

Characterization of the *Arabidopsis thaliana* Interactome Targeted by Viruses

Guillermo Rodrigo
Javier Carrera
Virginia Ruiz-Ferrer
Francisco J. del Toro
César Llave, et al.

SFI WORKING PAPER: 2011-10-049

SFI Working Papers contain accounts of scientific work of the author(s) and do not necessarily represent the views of the Santa Fe Institute. We accept papers intended for publication in peer-reviewed journals or proceedings volumes, but not papers that have already appeared in print. Except for papers by our external faculty, papers must be based on work done at SFI, inspired by an invited visit to or collaboration at SFI, or funded by an SFI grant.

©NOTICE: This working paper is included by permission of the contributing author(s) as a means to ensure timely distribution of the scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the author(s). It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author's copyright. These works may be reposted only with the explicit permission of the copyright holder.

www.santafe.edu



SANTA FE INSTITUTE

CHARACTERIZATION OF THE *ARABIDOPSIS THALIANA* INTERACTOME TARGETED BY VIRUSES

Guillermo Rodrigo^{*1}, Javier Carrera^{*1,2}, Virginia Ruiz-Ferrer^{‡,3}, Francisco J. del Toro⁴, César Llave⁴,
Olivier Voinnet^{§,3}, and Santiago F. Elena^{#,1,5}

¹Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas - Universidad Politécnica de Valencia, Valencia, Spain; ²ITACA, Universidad Politécnica de Valencia, Valencia, Spain; ³Institut de Biologie Moléculaire des Plantes, CNRS, Strasbourg, France; ⁴Centro de Investigaciones Biológicas, CSIC, Madrid, Spain; ⁵Santa Fe Institute, Santa Fe NM, USA; ^{*}Equal contributors; [‡]Present address: Centro de Investigaciones Biológicas, CSIC, Madrid, Spain; [§]Present address: Institute of Agricultural Sciences, ETH Zürich, Switzerland; [#]Corresponding author SFE: santiago.elena@csic.es.

Understanding the mechanisms by which plants trigger host defenses in response to viruses has been a challenging problem owing to the multiplicity of factors and complexity of interactions involved. The advent of genomic techniques, however, has opened the possibility to grasp a global picture of the interaction. Here, we used *Arabidopsis thaliana* to identify and compare genes that are differentially regulated upon infection with seven distinct (+)ssRNA and one ssDNA plant viruses. In a first approach, we established lists of genes differentially affected by each virus and compared their involvement in biological functions and metabolic processes. We found that phylogenetically-related viruses significantly alter the expression of similar genes and that viruses naturally infecting *Brassicaceae* display a greater overlap in the plant response. In a second approach, virus-regulated genes were contextualized using transcriptional and protein-protein interaction networks of *A. thaliana*. Our results confirm that host cells undergo significant reprogramming of their transcriptome during infection, which is possibly a central requirement for the mounting of host defenses. We uncovered a general mode of action in which perturbations preferentially affect genes that are highly connected, central and organized in modules. This indicates redundancy in the information transmission from virus effectors to immune response genes and allowed us to suggest that this over-triggering of hub proteins is a mechanism to confer robustness for expression of host defenses.

Introduction

For over decades, plant molecular virology has been overly focused on the pathogen itself, studying their individual genes and products, and their local effects on certain regulatory pathways related to antiviral responses. However, with the arrival of modern genomic tools allowing for high-throughput screenings, we can now tackle the problem of the plant host-virus interaction from a systemic perspective that would allow us reaching a deeper understanding on how host and virus genotypes, environmental effects and stochasticity interplay in determining the pathological outcome of an infection. Viral infections typically alter host physiology, notably by diverting almost all cellular resources for the production of virus-specific components, and by actively suppressing host defenses [1, 2]. As a response to infection, hosts compensate by over- or under-expressing certain cellular

pathways, and deploying specific antiviral measures. Collectively, these alterations determine the type and strength of symptoms displayed by infected organisms as well as the virulence of the infection. Much effort has gone into identifying individual cellular traits that may change as a consequence of viral infection [3] and this has greatly benefited from the contemporary development of genome-wide investigation technologies and their successful application to plant diseases research [4, 5]. These technologies have further demonstrated great potential in providing insights into multidimensional networks of plant-virus interactions [4, 6], notably by allowing combined analyses at the host transcriptome and proteome levels, as was recently shown for an animal virus such as HIV-1 [7].

Based on the above, it has been anticipated that a systems biology approach to infections should allow the identification of universal principles and features of host-

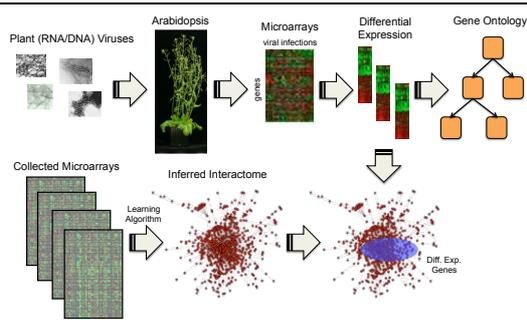


Figure 1. Overview of the Systems Biology approach we followed to study the viral infection in plants. We considered *A. thaliana* as model host. Microarray data from several infection experiments with viruses were collected to analyze the differentially expressed genes, and to perform functional analyses by harnessing GO annotations. In addition, by taking advantage of large databases of expression profiles derived from transcriptional perturbations, the global host interactome could be as a first approach unveiled by applying learning algorithms. The differential expression was then contextualized within the inferred network.

virus interactions, as opposed to scrutinizing many specific aspects of any given viral infection [8 - 10]. Such generic principles may indeed prove more predictive of the outcome of viral diseases and therefore, more efficient in the prophylaxis, diagnosis, and even treatment of such diseases. In a network approach, viral pathogenesis can be viewed as the expression of new constraints imposed by the virus upon the cellular interactome: while the host initiates a reprogramming of its genetic profile to activate the immune system to counteract the infection effects, replication and suppression of host defenses by viruses entail the manipulation of molecular connections that ultimately result in the misregulation and/or silencing of genes that trigger defense functions, and eventually in the emergence of new topological properties of the host interactome. Thus, understanding the bases for such modifications is crucial to acquire a systemic view of the infection process [1, 11, 12]. One of the main goals to this end would be the identification of the virus effectors (i.e., the targets of the viral proteins). Instead, herein we focus on the study of the mechanisms by which the host canalizes these virus effectors to trigger the global defense system. We propose a reverse-engineering approach by which we analyze the genetic profile of the cell upon viral infection and contextualize this information onto the host interaction network.

Analyses of interaction networks have already uncovered global, dynamic features that relate directly to biological properties [13]. For example, proteins with a large number of interactions within a network, also referred to as ‘hubs’, have a higher impact on multiple phenotypic traits (pleiotropy) than loosely connected proteins. Moreover, proteins essential for survival are highly clustered [14]. Hub proteins can be further partitioned into those that function in a specific biological module and those that connect different modules. The existence of such hub proteins generates two interesting properties in

networks. First, the network is scale-free in that the number of connections per node (i.e., its connectivity degree) probability distribution follows asymptotically a power-law. Second, the network presents the characteristic of small-worlds, in which the average number of intermediary nodes connecting any random pair is small [15]. These two properties confer robustness against random perturbations in the network, but at the cost of strong sensitivity to attacks directed against hubs [16]. Although plant viruses usually encode for few proteins, the genetic profile of the host after viral infection presents hundreds and even thousands of significant changes. A plausible explanation for this scenario is that virus effectors are highly connected proteins that spread the signal, and additionally interact in a short downstream pathway with the immune response genes. Of relevance, very recently it has been experimentally shown that bacterial effectors in *Arabidopsis thaliana* are hubs and canalize the signal onto the regulators of the global immune system [27]. Interestingly, these results are in concordance with those from previous studies with *Epstein-Barr virus* [35], *Hepatitis C virus* [36], *Influenza A virus* H1N1 [37], and other viral and bacterial pathogens of mammals [8, 38]. These studies have shown that viral proteins preferentially target hub proteins in the human interactome. Herein, by assuming that virus effectors are hubs, we investigate whether this information is propagated following the same scale of the plant interactome.

Microarray-based functional genomics, which provides a global view of transcriptional changes in host cells, has been the most commonly used method to study global changes during plant-virus interactions [4, 17 - 24]. However, the comparison of results obtained in distinct experiments involving different viruses is both complex and challenging; it has not been attempted in a systematic manner. Here, we present the results of a meta-analysis (Fig. 1) of microarray data gathered from infections of the same host plant, *A. thaliana*, by seven plant RNA viruses belonging to four taxonomic families (*Tobacco etch potyvirus*, *Turnip mosaic potyvirus*, *Plum pox potyvirus*, *Tobacco mosaic tobamovirus*, *Tobacco rattle tobnavirus*, *Turnip crinkle carmovirus*, and a laboratory-evolved strain of *Tobacco etch potyvirus*) and one DNA geminivirus (*Cabbage leaf curl geminivirus*). Using the same methodology, we first identified lists of genes that were up- and down-regulated, together with the sets of biological functions (gene ontology, GO) and metabolic pathways over-represented among them. These changes were then compared among the different virus infections, uncovering unexpected correlations within virus-specific phyla. In a second strategy, we explored these lists from a global network perspective, by mapping the altered genes onto different interactome models of the common host *A. thaliana*. This global computational approach unraveled a generic mode of interference by plant viruses, whereby perturbations incurred to the host interactome preferentially affect genes that are highly connected, central and form modules.

Results and Discussion

Genetic profile targeted by plant viruses

Using transcriptomic data (steady-state RNA levels) extracted from 8 distinct virus infections on the model plant *A. thaliana*, we identified lists of genes with altered expression levels, referred herein to as ‘virus-regulated genes’ (or VRGs). These set of genes involves those genes that are directly or indirectly regulated by the virus and that are differentially expressed when the virus infects the cell. Those VRGs were then used to establish both general and specific genetic profiles associated to the pathogens of interest (supplementary file S1, Supplementary Material online). We found that among the > 22,000 genes inspected, a set of 5296 VRGs (2646 over- and 2650 under-expressed, respectively) is altered by at least one of the eight viruses studied. This VRG set may thus be used to reflect the global plant response to any viral infection. We found that the number of VRGs shared by more than one virus declines exponentially (Figs. S1a and S1b). Seven VRGs were found up-regulated in common by six viruses, of which, surprisingly, six play a role in cell migration (*At3g57260*, *At5g10380*, *At3g14990*, *At3g28510*, *At5g52640*, and *At4g24690*) and one (*At1g75040*) encodes a PR-5 thaumatin-like protein, factors known for their involvement in pathogens responses. While no single VRG was identified in common among the eight infections, one VRG was systematically up-regulated by seven viruses (i.e., all except PPV) and found to encode an aspartyl protease involved, again, in cell migration in the diencephalon (*At5g10760*). Three VRGs were down-regulated by six viruses, two of which correspond to different subunits of the NADPH dehydrogenase complex (*At1g18730* and *At5g58260*).

