bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Widespread targeting of development-related host transcription factors by phytoplasma effectors

Miguel Correa Marrero[1], Sylvain Capdevielle[2], Weijie Huang[2], Marco Busscher[3,4], Jacqueline Busscher-Lange[3,4], Dick de Ridder[1], Aalt D.J. van Dijk[1,5], Saskia A. Hogenhout[2], and Richard G.H Immink[3,4]

[1]Bioinformatics Group, Wageningen University and Research, 6708 PB Wageningen, the Netherlands
[2]Department of Crop Genetics, John Innes Centre, Norwich Research Park, Norwich, Norfolk, NR4 7UH, United Kingdom
[3]Laboratory of Molecular Biology, Wageningen University and Research, 6708 PB Wageningen, The Netherlands
[4]Bioscience, Wageningen Plant Research, Wageningen University and Research, 6708 PB Wageningen, The Netherlands
[5]Biometris, Wageningen University and Research, 6708 PB Wageningen, the Netherlands

February 13, 2020

1 Abstract

Phytoplasmas are pathogenic bacteria that reprogram plant host development in order to attract their insect vectors to disseminate. Previous studies have characterized a few different phytoplasma effector proteins that supress specific transcription factors. However, these are only a small fraction of the potential effectors used by phytoplasmas, meaning that the molecular mechanisms through which phytoplasmas manipulate their hosts are largely uncharacterized. To obtain further insights into the phytoplasma infection mechanisms, we generated a protein-protein interaction network between a broad set of phytoplasma effectors and a large collection of *Arabidopsis thaliana* transcription factors and transcriptional regulators. We found widespread, but unique, interactions with host transcription factors by phytoplasma effectors, especially those related to developmental processes. In particular, many unrelated effectors target TCP transcription factors, which play roles in plant development

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and immunity. Comparison with other host-pathogen protein interaction networks shows that phytoplasma effectors have unusual targets, and indicates that phytoplasmas have evolved a unique and unusual infection strategy. This study provides a rich and solid data source that can be used to predict functional effects of individual effectors and as a guide for detailed studies of individual effectors in the future, as well as insights into the underlying molecular mechanisms of phytoplasma infection.

2 Significance statement

This work shows that the effectors of phytoplasma, a bacterial plant pathogen, show pervasive interactions with development-related host transcription factors, providing a way to take over plant growth and development in favor of the pathogen and its insect vector. The obtained comprehensive protein interaction network and showcase of the potential biological consequences of a selected effector advance our understanding of phytoplasma-host interactions and provide guidance for further study.

3 Introduction

Amongst the vast amount of plant pathogenic microorganisms, some have the intriguing ability to manipulate the development of their hosts in order to increase their own fitness. A remarkable, well-known example is that of the rust fungus *Puccinia monoica*, which alters the morphology of its host to create pseudoflowers covered by fungal spermatogonia. These pseudoflowers successfully attract pollinators, which then spread the fungal reproductive cells (Roy, 1993). The complex and diverse molecular underpinnings through which pathogens modify host development have been studied in recent years (Le Fevre *et al.*, 2015), but we are far from a complete understanding of most mechanisms.

The phytoplasmas, obligate bacterial plant pathogens, represent another example of a pathogen partially taking over plant development, and manipulating development of their plant hosts via processes that are yet to be fully understood. Phytoplasmas have a life cycle that alternates between plants and specific insect herbivores, such as leafhoppers (Weintraub and Beanland, 2006), that propagate them to new plant hosts. By feeding on the phloem of infected plants, the insect vectors become carriers, and once their salivary glands have been colonized, they are able to transmit the pathogen to other plants they feed on (Sugio *et al.*, 2011a). The infected plants display an altered morphology. For example, the Aster Yellows phytoplasma strain Witches' Broom (AY-WB) induces phyllody (conversion of flowers into leaf-like structures), virescence (green coloration of non-green floral tissue) and witches' brooms (increased proliferation of stems, branches and leaves) (MacLean *et al.*, 2011; Sugio *et al.*, 2011b). These alterations make the infected plants more attractive to the insect vectors (MacLean *et al.*, 2014; Orlovskis and Hogenhout, 2016). Additionally, infected

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

plants are often sterile, and thus they serve the sole purpose of propagating the bacteria. While this phenomenon alone makes phytoplasmas very interesting pathogens, they are not merely a scientific curiosity, but are also of socioeconomic importance: they have a broad host range (Firrao *et al.*, 2007), and their outbreaks cause significant yield losses in numerous crops of high economic importance, such as wheat, grapes and coconuts (Strauss, 2009; Marcone, 2019).

Plant pathogens produce effectors, proteins to hijack the host cell for improving the pathogen's fitness in the host by different means (e.g. by modulating the host immune response) (Mattoo *et al.*, 2007). Three different phytoplasma effectors, SAP54, SAP11 and TENGU, have been characterized, shedding light on the molecular mechanisms through which phytoplasmas manipulate their hosts. SAP54 mediates degradation of MIKC MADS-box transcription factors involved in floral development and induces phyllody (MacLean *et al.*, 2014). SAP11 binds and destabilizes certain TCP transcription factors involved in axillary meristem outgrowth, leaf shape determination and jasmonate signalling, leading to an altered morphology and decreased plant defense (Sugio *et al.*, 2011b; Chang *et al.*, 2018; Wang *et al.*, 2018). Finally, TENGU causes dwarfism, witches' broom and arrested flower development. Although its molecular mechanism has not been clearly identified, this effector downregulates expression of auxin response factors ARF6 and ARF8, leading to decreased jasmonate biosynthesis (Minato *et al.*, 2014). These alterations caused by individual effectors reflect the symptoms caused by phytoplasma infections, and, overall, suggest a trend of targeting plant transcription factors in order to exploit the host.

However, these three effectors are only a small part of the potential effector arsenal of phytoplasmas. Sequencing of phytoplasma genomes has allowed the identification of many more candidate effectors that the pathogen might use to colonize its hosts (Bai *et al.*, 2009), some of which are preferentially expressed in the plant and others in the insect (MacLean *et al.*, 2011). Given the precedent of targeting host transcription factors in order to manipulate plant development, we asked how prevalent this phenomenon is across phytoplasma effectors expressed in plants. In order to identify interactions between effectors and host transcription factors, we performed large-scale yeast-two hybrid (Y2H) assays, testing interactions between 21 phytoplasma effectors and a comprehensive library of *Arabidopsis thaliana* transcription factors and transcriptional regulators (Pruneda-Paz *et al.*, 2014). Such an approach is useful to obtain a systems-level view of host-pathogen interactions (Tripathi *et al.*, 2019; Rodriguez *et al.*, 2019), which is not yet well-developed in phytoplasma.

