salicylic acid
AT2G01760	AT1G35560	AT3G24520	Cat I	cytokinin	abscisic acid
AT2G01760	AT1G35560	AT5G51760	Cat I	cytokinin	abscisic acid
AT2G01760	AT1G35560	AT5G62000	Cat I	cytokinin	abscisic acid, auxin, brassinosteroid
AT2G01760	AT1G35560	AT5G55170	Cat I	cytokinin	salicylic acid
AT2G01760	AT1G35560	AT4G36540	Cat I	cytokinin	brassinosteroid
AT2G01760	AT1G35560	AT1G80340	Cat I	cytokinin	gibberellin
AT2G25490	AT1G35560	AT2G26980	Cat I	ethylene	abscisic acid, cytokinin
AT2G25490	AT1G35560	AT3G29350	Cat I	ethylene	cytokinin
AT2G25490	AT1G35560	AT2G04550	Cat II	ethylene	abscisic acid, auxin, ethylene
AT2G25490	AT1G35560	AT1G14280	Cat I	ethylene	auxin
AT2G25490	AT1G35560	AT1G10210	Cat I	ethylene	auxin
AT2G25490	AT1G35560	AT1G18350	Cat I	ethylene	auxin, salicylic acid
AT2G25490	AT1G35560	AT3G15150	Cat I	ethylene	cytokinin
AT2G25490	AT1G35560	AT5G05440	Cat I	ethylene	abscisic acid
AT2G25490	AT1G35560	AT4G29810	Cat I	ethylene	auxin
AT2G25490	AT1G35560	AT3G45640	Cat I	ethylene	abscisic acid, salicylic acid
AT2G25490	AT1G35560	AT3G24520	Cat I	ethylene	abscisic acid
AT2G25490	AT1G35560	AT5G51760	Cat I	ethylene	abscisic acid
AT2G25490	AT1G35560	AT3G57040	Cat I	ethylene	cytokinin
AT2G25490	AT1G35560	AT5G62000	Cat I	ethylene	abscisic acid, auxin, brassinosteroid
AT2G25490	AT1G35560	AT5G55170	Cat I	ethylene	salicylic acid
AT2G25490	AT1G35560	AT4G36540	Cat I	ethylene	brassinosteroid
AT2G25490	AT1G35560	AT1G80340	Cat I	ethylene	gibberellin
AT2G26980	AT1G35560	AT4G08150	Cat I	abscisic acid, cytokinin	gibberellin
AT2G26980	AT1G35560	AT3G29350	Cat II	abscisic acid, cytokinin	cytokinin
AT2G26980	AT1G35560	AT2G04550	Cat II	abscisic acid, cytokinin	abscisic acid, auxin, ethylene
AT2G26980	AT1G35560	AT1G14280	Cat I	abscisic acid, cytokinin	auxin
AT2G26980	AT1G35560	AT1G28360	Cat I	abscisic acid, cytokinin	ethylene
AT2G26980	AT1G35560	AT1G10210	Cat I	abscisic acid, cytokinin	auxin
AT2G26980	AT1G35560	AT1G18350	Cat I	abscisic acid, cytokinin	auxin, salicylic acid
AT2G26980	AT1G35560	AT3G15150	Cat II	abscisic acid, cytokinin	cytokinin
AT2G26980	AT1G35560	AT5G05440	Cat II	abscisic acid, cytokinin	abscisic acid
AT2G26980	AT1G35560	AT4G29810	Cat I	abscisic acid, cytokinin	auxin
AT2G26980	AT1G35560	AT3G45640	Cat II	abscisic acid, cytokinin	abscisic acid, salicylic acid
AT2G26980	AT1G35560	AT3G24520	Cat II	abscisic acid, cytokinin	abscisic acid
AT2G26980	AT1G35560	AT5G51760	Cat II	abscisic acid, cytokinin	abscisic acid
AT2G26980	AT1G35560	AT3G57040	Cat II	abscisic acid, cytokinin	cytokinin
AT2G26980	AT1G35560	AT5G62000	Cat II	abscisic acid, cytokinin	abscisic acid, auxin, brassinosteroid
AT2G26980	AT1G35560	AT5G55170	Cat I	abscisic acid, cytokinin	salicylic acid
AT2G26980	AT1G35560	AT4G36540	Cat I	abscisic acid, cytokinin	brassinosteroid
AT2G26980	AT1G35560	AT1G80340	Cat I	abscisic acid, cytokinin	gibberellin
AT4G08150	AT1G35560	AT3G29350	Cat I	gibberellin	cytokinin
AT4G08150	AT1G35560	AT2G04550	Cat I	gibberellin	abscisic acid, auxin, ethylene
AT4G08150	AT1G35560	AT1G14280	Cat I	gibberellin	auxin
AT4G08150	AT1G35560	AT1G28360	Cat I	gibberellin	ethylene
AT4G08150	AT1G35560	AT1G10210	Cat I	gibberellin	auxin
AT4G08150	AT1G35560	AT1G18350	Cat I	gibberellin	auxin, salicylic acid
AT4G08150	AT1G35560	AT3G15150	Cat I	gibberellin	cytokinin
AT4G08150	AT1G35560	AT5G05440	Cat I	gibberellin	abscisic acid
AT4G08150	AT1G35560	AT4G29810	Cat I	gibberellin	auxin
AT4G08150	AT1G35560	AT3G45640	Cat I	gibberellin	abscisic acid, salicylic acid
AT4G08150	AT1G35560	AT3G24520	Cat I	gibberellin	abscisic acid
AT4G08150	AT1G35560	AT5G51760	Cat I	gibberellin	abscisic acid
AT4G08150	AT1G35560	AT3G57040	Cat I	gibberellin	cytokinin
AT4G08150	AT1G35560	AT5G62000	Cat I	gibberellin	abscisic acid, auxin, brassinosteroid
AT4G08150	AT1G35560	AT5G55170	Cat I	gibberellin	salicylic acid
AT4G08150	AT1G35560	AT4G36540	Cat I	gibberellin	brassinosteroid
AT3G29350	AT1G35560	AT2G04550	Cat I	cytokinin	abscisic acid, auxin, ethylene
AT3G29350	AT1G35560	AT1G14280	Cat I	cytokinin	auxin
AT3G29350	AT1G35560	AT1G28360	Cat I	cytokinin	ethylene
AT3G29350	AT1G35560	AT1G10210	Cat I	cytokinin	auxin
AT3G29350	AT1G35560	AT1G18350	Cat I	cytokinin	auxin, salicylic acid
AT3G29350	AT1G35560	AT5G05440	Cat I	cytokinin	abscisic acid
AT3G29350	AT1G35560	AT4G29810	Cat I	cytokinin	auxin
AT3G29350	AT1G35560	AT3G45640	Cat I	cytokinin	abscisic acid, salicylic acid
AT3G29350	AT1G35560	AT3G24520	Cat I	cytokinin	abscisic acid
AT3G29350	AT1G35560	AT5G51760	Cat I	cytokinin	abscisic acid
AT3G29350	AT1G35560	AT5G62000	Cat I	cytokinin	abscisic acid, auxin, brassinosteroid

Table A.38 continued from previous page

Interactor A	Interactor B	Interactor C	Type	Annotation A	Annotation C
AT3G29350	AT1G35560	AT5G55170	Cat I	cytokinin	salicylic acid
AT3G29350	AT1G35560	AT4G36540	Cat I	cytokinin	brassinosteroid
AT3G29350	AT1G35560	AT1G80340	Cat I	cytokinin	gibberellin
AT2G04550	AT1G35560	AT1G14280	Cat II	abscisic acid, auxin, ethylene	auxin
AT2G04550	AT1G35560	AT1G28360	Cat II	abscisic acid, auxin, ethylene	ethylene
AT2G04550	AT1G35560	AT1G10210	Cat II	abscisic acid, auxin, ethylene	auxin
AT2G04550	AT1G35560	AT1G18350	Cat II	abscisic acid, auxin, ethylene	auxin, salicylic acid
AT2G04550	AT1G35560	AT3G15150	Cat I	abscisic acid, auxin, ethylene	cytokinin
AT2G04550	AT1G35560	AT5G05440	Cat II	abscisic acid, auxin, ethylene	abscisic acid
AT2G04550	AT1G35560	AT4G29810	Cat II	abscisic acid, auxin, ethylene	auxin
AT2G04550	AT1G35560	AT3G45640	Cat II	abscisic acid, auxin, ethylene	abscisic acid, salicylic acid
AT2G04550	AT1G35560	AT3G24520	Cat II	abscisic acid, auxin, ethylene	abscisic acid
AT2G04550	AT1G35560	AT5G51760	Cat II	abscisic acid, auxin, ethylene	abscisic acid
AT2G04550	AT1G35560	AT3G57040	Cat I	abscisic acid, auxin, ethylene	cytokinin
AT2G04550	AT1G35560	AT5G62000	Cat II	abscisic acid, auxin, ethylene	abscisic acid, auxin, brassinosteroid
AT2G04550	AT1G35560	AT5G55170	Cat I	abscisic acid, auxin, ethylene	salicylic acid
AT2G04550	AT1G35560	AT4G36540	Cat I	abscisic acid, auxin, ethylene	brassinosteroid
AT2G04550	AT1G35560	AT1G80340	Cat I	abscisic acid, auxin, ethylene	gibberellin
AT1G14280	AT1G35560	AT1G28360	Cat I	auxin	ethylene
AT1G14280	AT1G35560	AT1G18350	Cat II	auxin	auxin, salicylic acid
AT1G14280	AT1G35560	AT3G15150	Cat I	auxin	cytokinin
AT1G14280	AT1G35560	AT5G05440	Cat I	auxin	abscisic acid
AT1G14280	AT1G35560	AT3G45640	Cat I	auxin	abscisic acid, salicylic acid
AT1G14280	AT1G35560	AT3G24520	Cat I	auxin	abscisic acid
AT1G14280	AT1G35560	AT5G51760	Cat I	auxin	abscisic acid
AT1G14280	AT1G35560	AT3G57040	Cat I	auxin	cytokinin
AT1G14280	AT1G35560	AT5G62000	Cat II	auxin	abscisic acid, auxin, brassinosteroid
AT1G14280	AT1G35560	AT5G55170	Cat I	auxin	salicylic acid
AT1G14280	AT1G35560	AT4G36540	Cat I	auxin	brassinosteroid
AT1G14280	AT1G35560	AT1G80340	Cat I	auxin	gibberellin
AT1G28360	AT1G35560	AT1G10210	Cat I	ethylene	auxin
AT1G28360	AT1G35560	AT1G18350	Cat I	ethylene	auxin, salicylic acid
AT1G28360	AT1G35560	AT3G15150	Cat I	ethylene	cytokinin
AT1G28360	AT1G35560	AT5G05440	Cat I	ethylene	abscisic acid
AT1G28360	AT1G35560	AT4G29810	Cat I	ethylene	auxin
AT1G28360	AT1G35560	AT3G45640	Cat I	ethylene	abscisic acid, salicylic acid
AT1G28360	AT1G35560	AT3G24520	Cat I	ethylene	abscisic acid
AT1G28360	AT1G35560	AT5G51760	Cat I	ethylene	abscisic acid
AT1G28360	AT1G35560	AT3G57040	Cat I	ethylene	cytokinin
AT1G28360	AT1G35560	AT5G62000	Cat I	ethylene	abscisic acid, auxin, brassinosteroid
AT1G28360	AT1G35560	AT5G55170	Cat I	ethylene	salicylic acid
AT1G28360	AT1G35560	AT4G36540	Cat I	ethylene	brassinosteroid
AT1G28360	AT1G35560	AT1G80340	Cat I	ethylene	gibberellin
AT1G10210	AT1G35560	AT1G18350	Cat II	auxin	auxin, salicylic acid
AT1G10210	AT1G35560	AT3G15150	Cat I	auxin	cytokinin
AT1G10210	AT1G35560	AT5G05440	Cat I	auxin	abscisic acid
AT1G10210	AT1G35560	AT3G45640	Cat I	auxin	abscisic acid, salicylic acid
AT1G10210	AT1G35560	AT3G24520	Cat I	auxin	abscisic acid
AT1G10210	AT1G35560	AT5G51760	Cat I	auxin	abscisic acid
AT1G10210	AT1G35560	AT3G57040	Cat I	auxin	cytokinin
AT1G10210	AT1G35560	AT5G62000	Cat II	auxin	abscisic acid, auxin, brassinosteroid
AT1G10210	AT1G35560	AT5G55170	Cat I	auxin	salicylic acid
AT1G10210	AT1G35560	AT4G36540	Cat I	auxin	brassinosteroid
AT1G10210	AT1G35560	AT1G80340	Cat I	auxin	gibberellin
AT1G18350	AT1G35560	AT3G15150	Cat I	auxin, salicylic acid	cytokinin
AT1G18350	AT1G35560	AT5G05440	Cat I	auxin, salicylic acid	abscisic acid
AT1G18350	AT1G35560	AT4G29810	Cat II	auxin, salicylic acid	auxin
AT1G18350	AT1G35560	AT3G45640	Cat II	auxin, salicylic acid	abscisic acid, salicylic acid
AT1G18350	AT1G35560	AT3G24520	Cat I	auxin, salicylic acid	abscisic acid
AT1G18350	AT1G35560	AT5G51760	Cat I	auxin, salicylic acid	abscisic acid
AT1G18350	AT1G35560	AT3G57040	Cat I	auxin, salicylic acid	cytokinin
AT1G18350	AT1G35560	AT5G62000	Cat II	auxin, salicylic acid	abscisic acid, auxin, brassinosteroid
AT1G18350	AT1G35560	AT5G55170	Cat II	auxin, salicylic acid	salicylic acid
AT1G18350	AT1G35560	AT4G36540	Cat I	auxin, salicylic acid	brassinosteroid
AT1G18350	AT1G35560	AT1G80340	Cat I	auxin, salicylic acid	gibberellin
AT3G15150	AT1G35560	AT5G05440	Cat I	cytokinin	abscisic acid
AT3G15150	AT1G35560	AT4G29810	Cat I	cytokinin	auxin
AT3G15150	AT1G35560	AT3G45640	Cat I	cytokinin	abscisic acid, salicylic acid
AT3G15150	AT1G35560	AT3G24520	Cat I	cytokinin	abscisic acid
AT3G15150	AT1G35560	AT5G51760	Cat I	cytokinin	abscisic acid
AT3G15150	AT1G35560	AT5G62000	Cat I	cytokinin	abscisic acid, auxin, brassinosteroid
AT3G15150	AT1G35560	AT5G55170	Cat I	cytokinin	salicylic acid
AT3G15150	AT1G35560	AT4G36540	Cat I	cytokinin	brassinosteroid
AT3G15150	AT1G35560	AT1G80340	Cat I	cytokinin	gibberellin
AT5G05440	AT1G35560	AT4G29810	Cat I	abscisic acid	auxin
AT5G05440	AT1G35560	AT3G45640	Cat II	abscisic acid	abscisic acid, salicylic acid
AT5G05440	AT1G35560	AT3G57040	Cat I	abscisic acid	cytokinin
AT5G05440	AT1G35560	AT5G62000	Cat II	abscisic acid	abscisic acid, auxin, brassinosteroid
AT5G05440	AT1G35560	AT5G55170	Cat I	abscisic acid	salicylic acid
AT5G05440	AT1G35560	AT4G36540	Cat I	abscisic acid	brassinosteroid
AT5G05440	AT1G35560	AT1G80340	Cat I	abscisic acid	gibberellin
AT4G29810	AT1G35560	AT3G45640	Cat I	auxin	abscisic acid, salicylic acid
AT4G29810	AT1G35560	AT3G24520	Cat I	auxin	abscisic acid