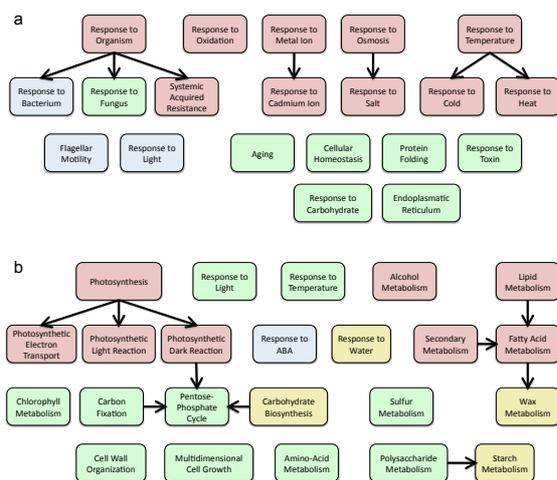


Figure 2. Functional analysis. (a) Over- and (b) under-expressed VRGs representing biological processes. In palid red, VRGs present in at least five of the total eight viral infections (unspecific viral response); in palid blue, VRGs in at least three of the four potyviral infections; in palid green, VRGs in at least three of the four *Brassica*-infecting viral infections; in palid yellow, common VRGs for *Potyvirus* and *Brassica*-infecting viruses.

Not surprisingly, infections by the two different strains of TEV under study share the largest number of VRGs (197 over- and 282 under-expressed genes, respectively), although this may probably reflect, to some extent, homogeneity in experimental procedures. In the overlapping set, over-expressed genes principally have roles in response to stress (e.g., fungal resistance TIR-NB-LRR protein *At1g56510*, transcription factor *At1g22070*, U-box-domain-containing E3 ubiquitin ligase *At3g11840* that acts as a negative regulator of immune responses, or the aforementioned *At1g75040*), transport (e.g., the mitochondrial inner membrane translocase *At1g20350*, the high-affinity ammonium transporter *At2g38290*, or the glycolipid transfer protein *At4g39670*), transcription (e.g., the Myb-like transcription factor *At1g25550*, or the C2H2-type zinc finger *At3g46080*), and protein metabolism (e.g., the chaperone DnaJ-domain *At1g56300*, or the eukaryotic aspartyl protease *At5g10760*). The overlapping set of under-expressed genes is mostly composed of factors involved in basic metabolic and cellular processes (e.g., the member of the R2R3 factor *At1g18710*, the enzyme *At1g03630* that is NADPH- and light-dependent, or the α/β -hydrolase *At1g10740*).

Interestingly, a set of 27 VRGs was significantly over-expressed upon infections by the three viruses that naturally infect hosts from the *Brassicaceae* family (TuMV, TCV and CaLCuV) and by the TEV laboratory strain, which has been experimentally adapted to *A. thaliana* (TEV-*At17*); hereafter, we will refer to this set of four viruses as *Brassica*-infecting viruses. A common feature of these VRGs is that all of them play roles in stress response, including, among others, the disulfide isomerase *At1g21750* implicated in the regulation of apoptosis during endoplasmic reticulum stress as well as in osmotic stress. The set also includes the homolog of mammalian Bax inhibitor 1, *At5g47120*, which functions as an attenuator of biotic and abiotic stress-associated cell death, and the cytosolic heat shock protein *At5g52640*. The list further comprises several genes involved in signal transduction, such as the BAK1-interacting receptor-like kinase *At5g48380* that regulates multiple signaling routes for plant resistance, or the ATP binding kinase *At5g45800* involved in embryonic development. A set of 22 VRGs was also under-expressed in common, in plants infected by the *Brassica*-infecting viruses. This list includes, as in the afore-mentioned study of the two TEV strains, genes involved in central metabolic and cellular processes.

Next, we sought to establish an overall comparison of the lists of VRGs identified from any of the eight viruses included in the analysis. To do so, we computed similarity scores among all pairs of lists, and constructed a dendrogram to visualize which viruses showed more closely related lists (supplementary Fig. S2, Supplementary Material online). The eight viruses do not represent independent draws from a population; rather, some are phylogenetically related. It was therefore important to test whether the above overlap in VRGs reflected taxonomic correlations. In other words, do closely phylogenetically

related viruses tend to share a higher number of VRGs, and does this overlap reduce as phylogenetic distance between viruses increases? To address this issue, we first used an alignment of the replicase genes from the eight viruses (the replicase-associated protein in the case of the geminivirus) to construct a maximum-likelihood phylogenetic tree (using the WAG + Γ model of amino acid substitutions and evaluating the significance of tree topology by 1000 bootstrap replicates; supplementary Fig. S3, Supplementary Material online). Next we computed a congruency index [25] measuring the overlap between the tree topology obtained from the VRG similarity matrix (supplementary Fig. S2, Supplementary Material online), on the one hand, and the topology of the estimated phylogenetic tree (supplementary Fig. S3, Supplementary Material online), on the other. The congruence index ($I_{cong} = 1.4720$) was significantly larger than expected by mere chance ($P = 0.0052$), suggesting that the two topologies are indeed highly congruent. This result supports the hypothesis that the overlap between VRG lists reflects the taxonomic relationships among viruses: two closely related viruses (e.g., the potyviruses TEV and TuMV) tend to alter the expression of a similar set of genes, whereas two non-related viruses (e.g., TEV and TRV) tend to alter different subsets of genes. The various viruses included in the study have distinct replication, gene expression, movement, and RNA silencing-suppression strategies that should somehow impact transcriptomic profiles differently. It is likely, however, that these strategies may be more conserved between phylogenetically related viruses than among viruses with weak or no phylogenetic relationship and, hence, the above study most certainly already accounts for the differences and commonalities observed among virus phyla. This being said, convergent evolution in phylogenetically unrelated viruses may contribute to increase the overlap of VRG lists. For instance, potyviruses and carmoviruses employ overlapping RNA silencing-suppression strategies affecting the global metabolism of miRNAs, which may lead to a set of related host responses.

A potential weakness of the above meta-analyses of gene lists is that the different experiments not only differed in the methodological details and plant ecotypes, as described in the Materials and Methods section, but also in that different experiments took samples at different time points during the infection process and, in some cases, different tissues were also sampled. Ecotype-specific, time-dependent and tissue-specific responses to viral infection may turn on/off different subsets of genes [20, 26] and thus may not receive a high enough score to be classified as VRG according to the stringent statistical criteria used in the study. To minimize as much as possible these potential problems, only data from leaves were included in the present analyses, although the possible effect of ecotype and sampling time may still exist. We performed several statistical analyses to assess this heterogeneity in the data (see Materials and Methods), and we concluded that differences in ecotype or in sampling

time would neither have a significant effect on the conclusions drawn from our meta-analyses.

Integrative functional analysis

Subsequently, we performed a functional analysis to map changes in gene expression onto regulations effecting global biological functions, thus establishing lists of ‘virus-regulated functions’ or VRFs (supplementary file S2, Supplementary Material online). Figs. S1c and S1d illustrate the number of over- and under-expressed VRFs found in common in several viral infections. Two over-expressed VRFs are common to all eight viruses and encompass stress responses to temperature and to pathogens. By contrast, under-expressed VRFs, were mostly found to encompass different metabolic and photosynthetic processes (Fig. 2). In order to set aside changes in VRFs reflecting direct consequences of viral infection from those of indirect effects resulting, for instance, from cross-talks between biotic and abiotic stress, we partitioned data sets into (i) the unspecific response (i.e., VRFs significantly over-represented in at least five of the eight viral infections), (ii) the specific response to *Brassica*-infecting viruses (i.e., VRFs significantly over-represented in at least three of the four infections by *Brassica*-infecting viruses, eliminating those terms included in the unspecific response), and (iii) the specific response to *Potyvirus* (i.e., VRFs significantly over-represented in at least three of the four infections by potyviruses, eliminating those terms included in the unspecific response). This partitioned analysis notably confirmed that activation of systemic acquired resistance (SAR), an innate and salicylate-based immune response to pathogens, is indeed a general mechanism triggered by the exposure of plants to viruses (Fig. 2). By contrast, down-regulation of polysaccharide metabolism, such as starch, appears as a specific response to *Brassica*-infecting viruses (Fig. 2). This possibly explains why *A. thaliana* mutants compromised in starch biosynthesis display less severe virus-induced symptoms compared to wild type plants [28]. As it can be further observed in Fig. 2, the specific response to *Brassica*-infecting viruses is larger than the unspecific one, suggesting that viruses impose additional constraints to their host through co-evolution to optimize their entire infectious cycle, by impelling the plant to introduce further gene reprogramming sentences, notably within the immune response.

Supplementary Fig. S4 (Supplementary Material online) displays the neighbor-joining dendrogram obtained from the similarity matrix computed from overlapping lists of significant VRFs obtained for the eight viruses. At the more stringent (i.e., 95%) bootstrap level, PPV, TMV and TRV form a distinctive cluster sharing several VRFs, which itself belongs to a significantly larger cluster that also includes the other three potyviruses (e.g., most of (+)ssRNA viruses) but excludes TCV. For those six viruses, over-expressed VRFs include responses to several abiotic stresses (e.g., temperature, osmotic, oxidative stresses) and defenses against pathogens, including SAR.

Under-expressed VRFs in this large cluster include fatty acid metabolism and photosynthesis (Fig. 2). CaLCuV is associated to a list of VRFs that differs significantly from that of the RNA viruses, as might be expected from the drastically distinct replication strategy of ssDNA viruses, which entails the reactivation of the host DNA replication machinery. Interestingly, TCV occupies an intermediate position in the dendrogram, close to the base of the cluster formed by the other (+)ssRNA viruses.