Overall, the resulting protein-protein interaction (PPI) network shows pervasive interactions of candidate phytoplasma effectors with plant host transcription factors, especially with those that regulate plant development. Furthermore, we find that many unrelated effectors interact with multiple TCP transcription factors, known to be targeted by diverse effector proteins of bacterial, fungal and oomycete pathogens (Mukhtar *et al.*, 2011; Weßling *et al.*, 2014). Compared to previously determined plant pathogen-host PPI networks, phytoplasma effectors more intensely target plant growth and development.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4 Materials and methods

4.1 Yeast-two hybrid assays

The genes encoding the 21 selected phytoplasma effector proteins were obtained by amplifying the genomic DNA of AY-WB phytoplasma-infected aster plants and cloned into pDONR207. The effector cDNAs were sub-cloned from pDONR207 into the Y2H bait vector pDEST32 by Gateway-based recombination. The resulting effector bait plasmids were transformed into yeast strain PJ69-4 mating type Alpha, followed by a test for autoactivation of reporter genes, as described previously (De Folter and Immink, 2011). Subsequently, matrix-based Y2H screenings were performed following the protocol in (De Folter and Immink, 2011). Each bait was screened against the *Arabidopsis* transcription factor collection, consisting of 1980 clones (Pruneda-Paz *et al.*, 2014). Protein interaction events were scored after incubation of yeast at 20 °C for 6 days. For the baits that did not exhibit autoactivation, incubation was done on SD medium lacking Leucine, Tryptophan, and Histidine and supplemented with either 1 or 5 mM 3-amino-1,2,4-triazole (3-AT), respectively. For SAP06 and SAP48, which showed autoactivation up to 15 mM 3-AT, the incubation was done on SD medium lacking Leucine, Tryptophan, and Histidine and supplemented with either 20 or 25 mM 3-amino-1,2,4-triazole (3-AT), or SD medium lacking Leucine, Tryptophan, and Adenine, respectively.

4.2 Host protein family annotation

We assigned each host protein to a particular family by first checking its assignment in PlantTFDB, a comprehensive database of plant transcription factors (Jin *et al.*, 2016). However, not all host proteins could be found in this database. The host library is based on proteins from different resources (Pruneda-Paz *et al.*, 2014), including PlnTFDB (Pérez-Rodríguez *et al.*, 2009), which is more general and contains not only transcription factors, but also other transcriptional regulators. Therefore, we used PlnTFDB to assign families to each host protein absent in PlantTFDB. In total, we were able to assign 1830 of the 1980 proteins to a unique family, 1434 found in PlantTFDB and 396 in PlnTFDB. The remaining proteins are simply marked as unannotated.

4.3 GO enrichment analyses

Gene Ontology (GO) enrichment analyses were performed using goatools v0.8.4 (Klopfenstein *et al.*, 2018). Genes annotated with a GO term were automatically annotated with the parent terms as well, and multiple testing correction was performed with the Benjamini-Hochberg (BH) method (false discovery rate threshold: 0.05) (Benjamini and Hochberg, 1995). 98% of the proteins in the host library were present in the GO annotation file. All of the analyses excluded terms inferred from electronic annotation (i.e. those with the evidence code IEA).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4.4 Effector comparison

In order to compare the interaction patterns of different effectors, we computed all pairwise Jaccard similarities (Jaccard, 1912). The Jaccard similarity is defined as the size of the intersection between sets (common interactions and non-interactions) divided by the size of the union of the sets (the whole set of interactions and non-interactions with host proteins that displayed at least one interaction in our Y2H assay). We also computed all possible pairwise sequence alignments (after removing signal peptides) using Clustal Omega v1.2.4 (Sievers *et al.*, 2011) and calculated the corresponding sequence similarities. Sequence similarity is defined as the proportion of aligned residues with a log-odds score greater than 0 in the BLOSUM62 substitution matrix (Henikoff and Henikoff, 1992). Finally, all dendrograms shown were constructed using average linkage hierarchical clustering.

4.5 Statistical analysis of the phytoplasma-*Arabidopsis* PPI network

As there might be an overall trend of effectors interacting predominantly with specific host protein families, we tested whether each host protein family is enriched in the PPI network with respect to the host protein library. We did this using Fisher's exact test. We performed multiple testing correction using the BH method (false discovery rate threshold: 0.05). Likewise, individual effectors might interact preferentially with certain host protein families with respect to the rest of the network. This was tested using the same procedure.

4.6 PPI network comparison

We compared the host-pathogen interaction networks presented in Mukhtar *et al.* (2011) and Weßling *et al.* (2014) to ours. For many years, the Y2H system has been the gold standard for interactomics studies (Brückner *et al.*, 2009). Nevertheless, variation, as well as false negative and false positive interactions, might be identified due to differences in the exact screening conditions, such as the yeast strain, vectors, incubation temperature and exact composition of selective media. For this reason, we focused on an overall comparison, rather than zooming in on unique individual cases. Furthermore, the data of these previous studies contains no information about the host protein splicing variants used. Therefore, we had to collapse different splicing variants in our network into one node. There is only one host protein within our PPI network with two different splicing variants, SEPALLATA4. One variant interacts with 5 different effectors, and the second one interacts with a subset of 3 of the same effectors. Therefore, ignoring different splicing variants has a small or negligible effect in the comparison.

We selected the 753 host proteins that were present in all Y2H assays. Within this subset, 528 proteins did not interact with any effector from any species and were removed. Likewise, 7 effectors that did not show interactions with the

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

remaining host proteins were removed. The final integrated network contains 118 effectors (19 from phytoplasma, 21 from *P. syringae*, 48 from *H. arabidopsidis* and 30 from *G. orontii*) and 225 host proteins. We performed a pairwise comparison of host protein degree per species using this network, taking into account host proteins that are relatively highly targeted by phytoplasma ($\geq$4 interactions; this is the top 7% of host proteins targeted by phytoplasma) but not in other species.