Table A.38 continued from previous page

Interactor A	Interactor B	Interactor C	Type	Annotation A	Annotation C
AT4G29810	AT1G35560	AT5G51760	Cat I	auxin	abscisic acid
AT4G29810	AT1G35560	AT3G57040	Cat I	auxin	cytokinin
AT4G29810	AT1G35560	AT5G62000	Cat II	auxin	abscisic acid, auxin, brassinosteroid
AT4G29810	AT1G35560	AT5G55170	Cat I	auxin	salicylic acid
AT4G29810	AT1G35560	AT4G36540	Cat I	auxin	brassinosteroid
AT4G29810	AT1G35560	AT1G80340	Cat I	auxin	gibberellin
AT3G45640	AT1G35560	AT3G24520	Cat II	abscisic acid, salicylic acid	abscisic acid
AT3G45640	AT1G35560	AT5G51760	Cat II	abscisic acid, salicylic acid	abscisic acid
AT3G45640	AT1G35560	AT3G57040	Cat I	abscisic acid, salicylic acid	cytokinin
AT3G45640	AT1G35560	AT5G62000	Cat II	abscisic acid, salicylic acid	abscisic acid, auxin, brassinosteroid
AT3G45640	AT1G35560	AT5G55170	Cat II	abscisic acid, salicylic acid	salicylic acid
AT3G45640	AT1G35560	AT4G36540	Cat I	abscisic acid, salicylic acid	brassinosteroid
AT3G45640	AT1G35560	AT1G80340	Cat I	abscisic acid, salicylic acid	gibberellin
AT3G24520	AT1G35560	AT3G57040	Cat I	abscisic acid	cytokinin
AT3G24520	AT1G35560	AT5G62000	Cat II	abscisic acid	abscisic acid, auxin, brassinosteroid
AT3G24520	AT1G35560	AT5G55170	Cat I	abscisic acid	salicylic acid
AT3G24520	AT1G35560	AT4G36540	Cat I	abscisic acid	brassinosteroid
AT3G24520	AT1G35560	AT1G80340	Cat I	abscisic acid	gibberellin
AT5G51760	AT1G35560	AT3G57040	Cat I	abscisic acid	cytokinin
AT5G51760	AT1G35560	AT5G62000	Cat II	abscisic acid	abscisic acid, auxin, brassinosteroid
AT5G51760	AT1G35560	AT5G55170	Cat I	abscisic acid	salicylic acid
AT5G51760	AT1G35560	AT4G36540	Cat I	abscisic acid	brassinosteroid
AT5G51760	AT1G35560	AT1G80340	Cat I	abscisic acid	gibberellin
AT3G57040	AT1G35560	AT5G62000	Cat I	cytokinin	abscisic acid, auxin, brassinosteroid
AT3G57040	AT1G35560	AT5G55170	Cat I	cytokinin	salicylic acid
AT3G57040	AT1G35560	AT4G36540	Cat I	cytokinin	brassinosteroid
AT3G57040	AT1G35560	AT1G80340	Cat I	cytokinin	gibberellin
AT5G62000	AT1G35560	AT5G55170	Cat I	abscisic acid, auxin, brassinosteroid	salicylic acid
AT5G62000	AT1G35560	AT4G36540	Cat II	abscisic acid, auxin, brassinosteroid	brassinosteroid
AT5G62000	AT1G35560	AT1G80340	Cat I	abscisic acid, auxin, brassinosteroid	gibberellin
AT5G55170	AT1G35560	AT4G36540	Cat I	salicylic acid	brassinosteroid
AT5G55170	AT1G35560	AT1G80340	Cat I	salicylic acid	gibberellin
AT4G36540	AT1G35560	AT1G80340	Cat I	brassinosteroid	gibberellin
AT1G01360	AT1G17550	AT2G40330	Cat II	abscisic acid	abscisic acid, jasmonic acid
AT2G40330	AT1G17550	AT5G53160	Cat II	abscisic acid, jasmonic acid	abscisic acid
AT2G40330	AT1G17550	AT5G05440	Cat II	abscisic acid, jasmonic acid	abscisic acid
AT2G40330	AT1G17550	AT4G27920	Cat II	abscisic acid, jasmonic acid	abscisic acid
AT2G40330	AT1G17550	AT4G01026	Cat II	abscisic acid, jasmonic acid	abscisic acid
AT1G25490	AT1G21690	AT1G77470	Cat I	abscisic acid, auxin, ethylene	salicylic acid
AT2G02560	AT3G21175	AT5G11270	Cat I	auxin	abscisic acid, jasmonic acid
AT2G02560	AT3G21175	AT4G37470	Cat I	auxin	karrikin
AT5G11270	AT3G21175	AT4G37470	Cat I	abscisic acid, jasmonic acid	karrikin
AT5G47100	AT2G25090	AT4G37260	Cat II	abscisic acid	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT5G47100	AT2G25090	AT1G31880	Cat II	abscisic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT5G47100	AT2G25090	AT1G25490	Cat II	abscisic acid	abscisic acid, auxin, ethylene
AT4G37260	AT2G25090	AT1G31880	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G37260	AT2G25090	AT1G25490	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, auxin, ethylene
AT1G31880	AT2G25090	AT1G25490	Cat II	abscisic acid, auxin, brassinosteroid, cytokinin	abscisic acid, auxin, ethylene
AT2G01570	AT2G42880	AT4G37260	Cat II	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT2G01570	AT2G42880	AT1G31880	Cat II	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT2G01570	AT2G42880	AT2G37630	Cat II	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	auxin, gibberellin, jasmonic acid, salicylic acid
AT4G37260	AT2G42880	AT1G31880	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G37260	AT2G42880	AT1G14920	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT4G37260	AT2G42880	AT3G03450	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT1G31880	AT2G42880	AT1G14920	Cat II	abscisic acid, auxin, brassinosteroid, cytokinin	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT1G31880	AT2G42880	AT3G03450	Cat II	abscisic acid, auxin, brassinosteroid, cytokinin	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT1G31880	AT2G42880	AT2G37630	Cat II	abscisic acid, auxin, brassinosteroid, cytokinin	auxin, gibberellin, jasmonic acid, salicylic acid
AT1G14920	AT2G42880	AT2G37630	Cat II	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	auxin, gibberellin, jasmonic acid, salicylic acid

Table A.38 continued from previous page

Interactor A	Interactor B	Interactor C	Type	Annotation A	Annotation C
AT3G03450	AT2G42880	AT2G37630	Cat II	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	auxin, gibberellin, jasmonic acid, salicylic acid
AT2G01570	AT3G14720	AT4G37260	Cat II	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT1G04100	AT3G61830	AT3G23050	Cat II	auxin	abscisic acid, auxin, jasmonic acid
AT1G04100	AT3G61830	AT5G13930	Cat II	auxin	auxin, jasmonic acid
AT1G04250	AT3G61830	AT3G23050	Cat II	auxin	abscisic acid, auxin, jasmonic acid
AT1G04250	AT3G61830	AT5G13930	Cat II	auxin	auxin, jasmonic acid
AT1G04550	AT3G61830	AT5G13930	Cat II	auxin	auxin, jasmonic acid
AT1G51950	AT3G61830	AT3G23050	Cat II	auxin	abscisic acid, auxin, jasmonic acid
AT1G51950	AT3G61830	AT5G13930	Cat II	auxin	auxin, jasmonic acid
AT3G15540	AT3G61830	AT5G13930	Cat II	auxin	auxin, jasmonic acid
AT3G16500	AT3G61830	AT5G13930	Cat II	auxin	auxin, jasmonic acid
AT3G23050	AT3G61830	AT5G13930	Cat II	abscisic acid, auxin, jasmonic acid	auxin, jasmonic acid
AT5G25890	AT3G61830	AT5G13930	Cat II	auxin	auxin, jasmonic acid
AT2G44050	AT2G24570	AT1G31880	Cat I	jasmonic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT1G31880	AT1G69810	AT2G25000	Cat I	abscisic acid, auxin, brassinosteroid, cytokinin	salicylic acid
AT1G31880	AT1G69560	AT5G48870	Cat II	abscisic acid, auxin, brassinosteroid, cytokinin	abscisic acid
AT4G08150	AT1G15200	AT4G37260	Cat I	gibberellin	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT1G04250	AT1G02690	AT5G01600	Cat I	auxin	cytokinin, salicylic acid
AT1G04550	AT1G02690	AT5G01600	Cat I	auxin	cytokinin, salicylic acid
AT5G01600	AT1G02690	AT4G00120	Cat I	cytokinin, salicylic acid	auxin
AT4G37260	AT3G13672	AT2G44050	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT1G66160	AT1G31880	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G37260	AT1G66160	AT5G08130	Cat I	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	brassinosteroid
AT1G04250	AT4G16143	AT4G37260	Cat II	auxin	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT1G04250	AT4G16143	AT5G01600	Cat I	auxin	cytokinin, salicylic acid
AT1G04250	AT4G16143	AT4G14720	Cat I	auxin	jasmonic acid
AT1G04250	AT4G16143	AT2G44840	Cat I	auxin	ethylene
AT4G37260	AT4G16143	AT5G01600	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	cytokinin, salicylic acid
AT4G37260	AT4G16143	AT4G14720	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT4G16143	AT2G44840	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	ethylene
AT5G01600	AT4G16143	AT4G14720	Cat I	cytokinin, salicylic acid	jasmonic acid
AT5G01600	AT4G16143	AT2G44840	Cat I	cytokinin, salicylic acid	ethylene
AT4G14720	AT4G16143	AT2G44840	Cat I	jasmonic acid	ethylene
AT1G04240	AT4G02150	AT4G37260	Cat II	auxin	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT1G04240	AT4G02150	AT5G01600	Cat I	auxin	cytokinin, salicylic acid
AT1G04240	AT4G02150	AT2G45820	Cat I	auxin	ethylene
AT1G04250	AT4G02150	AT4G37260	Cat II	auxin	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT1G04250	AT4G02150	AT5G01600	Cat I	auxin	cytokinin, salicylic acid
AT1G04250	AT4G02150	AT2G45820	Cat I	auxin	ethylene
AT4G37260	AT4G02150	AT5G01600	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	cytokinin, salicylic acid
AT4G37260	AT4G02150	AT2G45820	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	ethylene
AT5G01600	AT4G02150	AT2G45820	Cat I	cytokinin, salicylic acid	ethylene
AT4G37260	AT4G25670	AT3G11410	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid
AT4G37260	AT3G50670	AT2G44050	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT3G50670	AT1G23860	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	ethylene
AT4G37260	AT3G50670	AT3G61860	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	ethylene
AT2G44050	AT3G50670	AT1G23860	Cat I	jasmonic acid	ethylene
AT2G44050	AT3G50670	AT3G61860	Cat I	jasmonic acid	ethylene
AT4G37260	AT3G05640	AT2G44050	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT2G33430	AT1G24590	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	auxin, brassinosteroid
AT4G37260	AT2G33430	AT3G11410	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid
AT1G24590	AT2G33430	AT3G11410	Cat I	auxin, brassinosteroid	abscisic acid
AT4G37260	AT3G60360	AT2G44050	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT3G18210	AT1G22070	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	salicylic acid
AT4G37260	AT3G18210	AT3G09370	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	auxin, ethylene, jasmonic acid, salicylic acid
AT1G22070	AT3G18210	AT3G09370	Cat II	salicylic acid	auxin, ethylene, jasmonic acid, salicylic acid
AT4G37260	AT3G06430	AT4G32570	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT1G51580	AT4G32570	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid

Table A.38 continued from previous page

Interactor A	Interactor B	Interactor C	Type	Annotation A	Annotation C
AT4G14560	AT3G18140	AT4G37260	Cat II	auxin	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT4G14560	AT3G18140	AT4G32570	Cat I	auxin	jasmonic acid
AT4G37260	AT3G18140	AT4G32570	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT3G55150	AT4G32570	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT4G15770	AT4G32570	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT4G15770	AT4G25470	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, ethylene, jasmonic acid, salicylic acid
AT4G32570	AT4G15770	AT4G25470	Cat II	jasmonic acid	abscisic acid, ethylene, jasmonic acid, salicylic acid
AT4G37260	AT1G07210	AT4G00120	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	auxin
AT4G37260	AT1G51100	AT1G31880	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G37260	AT1G06460	AT1G31880	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G37260	AT3G49810	AT1G31880	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G37260	AT3G57720	AT4G32570	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT4G02770	AT1G31880	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G37260	AT1G67090	AT4G32570	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT2G43370	AT3G61860	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	ethylene
AT4G37260	AT4G28860	AT3G05420	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	ethylene, jasmonic acid
AT5G67300	AT4G28860	AT3G05420	Cat II	abscisic acid, auxin, ethylene, gibberellin, jasmonic acid, salicylic acid	ethylene, jasmonic acid
AT4G37260	AT1G09810	AT1G31880	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G37260	AT1G54200	AT1G31880	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G37260	AT3G02460	AT4G32570	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT3G48510	AT1G31880	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G37260	AT4G18830	AT1G31880	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G37260	AT1G10650	AT1G31880	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G37260	AT1G10650	AT4G32570	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT1G10650	AT2G44840	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	ethylene
AT1G31880	AT1G10650	AT4G32570	Cat I	abscisic acid, auxin, brassinosteroid, cytokinin	jasmonic acid
AT1G31880	AT1G10650	AT2G44840	Cat I	abscisic acid, auxin, brassinosteroid, cytokinin	ethylene
AT4G32570	AT1G10650	AT2G44840	Cat I	jasmonic acid	ethylene
AT1G01360	AT1G21600	AT4G37260	Cat II	abscisic acid	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT1G01360	AT1G21600	AT4G32570	Cat I	abscisic acid	jasmonic acid
AT4G37260	AT1G21600	AT4G32570	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G32570	AT4G22720	AT2G39760	Cat I	jasmonic acid	ethylene
AT1G25490	AT3G14080	AT3G53250	Cat II	abscisic acid, auxin, ethylene	auxin
AT5G01600	AT1G68590	AT1G24590	Cat I	cytokinin, salicylic acid	auxin, brassinosteroid
AT4G16420	AT3G06790	AT5G67300	Cat II	auxin, cytokinin	abscisic acid, auxin, ethylene, gibberellin, jasmonic acid, salicylic acid
AT4G16420	AT3G06790	AT5G17490	Cat I	auxin, cytokinin	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT5G67300	AT3G06790	AT5G17490	Cat II	abscisic acid, auxin, ethylene, gibberellin, jasmonic acid, salicylic acid	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT1G25490	AT5G57950	AT5G60890	Cat I	abscisic acid, auxin, ethylene	jasmonic acid
AT5G67300	AT4G00180	AT5G08130	Cat I	abscisic acid, auxin, ethylene, gibberellin, jasmonic acid, salicylic acid	brassinosteroid
AT4G32570	AT1G29260	AT1G04710	Cat I	jasmonic acid	abscisic acid
AT4G32570	AT1G29260	AT2G33150	Cat II	jasmonic acid	abscisic acid, jasmonic acid
AT1G04710	AT1G29260	AT2G33150	Cat II	abscisic acid	abscisic acid, jasmonic acid
AT1G31880	AT1G76080	AT5G67300	Cat II	abscisic acid, auxin, brassinosteroid, cytokinin	abscisic acid, auxin, ethylene, gibberellin, jasmonic acid, salicylic acid
AT1G31880	AT1G76080	AT5G19000	Cat I	abscisic acid, auxin, brassinosteroid, cytokinin	ethylene
AT5G67300	AT1G76080	AT5G19000	Cat II	abscisic acid, auxin, ethylene, gibberellin, jasmonic acid, salicylic acid	ethylene
AT2G44050	AT1G51090	AT4G00120	Cat I	jasmonic acid	auxin
AT4G08150	AT3G56270	AT1G25490	Cat I	gibberellin	abscisic acid, auxin, ethylene
AT4G08150	AT3G56270	AT1G24590	Cat I	gibberellin	auxin, brassinosteroid