As in the previous section, we tested if the dendrogram topology shown in supplementary Fig. S4 (Supplementary Material online) was congruent with the phylogenetic history of the viruses (supplementary Fig. S3, Supplementary Material online). In this case, the congruence index ($I_{cong} = 1.2267$) did not significantly differ from what was expected by chance ($P = 0.1005$), thus suggesting that both topologies are not highly congruent; in other words, that the set of VRFs altered by two related viruses is similar to the one altered by two non-related viruses. At first, this result may be seen as contradicting the previous one, obtained by comparing lists of VRGs. Broad GO terms, however, encompass a multitude of genes and it may well be that different viruses affect the same VRF by modifying the expression of different target genes. Consistent with this idea, TEV and TRV infections were both associated with a significant over-representation of the GO term ‘stress response’. In both cases, the number of VRGs connected to this specific GO is similar (108 for TEV and 93 for TRV); yet only four genes are affected in common between the two viruses. Therefore, this analysis reveals that comparable trends in the global reprogramming of cellular functions might be achieved via highly dissimilar gene expression changes induced by distinct viruses.

Metabolic pathways targeted by viruses

Next, a metabolic pathway analysis was conducted to uncover the biochemical networks that were over- and under-expressed by viral infections. To that aim, we represented the global metabolic map of *A. thaliana* infected with a given virus by overlaying the expression of the corresponding enzymes. To contextualize the information of all viruses in a single picture, we constructed, on the one hand, a list of genes that characterize the unspecific virus-plant interaction (VRGs in at least five of the eight viral infections studied), and, on the other hand, a list of genes that depict the interaction between *Brassica*-infecting viruses and *A. thaliana* (VRGs in at least three of the four infections by *Brassica*-infecting viruses, eliminating those genes included in the unspecific response). The results presented in Fig. 3 show that the metabolic pathways over-represented as part of the unspecific response to viruses notably included cellulose biosynthesis -required for cell wall integrity- and nitrogen fixation. In addition, *Brassica*-infecting viruses were found to distinctively induce the biosynthesis of cytokines, a family of hormones central to plant development and growth, also involved in detoxification [29]. *Brassica*-

infecting viruses also specifically induced the Rubisco shunt, which is a more efficient converter of carbohydrates into acetyl-CoA than glycolysis [30]. The unspecific gene down-regulation response as a part of the viral reprogramming of central metabolism was found to affect the Calvin cycle and glycolysis. This observation is consistent with the previously proposed idea [31] that viruses impel the plant to redirect resources towards immune systems and, in particular, biotic stress responses, to the detriment of developmental processes. As a result, increased tolerance to viral infections might be achieved. Biosynthesis of starch -the major energy reservoir in the cell- photorespiration, and fatty acid biosynthesis are some of the biochemical routes that were found to be specifically down-regulated by *Brassica*-infecting viruses. In addition, *Brassica*-infecting viruses induce symptoms such as chlorosis and spotting, leaf curling and overall dwarfism, whereas, apart from a minor delay in growth, the other viruses progress asymptotically. All together suggests that viruses more adapted to their hosts impose a more stringent rerouting of the plant metabolism, which may result in more severe symptoms and/or increased viral virulence.

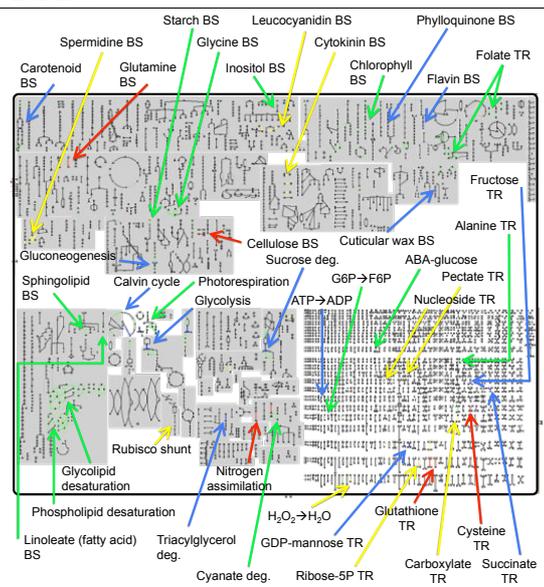


Figure 3. Metabolic map of *A. thaliana* highlighting the reactions altered by the unspecific viral response (VRGs in at least five of the total eight viral infections) and the specific *Brassica*-infecting virus response (VRGs in at least three of the four infections by *Brassica*-infecting viruses). Red (unspecific) and yellow (specific to *Brassica*-infecting viruses) reactions are over-expressed, whereas blue (unspecific) and green (*Brassica*-infecting) reactions are under-expressed. BS means biosynthesis, and TR, transformations.

Viruses preferentially alter highly connected central genes

Next we focused on the impact of viruses on two different predicted interactomes of *A. thaliana*: a transcriptional regulatory network (TRN) and a protein-protein interaction network (PPIN). A second TRN (TRN2; see Material and Methods) and the graphical Gaussian model of interaction network (GGIN; Ma et al. 2007; supplementary file S3,

Supplementary Material online) were also added to consolidate this study. Both TRN and PPIN have the properties of scale-free and small-worlds, the two major characteristic properties of real biological networks [32 - 34]. A network is named scale-free if a node selected randomly has a number of links (i.e., degree or connectivity degree) that follows a specific mathematical function referred to as the power law, implying that the network has no characteristic scale and is self-similar. A network has the property of a small-world if any two nodes are connected by a small number of edges; mathematically, this number should grow as the logarithm of the number of nodes in the network.

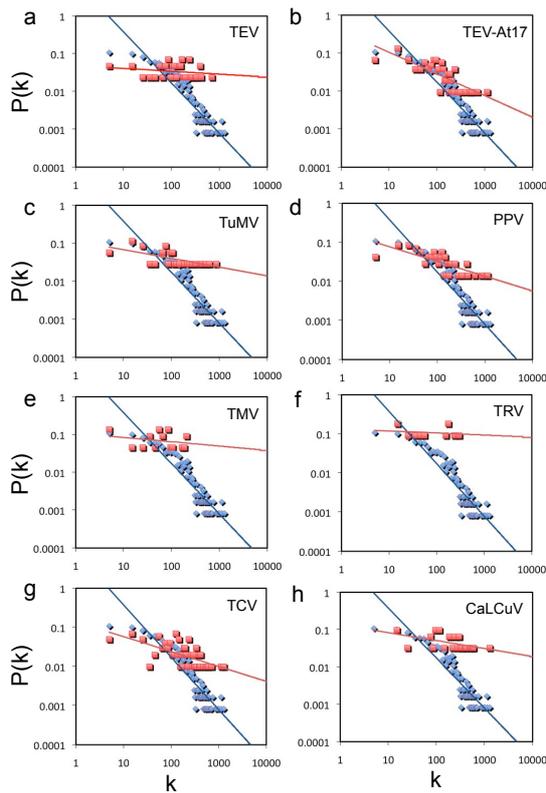


Figure 4. Outgoing connectivity distributions, contextualized in the TRN, for the VRGs (red) and the whole interactome (blue).

First, we analyzed the connectivity degree distribution for the VRGs as compared to the global set of genes. Roughly, if those genes were located in the periphery of the network, their connectivity would be expected to be smaller than if they were central, since the interactome is scale-free (supplementary Fig. S5, Supplementary Material online). As the TRN is a directed interactome, we focused on the outgoing connectivity, that is, the number of edges that leave from a given node to connect other nodes in the network. In Fig. 4 we show the connectivity distributions for all viral infections. Table 1 summarizes the value of the power-law exponent that better fits this particular distribution as well as the average connectivity. To statistically assess differences between the VRGs and the global set of genes, we used *t*-tests for differences in slopes and Mann-Whitney tests for differences in the location of the high-degree genes within

the distributions. No significant differences were found between the incoming connectivity distributions of VRGs and the one characterizing the whole interactome (supplementary Fig. S6, Supplementary Material online). Fig. 5 shows the corresponding connectivity distributions using the PPIN. We found that, in all cases, i.e. whether considering the TRN, the PPIN or both, the slope of the power-law distributions was significantly smaller than the slope estimated for the whole interactome. These results were also confirmed by analyzing the TRN2 (supplementary Fig. S7, Supplementary Material online) but not upon analysis of the GGIN (supplementary Fig. S8, Supplementary Material online). Apart of that *Brassica*-infecting viruses affect on average more genes, these viruses also manipulate more interactions, irrespective to the network model (one-tailed *t*-test, $P < 0.05$). However, we did not find any significant difference among viruses in terms of the average connectivities and power-law distribution exponents. We conclude that a smaller slope of the power-law distribution is a general trend characterizing the VRGs, indicating that viral infection preferentially alters the expression of highly connected genes (hubs) rather than random genes within the whole network. This could reflect a cellular response to gain robustness against the manipulation of the host by viruses.

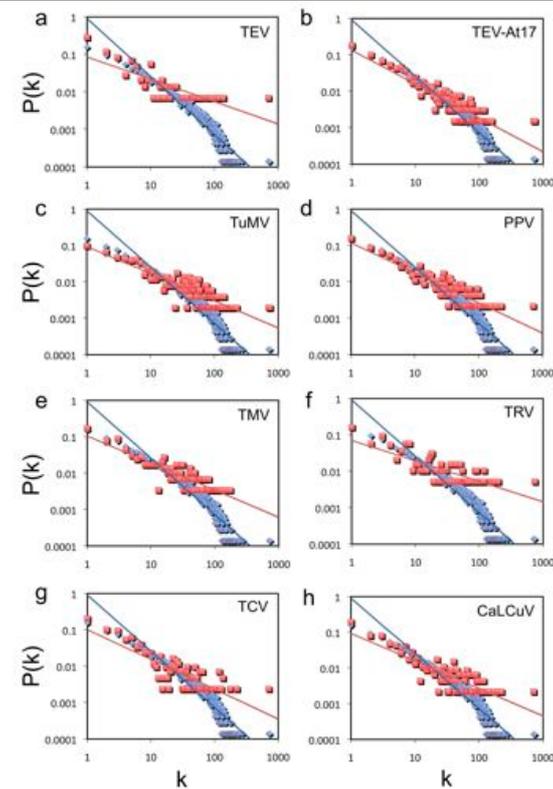


Figure 5. Connectivity distributions, contextualized in the PPIN, for the VRGs (red) and the whole interactome (blue).