4.7 TCP phylogenetic tree construction

We constructed a phylogenetic tree of all TCP transcription factors in *A. thaliana* using their full-length sequences with the ete toolkit (v 3.1.1) (Huerta-Cepas *et al.*, 2016). The best model from JTT, WAG, VT, LG and mtREV was chosen using ProtTest (Abascal *et al.*, 2005), and a maximum likelihood tree was built using PhyML (Guindon *et al.*, 2010). Reliability of branching was assessed with 100 bootstrap replicates.

4.8 Generation of SAP06 transgenic plants

A synthetic plant codon-optimized ORF was created for AY-WB SAP06 by GenScript Biotech (New Jersey, US) and subcloned by Gateway-based cloning into the entry vector pDONR207. The obtained donor vector was sequenced to confirm that the SAP06 ORF was correct and subsequently, a plant expression vector was generated by Gateway-based recombination. In this reaction, SAP06 was transferred into the destination vector pB7WG2 (Karimi *et al.*, 2002), containing a CaMV 35S promoter-driven expression cassette. The obtained expression vector was transferred to *Agrobacterium tumefaciens* strain C58C1 and transformed into *Arabidopsis* Col0 by the floral dip method (Clough and Bent, 1998). Primary transformants were obtained upon selective germination of seeds harvested from the transformed plants on agar plates containing 10 mg/l phosphinotrycin (PPT). From these individual lines, six 3:1 segregating lines were selected based on selective germination on PPT medium, followed by selection of homozygous lines in the next generation. Subsequenty, qRT-PCR was used to select a transgenic line with a high ectopic expression level (SAP068) and a line with an intermediate expression level (SAP064). Primers used for this assay are PZN1326 'TGATGGTTGCTATCTCTAACACT' and PZN1327 'GGCTTGTTCTGGTAGTTTCTTCT'. For all further phenotyping experiments, fresh seeds of these two lines were used.

4.9 Phenotyping of SAP06 ectopic expression lines

For the phenotyping of seed dormancy, seeds of the two selected SAP06 ectopic expression lines and of wild type Col0 were sown on blue filter paper wetted with MQ water, in 10 cm Petri dishes, followed by scoring of the percentage of germination after 6 days of incubation at either 21 °C or 25 °C. The remaining seeds were stored at room temperature (20-21 °C) and sown in a similar way

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

after one and two weeks of storage. Germination of these batches was scored as well at 6 days after start of the germination assay. Since all seeds were germinating after two weeks of storage, we stopped scoring for dormancy at this point.

For the characterization of potential effects on vegetative development, seeds of the two SAP06 ectopic expression lines and of Col0 wild type were sown on wetted filter paper and stratified for three days at 4 °C. Subsequently, seeds were sown on Rockwool plugs. Twenty seedlings of each genetic background were grown under long day conditions (16/8, light/dark) at 21 °C. Three, four and five weeks after germination, pictures were taken of representative individual plants of each genetic background.

5 Results & discussion

5.1 Widespread interactions between phytoplasma effectors and *A. thaliana* transcription factors

To obtain a better understanding of the molecular mechanisms through which phytoplasmas infect and manipulate their host, we performed a large-scale matrix-based Y2H screening to find putative interaction partners of 21 phytoplasma effectors in an existing library containing 1980 *A. thaliana* transcription factors and transcriptional regulators (Pruneda-Paz *et al.*, 2014). The library contains a broad and diverse set of host proteins containing 95 different families, with no clearly predominating group (Fig. S1).

Although *in planta* activity as effectors has not been validated for all of the 21 putative effector proteins, this activity has been suggested based on various criteria (MacLean *et al.*, 2011). Therefore, we will hereafter refer to these putative effectors as effectors. 17 of these effector proteins were chosen from phytoplasma strain AY-WB based on significant up-regulation of the candidate genes encoding for these effectors in *A. thaliana* compared to expression levels of these phytoplasma genes in the vector *M. quadrilineatus* (MacLean *et al.*, 2011). The other four selected SAP proteins represent orthologs of the previously characterized AY-WB SAP11 and SAP54 proteins from other phytoplasma isolates. Initially, an auto-activation test was performed for the 21 effector protein used as bait proteins. Autoactivation was only found for SAP06 and SAP48 on selective SD medium supplemented with up to 15 mM 3-amino-1,2,4-triazole (3-AT). For this reason, these baits were screened in the Y2H assay with the more selective Adenine reporter and on medium supplemented with higher concentrations of 3AT.

For the previously studied AY-WB SAP11 and SAP54 proteins we identified various new interactions besides the already known transcription factor targets. We identified, for example, known interactions between SAP11 and TCP2, TCP7 and TCP13 (Sugio *et al.*, 2011b, 2014). Likewise, the screen identifies interactions between SAP54 and different MIKC MADS-box proteins involved in floral organ specification and determination of flowering time (such