Table A.38 continued from previous page

Interactor A	Interactor B	Interactor C	Type	Annotation A	Annotation C
AT4G08150	AT3G56270	AT3G11410	Cat I	gibberellin	abscisic acid
AT1G25490	AT3G56270	AT1G24590	Cat II	abscisic acid, auxin, ethylene	auxin, brassinosteroid
AT1G25490	AT3G56270	AT3G11410	Cat II	abscisic acid, auxin, ethylene	abscisic acid
AT1G24590	AT3G56270	AT3G11410	Cat I	auxin, brassinosteroid	abscisic acid
AT4G08150	AT2G32840	AT4G16420	Cat I	gibberellin	auxin, cytokinin
AT4G08150	AT2G32840	AT4G00120	Cat I	gibberellin	auxin
AT4G16420	AT2G32840	AT4G00120	Cat II	auxin, cytokinin	auxin
AT4G08150	AT2G01620	AT1G66370	Cat I	gibberellin	jasmonic acid
AT3G11410	AT3G07090	AT4G38630	Cat II	abscisic acid	abscisic acid, auxin, cytokinin
AT4G32570	AT1G16705	AT2G45820	Cat I	jasmonic acid	ethylene
AT1G31880	AT1G12120	AT2G39760	Cat I	abscisic acid, auxin, brassinosteroid, cytokinin	ethylene
AT4G32570	AT3G57290	AT5G08130	Cat I	jasmonic acid	brassinosteroid
AT4G32570	AT1G05410	AT1G25490	Cat I	jasmonic acid	abscisic acid, auxin, ethylene
AT4G32570	AT2G04630	AT1G06180	Cat II	jasmonic acid	gibberellin, jasmonic acid, salicylic acid
AT4G32570	AT4G12100	AT1G22070	Cat I	jasmonic acid	salicylic acid
AT4G32570	AT4G17680	AT3G23240	Cat II	jasmonic acid	ethylene, jasmonic acid
AT1G28360	AT2G44740	AT4G32570	Cat I	ethylene	jasmonic acid
AT4G32570	AT1G11810	AT2G33860	Cat I	jasmonic acid	auxin
AT4G32570	AT3G02000	AT1G22070	Cat I	jasmonic acid	salicylic acid
AT4G32570	AT3G02000	AT5G65210	Cat I	jasmonic acid	salicylic acid
AT4G32570	AT4G02485	AT5G04190	Cat I	jasmonic acid	auxin
AT1G31880	AT1G56450	AT4G32570	Cat I	abscisic acid, auxin, brassinosteroid, cytokinin	jasmonic acid
AT1G28360	AT1G77770	AT4G00120	Cat I	ethylene	auxin
AT1G18710	AT3G54130	AT3G19820	Cat I	jasmonic acid	brassinosteroid
AT1G18710	AT3G54130	AT5G05690	Cat I	jasmonic acid	brassinosteroid
AT1G04550	AT5G18580	AT3G18240	Cat I	auxin	ethylene
AT1G04550	AT5G18580	AT1G66370	Cat I	auxin	jasmonic acid
AT3G18240	AT5G18580	AT5G25890	Cat I	ethylene	auxin
AT3G18240	AT5G18580	AT1G66370	Cat I	ethylene	jasmonic acid
AT5G25890	AT5G18580	AT1G66370	Cat I	auxin	jasmonic acid
AT1G25490	AT1G11430	AT5G17490	Cat II	abscisic acid, auxin, ethylene	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT4G37260	AT5G14070	AT1G24590	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	auxin, brassinosteroid
AT4G37260	AT5G14070	AT5G17490	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT4G37260	AT5G14070	AT5G65210	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	salicylic acid
AT1G24590	AT5G14070	AT5G17490	Cat I	auxin, brassinosteroid	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT1G24590	AT5G14070	AT5G65210	Cat I	auxin, brassinosteroid	salicylic acid
AT5G17490	AT5G14070	AT5G65210	Cat II	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	salicylic acid
AT1G25490	AT2G35900	AT5G59220	Cat II	abscisic acid, auxin, ethylene	abscisic acid
AT1G09700	AT3G22490	AT5G59220	Cat II	abscisic acid, auxin, cytokinin	abscisic acid
AT2G44050	AT4G18650	AT5G17490	Cat II	jasmonic acid	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT1G18400	AT5G41920	AT2G01570	Cat I	brassinosteroid	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT1G18400	AT5G41920	AT4G32570	Cat I	brassinosteroid	jasmonic acid
AT1G18400	AT5G41920	AT1G19180	Cat I	brassinosteroid	auxin, jasmonic acid
AT2G01570	AT5G41920	AT4G32570	Cat II	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	jasmonic acid
AT2G01570	AT5G41920	AT1G19180	Cat II	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid	auxin, jasmonic acid
AT4G32570	AT5G41920	AT1G19180	Cat II	jasmonic acid	auxin, jasmonic acid
AT3G62100	AT5G61380	AT1G31880	Cat II	auxin	abscisic acid, auxin, brassinosteroid, cytokinin
AT3G62100	AT5G61380	AT2G22090	Cat I	auxin	abscisic acid
AT1G31880	AT5G61380	AT2G22090	Cat II	abscisic acid, auxin, brassinosteroid, cytokinin	abscisic acid
AT1G31880	AT5G16080	AT3G11410	Cat II	abscisic acid, auxin, brassinosteroid, cytokinin	abscisic acid
AT4G37260	AT5G16400	AT1G31880	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G08150	AT5G59730	AT4G37260	Cat I	gibberellin	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT4G08150	AT5G59730	AT5G25890	Cat I	gibberellin	auxin
AT4G37260	AT5G59730	AT4G32570	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT5G59730	AT5G25890	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	auxin
AT4G32570	AT5G59730	AT5G25890	Cat I	jasmonic acid	auxin
AT1G04550	AT5G62520	AT4G37260	Cat II	auxin	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT1G04550	AT5G62520	AT2G44050	Cat I	auxin	jasmonic acid
AT1G04550	AT5G62520	AT5G07690	Cat I	auxin	ethylene, gibberellin, jasmonic acid, salicylic acid

Table A.38 continued from previous page

Interactor A	Interactor B	Interactor C	Type	Annotation A	Annotation C
AT4G37260	AT5G62520	AT2G44050	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT5G62520	AT5G07690	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	ethylene, gibberellin, jasmonic acid, salicylic acid
AT2G44050	AT5G62520	AT5G07690	Cat II	jasmonic acid	ethylene, gibberellin, jasmonic acid, salicylic acid
AT4G37260	AT5G61230	AT4G00120	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	auxin
AT4G16420	AT5G49210	AT4G37260	Cat II	auxin, cytokinin	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid
AT4G16420	AT5G49210	AT1G24590	Cat II	auxin, cytokinin	auxin, brassinosteroid
AT4G37260	AT5G49210	AT1G24590	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	auxin, brassinosteroid
AT4G37260	AT5G22890	AT2G44050	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT5G22890	AT1G31880	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT2G44050	AT5G22890	AT1G31880	Cat I	jasmonic acid	abscisic acid, auxin, brassinosteroid, cytokinin
AT4G37260	AT5G48335	AT1G24590	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	auxin, brassinosteroid
AT4G37260	AT5G05790	AT4G32570	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	jasmonic acid
AT4G37260	AT5G05790	AT1G24590	Cat II	abscisic acid, auxin, ethylene, jasmonic acid, salicylic acid	auxin, brassinosteroid
AT4G32570	AT5G05790	AT1G24590	Cat I	jasmonic acid	auxin, brassinosteroid
AT1G31880	AT5G23130	AT3G03450	Cat II	abscisic acid, auxin, brassinosteroid, cytokinin	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT1G31880	AT5G65683	AT3G03450	Cat II	abscisic acid, auxin, brassinosteroid, cytokinin	abscisic acid, ethylene, gibberellin, jasmonic acid, salicylic acid
AT4G32570	AT5G07260	AT4G01370	Cat II	jasmonic acid	abscisic acid, ethylene, jasmonic acid, salicylic acid
AT4G32570	AT5G07260	AT2G43790	Cat II	jasmonic acid	abscisic acid, ethylene, jasmonic acid, salicylic acid
AT1G31880	AT5G41070	AT1G09700	Cat II	abscisic acid, auxin, brassinosteroid, cytokinin	abscisic acid, auxin, cytokinin
AT1G31880	AT5G18110	AT4G32570	Cat I	abscisic acid, auxin, brassinosteroid, cytokinin	jasmonic acid
AT2G25000	AT5G52650	AT1G76420	Cat I	salicylic acid	cytokinin
AT1G28360	AT5G24660	AT1G24590	Cat I	ethylene	auxin, brassinosteroid

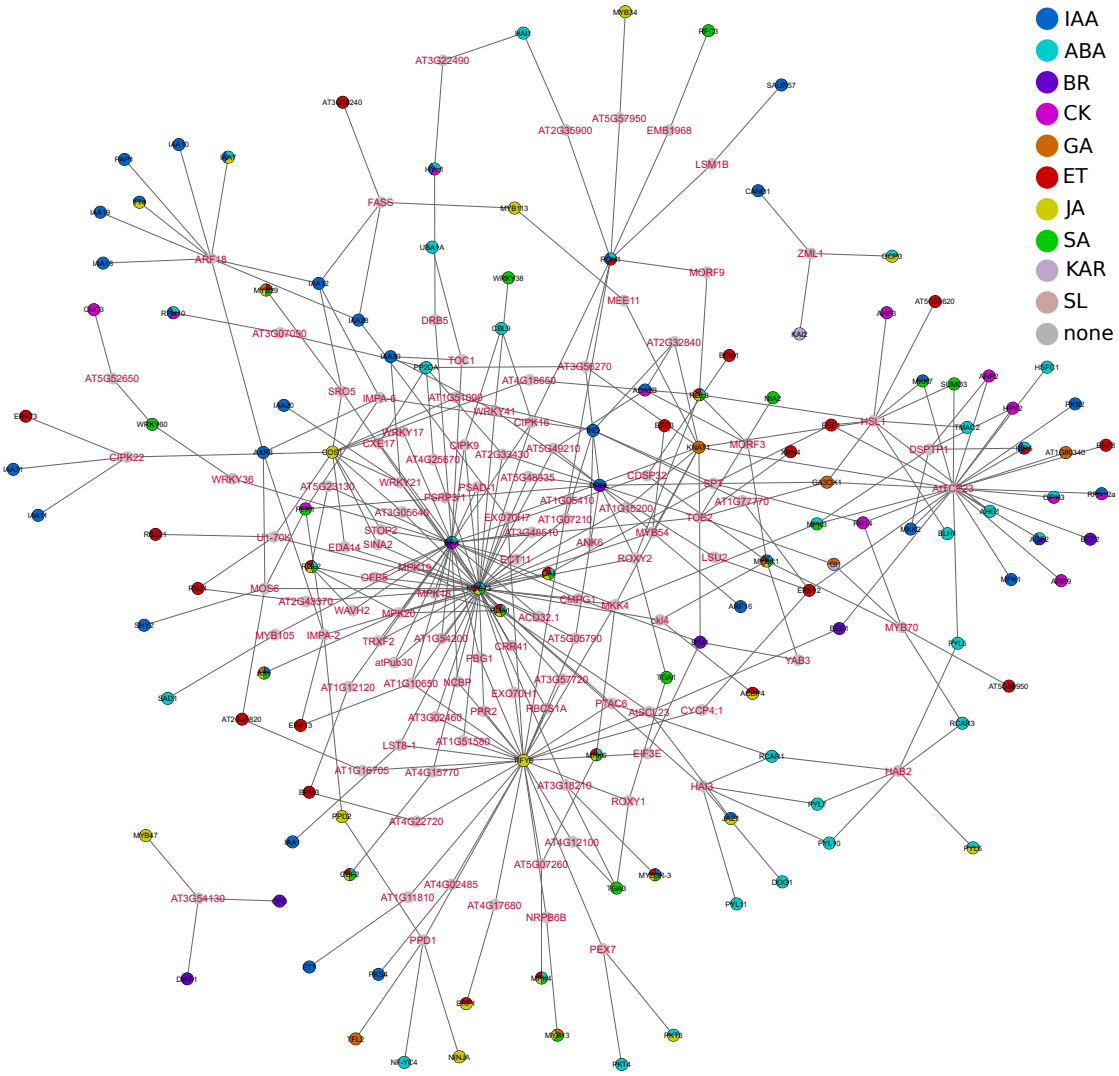


Figure A.7: Network representation of all three node crosstalks in $\mathrm{PhI}_{\mathrm{ext}}$.

A.17 TCP23 condition-dependent interactions

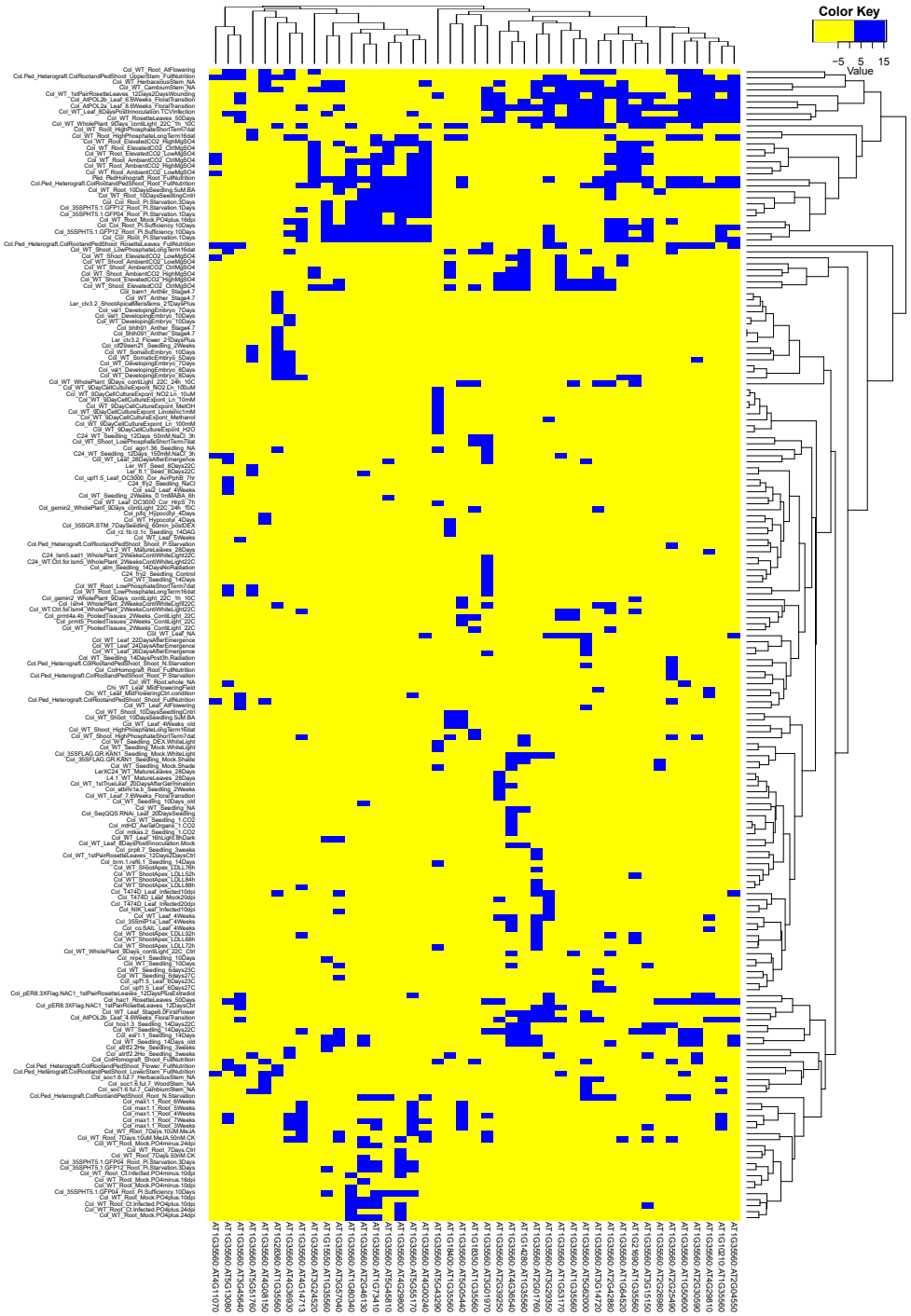


Figure A.8: Condition-dependent interactions of TCP23 complete

A.18 Kinase interactions

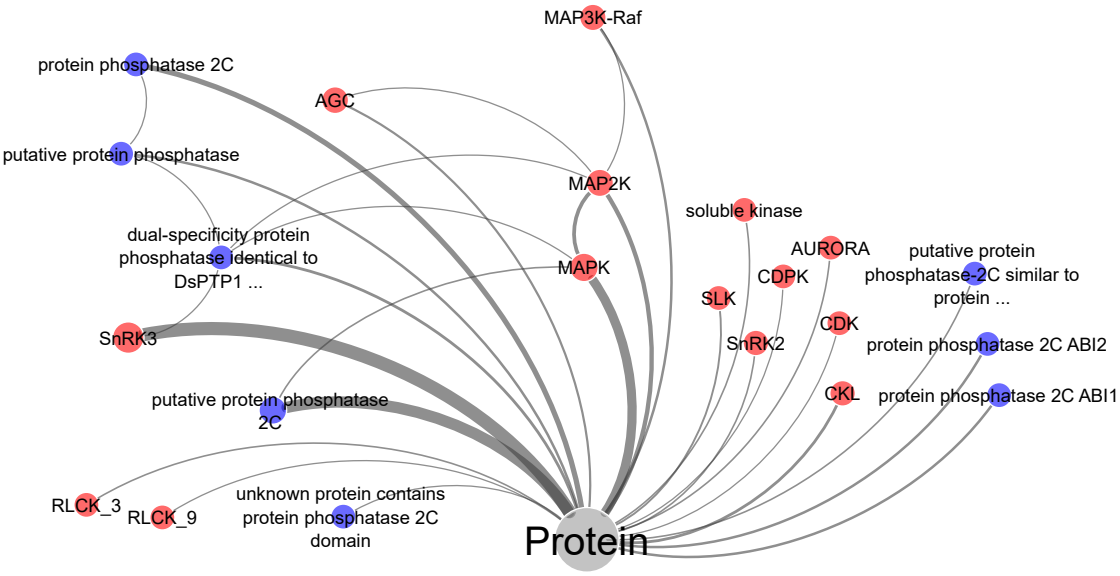


Figure A.9: Kinase and phosphatase family interactions in combined network $\mathrm{PhI}_{\mathrm{ext}}$.