A more global scale analysis involved calculation of the betweenness centrality distribution, understood as the relative number of shortest paths traversing a given gene. Table 1 summarizes the values of the average betweenness for these sets of VRGs. For each sub-network generated with the VRGs, we computed the total number of

Table 1: Summary of topological properties of the differentially expressed VRGs from several viral infections contextualized in the *A. thaliana* TRN and PPIN. We show the number of VRGs and VRFs (over- and under-expressed), the number of interactions (edges) manipulated by the virus, the power-law distribution exponent (for connectivity γ), the average connectivity ($\langle k \rangle$), and the average betweenness ($\langle b \rangle$). We also show the P -value for the tests comparing the shape and location of the VRGs distributions with respect to the corresponding whole interactome (^aStudent t -test, ^bMann-Whitney U -test).

	VRGs		VRFs		TRN			PPIN			
	Over	Under	Over	Under	Edges	γ (P^a)	$\langle k \rangle$ (P^b)	Edges	γ (P^a)	$\langle k \rangle$ (P^b)	$\langle b \rangle \times 10^{-4}$ (P^b)
TEV	356	322	35	41	1275	0.07 ($<10^{-4}$)	162 (0.02)	64	0.59 ($<10^{-4}$)	18 (0.18)	9.56 (10^{-4})
TEV-At17	950	1441	32	90	2850	0.57 ($<10^{-4}$)	115 (0.23)	881	0.92 ($<10^{-4}$)	22 (0.03)	5.72 ($<10^{-4}$)
TuMV	754	390	29	30	1034	0.23 ($<10^{-4}$)	172 ($2 \cdot 10^{-4}$)	1665	0.74 ($<10^{-4}$)	34 ($<10^{-4}$)	8.63 ($<10^{-4}$)
PPV	747	740	98	8	945	0.37 ($<10^{-4}$)	153 ($5 \cdot 10^{-3}$)	535	0.82 ($<10^{-4}$)	24 (0.02)	6.46 ($<10^{-4}$)
TMV	498	225	62	0	67	0.11 ($<10^{-4}$)	76 (1)	214	0.74 ($<10^{-4}$)	22 (0.15)	3.35 (0.40)
TRV	215	284	14	26	82	0.05 ($<10^{-4}$)	111 (0.45)	154	0.56 ($<10^{-4}$)	26 ($5 \cdot 10^{-4}$)	8.20 ($<10^{-4}$)
TCV	708	846	91	70	4328	0.38 ($<10^{-4}$)	188 ($<10^{-4}$)	364	0.81 ($<10^{-4}$)	19 (0.04)	5.50 ($3 \cdot 10^{-4}$)
CaLCuV	454	732	66	107	2117	0.21 ($<10^{-4}$)	255 ($<10^{-4}$)	664	0.77 ($<10^{-4}$)	24 ($<10^{-4}$)	6.14 ($<10^{-4}$)
Interactome	-	-	-	-	139,440	1.33 (-)	114 (-)	72,266	1.54 (-)	20 (-)	3.43 (-)

shortest paths present in it, the distribution of path lengths, and their characteristic length (supplementary Figs. S9, S10 and S11, respectively; Supplementary Material online). Since PPIN represents the case of an undirected graph, we restricted the analysis to this interactome to evaluate the betweenness of the VRGs. We found that, as occurred with the connectivity at the local level, the VRGs were significantly central for seven out of eight viruses, with average betweenness centrality values significantly greater than observed for the whole interactome (supplementary Fig. S12a, Supplementary Material online; Mann-Whitney test $P < 0.05$). TMV was the exception to this rule. We also found that betweenness and connectivity are significantly positively correlated (supplementary Fig. S12b, Supplementary Material online; Spearman $\rho = 0.8885$, 10286 d.f., $P < 0.0001$, releasing the isolated nodes), despite the high variability of betweenness at low connectivity values, a characteristic of hierarchical networks.

With our study, we do not explicitly prove a physical interaction between hubs and viral proteins, but we demonstrate that the hubs of the plant interactome are over/under-expressed after viral infection. These hubs mediate the immune response to produce large changes in the genetic profile, but they are redundant in this process. Furthermore, Uetz et al. [39] found that, whereas the PPINs of *Herpesviruses* lack scale-free and small-world properties, the topology of such viral networks completely change from a highly coupled module to a more typical scale-free network of interacting submodules when integrating the interactions with human proteins. Available data from yeast two-hybrid experiments [40 - 44] allowed us to further infer a PPIN for *Potyvirus* (supplementary Fig. S13, Supplementary Material online). This inferred PPIN shows that all 11 potyviral-encoded proteins are

highly connected and that, similar to the herpesviruses case, the underlying network is not scale-free.

Modular organization of virus-altered genes

We further analyzed the subnetworks generated from the VRGs according to the different interactomes by focusing on two important topological properties: clustering and modularity. Noteworthy, the relatively smaller lists of genes available for TMV and TRV infections (owing to the type of microarray used) are to be treated with caution, as the corresponding analyses for the two viruses may reflect non-significant results. In Figs. 6a and 6b, we show the clustering coefficient for all subnetworks. According to the TRN, almost all subnetworks are tightly clustered (also validated for the TRN2 and GGIN, supplementary Fig. S14, Supplementary Material online). However, only the subnetworks for two viruses (TuMV and CaLCuV) are clustered according to the PPIN, suggesting that a disruption of the transcriptional organization could be more advantageous for the two viruses. To further support this clustering analysis, we computed the assortativity coefficients that characterize those subnetworks, in this case only focusing on transcription factors (supplementary Table S1, Supplementary Material online). Assortativity refers to a preference for a network's node to attach to others that are similar (assortative) or different (disassortative) in some way. We found that TRN-based subnetworks are essentially assortative, whereas PPIN-based ones are mostly disassortative. Thus, although biological networks tend to show a disassortative pattern - a likely consequence of the scale-free topology - regulations between transcription factor-related VRGs behave like in social networks, in which master regulators tend to be connected among them [45]. We further studied the modularity properties of those subnetworks, understood as

a decomposition metric based on the number of connected components present in the subnetwork (supplementary Fig. S15, Supplementary Material online), which is a less stringent clustering parameter. Remarkably, we found that in both interactomes, VRGs display a modular arrangement (Figs. 6c and 6d) for all infections except with TRV and TMV (a possible reflection of the much more modest number of genes analyzed) and the same result was obtained using the TRN2 and GGIN (supplementary Fig. S16, Supplementary Material online). We conclude that, in general, virus effectors tend to manipulate genes which are linked among them and that belong to the same network modules rather than randomly interacting with genes sparsely distributed in the network.

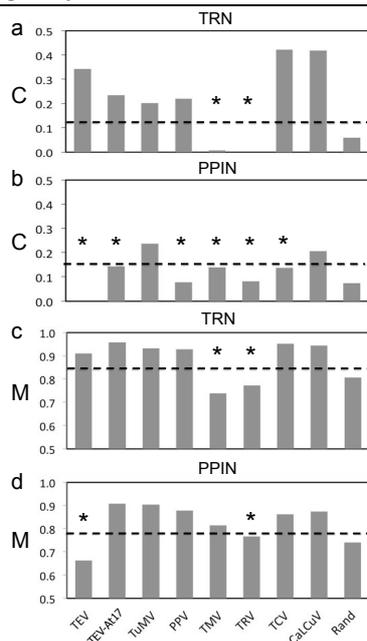


Figure 6. (a, b) Clustering (C) and (c, d) modularity (M) coefficients for the subnetworks generated by the VRGs, contextualized in the TRN and PPIN. *Rand* indicates the average value for random subnetworks (100 replicates). Asterisks denote non-significant values following a one-tailed z -test. Horizontal dashed lines represent the cutoff value for statistical significance.

Conclusions

The results of our meta-analyses combining transcriptomic data gathered for eight different viruses all infecting a common host, *A. thaliana*, confirm that host cells undergo significant reprogramming of their transcriptome during infection, which is likely a central requirement for turning on host defenses. Rather than focusing on the details of each virus infection, however, our study was designed to uncover generic features defining either the host response to, or the targets manipulated by, the various viruses tested. We found that the overlap in the lists of genes whose expression is altered upon infection (VRGs) decreases as the phylogenetic distance between the viruses increases, thus suggesting that related viruses may interact with similar host components, whereas non-related viruses may manipulate different targets. This association at the VRG

level does not hold, however, at the level of altered, global biological functions (VRF), thus suggesting that a common set of overall functional responses to infection may result from the manipulation of sometimes drastically different target genes. One caveat of the meta-analysis studies such as the one reported here, however, is that they are conservative in design. They identify responses that are strong enough to be detected against the intrinsically high noise level as a consequence of the diversity of viral systems and microarray platforms used in the original studies that served as the basis for the present one. While reductionism through single-cell transcriptome analyses has been successfully employed in virus-infected mammalian cell cultures [9] and in plant protoplasts [24], studying *in vivo* virus-host interactions obviously adds many layers of complexity and variability, which are clearly reflected here. Nonetheless, our study shows that such complexity does not, *a priori*, constitute an insurmountable obstruction to the discovery of generic patterns associated to plant viral infections.

Our study points out that VRGs are, in general, more highly connected, central and modular than expected by chance. This result agrees with the fact that viral proteins preferentially interact with hub regulator genes [27, 35 - 39], although VRGs not necessarily entail virus effectors. Probably as a plant strategy, through hub genes the signal can be disseminated at large to change the whole genetic profile. Then, even a small number of viral proteins can affect a considerable number of host genes. In the case of *Potyvirus*s, 11 mature proteins provoke significant changes in expression in about a thousand of host genes. That more hub genes (both from TRN and PPIN) than expected by chance were differentially expressed indeed reflects an effect of the virus over them, and also indicates that the information flow from virus effectors to immune response proteins is strengthened *ab initio* (lower slope in the power-law distribution). We therefore hypothesize that this over-triggering of hubs is a mechanism that confers robustness to the plant to express the immune system. Whether a virus deactivated a recognition pathway, redundant hubs would emerge to counteract this viral action. We have confirmed this observation for all the plant viruses included in our meta-analysis, thus uncovering a possible universal pattern in virus-host interactions. A second general pattern emerging from animal virus studies is that the topological properties of viral infections differ when considering only viral proteins (e.g., using yeast-two hybrids experiments) than when they are considered in the context of PPIN from the host cell [39]. Here, we have not been able to test this property owing to the lack of adequate information. A future challenge in plant virology research will be to combine data sets from yeast two-hybrids studies, transcriptomic experiments and carefully curated literature surveys, in order to reveal the specific interactions between plant and virus proteins and the effect of such interactions on viral PPIN topology. *In vitro*-reconstructed interactomes by themselves do not capture all the

biological features of the viral infection, so *in vivo* data, while adding more complexity, are essential for further studies.