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

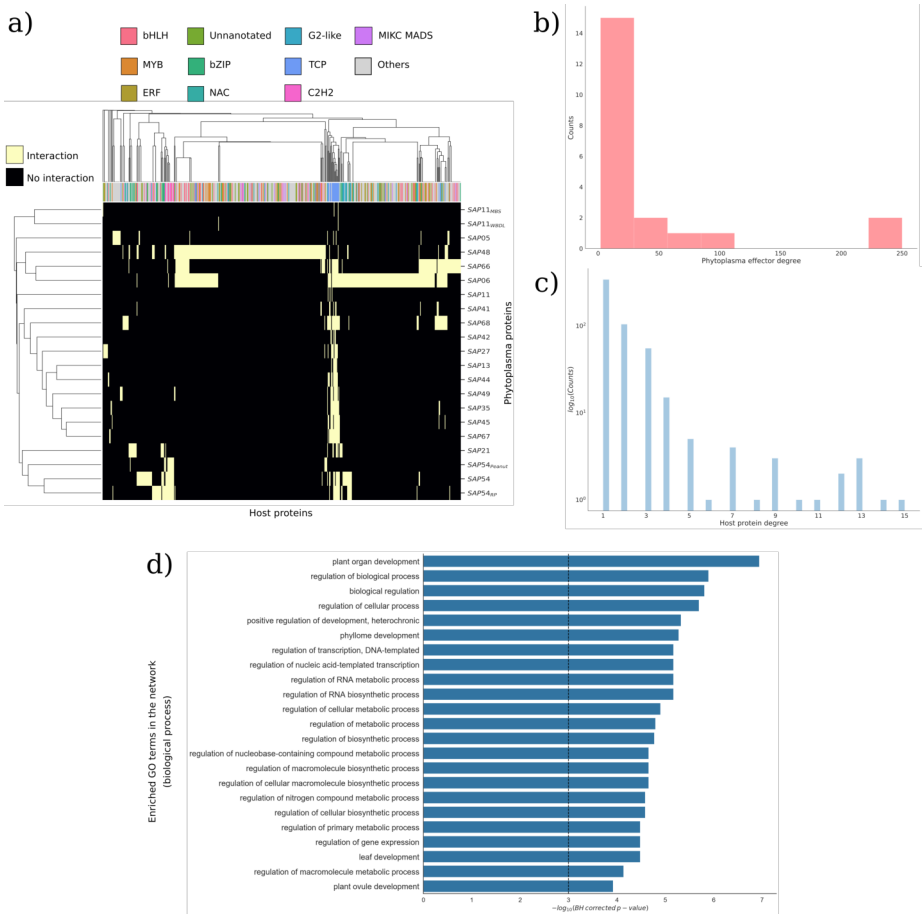

Figure 1: Overview of the obtained phytoplasma-*Arabidopsis* interaction network. (a) Clustermap of the interactions, grouping together host proteins (x-axis) and effectors (y-axis) with similar interaction patterns. Colors indicate family membership of the corresponding host proteins for the 10 most abundant families in the network. Clustering was performed using Jaccard distance as a metric. (b) Histogram of candidate effector degrees. (c) Histogram of host protein degrees (with counts in log scale). (d) Significantly enriched biological process GO terms in the interaction network, sorted from lowest (above) to highest (below) p-value. Vertical dashed line indicates the statistical significance threshold (adjusted $p < 0.05$).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

as SOC1, AP1, and SEP3) (MacLean *et al.*, 2014). For all other tested effector proteins, various interacting host proteins were identified, and the resulting pathogen-host PPI network contains 979 interactions, involving 536 (˜28%) of the screened host proteins (Fig. 1a). All of the effectors display at least two interactions, but the degree distribution is broad (Fig. 1b). Remarkably, two related effectors, SAP48 and SAP06, interact with a large number of host proteins (250 and 242, respectively) from many different families. Many of these host proteins do not interact with any other effector we have tested.

There is also large variation in the number of interactions shown by host proteins. Although ˜63% of the interacting host proteins only bind to one effector, some of them are clearly highly targeted (Fig. 1c). Many of these highly targeted host proteins belong to the TCP transcription factor family (Fig. 1a): the 10 host proteins with the highest degree are all TCPs. Other host proteins with a high degree include LBD15, a protein involved in shoot apical meristem development (Sun *et al.*, 2013), and two proteins of unknown function: a homeobox and a NAC transcription factor (AT4G03250.1 and AT3G12910.1, respectively). This high amount of interactions suggests an important role in phytoplasma infection.

We performed GO enrichment analysis to determine whether there is an overrepresentation of host proteins in the network that participate in particular functions with respect to the library. Aside from general, high-level terms, this revealed a highly significant enrichment in different *biological process* terms related to plant development (Fig. 1d), and in particular in terms related to the development of phyllomes (i.e., organs homologous to leaves or derived from leaves, such as flowers), which is consistent with known phytoplasma infection symptoms. Altogether, this indicates that the network captures biologically relevant interactions. Note that there is also an enrichment in terms related to e.g. transcription factor activity (Fig. S2). As the host protein library contains other transcriptional regulators than transcription factors, this enrichment indicates that, within the screened library, the effectors mostly interact with transcription factors.

5.2 Multiple unrelated effectors specifically target TCP transcription factors

To investigate whether phytoplasma effectors interact specifically with certain protein families, we evaluated whether each family is enriched in the network with respect to the library. We find that only one family, the TCP transcription factors, is enriched in the network (adjusted $p = 2.5 \times 10^{-2}$). Of the 24 TCPs present in the *A. thaliana* genome, which are all present in the library, 20 are found in the network; effectors we have not screened might interact with the remaining TCPs. Furthermore, these interactions are distributed over the whole TCP family, without particular enrichment of a specific class or subclass (Fig. 2a). TCPs form an ancient, plant-specific family of transcription factors, found from green algae to eudicots (Floyd and Bowman, 2007; Navaud *et al.*, 2007). Although TCPs were initially linked to plant growth and development