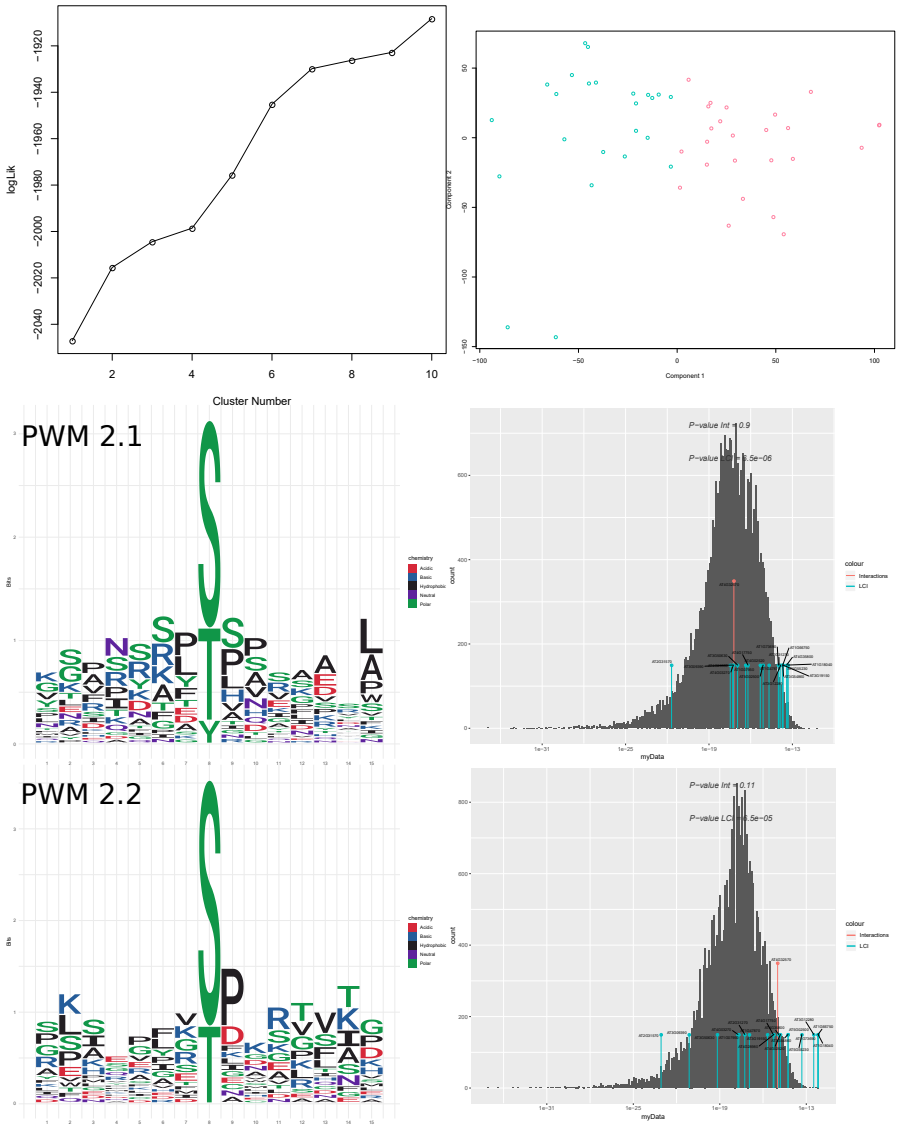


Figure A.10: CDK Results

Table A.39: PhI kinase group interactions

Group A	Group B	Interactions
Protein	Protein	339
SnRK3	Protein	31
putative protein phosphatase 2C	Protein	26
SnRK2	Protein	1
MAP2K	Protein	10
MAPK	Protein	21
MAPK	putative protein phosphatase 2C	2
MAPK	MAP2K	9
protein phosphatase 2C ABI1	Protein	5
CDPK	Protein	1
dual-specificity protein phosphatase identical to DsPTP1 ...	Protein	3
dual-specificity protein phosphatase identical to DsPTP1 ...	SnRK3	1
dual-specificity protein phosphatase identical to DsPTP1 ...	MAP2K	1
dual-specificity protein phosphatase identical to DsPTP1 ...	MAPK	1
putative protein phosphatase	Protein	3
putative protein phosphatase	dual-specificity protein phosphatase identical to DsPTP1 ...	1
SLK	Protein	1
protein phosphatase 2C	Protein	8
protein phosphatase 2C	putative protein phosphatase	1
MAP3K-Raf	Protein	2
MAP3K-Raf	MAP2K	1
AGC	Protein	1
AGC	MAP2K	1
protein phosphatase 2C ABI2	Protein	5

Table A.40: PhI kinase group sizes

Group	Size
AGC	2
CDPK	1
dual-specificity protein phosphatase identical to DsPTP1 ...	1
MAP2K	6
MAP3K-Raf	1
MAPK	8
Protein	207
protein phosphatase 2C	1
protein phosphatase 2C ABI1	1
protein phosphatase 2C ABI2	1
putative protein phosphatase	1
putative protein phosphatase 2C	5
SLK	1
SnRK2	1
SnRK3	14

Table A.41: PhI_{ext} kinase group interactions

Group A	Group B	Interactions
Protein	Protein	888
SnRK3	Protein	41
putative protein phosphatase 2C	Protein	34
SnRK2	Protein	2
MAP2K	Protein	11
MAPK	Protein	27
MAPK	putative protein phosphatase 2C	2
MAPK	MAP2K	9
protein phosphatase 2C ABI1	Protein	5
CDPK	Protein	1
dual-specificity protein phosphatase identical to DsPTP1 ...	Protein	6
dual-specificity protein phosphatase identical to DsPTP1 ...	SnRK3	1
dual-specificity protein phosphatase identical to DsPTP1 ...	MAP2K	1
dual-specificity protein phosphatase identical to DsPTP1 ...	MAPK	1
putative protein phosphatase	Protein	5
putative protein phosphatase	dual-specificity protein phosphatase identical to DsPTP1 ...	1
SLK	Protein	3
protein phosphatase 2C	Protein	14
protein phosphatase 2C	putative protein phosphatase	1
putative protein phosphatase-2C similar to protein ...	Protein	2
RLCK_3	Protein	2
CKL	Protein	6
AURORA	Protein	2
RLCK_9	Protein	1
AGC	Protein	4
AGC	MAP2K	1
MAP3K-Raf	Protein	5
MAP3K-Raf	MAP2K	1
protein phosphatase 2C ABI2	Protein	5
unknown protein contains protein phosphatase 2C domain	Protein	1
soluble kinase	Protein	2
CDK	Protein	1

Table A.42: PhI_{ext} kinase group sizes

Group	Size
AGC	5
AURORA	1
CDK	1
CDPK	1
CKL	2
dual-specificity protein phosphatase identical to DsPTP1 ...	1
MAP2K	6
MAP3K-Raf	4
MAPK	9
Protein	631
protein phosphatase 2C	1
protein phosphatase 2C ABI1	1
protein phosphatase 2C ABI2	1
putative protein phosphatase	1
putative protein phosphatase-2C similar to protein ...	1
putative protein phosphatase 2C	7
RLCK_3	1
RLCK_9	1
SLK	1
SnRK2	2
SnRK3	16
soluble kinase	1
unknown protein contains protein phosphatase 2C domain	1

Table A.43: Results for all interacting proteins of all kinases in PhI_{ext}.

Locus	P-Value	PWM	Status	Locus	P-Value	PWM	Status
AGC (4 cluster)				MAPK (7 cluster)			
AT1G31880	1,96E-01	-	new	AT1G51950	2,10E-01	-	new
AT1G18350	3,14E-01	-	new	AT2G01570	3,78E-02	1	new
AT4G17490	1,40E-01	-	new	AT3G23610	3,57E-03	6	new
CDK (2 cluster)				AT1G35560	6,89E-02	-	new
AT4G32570	4,03E-02	2	new	AT4G37260	7,77E-02	-	new
CDPK (7 cluster)				AT1G07340	3,91E-01	-	new
AT5G17690	2,65E-02	3	new	AT2G30020	1,60E-03	4	new
SnRK2 (6 Cluster)				AT4G29810	7,10E-02	-	known
AT2G36270	2,55E-04	4	known	AT3G03450	1,23E-01	-	known
AT5G39360	1,31E-01	-	new	AT1G51660	7,73E-03	1	known
SnRK3 (4 Cluster)				AT4G26070	1,54E-01	-	new
AT2G23290	3,71E-01	-	new	AT1G31880	1,31E-01	-	new
AT5G47100	7,30E-05	3	known	AT1G14920	3,67E-01	-	new
AT1G53910	3,64E-01	-	new	AT5G56580	2,19E-01	-	new
AT3G17600	8,06E-03	3	new	AT5G40440	1,55E-01	-	known
AT4G28640	1,01E-02	2	new	AT3G63210	3,74E-01	-	new
AT3G23610	1,53E-01	-	new	AT2G37630	2,22E-01	-	new
AT1G35560	3,71E-01	-	new	AT5G14600	9,10E-02	-	new
AT4G37260	2,00E-01	-	new	AT5G07260	7,70E-01	-	new
AT2G44050	1,26E-01	-	new	AT2G18730	1,65E-01	-	new
AT1G68550	5,35E-02	-	new	AT1G12860	2,50E-02	5	new
AT1G31880	5,99E-01	-	new	AT1G08780	8,96E-01	-	new
AT5G25190	1,42E-01	-	new	MAPK2K (3 Cluster)			
AT1G25490	4,22E-01	-	new	AT4G08150	2,04E-01	-	new
AT3G25890	6,68E-02	-	new	AT3G23610	2,45E-01	-	new
AT1G32640	5,69E-03	1	new	AT1G35560	2,69E-02	1	new
AT4G17615	3,65E-05	3	known	AT4G37260	5,48E-01	-	new
AT3G22810	4,07E-02	1	new	AT4G32010	1,54E-01	-	new
AT1G72360	2,03E-01	-	new	AT1G07340	1,30E-01	-	new
AT3G04300	1,99E-01	-	new	AT4G01370	7,30E-05	3	known
SLK (6 Cluster)				AT1G31880	3,14E-03	1	new
AT4G18890	1,70E-02	4	new	AT3G45640	3,65E-05	1	known
AT5G65300	1,38E-02	5	new	AT2G34650	9,67E-02	-	new
AT1G35210	1,91E-02	3	new	AT4G32570	1,69E-01	-	new
MAP3K (3 Cluster)				AT1G10210	1,82E-04	3	known
AT2G23290	3,80E-02	1	new	AT5G58950	1,32E-01	-	new
AT4G37260	4,67E-03	3	new	AT2G43790	4,00E-05	3	known
AT4G29810	3,92E-02	1	known	AT1G24590	5,01E-01	-	new
AT1G31880	3,11E-01	-	new	AT2G18170	2,92E-04	2	known
AT1G22070	3,10E-01	-	new	AT3G50800	7,13E-01	-	new
AT1G76680	3,80E-01	-	new				