Materials and Methods

Plant viruses

In this work, we studied the mode of action of seven positive-sense single-stranded RNA viruses (Baltimore's group IV) and of one virus whose genome is composed by a single-stranded circular ambisense DNA molecule (Baltimore's group II) on a common plant host, *A. thaliana*. The set of RNA viruses comprises three potyviruses (family *Potyviridae*), *Tobacco etch virus* (TEV), *Turnip mosaic virus* (TuMV) and *Plum pox virus* (PPV), one tobamovirus (family *Virgaviridae*), *Tobacco mosaic virus* (TMV), one tobnavirus (family *Virgaviridae*), *Tobacco rattle virus* (TRV), and one carmovirus (family *Tombusviridae*), *Turnip crinkle virus* (TCV). In addition, we considered a laboratory-evolved strain of TEV (TEV-*At17*), which was obtained after 17 serial passages in *A. thaliana* [22]. TEV-*At17* shows higher fitness and produces more severe symptoms in *A. thaliana* than the ancestral TEV strain. The ssDNA virus included in the study was the begomovirus (family *Geminiviridae*) *Cabbage leaf curl virus* (CaLCuV).

Transcriptomic data

TEV and TEV-*At17* expression data (two-color raw data, NCBI GEO accession GSE11088) were obtained from ecotype *Ler-0* plants 14 days post-inoculation (dpi) [21 - 22]. TuMV data (Affymetrix raw data, ArrayExpress accession e-mexp-509) were obtained 5 dpi from ecotype Col-0 plants [20]. These three data sets were normalized using the RMA method [46] for background correction and quantiles for array scaling, and the list of differentially expressed genes was obtained by performing a Limma test [47] with a correction for multiple testing using the false discovery rate (FDR) procedure [55] (adjusted $P < 0.05$). PPV data (Affymetrix preprocessed data, NCBI GEO accession GSE11217) were obtained 17 dpi from Col-0 plants [24]. In this case, data normalization was done using the Affymetrix MAS 5.0 software package, and the differential expression using a one-way ANOVA test with a correction for multiple testing using the FDR procedure (adjusted $P < 0.05$), followed by a fold-change criterion of 1.5 in z -score over all genes (averaging replicates). TMV data (two-color raw data, deposited in www.bio.puc.cl/labs/arce/index.html) were obtained from ecotype *Uk-4* plants 10 dpi [19], and normalized using the RMA method for background correction and quantiles for array scaling. The list of differentially expressed genes was obtained by performing a fold-change criterion of 1.96 in z -score over all genes (averaging replicates). TRV data (two-color raw data, NCBI GEO accession GSE15557) were measured 8 dpi from Col-0 leaves. TCV data (two-color raw data, NCBI GEO accession GSE29387) were quantified 10 dpi in Col-0 plants. These two data sets were normalized using the CATMA BGS procedure [48], and

the list of differentially expressed genes was obtained by performing a Limma test with FDR correction (adjusted $P < 0.05$). In addition, for TCV data, a fold-change criterion of 1.96 in z -score over all genes (averaging replicates) was applied. Finally, CaLCuV data (Affymetrix raw data, ArrayExpress accession E-ATMX-34) were collected from Col-0 plants 12 dpi [23]. These data were normalized using subtraction for background correction and LOWESS [49] for array scaling, and the list of differentially expressed genes was obtained by performing a mixed ANOVA test with a correction for multiple testing using the FDR procedure (adjusted $P < 0.05$). To perform the data normalization and to obtain the differentially expressed genes, we used the GEPAS tool [50], which is implemented within the BABELOMICS webserver [51]. We would like to notice at this point that the variability in the normalization methods comes from the lack of source data files that prevents incorporating them into a common platform, although we have intended to provide a homogeneous compendium as much as possible.

Validity of meta-analysis

The heterogeneity in the host ecotype used in different experiments (*Ler-0* for TEV and TEV-*At17*, *Uk-4* for TMV and Col-0 for the rest) and in the time at which samples were obtained (ranging from 5 to 17 dpi) may weaken the conclusions from a meta-analysis. Therefore, to evaluate the robustness of our results, we first tested for the effect of these two variables. First, for each gene that showed a significant alteration in its expression level in at least one of the eight viral infections (a total of 6546 genes), we sought whether the observed differences grouped according to the plant ecotype used in the experiments. For this, we classified experiments into two categories: those performed in Col-0 *versus* those not performed in Col-0. Only seven genes (*At1g14970*, *At1g50250*, *At1g78170*, *At2g16700*, *At2g20780*, *At3g45860*, and *At4g12520*) had expression levels that were significantly affected by the host ecotype (Mann-Whitney test). However, if a correction for multiple testing (FDR procedure; adjusted $P < 0.05$) was used, none of these seven genes remained significant. Second, for each of the 6546 altered genes, we sought whether expression levels classified according to sampling time. Sampling times were ranked into three categories (early, between 5 - 8 dpi; intermediate, between 10 - 12 dpi; and late, in the range 14 - 17 dpi). In this case, 54 genes showed a significant effect of the sampling time (Kruskal-Wallis test), although none of them remained significant after applying the more stringent FDR procedure. In addition to this test, we also sought for significant correlations between expression levels and sampling time (Pearson correlation coefficient). In this case, 96 genes showed a significant correlation (either positive or negative), although none of them remained so after the FDR correction. Therefore, we conclude that differences in ecotype or in sampling time would neither have a significant effect on the conclusions drawn from our

meta-analyses. Nonetheless, the conclusions from our meta-analyses should be taken conservatively.

Functional analysis

For each list of “virus-regulated genes”, or simply VRGs, (over- or under-expressed), we looked for the significant over-represented biological processes (GO terms between levels 3 and 9, referred in the text as VRFs – or virus-regulated functions–) within that list. The statistical significance was evaluated by means of a Fisher's exact test for 2x2 contingency tables with a correction for multiple testing using the FDR procedure (adjusted $P < 0.05$). To perform the functional analysis of the VRGs, we used the FatiGO tool [52], implemented in the BABELOMICS webserver [51]. At the metabolic level, we overlaid the expression data for each viral infection into the global metabolic map of *A. thaliana*. To visualize the map we used the TAIR AraCyc tool [53].

To quantitatively evaluate the similarity in the lists of VRGs and of the corresponding biological functions, we performed a hierarchical clustering analysis by constructing neighbor-joining dendrograms using the program NEIGHBOR from the PHYLIP v3.6 package (<http://evolution.genetics.washington.edu/phylip.html>).

The similarity matrix S was defined by using the metric $S_{ij} = 2N_{ij}/(N_i + N_j)$, where N_i and N_j are the total number of genes (or GO terms) whose expression is altered upon infection with virus i and j , and N_{ij} the number of genes (or GO terms) altered by both viruses i and j . Statistical significance of the different clusters was evaluated by bootstrapping the gene (or GO) lists (based on 1000 pseudoreplicates).

Genomic interaction networks

The transcriptional regulatory network (TRN) was previously published by Carrera et al. [32]. In short, the model was inferred using a reverse-engineering procedure, based on mutual information with a local significance (z -score computation) as estimator of the likelihood, for capturing coexpression patterns between transcription factors (TFs) and genes, and has optimal levels of confidence and coverage. This network contains 139,440 TF-gene interactions and involves 19,108 genes. For the protein-protein interaction network (PPIN), we used the release 2.0 of the *A. thaliana* predicted interactome available for downloading at TAIR (www.arabidopsis.org). This network consists of a set of 72,266 predicted interactions involving 7177 proteins, of which about 3000 interactions are experimentally confirmed (merging datasets from TAIR, IntAct-EBI, and BIND/BOND). In short, the prediction algorithm [33] began with the identification of orthologs of *A. thaliana* proteins in seven other species (*Escherichia coli*, *Saccharomyces pombe*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus*, and *Homo sapiens*) for which partial interactions existed. Next, an interaction was predicted to exist in *A. thaliana* if it was described for any of the seven species. Notice that TRN is

directed, whereas PPIN is undirected. In addition, we considered a smaller transcriptional network (TRN2), with high confidence and low coverage, that contains 18,446 TF-gene interactions and links 7108 genes [32], and the graphical Gaussian interaction network (GGIN) model previously published by Ma et al. [34] that contains 21,101 effective gene-to-gene interactions involving 6722 genes, where coexpression patterns between gene pairs were evaluated according to a conditional correlation.

Topological analysis

To analyze the impact of a viral infection in terms of genetic interactions, we studied the principal topological properties on the inferred networks: connectivity, clustering, connected components, shortest paths and modularity. For each VRG (up and down), we collected its connectivity degree and betweenness centrality, according to the global interactome. Differences in connectivity (k) and betweenness (b) among the VRGs and the total set of plant genes were analyzed by means of one-tailed Mann-Whitney U -tests ($P < 0.05$) considering the superior tails of the distributions (i.e., the genes satisfying $k > \langle k \rangle$ or $b > \langle b \rangle$). Furthermore, we performed linear regressions in the log-log space to obtain the critical exponents, γ , of the power-law degree distribution $P(k) \sim k^{-\gamma}$ and assessed the statistical significance of the inferred values using Student t -tests ($P < 0.0001$). In addition, for each virus we generated the corresponding subnetwork by selecting those VRGs, releasing those isolated nodes. Random subnetworks were also constructed to perform statistical significance tests. The clustering coefficient (C), the assortativity coefficient (A), the number of connected components (CC), and the shortest path (SP) distribution were computed to characterize the subnetworks. Here, A was defined as the slope of the linear regression between the connectivity of a node and the average connectivity of its neighbors, being a network assortative when $A > 0$. Moreover, we defined a modularity coefficient (M), less stringent than the clustering one, given by

$$M = 1 - \sum_c \frac{N_c \log(N/N_c)}{N \log N},$$

(*sensu* Shannon entropy in information theory) where N_c is the number of genes in the connected component c , and N the total number of genes in the subnetwork [54]. Accordingly, $M = 1$ in case of just one CC , whereas M tends to 0 as the number of CC increases. To assess the statistical significance of the results, we performed a one-tailed z -test with a confidence level of 95% ($z > 1.64$) over 100 random subnetworks.