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

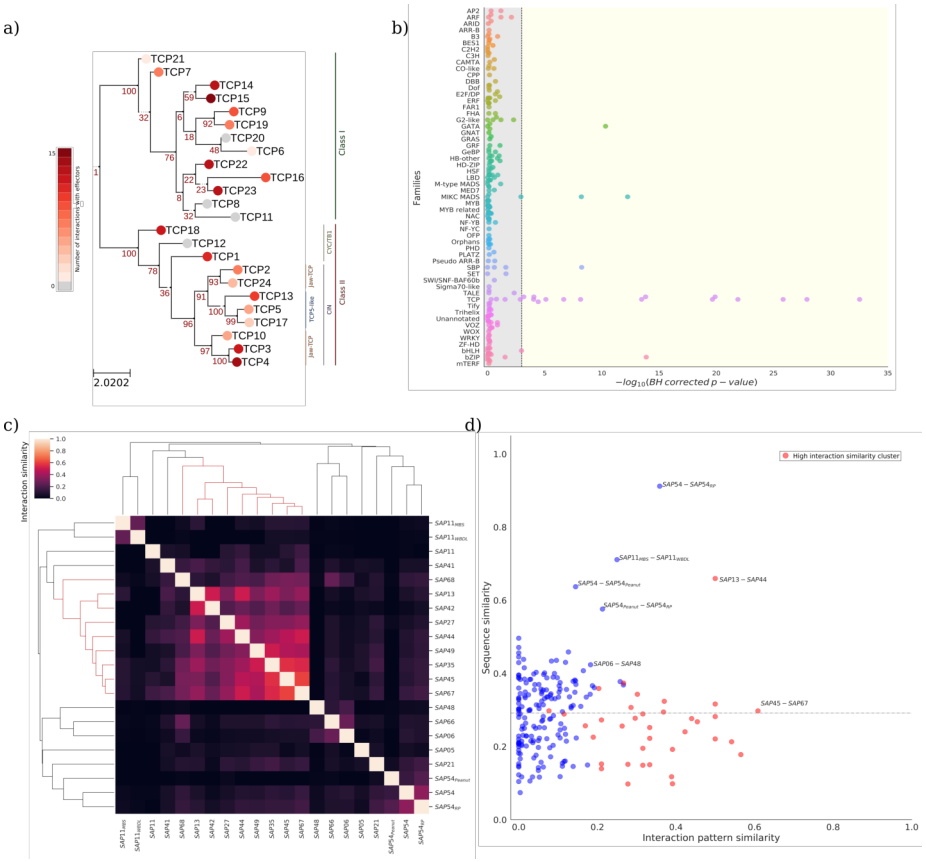

Figure 2: (a) Phylogenetic tree of all full-length TCP factors in *A. thaliana*. Leaf node color indicates number of interactions with phytoplasma effectors. Numbers at internal nodes indicate bootstrap branching support. Class assignment according to González-Grandío and Cubas (2015). (b) Overview of overrepresentation in interactions for different host protein families across individual effectors. Vertical dashed line indicates the statistical significance threshold (adjusted $p < 0.05$). (c) Clustermap of effector interaction similarities, measured by Jaccard distance. The cluster of TCP-specific effectors with high interaction similarities is highlighted in red in the dendrogram. (d) Relationship between effector interaction similarity and sequence similarity. Effectors belonging to the cluster of sequences with similar interaction patterns are marked in red; outliers and interesting examples are annotated with the pair they represent. Horizontal dashed line indicates median sequence similarity.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Martín-Trillo and Cubas, 2010), research in recent years has uncovered that they participate in the plant immune response as well (Lopez *et al.*, 2015; Li, 2015). These properties seem to make them important targets for pathogens. Some TCPs have previously been found to be targeted by the phytoplasma effector SAP11 (Sugio *et al.*, 2011b), and by evolutionarily distant pathogens as well (Mukhtar *et al.*, 2011; Weßling *et al.*, 2014).

To obtain a more precise picture, we tested whether individual effectors interact predominantly with specific host protein families. Strikingly, we find that 15 effectors (71%) are enriched in interactions with TCPs (Fig. 2b). This set includes SAP54 and $SAP54_{RP}$, whose interactions with TCPs had previously not been identified. The fact that there is a high level of functional redundancy within the TCPs (Danisman *et al.*, 2013), together with their interactions with numerous phytoplasma effectors, suggests that suppression of TCP activity is crucial for phytoplasmas, and that it can only be achieved effectively by targeting many TCPs with multiple effectors. Finally, as expected, the SAP54s are enriched in interactions with MIKC MADS-box proteins.

We were also interested in evaluating similarities in interaction patterns accross SAPs, which tend to be very low (median: 0.064). Clustering effectors by interaction similarity reveals, aside from the expected clusters of orthologous effectors from different strains, a clear cluster of nine effectors with similar interaction patterns (Fig. 2c, highlighted in red in the dendrogram), all of which showed an enrichment in interactions with TCPs. Interestingly, despite having similar interaction patterns, these effectors tend to have low sequence similarities, suggesting that they are phylogenetically unrelated (Fig. 2d). Furthermore, discriminant motif analysis (Bailey, 2011) on the clustered sequences, using effectors with no specificity for TCPs as a control, yields no statistically significant results. Thus, the binding site might be rather different across effectors. Overall, this indicates a certain degree of functional redundancy between these effectors, which could make it very difficult for the plant to mount a defense against these interactions. It is known that many pathogens produce mimetic molecules that resemble those of the host in order to exploit it (Elde and Malik, 2009; Via *et al.*, 2015), and it has been postulated that phytoplasma effectors mimic the structures of their targets in order to interact with them (Rümpler *et al.*, 2015). If this is the case, it could be difficult for targeted host proteins to evolve in a way that prevents interactions with effectors. Changes at the interaction interface that weaken interactions with an effector might also weaken important, physiological interactions. Additionally, multiple unrelated effectors with similar interaction patterns could still be able to interact with the host protein, rendering such adaptive changes moot. In this way, phytoplasmas could mount a very effective attack that hampers the evolution of a defense by setting an evolutionary trap in sequence space.

Interestingly, orthologous effectors from different strains tend to have somewhat low interaction similarities. We observe that interactions with MIKC MADS-box proteins responsible for floral organ identity, such as AP1 or SEP3, are conserved across different SAP54s, while interactions with MIKCs involved in flowering time are less conserved (Fig. S3a). Floral organ specification mech-

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

anisms are strongly conserved, while flowering time is linked to adaptation of a plant to its environment. Also, while interactions of SAP54 and $SAP54_{RP}$ with TCPs are highly conserved (Fig. S3b), no such interactions are detected for $SAP54_{Peanut}$. Altogether, this strongly suggests coevolution with orthologous host proteins, eventually leading divergence at the interaction interface.

5.3 Analysis of SAP06 interaction partners sucessfully predicts effector-induced phenotypes

In order to study whether interactions with a defined set of host proteins might be used to predict potential functional consequences of the effectors, we tested for over-representation of particular GO-terms in the effector-specific sets of host interactors (Suppl. table 1). For the highly reactive SAP06 protein, GO enrichment analysis of the interactors of SAP06 revealed enrichment in interactions with multiple proteins from different families involved in leaf growth and determination of the level of seed dormancy. Related to the latter biological process, the presence of multiple DELLA proteins, which are well-known repressors of gibberellic acid-induced seed germination (Ravindran and Kumar, 2019), stood out. Inspired by these findings, we tested whether ectopic expression of SAP06 has an effect on plant growth and seed dormancy.

We produced two *A. thaliana* lines with different levels of ectopic SAP06 expression. Phenotyping of these plants during the vegetative stage of development revealed stunted growth compared to Col0 control plants, and that this effect is proportional to the level of SAP06 expression (Fig. 3a).