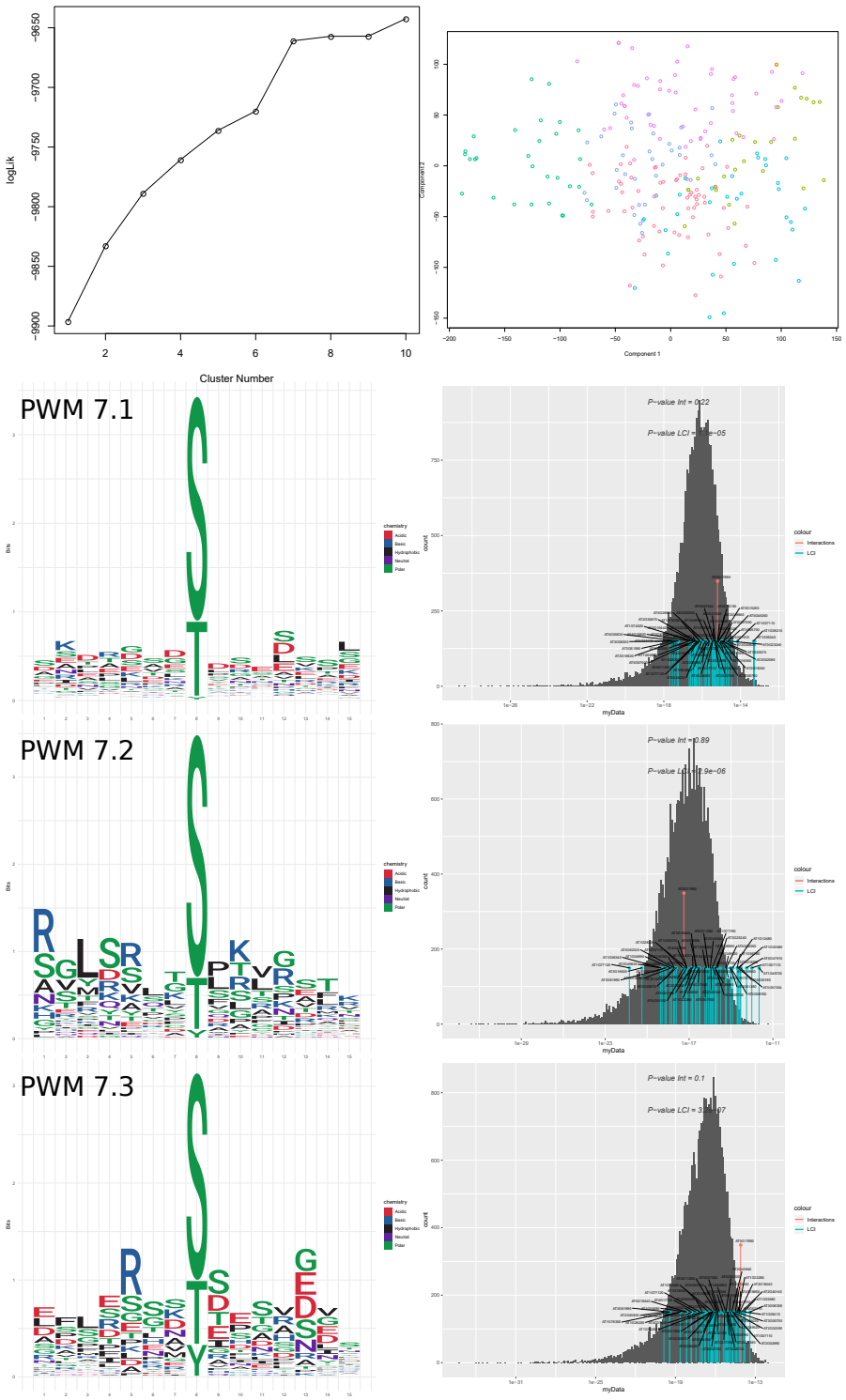


Figure A.11: CDPK Results

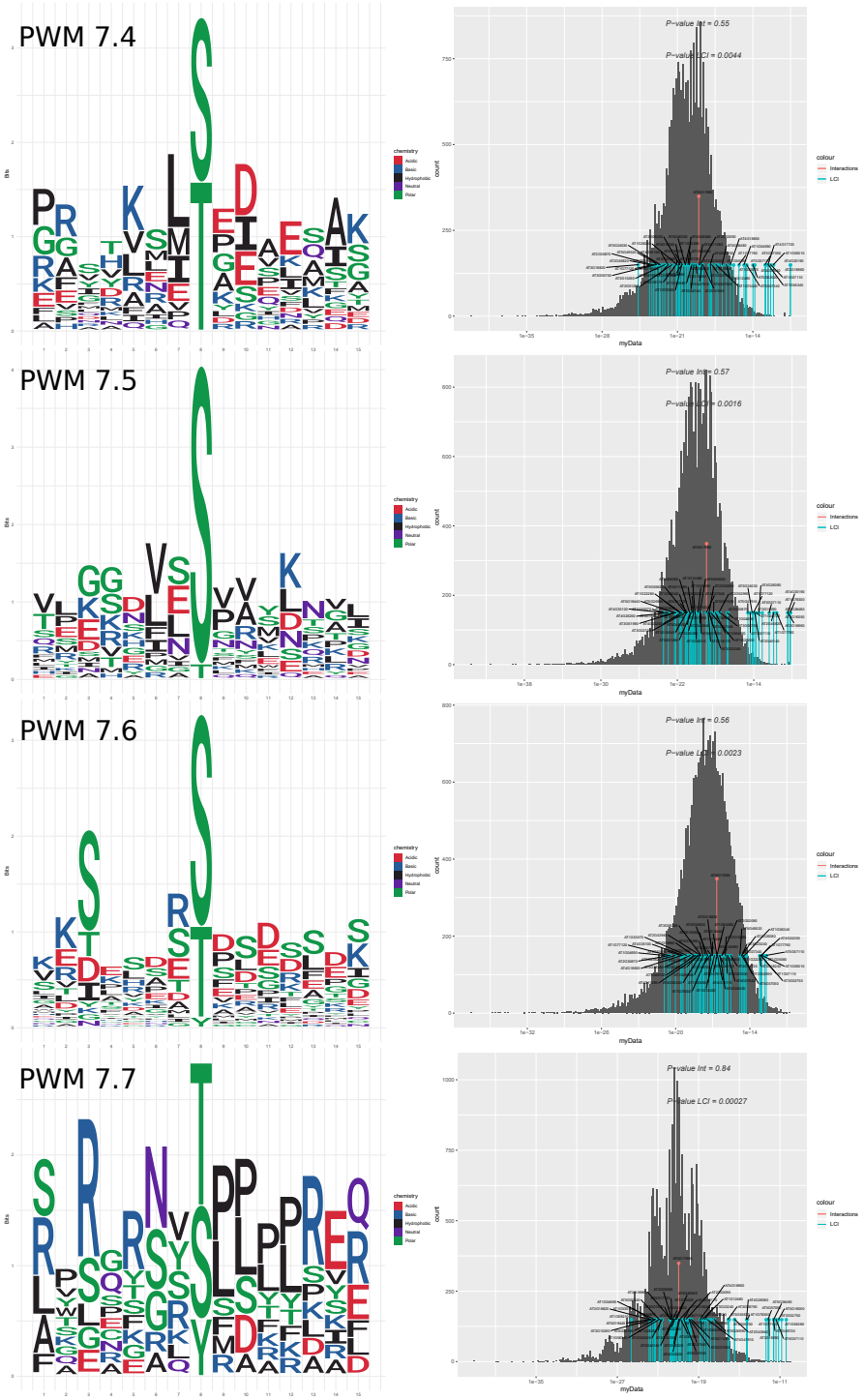


Figure A.12: CDPK Results continued

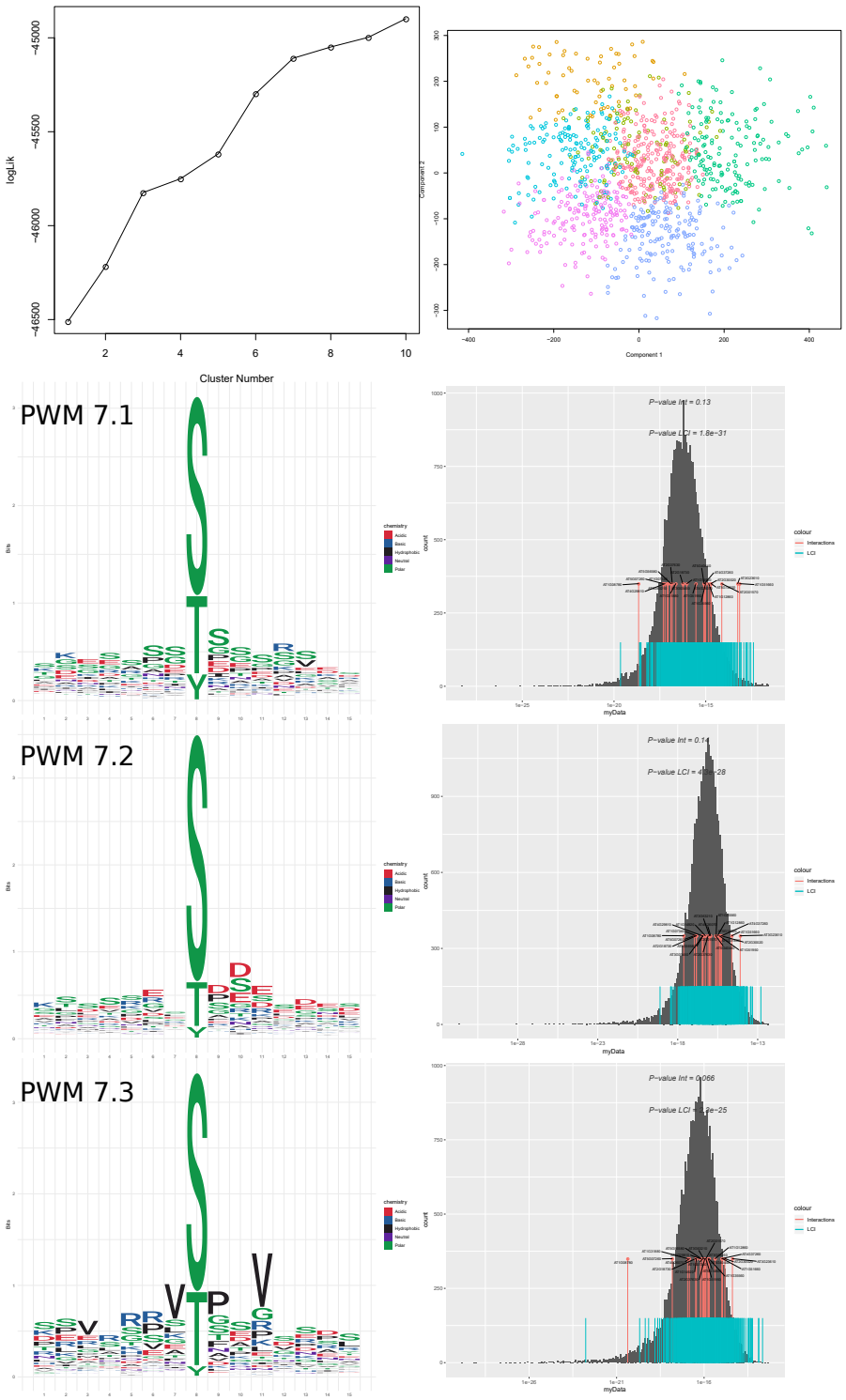


Figure A.13: MAPK Results

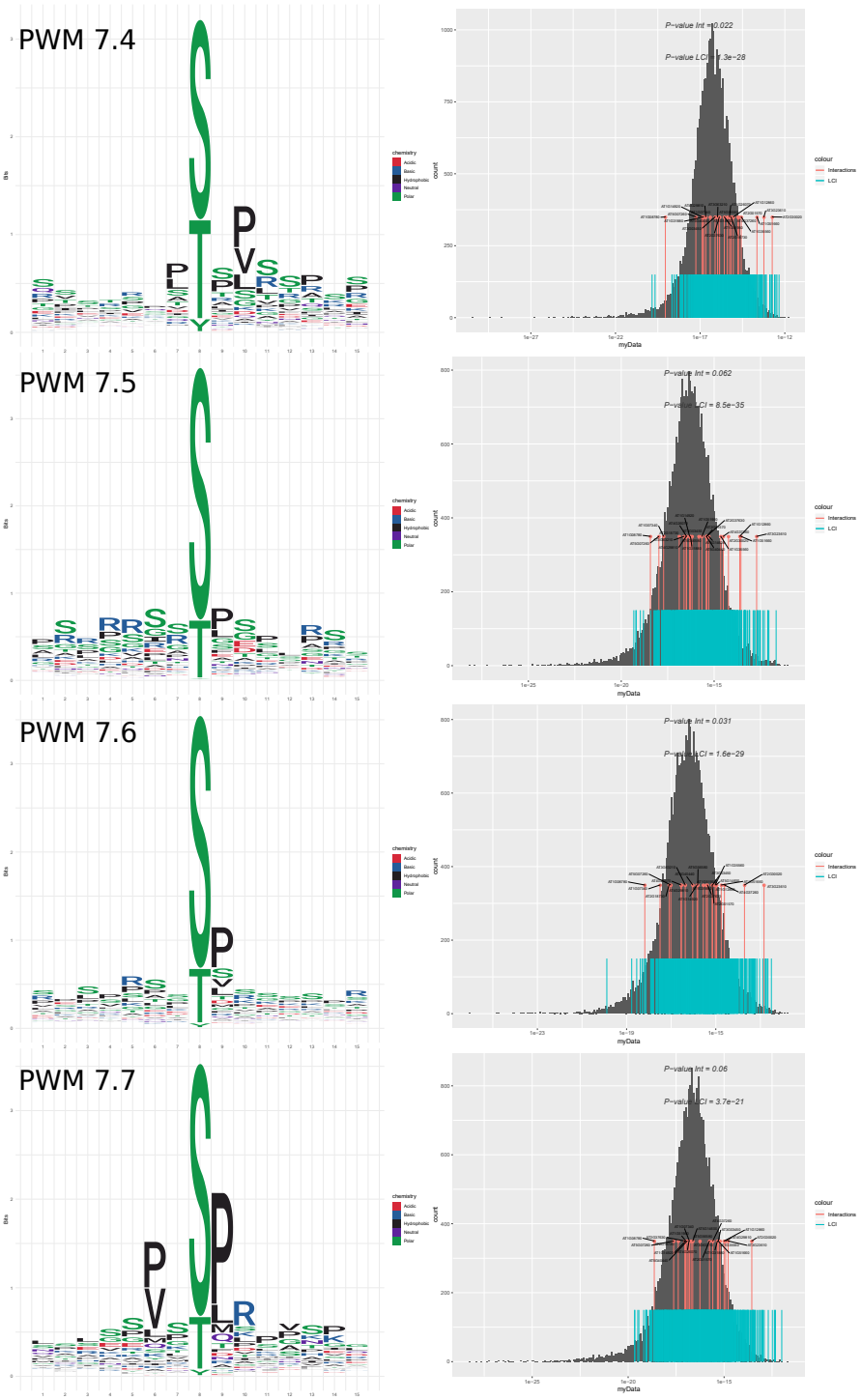


Figure A.14: MAPK Results continued

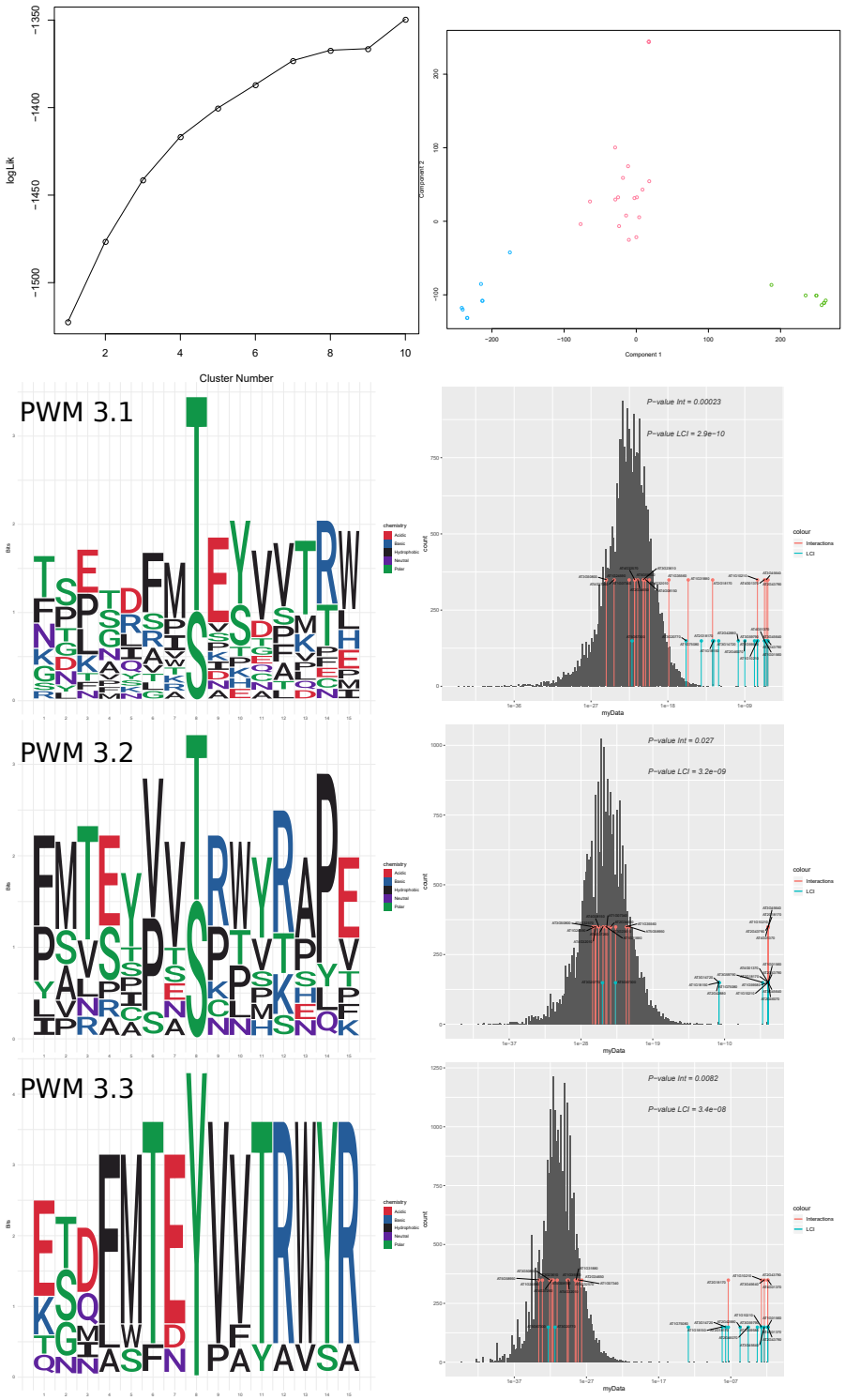


Figure A.15: MAP2K Results

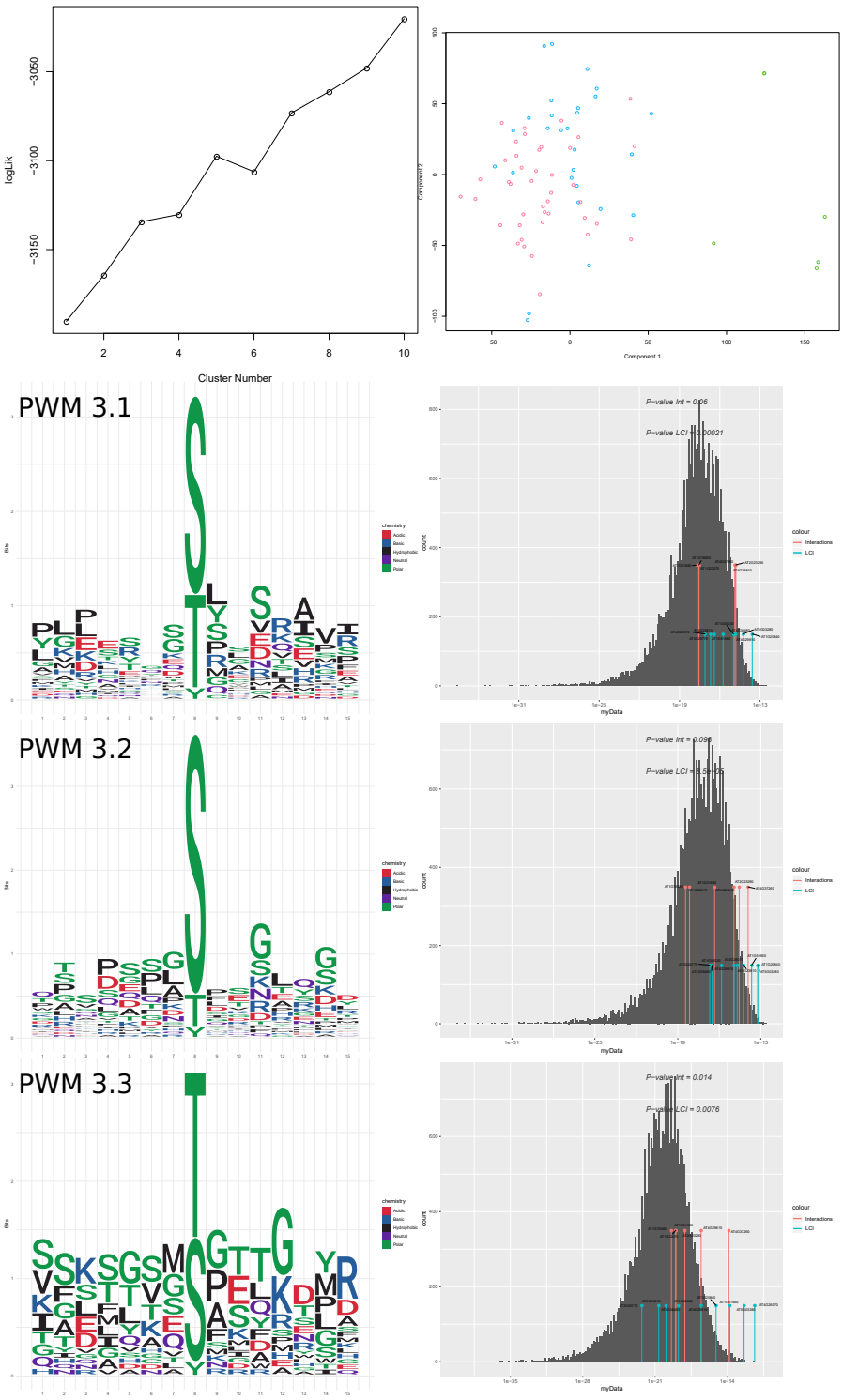


Figure A.16: MAP3K Results

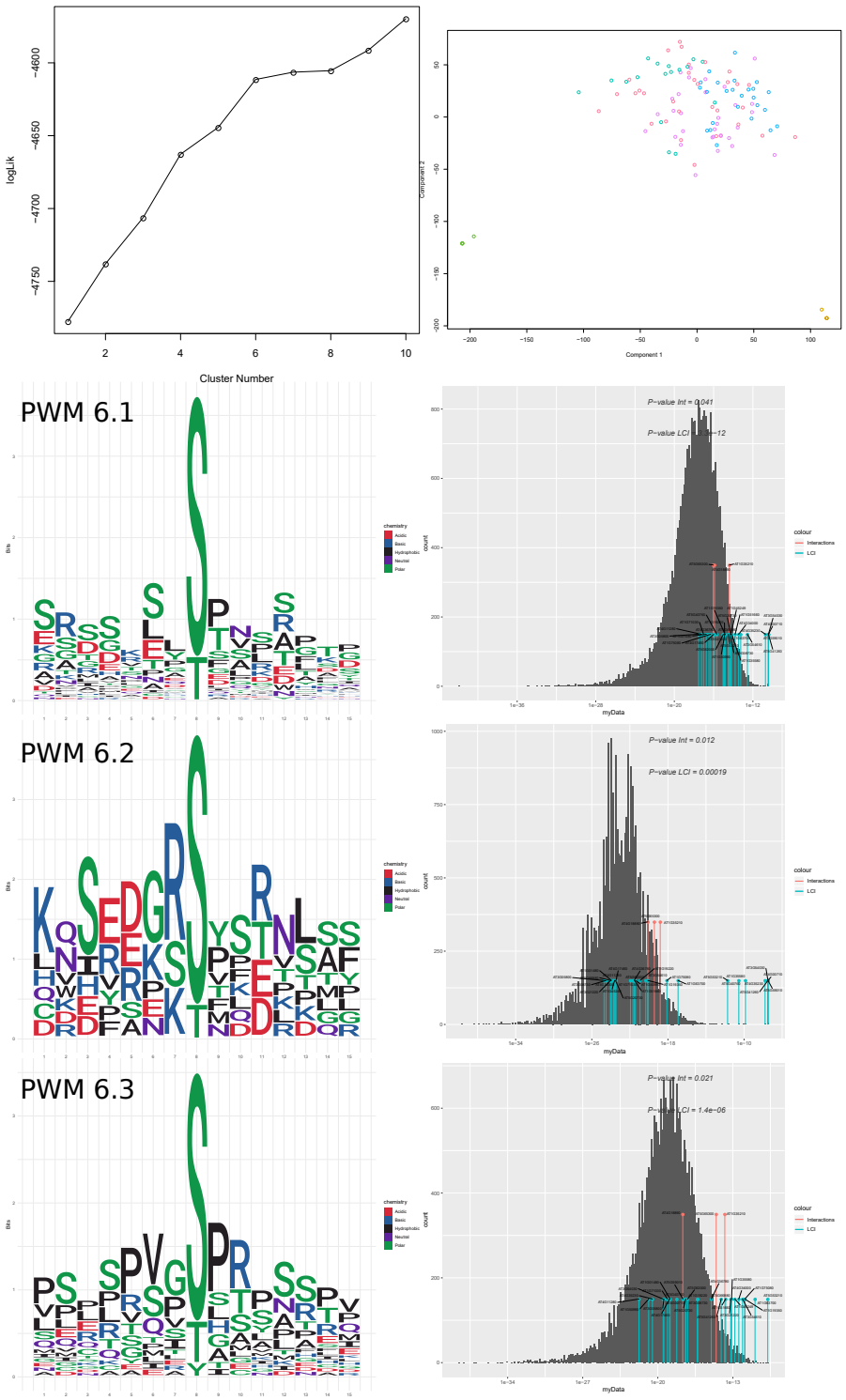


Figure A.17: SLK Results

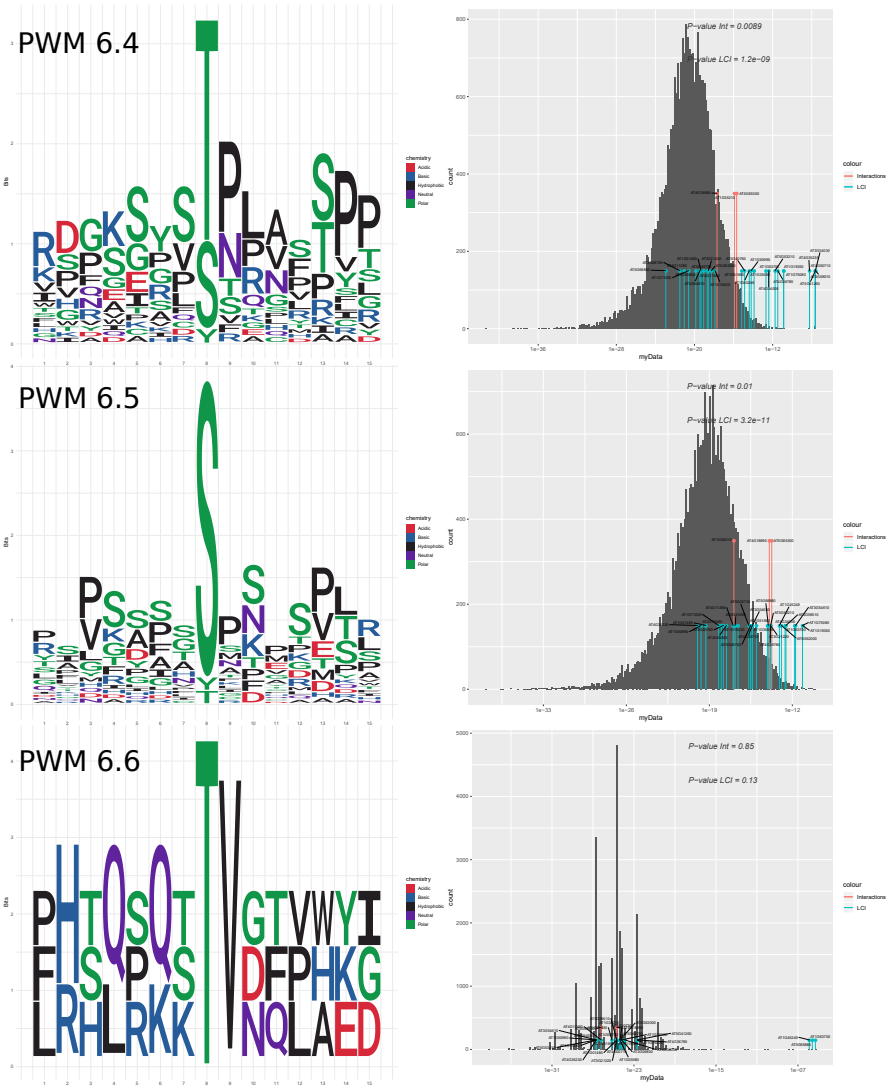


Figure A.18: SLK Results continued

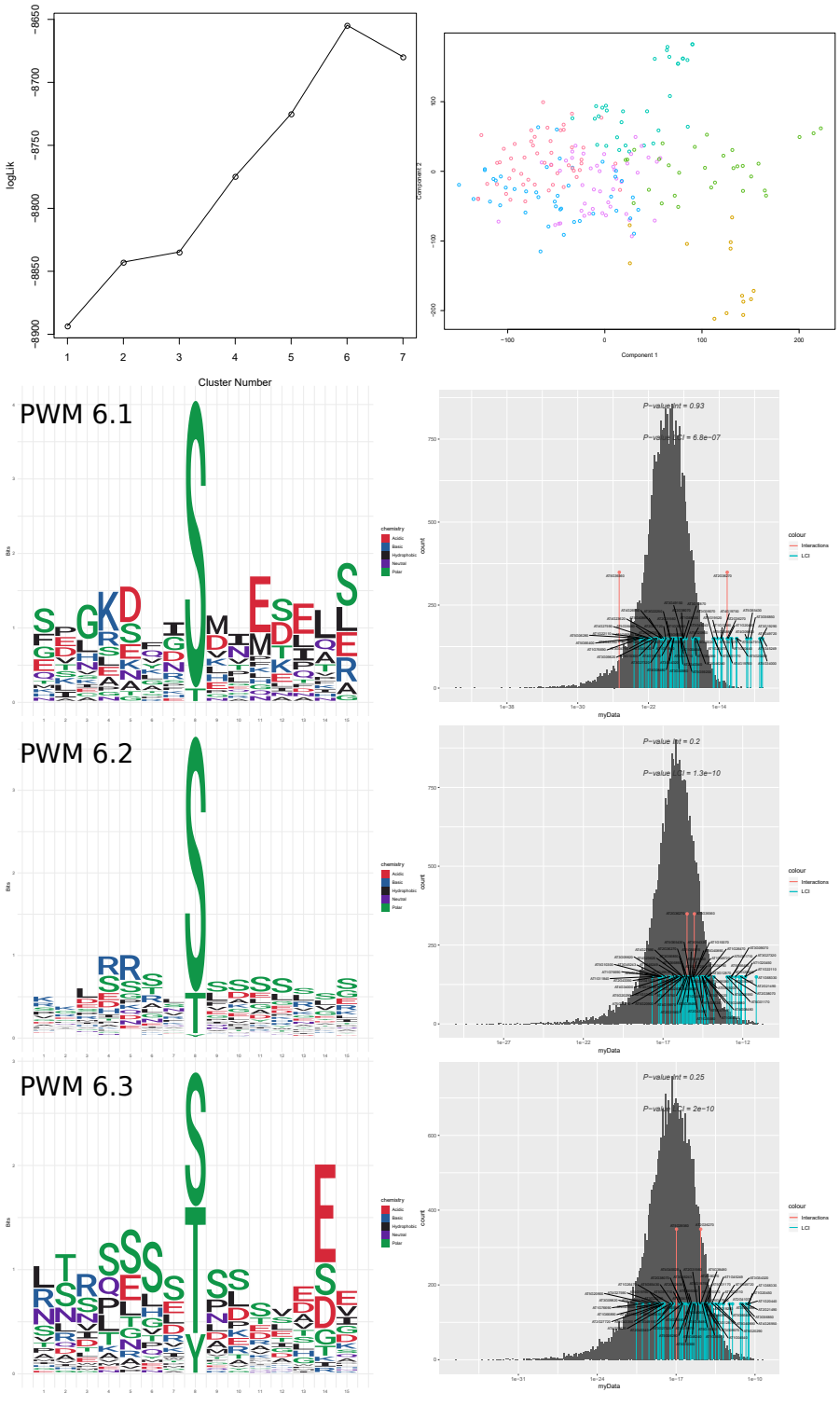


Figure A.19: SnRK2 Results

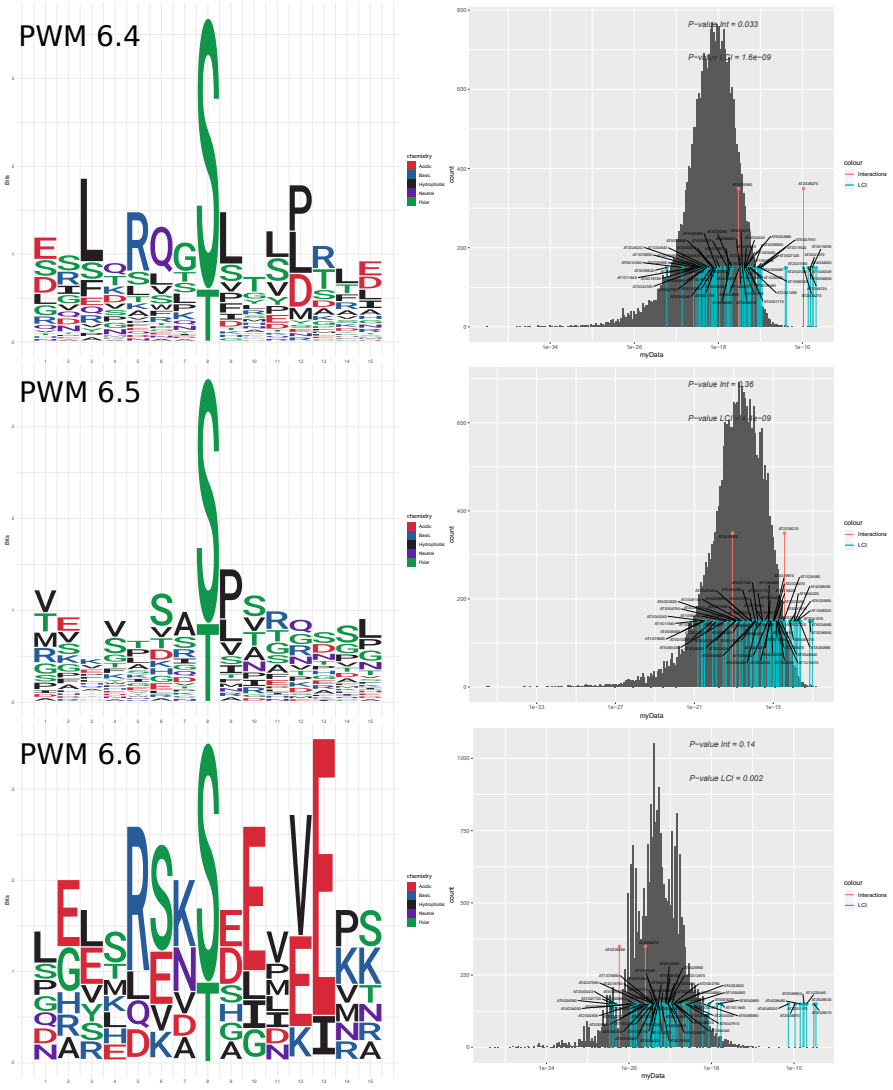


Figure A.20: SnRK2 Results continued

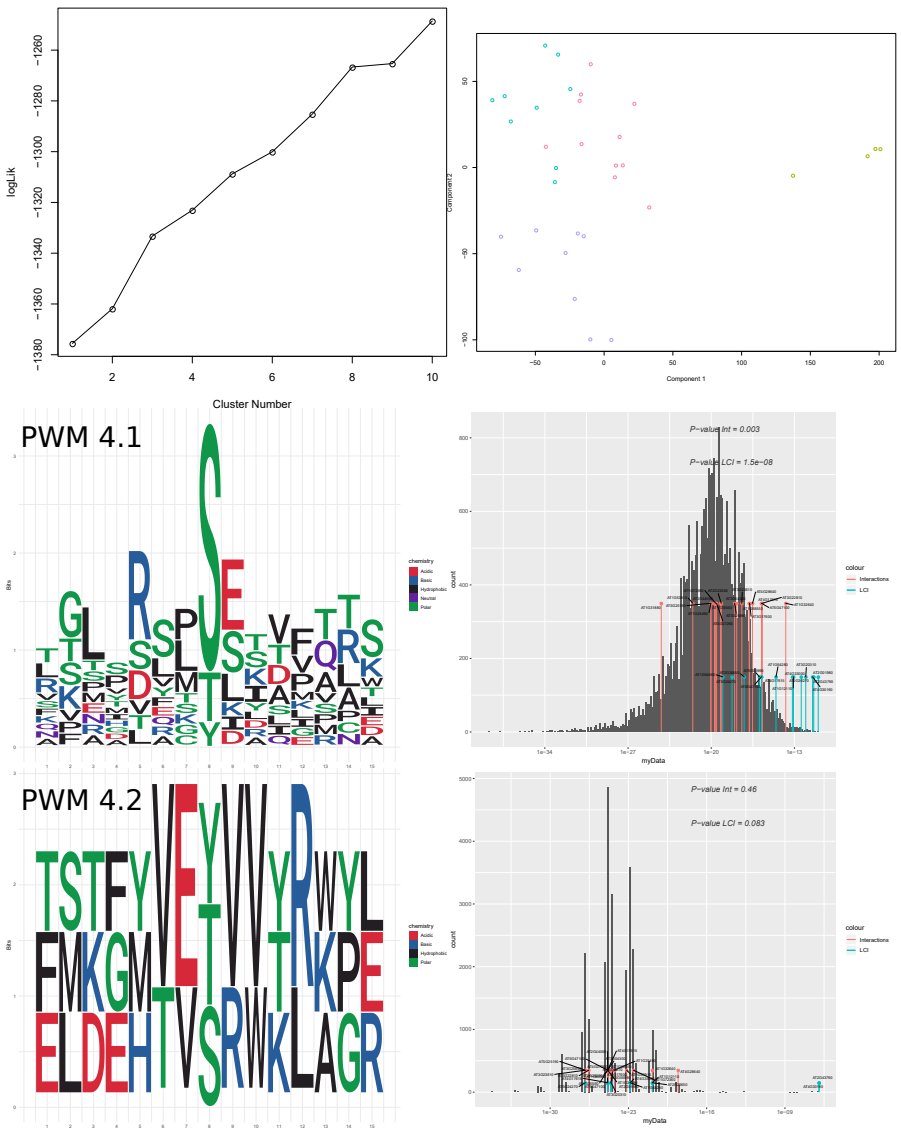


Figure A.21: SnRK3 Results

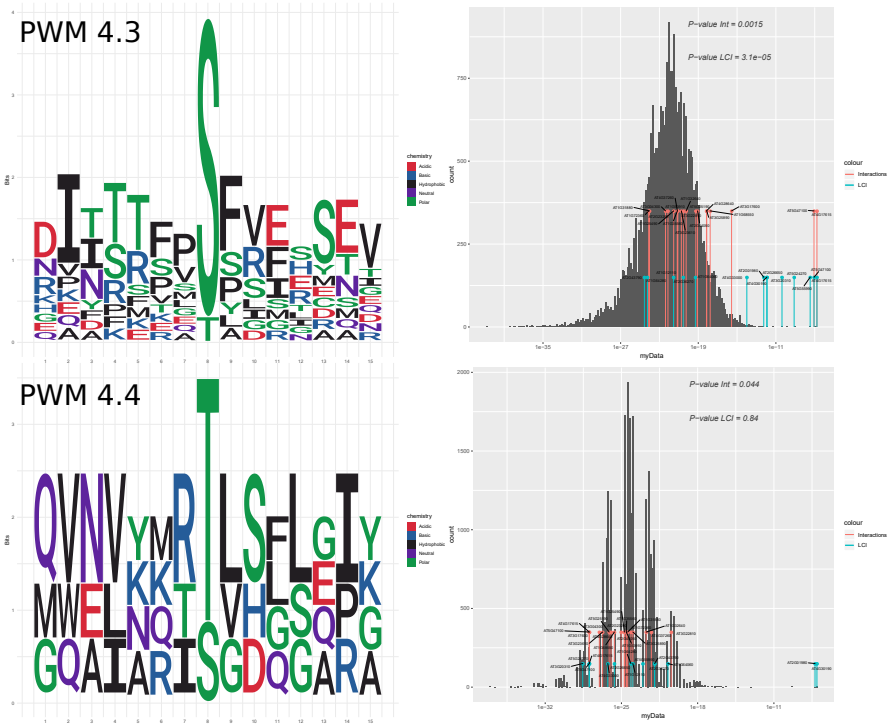


Figure A.22: SnRK3 Results continued

A.19 Repressor proteins

Table A.44: Repressor proteins manually curated from literature by Melina Altmann. Locus IDs are listed with the respective symbol and phytohormone pathway.

Hormone	Symbol	Locus ID	Hormone	Symbol	Locus ID
Cytokinin	ARR3	AT1G59940	Auxin	IAA1	AT4G14560
Cytokinin	ARR4	AT1G10470	Auxin	IAA2	AT3G23030
Cytokinin	ARR6	AT5G62920	Auxin	IAA3	AT1G04240
Cytokinin	ARR7	AT1G19050	Auxin	IAA4	AT5G43700
Cytokinin	ARR8	AT2G41310	Auxin	IAA5	AT1G15580
Cytokinin	ARR9	AT3G57040	Auxin	IAA6	AT1G52830
Cytokinin	ARR15	AT1G74890	Auxin	IAA7	AT3G23050
Cytokinin	ARR16	AT2G40670	Auxin	IAA8	AT2G22670
Cytokinin	ARR17	AT3G56380	Auxin	IAA9	AT5G65670
Cytokinin	AHP6	AT1G80100	Auxin	IAA10	AT1G04100
Ethylene	ERF3	AT1G50640	Auxin	IAA11	AT4G28640
Ethylene	ERF4	AT3G15210	Auxin	IAA12	AT1G04550
Ethylene	ERF7	AT3G20310	Auxin	IAA13	AT2G33310
Ethylene	ERF8	AT1G53170	Auxin	IAA14	AT4G14550
Ethylene	ERF9	AT5G44210	Auxin	IAA15	AT1G80390
Ethylene	ERF10	AT1G03800	Auxin	IAA16	AT3G04730
Ethylene	ERF11	AT1G28370	Auxin	IAA17	AT1G04250
Ethylene	ERF12	AT1G28360	Auxin	IAA18	AT1G51950
Ethylene	CTR1	AT5G03730	Auxin	IAA19	AT3G15540
Brassinosteroid	BZR1	AT1G75080	Auxin	IAA20	AT2G46990
Brassinosteroid	BZR2/BES1	AT1G19350	Auxin	IAA26	AT3G16500
Brassinosteroid	BIN2	AT4G18710	Auxin	IAA27	AT4G29080
Jasmonic acid	JAZ1	AT1G19180	Auxin	IAA28	AT5G25890
Jasmonic acid	JAZ2	AT1G74950	Auxin	IAA29	AT4G32280
Jasmonic acid	JAZ3	AT3G17860	Auxin	IAA30	AT3G62100
Jasmonic acid	JAZ4	AT1G48500	Auxin	IAA31	AT3G17600
Jasmonic acid	JAZ5	AT1G17380	Auxin	IAA32	AT2G01200
Jasmonic acid	JAZ6	AT1G72450	Auxin	IAA33	AT5G57420
Jasmonic acid	JAZ7	AT2G34600	Auxin	IAA34	AT1G15050