Acknowledgements

This work was supported by the Spanish Ministerio de Ciencia e Innovación (MICINN) grants BFU2009-06993 (S.F.E.) and BIO2006-13107 (C.L.) and by Generalitat Valenciana grant PROMETEO2010/016 (S.F.E.). G.R. is supported by a graduate fellowship from the Generalitat Valenciana (BFPI2007-160) and J.C. by a contract from MICINN grant TIN2006-12860. This work was conceived

and in part developed by SFE and JC while at residence at SFI.

References

1. Penga X, Chana EY, Lia Y, Diamonda DL, Kortha MJ, Katze MG (2009) Virus-host interactions: from systems biology to translational research. *Curr Opin Microbiol*, 12:432-438.
2. Dodds PN, Rathjen JP (2010) Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet*, 11:539-548.
3. Maule A, Leh V, Lederer C (2002) The dialogue between viruses and hosts in compatible interactions. *Curr Opin Plant Biol*, 5:279-284.
4. Whitham SA, Quan S, Chang HS, Cooper B, Estes B, Zhu T, Wang W, Hou YM (2003) Diverse RNA viruses elicit the expression of common sets of genes in susceptible *Arabidopsis thaliana* plants. *Plant J*, 33:271-283.
5. Bailer SM, Haas J (2009) Connecting viral with cellular interactomes. *Curr Opin Microbiol*, 12:453-459.
6. Whitham SA, Yang C, Goodin MM (2006) Global impact: elucidating plant responses to viral infection. *Mol Plant-Microb Interact*, 19:1207-1215.
7. MacPherson JI, Dickerson JE, Pinney JW, Robertson DL (2010) Patterns of HIV-1 protein interaction identify perturbed host-cellular subsystems. *PLoS Comput Biol*, 6:e1000863.
8. Jenner RG, Young RA (2005) Insights into host responses against pathogens from transcriptional profiling. *Nat Rev Microbiol*, 3:281-294.
9. Andeweg AC, Haagmans BL, Osterhaus ADME (2008) Virogenomics: the virus-host interaction revisited. *Curr Opin Microbiol*, 11:461-466.
10. Elena SF, Carrera J, Rodrigo G (2011) A systems biology approach to the evolution of plant-virus interactions. *Curr Opin Plant Biol*, 14:372-377.
11. Tan SL, Ganji G, Paepfer B, Proll S, Katze MG (2007) Systems biology and the host response to viral infection. *Nat Biotechnol*, 25:1383-1389.
12. De la Fuente A (2010) From 'differential expression' to 'differential networking' - identification of dysfunctional regulatory networks in diseases. *Trends Genet*, 26:326-333.
13. Albert R (2005) Scale-free networks in cell biology. *J Cell Sci*, 118:4947-4957.
14. Yu H, Braun P, *et al.* (2008) High-quality binary protein interaction map of the yeast interactome network. *Science*, 322:104-110.
15. Barabási AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet*, 5:101-113.
16. Albert R, Jeong H, Barabási AL (2000) Error and attack tolerance of complex networks. *Nature*, 406:378-382.
17. Golem S, Culver JN (2003) *Tobacco mosaic virus* induced alterations in the gene expression profile of *Arabidopsis thaliana*. *Mol Plant-Microb Interact*, 16:681-688.
18. Ishihara T, Sakurai N, Sekine KT, Hase S, Ikegami M, Shibata D, Takahashi H (2004) Comparative analysis of expressed sequence tags in resistant and susceptible ecotypes of *Arabidopsis thaliana* infected with *Cucumber mosaic virus*. *Plant Cell Physiol*, 45:470-480.
19. Espinoza C, Medina C, Somerville S, Arce-Johnson P (2007) Senescence-associated genes induced during compatible viral interactions with grapevine and *Arabidopsis*. *J Exp Bot*, 58:3197-3212.
20. Yang C, Guo R, Jie F, Nettleton D, Peng J, Carr T, Yeakley JM, Fan JB, Whitham SA (2007) Spatial analysis of *Arabidopsis thaliana* gene expression in response to *Turnip mosaic virus* infection. *Mol Plant-Microb Interact*, 20:358-370.
21. Agudelo-Romero P, Carbonell P, De la Iglesia F, Carrera J, Rodrigo G, Jaramillo A, Pérez-Amador MA, Elena SF (2008) Changes in the gene expression profile of *Arabidopsis thaliana* after infection with *Tobacco etch virus*. *Virol J*, 5:92.
22. Agudelo-Romero P, Carbonell P, Perez-Amador MA, Elena SF (2008) Virus adaptation by manipulation of host's gene expression. *PLoS ONE*, 3:e2397.
23. Ascencio-Ibáñez J, Sozzani R, Lee TJ, Chu TM, Wolfinger RD, Cella R, Hanley-Bowdoin L (2008) Global analysis of *Arabidopsis* gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. *Plant Physiol*, 148:436-454.
24. Babu M, Griffiths JS, Huang TS, Wang A (2008) Altered gene expression changes in *Arabidopsis* leaf tissues and protoplasts in response to *Plum pox virus* infection. *BMC Genomics*, 9:325.
25. De Vienne DM, Giraud T, Martin OC (2007) A congruence index for testing topological similarity between trees. *Bioinformatics*, 23:3119-3124.
26. Wise RP, Moscou MJ, Bogdanove AJ, Whitham SA (2007) Transcript profiling in host-pathogen interactions. *Annu Rev Phytopathol*, 45:329-369.
27. Mukhtar MS, *et al.* (2011) Independently Evolved Virulence Effectors Converge onto Hubs in a Plant Immune System Network. *Science*, 333:596-601.
28. Handford MG, Carr JP (2007) A defect in carbohydrate metabolism ameliorates symptom severity in virus-infected *Arabidopsis thaliana*. *J Gen Virol*, 88:337-341.
29. Hou B, Lim EK, Higgins GS, Bowles DJ (2004) N-glycosylation of cytokinins by glycosyltransferases of *Arabidopsis thaliana*. *J Biol Chem*, 279:47822-47832.
30. Schwender J, Goffman F, Ohlrogge JB, Shachar-Hill Y (2004) Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature*, 432:779-782.
31. Pagán I, Alonso-Blanco C, García-Arenal F (2008) Host responses in life-history traits and tolerance to

- virus infection in *Arabidopsis thaliana*. *PLoS Pathog*, 4:e1000124.
32. Carrera J, Rodrigo G, Jaramillo A, Elena SF (2009) Reverse-engineering the *Arabidopsis thaliana* transcriptional network under changing environmental conditions. *Genome Biol*, 10:R96.
 33. Geisler-Lee J, O'Toole N, Ammar R, Provart NJ, Millar AH, Geisler M (2007) A predicted interactome for *Arabidopsis thaliana*. *Plant Physiol*, 145:317-329.
 34. Ma S, Gong Q, Bohnert HJ (2007) An *Arabidopsis* gene network based on the graphical Gaussian model. *Genome Res*, 17:1614-1625.
 35. Calderwood MA, Venkatesan K, Xing L, Chase MR, Vazquez A, Holthaus AM, Ewence AE, Li N, Hirozane-Kishikawa T, Hill DE, Vidal M, Kieff, E, Johannsen E (2007) Epstein-Barr virus and virus human protein interaction maps. *Proc Natl Acad Sci USA*, 104:7606-7611.
 36. De Chassey B, Navratil V, Tafforeau L, Hietet MS, Aublin-Gex A, Agaugué S, Meinffren G, Pradezynski F, Faria BF, Chantier T, Le Breton M, Pellet J, Davoust N, Mangeot PE, Chaboud A, Penin F, Jacob Y, Vidalain PO, Vidal M, André P, Rabourdin-Combe C, Lotteau V (2008) Hepatitis C virus infection protein network. *Mol Syst Biol*, 4:230.
 37. Shapira SD, Gat-Viks I, Shum BOV, Dricot A, de Grace MM, Wu L, Gupta PB, Hao T, Silver SJ, Root DE, Hill DE, Regev A, Hacohen N (2009) A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell*, 139:1255-1267.
 38. Dyer MD, Murali TM, Sobral BW (2008) The landscape of human protein interacting with viruses and other pathogens. *PLoS Pathog*, 4:e32.
 39. Uetz P, Dong YA, Zeretzke C, Atzler C, Baiker A, Berger B, Rajagopala SV, Roupelieva M, Rose D, Fossum E, Haas J (2006) Herpesviral protein networks and their interaction with the human proteome. *Science*, 311:239-242.
 40. Choi IR, Stenger, DC, French R (2000) Multiple interactions among proteins encoded by the mite-transmitted *Wheat streak mosaic tritimovirus*. *Virology*, 267:185-198.
 41. Guo D, Rajamäki, M.L., Saarma, M., Valkonen JPT (2001) Towards a protein interaction map of potyviruses: protein interaction matrixes of two potyviruses based on the yeast two-hybrid system. *J Gen Virol*, 82:935-939.
 42. Kang SH, Lim WS, Kim KH (2004) A protein interaction map of *Soybean mosaic virus* strain G7H based on the yeast two-hybrid system. *Mol Cells*, 18:122-126.
 43. Lin L, Shi Y, Luo Z, Lu Y, Zheng H, Yan F, Chen J, Chen J, Adams MJ, Wu Y (2009) Protein-protein interactions in two potyviruses using the yeast two-hybrid system. *Virus Res*, 142:36-40.
 44. Shen WT, Wang MQ, Yan P, Gao L, Zhou P (2010) Protein interaction matrix of *Papaya ringspot virus* type P based on a yeast two-hybrid system. *Acta Virol*, 54:49-54.
 45. Redner S (2008) Networks: teasing out the missing links. *Nature*, 453:47-48.
 46. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4:249-264.
 47. Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statist Appl Genet Mol Biol*, 3:3.
 48. Allemeersch J, Durinck S, Vanderhaeghen R, Alard P, Maes R, Seeuws K, Bogaert T, Coddens K, Deschouwer K, van Hummelen P, Vuylsteke M, Moreau Y, Kwekkeboom J, Wijffes AH, May S, Beynon J, Hilson P, Kuiper MT (2005) Benchmarking the CATMA microarray. A novel tool for *Arabidopsis* transcriptome analysis. *Plant Physiol*, 137:588-601.
 49. Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc*, 74:829-836.
 50. Tárraga J, Medina I, Carbonell J, Huerta-Cepas J, Minguez P, Alloza E, Al-Shahrour F, Vegas-Azcárate S, Goetz S, Escobar P, García-García F, Conesa A, Montaner D, Dopazo J (2008) GEPAS, a web-based tool for microarray data analysis and interpretation. *Nucl Acids Res*, 36:W308-W314.
 51. Al-Shahrour F, Minguez P, Vaquerizas JM, Conde L, Dopazo J (2005) BABELOMICS: a suite of web tools for functional annotation and analysis of groups of genes in high-throughput experiments. *Nucl Acids Res*, 33:W460-W464.
 52. Al-Shahrour F, Minguez P, Tárraga J, Medina I, Alloza E, Montaner D, Dopazo J (2007) FatiGO+: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. *Nucl Acids Res*, 35:W91-W96.
 53. Mueller LA, Zhang P, Rhee SY (2003) AraCyc: A biochemical pathway database for *Arabidopsis*. *Plant Physiol*, 132:453-460.
 54. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J*, 27:379-423.
 55. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B*, 57:289-300.