Additionally, we found altered levels of seed dormancy in the SAP06 transgenic lines (Fig. 3b) proportional to the level of ectopic SAP06 expression. In light of the phytoplasma needs during the infection process and its life cycle, finding a seed dormancy phenotype upon ectopic expression of an effector is surprising. It is possible that SAP06 interacts with DELLA proteins to interfere with other functions they participate in, such as leaf growth and immunity and the balance and trade-off between these two processes (Daviere and Achard, 2016).

5.4 Phytoplasma has developed specific strategies to target host development

Prior research has determined PPI networks between *A. thaliana* proteins and the effectors of different, evolutionarily distant pathogens (Mukhtar *et al.*, 2011; Weßling *et al.*, 2014) (the bacterium *Pseudomonas syringae*, the fungus *Golovinomyces orontii* and the oomycete *Hyaloperonospora arabidopsidis*). These studies found that effectors produced by pathogens from different kingdoms target overlapping sets of plant proteins, indicating the existence of a conserved host-pathogen interface. This includes proteins involved in defense response, but also e.g. auxin and salycilic acid signaling. In order to determine whether phytoplasma fits this picture, we compared our PPI network to the networks

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

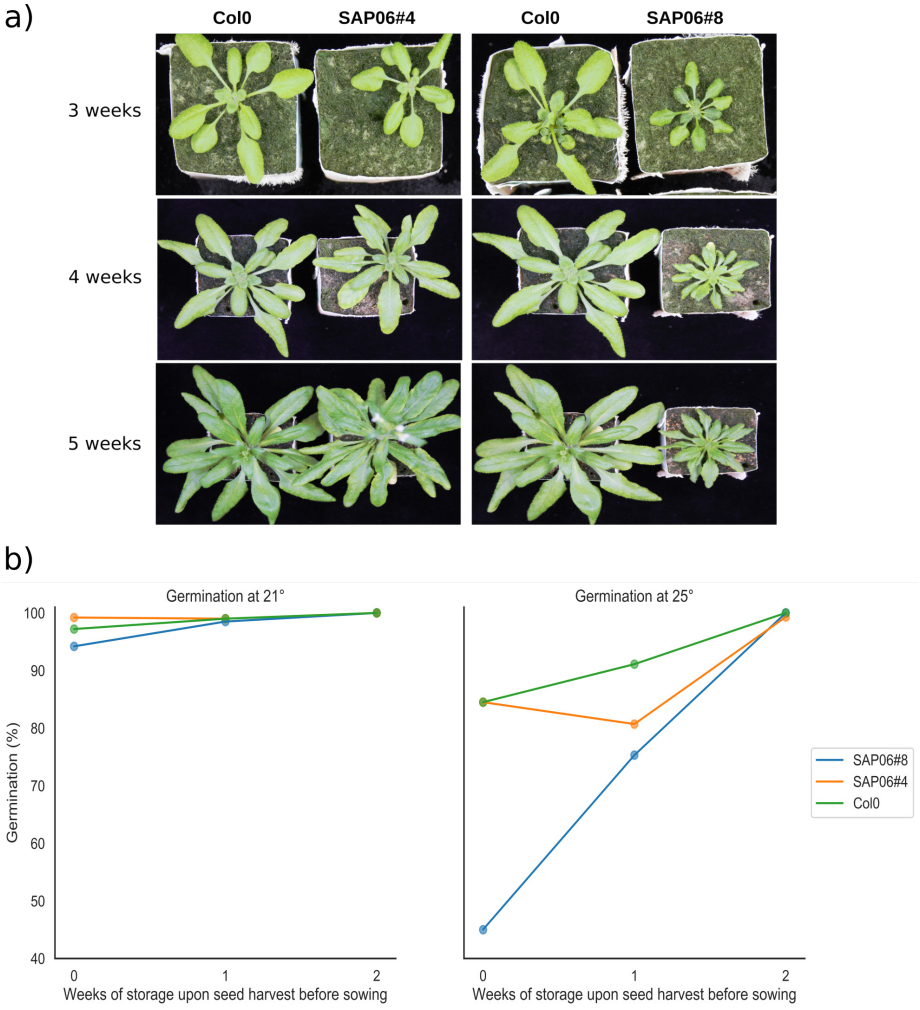

Figure 3: Phenotypic alterations upon ectopic expression of AY-WB SAP06 in *Arabidopsis* Col0 background. (a) Effect of ectopic SAP06 expression on growth and development in the vegetative stage, three, four and five weeks after germination. To the left, comparison between Col0 and a line with a medium level of SAP06 expression; to the right, comparison between Col0 and a line with a high level of SAP06 expression. (b) Effect of ectopic SAP06 expression on seed dormancy in the same transformed lines. To the left, germination percentage at 21°C upon direct sowing or storage of seeds for one or two weeks at room temperature. To the right, germination percentage over several weeks at 25°C.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