Jasmonic acid	JAZ8	AT1G30135	Gibberellic acid	GAI	AT1G14920
Jasmonic acid	JAZ9	AT1G70700	Gibberellic acid	RGA	AT2G01570
Jasmonic acid	JAZ10	AT5G13220	Gibberellic acid	RGL1	AT1G66350
Jasmonic acid	JAZ11	AT3G43440	Gibberellic acid	RGL2	AT3G03450
Jasmonic acid	JAZ12	AT5G20900	Gibberellic acid	RGL3	AT5G17490
Salicylic acid	NIMIN1	AT1G02450			
Salicylic acid	NIMIN2	AT3G25882			
Salicylic acid	NIMIN3	AT1G09415			

A.20 Natural variation data

Table A.45: All loci of PhI_{ext}, which have a value outside the 95 Percentile in one of the tested selection measures. A tick marks, if the respective locus show hints for non-neutral selection in the respective selection measure.

Locus	Symbol	Tajima's D	Fu and Li's D	Fu and Li's F	Ramos R2	Wall's B	wall's Q	Tajima's D.Value	Fu and Li's D.Value	Fu and Li's F.Value	Ramos R2.Value	Wall's B.Value	wall's Q.Value
AT1G25490	RCN1	✓			✓			-2.48	-7.16	-5.88	0.00	0.07	0.10
AT1G54490	XRN4	✓			✓			-2.54	-8.19	-6.29	0.00	0.00	0.00
AT1G77470	RFC3	✓			✓			-2.44	-5.65	-5.08	0.00	0.02	0.04
AT2G01760	RR14	✓			✓			0.86	-3.81	-1.76	0.08	0.01	0.02
AT2G03710	SEP4	✓			✓			-2.42	-5.02	-4.62	0.00	0.00	0.00
AT2G04890	SCL21	✓	✓	✓		✓		-2.54	1.88	-8.59	0.00	0.13	0.14
AT2G18730	DGK3	✓						0.49	-2.07	-1.00	0.07	0.00	0.00
AT2G21630	AT2G21630	✓			✓			0.56	-7.10	-3.39	0.08	0.01	0.02
AT2G25000	WRKY60	✓			✓			-2.48	-7.55	-6.18	0.00	0.06	0.07
AT2G46900	AT2G46900	✓	✓	✓	✓			-2.48	0.49	-8.06	0.00	0.01	0.03
AT3G23610	DSPTP1	✓			✓			0.59	-2.00	-1.00	0.08	0.00	0.00
AT3G53710	AGD6	✓			✓			1.05	-3.20	-1.23	0.09	0.01	0.02
AT3G57290	EIF3E	✓			✓			-2.42	-5.11	-4.71	0.00	0.02	0.04
AT4G01026	PYL7	✓			✓			1.09	-3.37	-1.59	0.09	0.05	0.07
AT4G15560	CLA1	✓						-2.49	-4.85	-4.25	0.00	0.08	0.15
AT4G20130	PTAC14	✓			✓			0.60	-2.61	-1.11	0.08	0.00	0.00
AT5G22480	AT5G22480	✓			✓			-2.42	-6.46	-5.56	0.00	0.00	0.00
AT5G41070	DRB5	✓						-2.43	-4.50	-4.06	0.00	0.04	0.07
AT5G42050	AT5G42050	✓			✓			-2.45	-7.04	-5.78	0.00	0.00	0.00
AT5G43820	AT5G43820	✓			✓			0.49	-3.33	-1.40	0.08	0.00	0.00
AT5G45680	FKBP13	✓		✓	✓			0.57	-1.17	-0.46	0.08	0.00	0.00
AT5G49620	MYB78	✓			✓			-2.44	-4.99	-4.57	0.00	0.00	0.00
AT5G56580	MKK6	✓			✓			1.32	-3.07	-1.21	0.09	0.00	0.00
AT5G59450	AT5G59450	✓			✓			-2.48	-7.41	-5.88	0.00	0.07	0.08
AT5G65683	WAVH2	✓	✓					-2.43	0.09	-6.83	0.00	0.07	0.10
AT1G05710	AT1G05710		✓	✓				-0.05	0.40	0.26	0.06	0.00	0.00
AT1G07430	HAI2		✓	✓				-2.07	-9.65	-7.04	0.01	0.07	0.08
AT1G09810	ECT11		✓					-1.82	-8.87	-6.57	0.02	0.00	0.00