Supplementary Material

File S1: Differentially expressed genes (up/down) for all viral infections and those genes shared by pairs of viruses. In addition, we provide the similarity matrix.

File S2: Over-represented GO terms (up/down) for all viral infections and those GO terms shared by pairs of viruses. In addition, we provide the similarity matrix.

File S3: List of genetic interactions for the different interactomes we considered in this work (TRN, TRN2, PPIN, and GGIN).

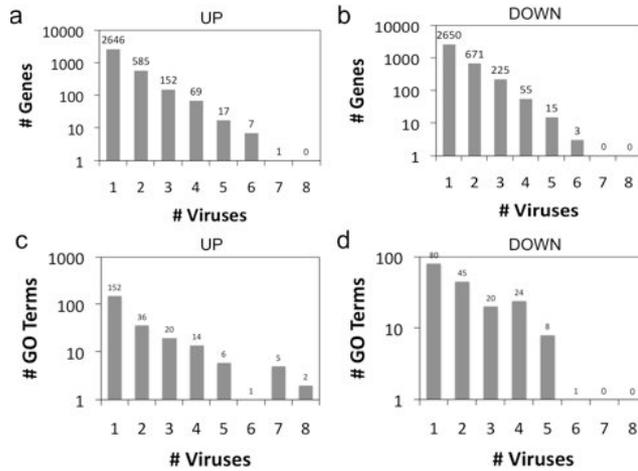


Figure S1: Distribution of genes up/down-expressed (a and b) and GO terms over/under represented (c and d) in *A. thaliana* after infection with the number of viruses indicated in the ordinates axis. The distributions are the result of comparing the differential patterns *a posteriori* between several viruses.

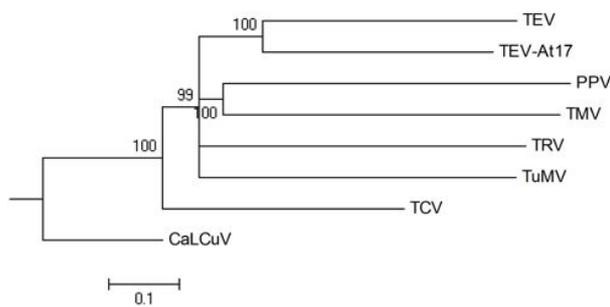


Figure S2: Neighbor-joining dendrogram constructed using the similarity matrix computed from the lists of differentially expressed genes. Bootstrap support values are reported next to each node.

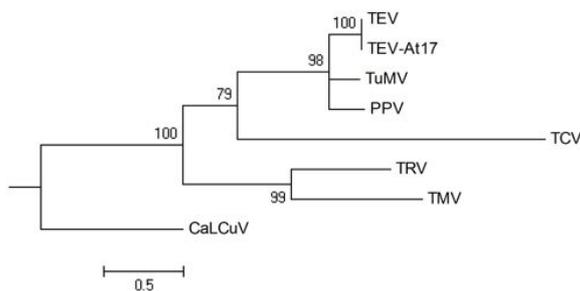


Figure S3: Maximum-likelihood phylogenetic tree constructed from the replicase genes of the seven RNA viruses included in the study. For CaLCuV, the Rep (replicase-associated protein) was used instead. The statistical quality of the different clusters was evaluated by bootstrap.

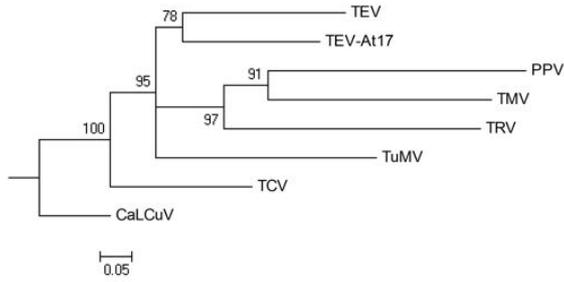


Figure S4: Neighbor-joining dendrogram constructed using the similarity matrix computed using the lists of over-represented GO terms.

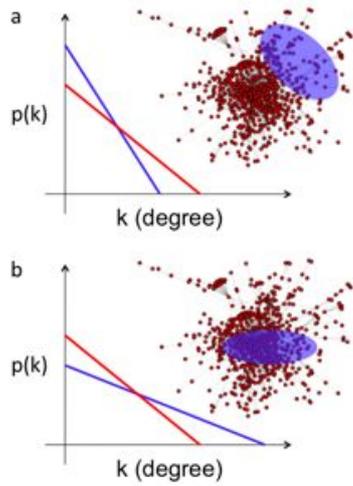


Figure S5: Predicted connectivity distribution for the whole plant interactome (red line) and the subnetwork generated by the total differentially (up/down) expressed VRGs (blue line). The set of VRGs can be in the periphery of the interactome if they have low connectivity (a) or in the core in they are highly connected (b).

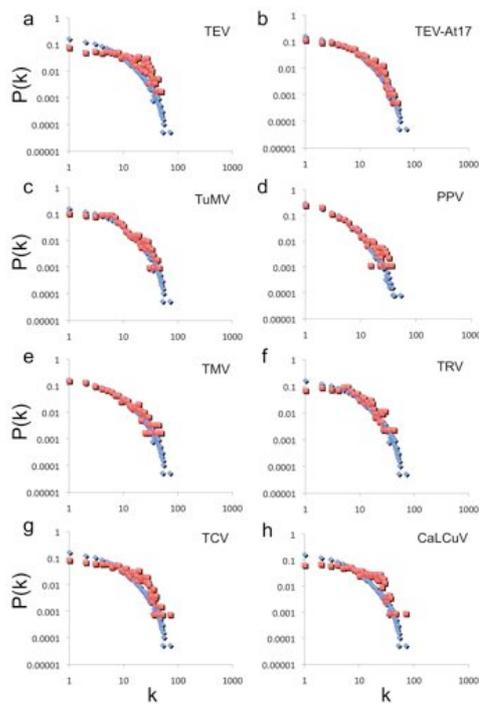


Figure S6: Incoming connectivity distribution, contextualized in the TRN interactome, for the VRGs (red), and the whole interactome (blue). The label *evoTEV* denotes *TEV-At17*.

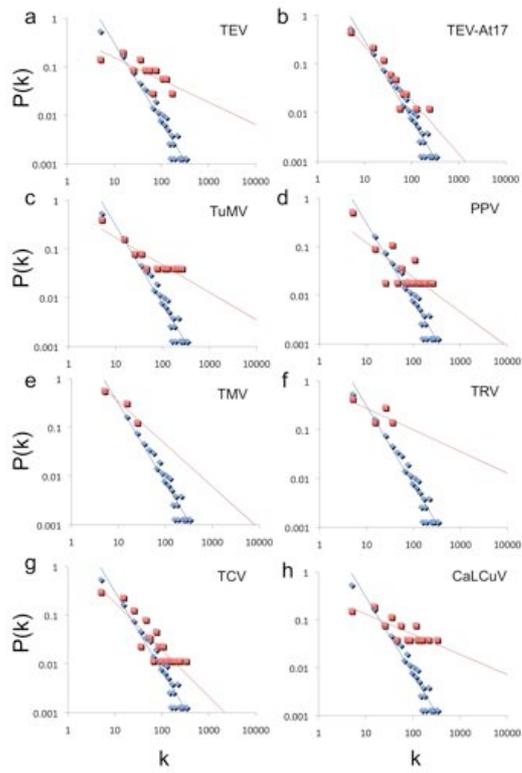


Figure S7: Outgoing connectivity distribution, contextualized in the TRN2 interactome, for the VRGs (red), and the whole interactome (blue).

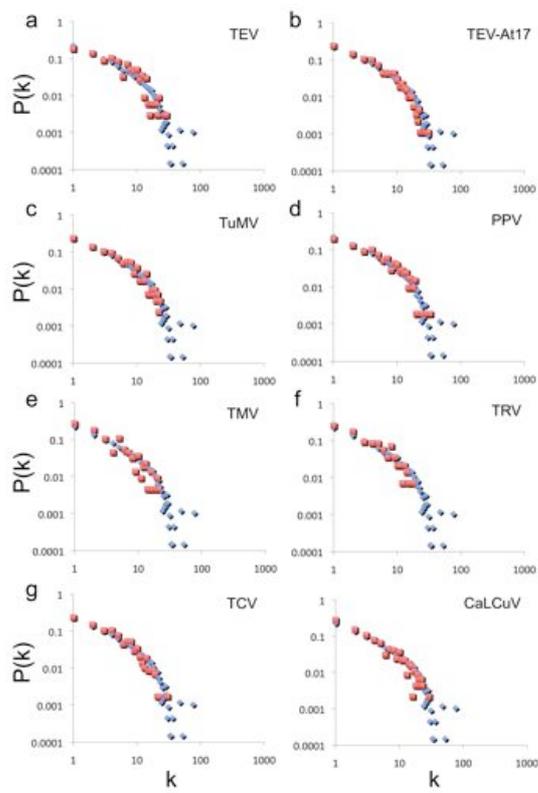


Figure S8: Connectivity distribution, contextualized in the GGIN interactome, for the VRGs (red), and the whole interactome (blue).