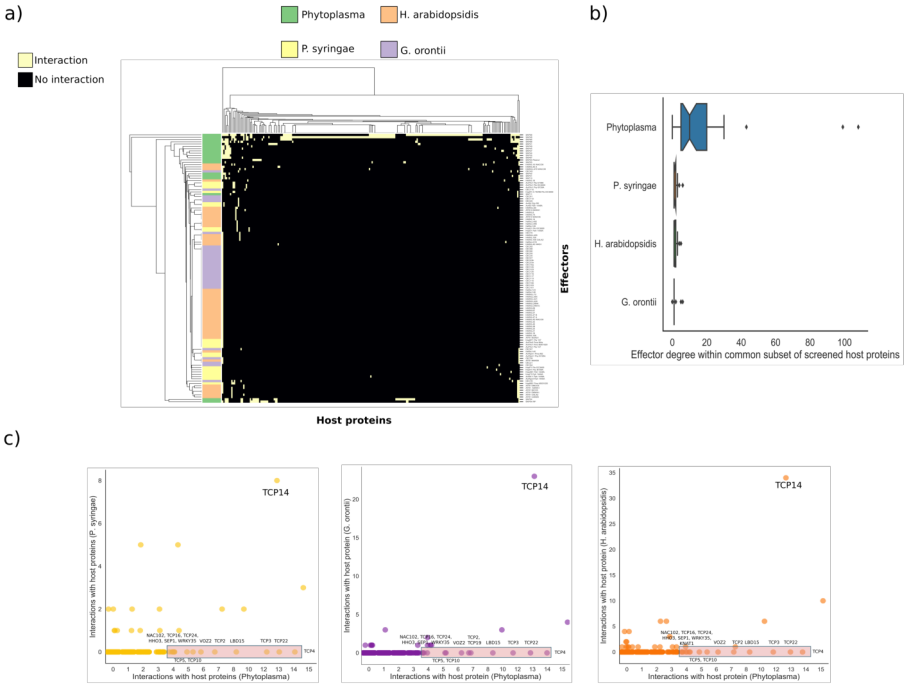

Figure 4: (a) Clustermap of interaction patterns in different host-pathogen PPI network within the subset of screened host proteins shared by all assays. Clustering was performed using Jaccard similarity as a metric. (b) Boxplot of the degree of effectors within the common host protein subset. Notches indicate the width of a 95% confidence interval of the median degree, determined by 1000 bootstrap resamples. (c) Comparison of host protein degrees between phytoplasma and *P. syringae* (left), *H. arabidopsidis* (center) and *G. orontii* (right). Host proteins highly targeted by phytoplasma but not by individual pathogens are highlighted.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

determined by Mukhtar *et al.* (2011) and Weßling *et al.* (2014) using the common subset of host proteins screened in all assays (Fig. S4a). Remarkably, a high number of interactions with TCP14, which promotes disease resistance (Yang *et al.*, 2017), is conserved across all pathogens. However, clustering of all effectors according to their interaction patterns separates most phytoplasma effectors from the rest (Fig. 4a). Clearly, phytoplasma effectors tend to have far more interactions with this subset of proteins that other species' effectors (Fig. 4a,b). This indicates that phytoplasma effectors diverge in their interaction patterns from effectors secreted by other pathogens.

We performed a careful, conservative pairwise comparison of host protein degree per species. We find that the set of proteins highly targeted by phytoplasma but not at all by other, individual pathogens, has little variation (Fig. 4c), and that there is a core of 12 proteins that remains constant. This highly targeted core is enriched in many terms related to development, especially its regulation and phyllome development (Fig. S4b). The core includes multiple TCPs (including all of the jaw-TCPs) and other development-related proteins from different families, such as SEP1 and LBD15. It also includes the flowering time regulator VOZ2 (Yasui *et al.*, 2012). We repeated the analysis using the larger subset of shared host proteins between our assay and the one performed with *G. orontii* effectors, reaching the same conclusions (Fig. S5). Although previous studies have found a certain level of enrichment in targeting development-related proteins in *H. arabidopsidis*, this focus on targeting development-related proteins is unprecedented. Furthermore, we find only one protein (TCP20, Fig. S5) that does not interact with a phytoplasma effector but interacts with effectors of a different pathogen. Put together, these observations indicate that phytoplasma has developed idiosyncratic molecular mechanisms that specifically target development in order to alter its host.

6 Conclusions

We have experimentally determined interactions between phytoplasma effectors and *A. thaliana* transcription factors and transcriptional regulators. Many of the assayed effectors have not been previously studied, and this work presents the first insights into their putative molecular mode-of-action. We find that interactions with transcription factors are pervasive amongst phytoplasma effectors. The set of interacting host proteins is enriched in transcription factors involved in development, especially that of phyllomes.

Remarkably, our analysis shows that many effectors interact with TCP transcription factors, almost all of which are targeted. This indicates that supressing the activity of TCPs, which are at the crossroads between development and immunity, is crucial for the infection strategy of phytoplasma. Furthermore, while many of these TCP-specific effectors have similar interaction patterns, they bear little sequence similarity to each other. This functional redundancy between very different sequences could make it difficult for the affected host proteins to evolve to avoid these interactions, as they are efffectively surrounded in

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sequence space. Finally, comparison with previous host-pathogen PPI networks shows that the infection strategy of phytoplasmas seems to require unusual targets involved in development.

We have shown that the determined network can be used to predict functional consequences of specific effectors, and our interaction data and findings can be further utilized to formulate mechanistic hypotheses regarding individual effectors. Additionally, further molecular data on phytoplasma can be integrated with the presented PPI network to yield further insights. Thus, this study serves as a stepping stone for a more global understanding of phytoplasmas and how they interact with and manipulate their hosts.

References

Abascal, F. *et al.* (2005). Prottest: selection of best-fit models of protein evolution. *Bioinformatics*, **21**(9), 2104–2105.

Bai, X. *et al.* (2009). Ay-wb phytoplasma secretes a protein that targets plant cell nuclei. *Molecular Plant-Microbe Interactions*, **22**(1), 18–30.

Bailey, T. L. (2011). Dreme: motif discovery in transcription factor chip-seq data. *Bioinformatics*, **27**(12), 1653–1659.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300.

Brückner, A. *et al.* (2009). Yeast two-hybrid, a powerful tool for systems biology. *International journal of molecular sciences*, **10**(6), 2763–2788.

Chang, S. H. *et al.* (2018). Alterations of plant architecture and phase transition by the phytoplasma virulence factor sap11. *Journal of experimental botany*.

Clough, S. J. and Bent, A. F. (1998). Floral dip: a simplified method for agrobacterium-mediated transformation of arabidopsis thaliana. *The plant journal*, **16**(6), 735–743.

Danisman, S. *et al.* (2013). Analysis of functional redundancies within the arabidopsis tcp transcription factor family. *Journal of experimental botany*, **64**(18), 5673–5685.

Daviere, J.-M. and Achard, P. (2016). A pivotal role of dellas in regulating multiple hormone signals. *Molecular plant*, **9**(1), 10–20.

De Folter, S. and Immink, R. G. (2011). Yeast protein–protein interaction assays and screens. In *Plant Transcription Factors*, pages 145–165. Springer.

Elde, N. C. and Malik, H. S. (2009). The evolutionary conundrum of pathogen mimicry. *Nature Reviews Microbiology*, **7**(11), 787.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Firrao, G. *et al.* (2007). Phytoplasmas: genetics, diagnosis and relationships with the plant and insect host. *Frontiers in bioscience: a journal and virtual library*, **12**, 1353–1375.