AT1G10760	AT1G10760		✓					-1.84	-9.46	-6.25	0.02	0.01	0.02
AT1G30880	AT1G30880		✓	✓			✓	-1.45	0.97	-0.13	0.02	0.10	0.20
AT1G32640	MYC2		✓	✓				-2.40	0.59	-7.37	0.01	0.12	0.14
AT1G76890	GT2		✓	✓				-2.34	0.46	-7.53	0.00	0.03	0.06
AT2G35940	BLH1		✓	✓				-2.39	0.42	-7.82	0.00	0.00	0.00
AT3G56090	FER3		✓	✓				0.02	0.21	0.16	0.06	0.00	0.00
AT4G14160	AT4G14160		✓					-2.42	-8.89	-6.63	0.00	0.00	0.00
AT4G18880	HSF A4A		✓	✓				-2.26	-9.30	-7.05	0.01	0.03	0.06
AT4G30080	ARF16		✓	✓				-2.29	-9.69	-7.18	0.01	0.05	0.09
AT4G31800	WRKY18		✓			✓	✓	-2.29	0.40	-1.20	0.01	0.59	0.59
AT5G07690	MYB29		✓	✓		✓	✓	-2.03	7.03	0.21	0.02	0.51	0.53
AT5G12190	AT5G12190		✓	✓				-1.20	0.17	-0.41	0.02	0.00	0.00
AT5G25340	AT5G25340		✓	✓				-2.04	-9.10	-6.90	0.01	0.00	0.00
AT5G25510	AT5G25510		✓					-1.96	-9.28	-6.72	0.01	0.02	0.05
AT5G44420	PDF1.2		✓	✓				-0.95	0.47	-0.09	0.03	0.00	0.00
AT5G62000	ARF2		✓	✓				-2.27	-9.81	-7.00	0.01	0.03	0.06
AT1G11430	MORF9			✓		✓	✓	-2.06	-8.47	-6.94	0.02	0.20	0.23
AT3G23240	ERF1			✓				0.44	-0.86	-0.33	0.07	0.06	0.12
AT4G18800	RABA1d			✓				0.36	-0.76	-0.33	0.07	0.00	0.00
AT4G22220	ISU1			✓				-1.10	0.03	-0.57	0.03	0.00	0.00
AT5G17300	RVE1			✓				0.25	-1.13	-0.56	0.07	0.01	0.03
AT5G18020	SAUR20			✓				-0.22	-0.71	-0.62	0.06	0.05	0.10
AT5G25190	ESE3			✓				-0.15	-0.75	-0.61	0.06	0.05	0.09
AT1G77200	AT1G77200				✓			-2.26	-5.73	-5.18	0.00	0.00	0.00
AT2G22880	AT2G22880				✓	✓	✓	-2.34	-3.49	-3.64	0.00	0.12	0.21
AT2G45820	AT2G45820				✓			-2.21	-5.25	-4.89	0.00	0.00	0.00
AT2G46130	WRKY43				✓			-2.20	-3.46	-3.59	0.00	0.00	0.00
AT3G18960	AT3G18960				✓			-2.30	-4.67	-4.44	0.00	0.00	0.00
AT3G29370	P1R3				✓			-2.28	-5.02	-4.66	0.00	0.03	0.07
AT3G45640	MPK3				✓			-2.39	-2.96	-3.29	0.00	0.02	0.04
AT3G50800	AT3G50800				✓			-2.13	-2.58	-2.94	0.00	0.00	0.00
AT3G61860	RS31				✓			-2.19	-5.03	-4.71	0.00	0.00	0.00
AT5G19000	BPM1				✓			-2.38	-6.75	-5.75	0.00	0.00	0.00
AT5G50180	AT5G50180				✓			-2.23	-5.01	-4.69	0.00	0.04	0.08
AT5G52010	AT5G52010				✓			-2.27	-6.61	-5.86	0.00	0.00	0.00
AT5G65210	TGA1				✓			-2.31	-6.02	-5.36	0.00	0.00	0.00
AT1G72450	JAZ6					✓	✓	-2.36	-8.55	-6.80	0.01	0.20	0.22
AT3G53250	SAUR57					✓	✓	-1.59	-1.18	-1.62	0.00	0.16	0.28
AT3G57720	AT3G57720					✓		-1.82	-6.63	-5.19	0.02	0.13	0.18
AT4G17615	CBL1					✓	✓	-2.20	-3.28	-3.44	0.00	0.18	0.32
AT4G19200	AT4G19200					✓		-1.88	-5.79	-5.06	0.01	0.13	0.17
AT4G39540	SK2					✓		-2.26	-5.20	-4.73	0.00	0.14	0.17
AT5G48870	SAD1					✓	✓	-1.92	-6.09	-5.52	0.02	0.42	0.50
AT5G57860	AT5G57860					✓	✓	-1.49	-0.44	-0.99	0.00	0.25	0.40