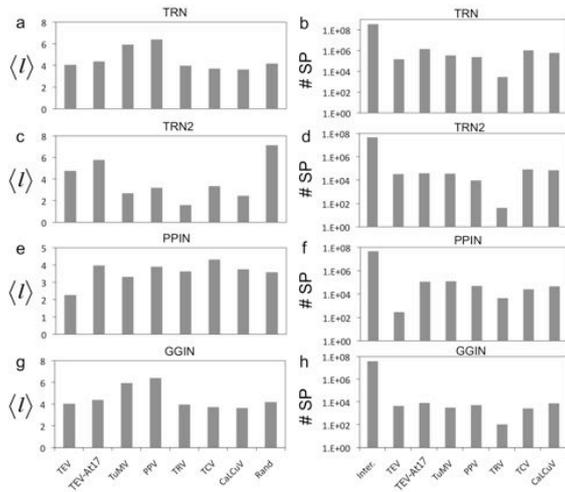


Figure S9: Shortest path average (a, c, e, and g) and total number of shortest paths (b, d, f, and h) for the subnetworks generated by the differentially expressed genes from several viral infections contextualized in different interactomes. *Rand* indicates the average value of randomly selected gene lists. *Inter* stands for interactome. For the TRN and TRN2 interactomes, we considered undirected edges; otherwise, the number of shortest paths is very low.

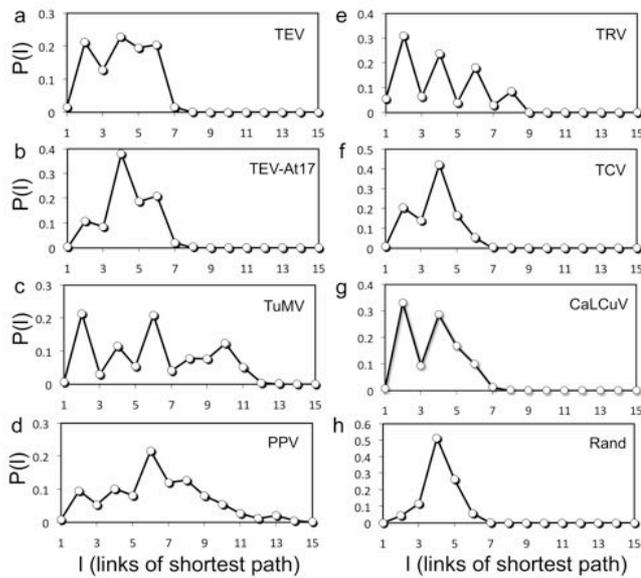


Figure S10: Shortest path distribution, contextualized in the TRN interactome, for the subnetwork generated by the differentially expressed genes after viral infection (a, b, c, d, e, f, and g), and by lists of randomly selected genes (h). For this interactome, we considered undirected edges; otherwise, the number of shortest paths is very low.

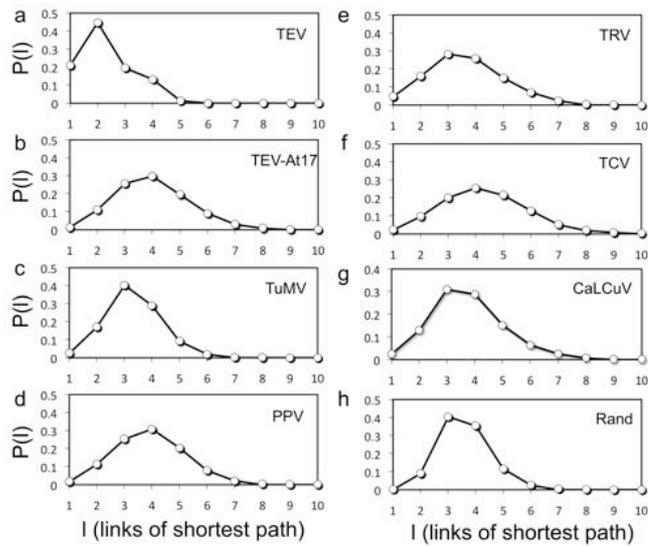


Figure S11: Shortest path distribution, contextualized in the PPIN interactome, for the subnetwork generated by the differentially expressed genes after viral infection (a, b, c, d, e, f, and g), and by lists of randomly selected genes (h).

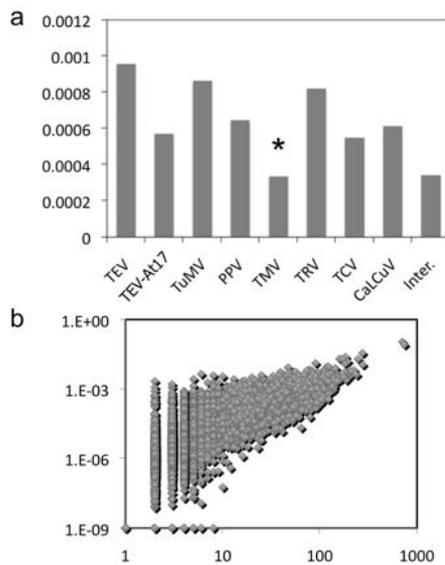


Figure S12: (a) Average betweenness, contextualized in the PPIN interactome, for the differentially expressed genes after viral infection and the whole interactome. Asterisk denotes non-significant value according to a Mann-Whitney U -test. (b) Scatter plot of betweenness centrality and connectivity degree for the whole PPIN interactome. *Inter* corresponds to the betweenness computed to the PPIN interactome.

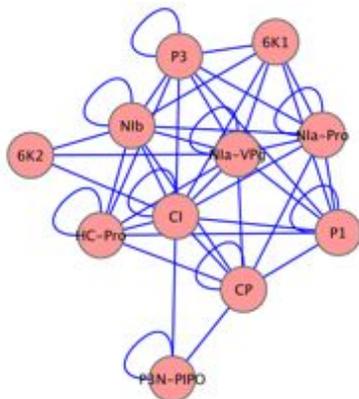


Figure S13: Protein-protein interaction network of *Potyviruses* inferred from empirical data gathered by different authors using the yeast two-hybrid system. The parameters describing the network are: clustering coefficient 0.8713, network diameter 2, shortest path 110, characteristic path length 1.345, average number of neighbors 6.545, number of edges 45, network density 0.655, and number of self-loops 9. The 11 potyviral proteins are: P1 (trypsin-like serine proteinase), HC-Pro (helper-component during aphid transmission, RNA-silencing suppressor, and papain-like cysteine proteinase), P3 (pathogenicity determinant), P3N-PIPO (movement protein), 6K1 (unknown function), CI (ATPase/RNA helicase and cell-to-cell movement), 6K2 (anchoring replication complexes to membranes), Nla-VPg (5'-linked protein involved in genome replication), Nla-Pro (trypsin-like serine proteinase), NIB (replicase), and CP (coat protein).

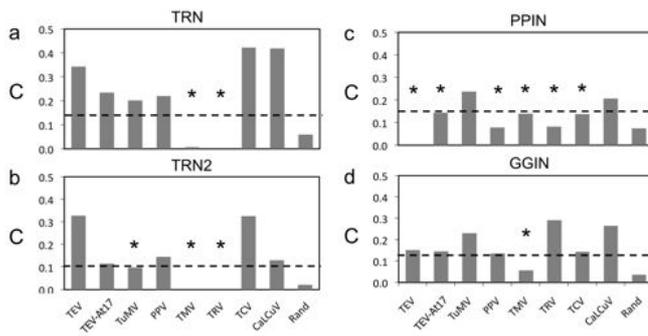


Figure S14: Clustering coefficient (C) for the subnetworks generated by the differentially expressed genes from several viral infections, contextualized in different *A. thaliana* interactomes. *Rand* indicates the average value of randomly selected gene lists (100 replicates). Asterisks denote non-significant values following a one-tailed z-test. Horizontal dashed lines represent the cutoff value for statistical significance.

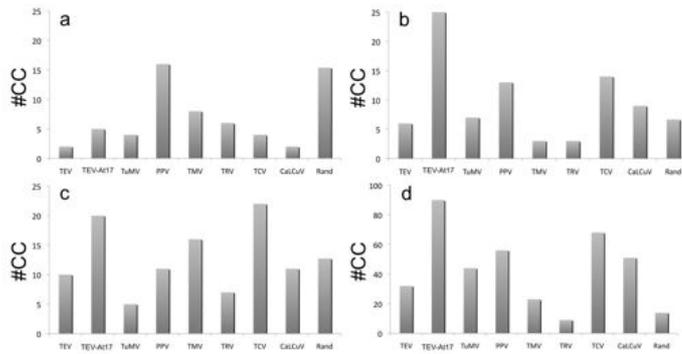


Figure S15: Number of connected components (CC) for the subnetworks generated by the differentially expressed genes from several viral infections, contextualized in *A. thaliana* (a) TRN, (b) TRN2, (c) PPIN, and (d) GGIN interactome. *Rand* indicates the average value of randomly selected gene lists (100 replicates).

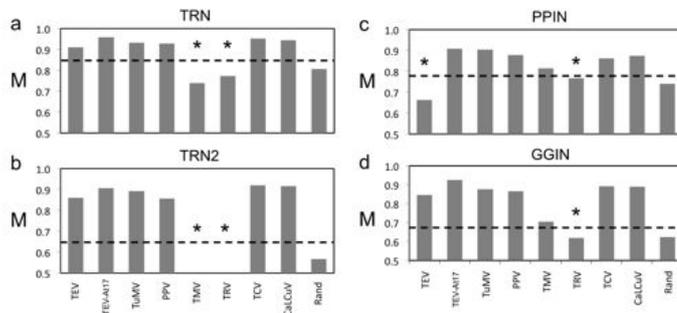


Figure S16: Modularity coefficient (M) for the subnetworks generated by the differentially expressed genes from several viral infections, contextualized in different *A. thaliana* interactomes. *Rand* indicates the average value of randomly selected gene lists (100 replicates). Asterisks denote non-significant values following a one-tailed z-test. Horizontal dashed lines represent the cutoff value for statistical significance.

Table S1: Assortativity coefficients (A) for the virus-associated subnetworks and the interactomes TRN_{TF} (only considering transcription factors) and PPIN.

	A (P -value) in TRN_{TF}	A (P -value) in PPIN
TEV	0.2914 (0.0002)	-0.6163 (0.0341)
TEV- <i>At17</i>	0.1343 (< 0.0001)	-0.2783 (0.0008)
TuMV	0.0700 (0.1287)	0.1058 (0.0397)
PPV	0.1810 (0.0101)	-0.2185 (0.0008)
TMV	0.1200 (0.3188)	-1.6168 (0.0004)

TRV	-0.5422 (0.0040)	-0.0843 (0.4485)
TCV	0.1929 (0.0012)	-0.2765 (0.0199)
CaLCuV	0.1591 (0.0225)	-0.0122 (0.8858)
Interactome	0.1692 (0.0022)	-0.0140 (0.3012)