Floyd, S. K. and Bowman, J. L. (2007). The ancestral developmental tool kit of land plants. *International journal of plant sciences*, **168**(1), 1–35.

González-Grandío, E. and Cubas, P. (2015). Tcp transcription factors: Evolution, structure, and biochemical function. In *Plant Transcription Factors*, pages 139–151. Elsevier.

Guindon, S. *et al.* (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of phyml 3.0. *Systematic biology*, **59**(3), 307–321.

Henikoff, S. and Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences*, **89**(22), 10915–10919.

Huerta-Cepas, J. *et al.* (2016). Ete 3: reconstruction, analysis, and visualization of phylogenomic data. *Molecular biology and evolution*, **33**(6), 1635–1638.

Jaccard, P. (1912). The distribution of the flora in the alpine zone. 1. *New phytologist*, **11**(2), 37–50.

Jin, J. *et al.* (2016). Planttfdb 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic acids research*, page gkw982.

Karimi, M. *et al.* (2002). Gateway^{TM} vectors for agrobacterium-mediated plant transformation. *Trends in plant science*, **7**(5), 193–195.

Klopfenstein, D. *et al.* (2018). Goatools: A python library for gene ontology analyses. *Scientific reports*, **8**(1), 10872.

Le Fevre, R. *et al.* (2015). Modulation of host cell biology by plant pathogenic microbes. *Annual review of cell and developmental biology*, **31**, 201–229.

Li, S. (2015). The arabidopsis thaliana tcp transcription factors: a broadening horizon beyond development. *Plant signaling & behavior*, **10**(7), e1044192.

Lopez, J. A. *et al.* (2015). Tcp three-way handshake: linking developmental processes with plant immunity. *Trends in plant science*, **20**(4), 238–245.

MacLean, A. M. *et al.* (2011). Phytoplasma effector sap54 induces indeterminate leaf-like flower development in arabidopsis plants. *Plant Physiology*, pages pp–111.

MacLean, A. M. *et al.* (2014). Phytoplasma effector sap54 hijacks plant reproduction by degrading mads-box proteins and promotes insect colonization in a rad23-dependent manner. *PLoS biology*, **12**(4), e1001835.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Marcone, C. (2019). Candidatus phytoplasma asteris (yellow disease phytoplasmas). `https://www.cabi.org/isc/datasheet/7642`.

Martín-Trillo, M. and Cubas, P. (2010). Tcp genes: a family snapshot ten years later. *Trends in plant science*, **15**(1), 31–39.

Mattoo, S. *et al.* (2007). Interactions of bacterial effector proteins with host proteins. *Current opinion in immunology*, **19**(4), 392–401.

Minato, N. *et al.* (2014). The phytoplasmal virulence factor tengu causes plant sterility by downregulating of the jasmonic acid and auxin pathways. *Scientific reports*, **4**, 7399.

Mukhtar, M. S. *et al.* (2011). Independently evolved virulence effectors converge onto hubs in a plant immune system network. *science*, **333**(6042), 596–601.

Navaud, O. *et al.* (2007). Tcp transcription factors predate the emergence of land plants. *Journal of molecular evolution*, **65**(1), 23–33.

Orlovskis, Z. and Hogenhout, S. A. (2016). A bacterial parasite effector mediates insect vector attraction in host plants independently of developmental changes. *Frontiers in plant science*, **7**, 885.

Pérez-Rodríguez, P. *et al.* (2009). Plntfdb: updated content and new features of the plant transcription factor database. *Nucleic acids research*, **38**(suppl_1), D822–D827.

Pruneda-Paz, J. L. *et al.* (2014). A genome-scale resource for the functional characterization of arabidopsis transcription factors. *Cell reports*, **8**(2), 622–632.

Ravindran, P. and Kumar, P. P. (2019). Regulation of seed germination: The involvement of multiple forces exerted via gibberellic acid signaling. *Molecular plant*, **12**(1), 24–26.

Rodriguez, P. A. *et al.* (2019). Systems biology of plant microbiome interactions. *Molecular plant*.

Roy, B. (1993). Floral mimicry by a plant pathogen. *Nature*, **362**(6415), 56.

Rümpler, F. *et al.* (2015). Did convergent protein evolution enable phytoplasmas to generate 'zombie plants'? *Trends in plant science*, **20**(12), 798–806.

Sievers, F. *et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular systems biology*, **7**(1), 539.

Strauss, E. (2009). Phytoplasma research begins to bloom.

Sugio, A. *et al.* (2011a). Diverse targets of phytoplasma effectors: from plant development to defense against insects. *Annual review of phytopathology*, **49**, 175–195.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.13.946517; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sugio, A. *et al.* (2011b). Phytoplasma protein effector sap11 enhances insect vector reproduction by manipulating plant development and defense hormone biosynthesis. *Proceedings of the National Academy of Sciences*, page 201105664.

Sugio, A. *et al.* (2014). The small phytoplasma virulence effector sap11 contains distinct domains required for nuclear targeting and cin-tcp binding and destabilization. *New Phytologist*, **202**(3), 838–848.

Sun, X. *et al.* (2013). Arabidopsis asl11/lbd15 is involved in shoot apical meristem development and regulates wus expression. *Planta*, **237**(5), 1367–1378.

Tripathi, L. P. *et al.* (2019). Network-based analysis of host-pathogen interactions.

Via, A. *et al.* (2015). How pathogens use linear motifs to perturb host cell networks. *Trends in biochemical sciences*, **40**(1), 36–48.

Wang, N. *et al.* (2018). Phytoplasma effector swp1 induces witches' broom symptom by destabilizing the tcp transcription factor branched1. *Molecular plant pathology*.

Weintraub, P. G. and Beanland, L. (2006). Insect vectors of phytoplasmas. *Annu. Rev. Entomol.*, **51**, 91–111.

Weßling, R. *et al.* (2014). Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell host & microbe*, **16**(3), 364–375.

Yang, L. *et al.* (2017). Pseudomonas syringae type iii effector hopbb1 promotes host transcriptional repressor degradation to regulate phytohormone responses and virulence. *Cell host & microbe*, **21**(2), 156–168.

Yasui, Y. *et al.* (2012). The phytochrome-interacting vascular plant one–zinc finger1 and voz2 redundantly regulate flowering in arabidopsis. *The Plant Cell*, pages tpc–112.