Table A.45 continued from previous page

Locus	Symbol	Tajima's D	Fu and Li's D*	Fu and Li's F*	Ramos R2	Wall's B	Wall's Q	Tajima's D	Fu and Li's D*	Fu and Li's F*	Ramos R2	Wall's B	Wall's Q
AT3G62550	AT3G62550						✓	-0.79	-1.75	-1.70	0.03	0.12	0.22
AT5G14600	AT5G14600						✓	-1.74	-1.99	-2.29	0.02	0.11	0.20

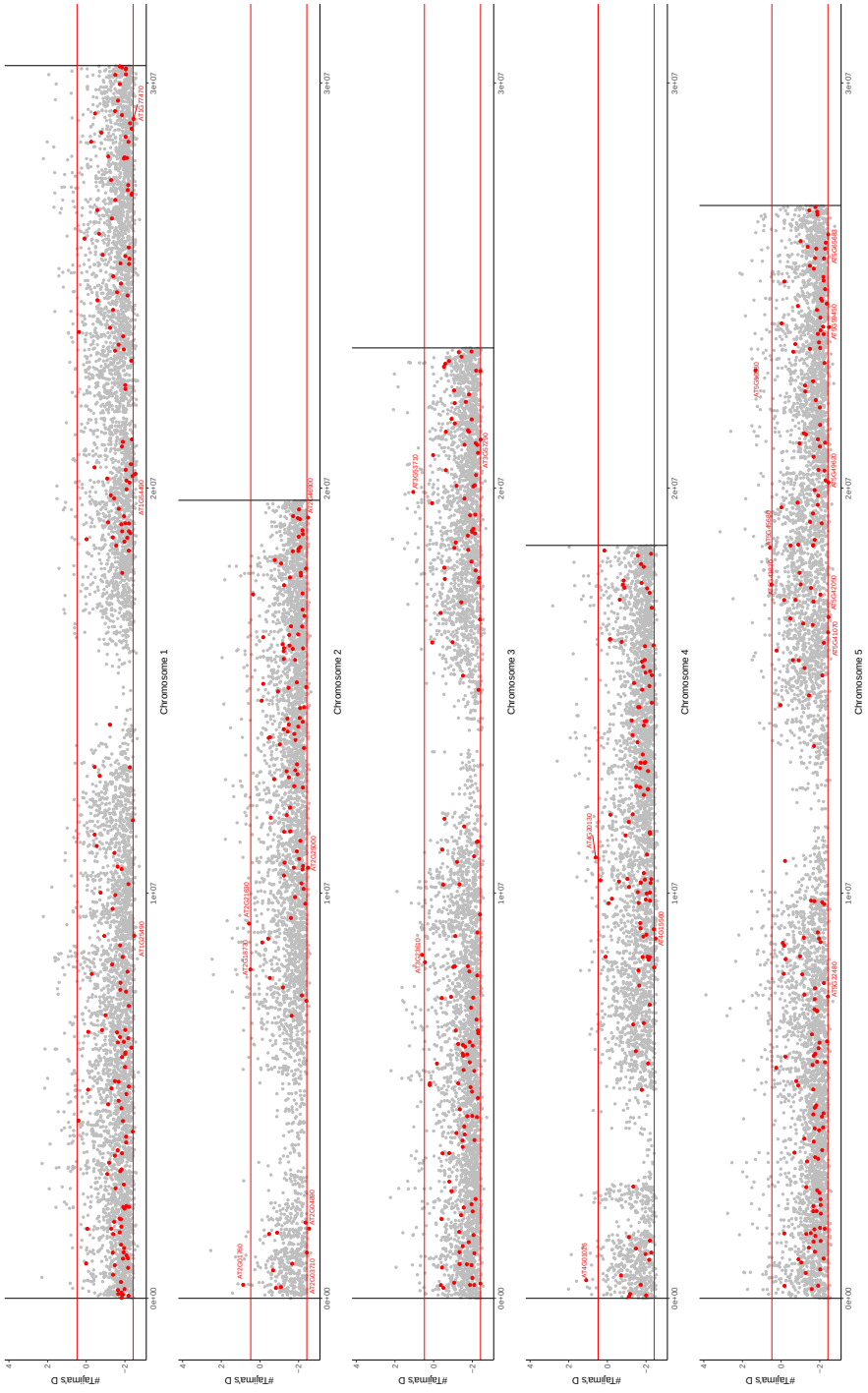


Figure A.23: Tajima's D values of coding sequences of Arabidopsis thaliana primary gene models. Proteins contained in $\mathrm{PhI}_{\mathrm{ext}}$ are in red, others are in gray. Red lines mark 95 Percentile of Tajima's D values. Proteins in $\mathrm{PhI}_{\mathrm{ext}}$ are labeled with locus ID.

A.21 PPIN2 network

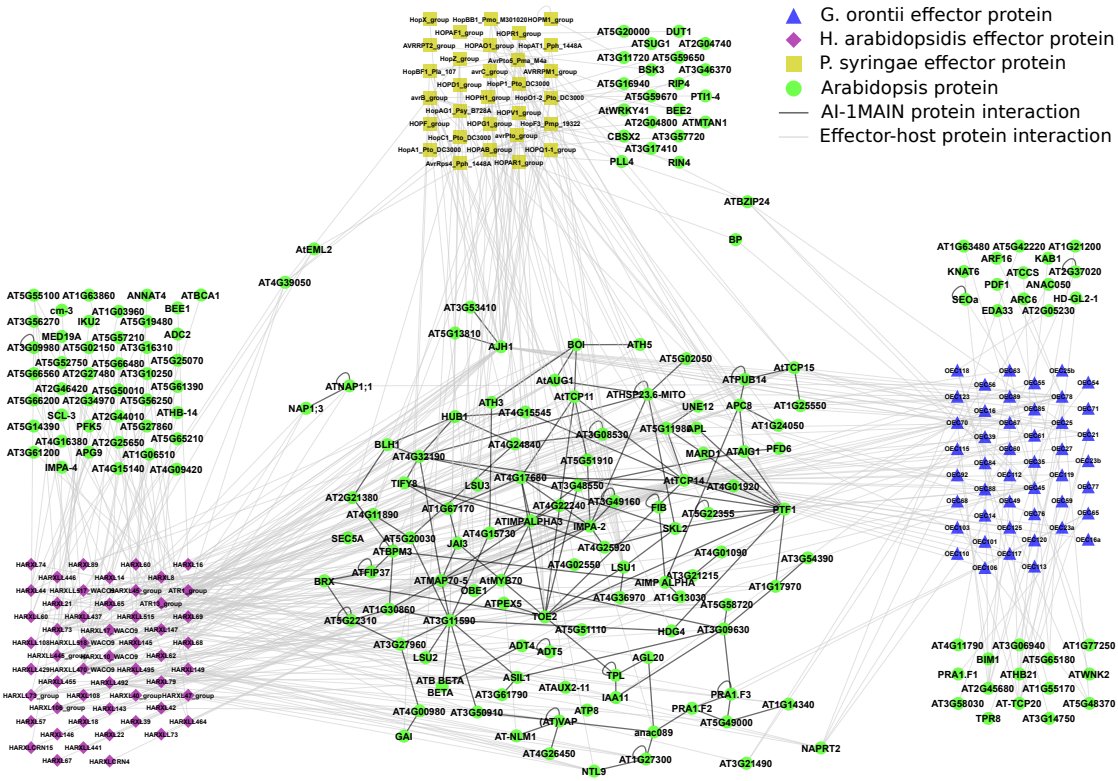


Figure A.24: Merged effector - host protein interactions from *G. orontii* effectors (blue), *H. arabidopsidis* effectors (violet), and *P. syringae* (yellow). Effector - host interactions of *H. arabidopsidis* and *P. syringae* are from [139]. Interactions between host proteins are from AI-1_{MAIN} [11].

A.22 PPIN2 phenotyping results

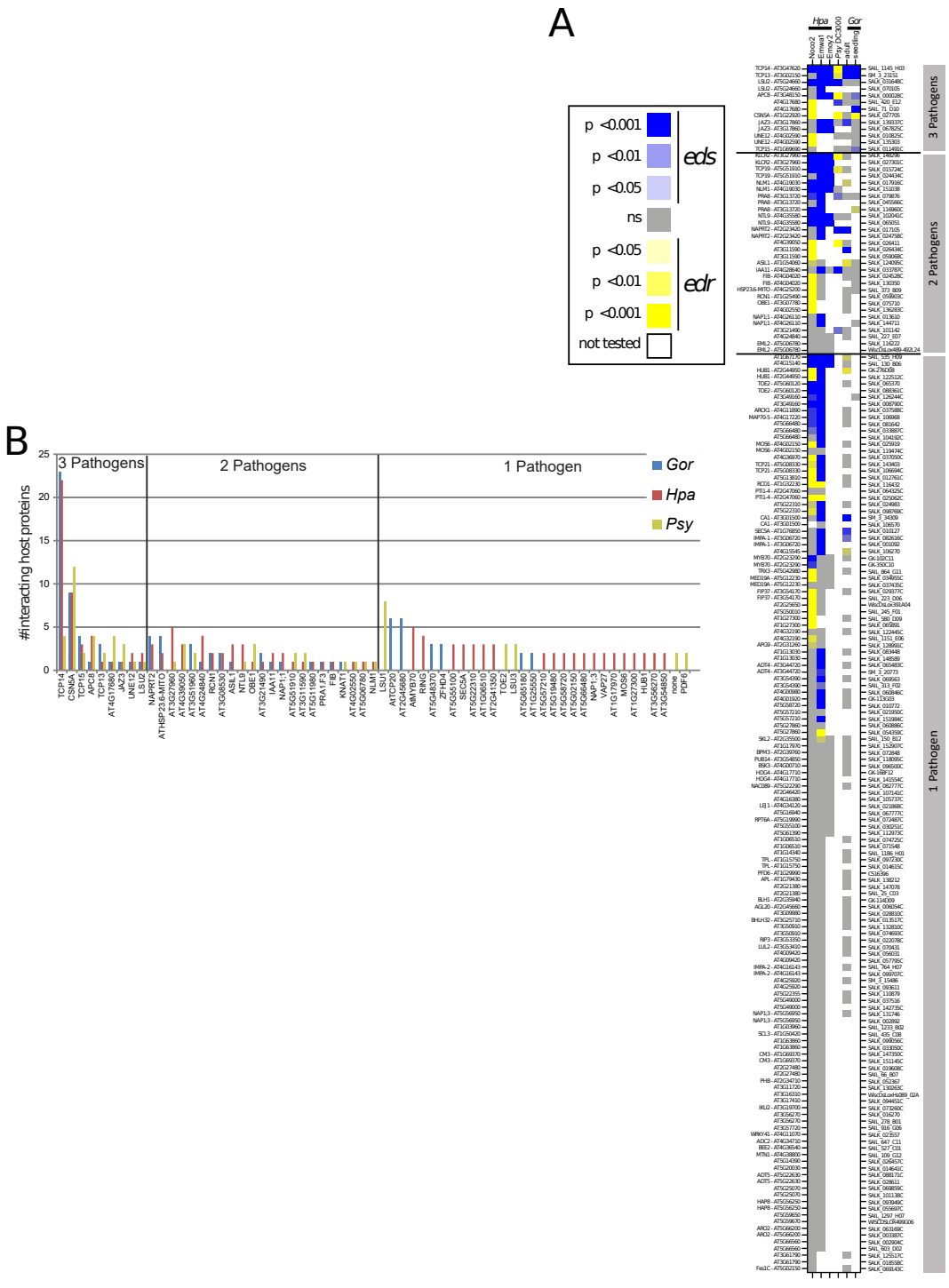


Figure A.25: A) Phenotyping results of all T-DNA lines. Heatmap shows P-value for all tested T-DNA lines against three different *Hyaloperonospora arabidopsidis* strains (Hpa), for *Pseudomonas syringae* (Psy) and for *Golovinomyces orontii* tested on seedlings and adult plants. B) Number of effector proteins binding to host proteins.

A.23 PPIN2 natural variation

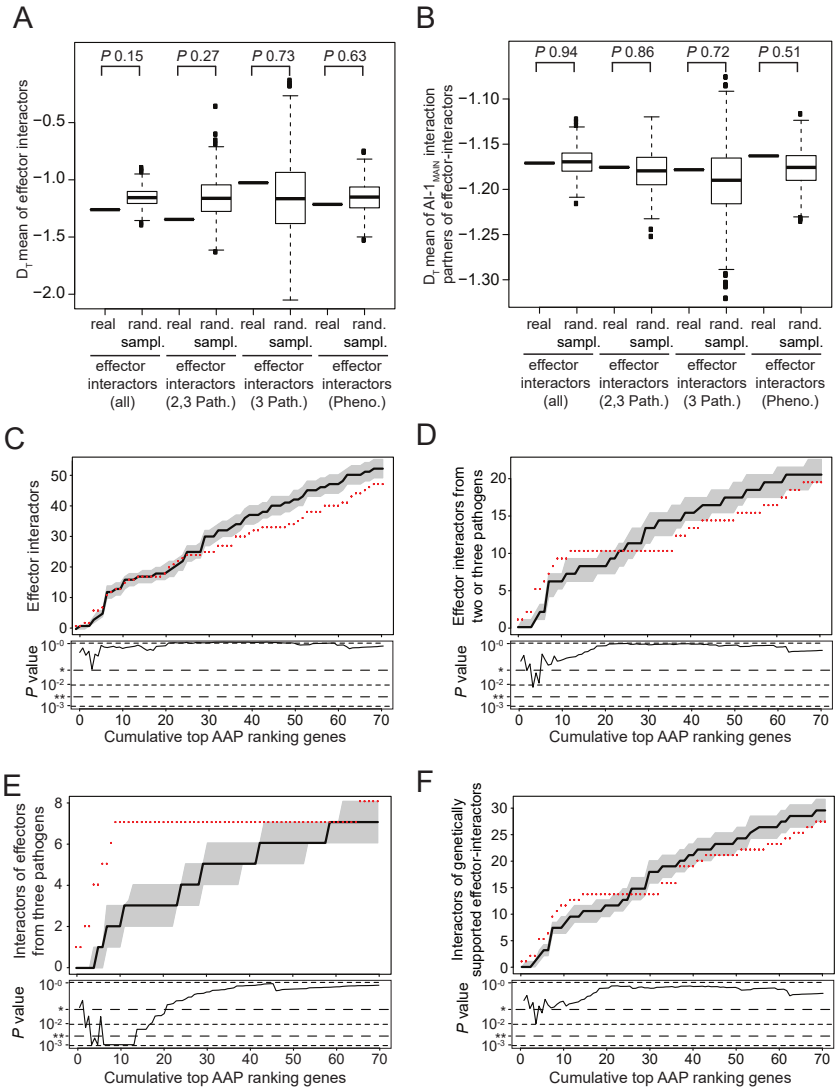


Figure A.26: A) Comparison Tajima's D (D_T) mean values of all effector interactors (all), effector interactors with interspecies convergence of 2 or 3 pathogens (2, 3 Path.), effector interactors with interspecies convergence of 3 pathogens (3 Path.), and effector interactors which show a phenotype in infection assay (Pheno.) against distribution of means values from 1,000 random samples of AI-1_{MAIN}. B) As in A, but comparison of *Arabidopsis* proteins interacting with effector targets. C) Cumulative number of top ranking AAP proteins interacting with effector interacting proteins in AI-1_{MAIN} compared to observed number in 1,000 randomly rewired networks. Y-axis shows number of effector interactors interacting with the number of top AAP ranking genes shown on x-axis. Red dotted line represents real data determined from AI-1_{MAIN}, black line shows mean value from 1,000 degree-preserved randomly rewired AI-1_{MAIN} networks, gray area shows value between 25^{th} and 75^{th} percentile. The lower panel provides the experimentally determined p-value (* 0.05; ** 0.005) for each number of considered top ranking AAP genes. D) As in C, but number of interacting effector proteins, targeted by at least two pathogens are determined per number of top ranking AAP genes. E) As in C, but number of interacting effector proteins, targeted by three pathogens are determined per number of top ranking AAP genes. F) As in C, but number of interacting effector proteins having an immune phenotype in infection assays are determined per number of top ranking AAP genes.