A comprehensive physical interaction map between the Turnip mosaic potyvirus and Arabidopsis thaliana proteomes

Fernando Martínez
CSIC
Christina Toft
CSIC
Julia Hillung
CSIC
José L. Carrasco
CSIC https://orcid.org/0000-0002-5905-8042
Silvia Giménez-Santamarina
CSIC https://orcid.org/0000-0001-8021-4348
Lynne Yenush
UPV https://orcid.org/0000-0001-8589-7002
Guillermo Rodrigo
CSIC
Santiago Elena (santiago.elena@csic.es)
CSIC

Article

Keywords: co-evolution, high-throughput screening, plant virus, protein network, systems biology, virus-host interaction.

Posted Date: May 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-149993/v2

A comprehensive physical interaction map between the *Turnip mosaic potyvirus* and *Arabidopsis thaliana* proteomes

Fernando Martínez[1,†], Christina Toft[1], Julia Hillung[1], José L. Carrasco[1], Silvia M. Giménez-Santamarina[2], Lynne Yenush[2], Guillermo Rodrigo[1,†,*], and Santiago F. Elena[1,3,*]

[1]Instituto de Biología Integrativa de Sistemas (I[2]SysBio), CSIC - Universitat de València, 46980 Paterna, Spain

[2]Instituto de Biología Molecular y Celular de Plantas (IBMCP), CSIC - Universitat Politècnica de València, 46022 València, Spain

[3]The Santa Fe Institute, Santa Fe NM87501, USA

[†]Equal contribution

[*]Correspondence: guillermo.rodrigo@csic.es (G.R.) and santiago.elena@csic.es (S.F.E.)

Abstract

Viruses are obligate intracellular parasites that have co-evolved with their hosts to establish an intricate network of protein-protein interactions. Yet, the systems-level mode of action of plant viruses remains poorly understood. Here, we followed a high-throughput yeast two-hybrid screening to identify 378 novel virus-host protein-protein interactions between *Turnip mosaic virus* (TuMV), a representative plant RNA virus, and *Arabidopsis thaliana*, one of its natural hosts. We found the RNA-dependent RNA polymerase NIb as the virus protein with the largest number of contacts. We verified a subset of 25 selected interactions *in planta* by bimolecular fluorescence complementation assays. We then constructed a comprehensive network comprising 399 TuMV-*A. thaliana* interactions to perform, together with intravirus and intrahost connections, detailed computational analyses. In particular, we found that the host proteins targeted by the virus participate in a higher number of infection-related functions, are more connected and have an increased capacity to spread information throughout the cell proteome, display higher expression levels, and have been subject to stronger purifying selection than expected by chance. Overall, our results provide a comprehensive mechanistic description of a plant virus-host interplay, with potential impact on disease etiology, and reveal that plant and animal viruses share fundamental features in their mode of action.

Keywords: co-evolution; high-throughput screening; plant virus; protein network; systems biology; virus-host interaction.

Viral infections represent an important threat in our global society, not only because some viruses (established or emergent) can greatly compromise human health [1], but because some others jeopardize intensive agriculture and livestock production [2], both which are of special relevance in the current scenario of climate change [3]. A correct understanding of the different molecular mechanisms that contribute to the burden imposed by viral infections is instrumental for the successful development of treatments and corrective measures. In this regard, the identification of host factors that directly interact with the virus (*i.e.*, with its genome and proteome) has long been one of the central goals of molecular virology [4]. However, with the advent of high-throughput omics techniques [5], a fundamental research line has emerged by moving from the identification and thorough characterization of individual host factors to the simultaneous experimental determination and network-based computational analyses of hundreds of virus-host interactions [6,7]. Certainly, by defining how the virus perturbs the host proteome in terms of function and structure, we can better understand the infection process and its consequences for host physiology in a systemic manner. However, the field of systems virology still requires much work in order to grasp the yet unknown mode of action of many plant and animal viruses that appear to challenge lifestyle improvement and sustainable development of our society.

Viruses exhibit, through a limited number of proteins (especially limited in the case of RNA viruses) [8], multiple contact points with the host proteome, *i.e.*, virus-host protein-protein interactions. Indeed, this is the combined result of, on one hand, their necessity to manipulate diverse cellular pathways to create a favorable environment for their replication (either by sequestering resources for their own benefit or by interfering with the host immune responses) and, on the other hand, the ability of the host cell receptors to sense foreign elements to then act accordingly. Considerable progress has

been made over the last years to generate detailed, high quality virus-host maps in the case of human viruses [9-13]. Furthermore, integrative approaches have identified general and specific molecular mechanisms employed by different human viruses [14,15] and, together with additional omics data, have even been used to predict phenotypic outcomes of infection [16]. Certainly, all these developments take advantage of the continuous elaboration of an accurate large-scale map of human protein-protein interactions [17], which despite its undeniable incompleteness appears to be very useful to recognize disease-associated modules [18].

According to previous network-based analyses, viral targets tend to be more connected with other host proteins than expected by chance [9,14]. This means that viral proteins interact with some hub proteins of the host interactome, which are otherwise difficult to reach by random attacks [19], since biological networks are scale-free (*i.e.*, they are described by a power-law distribution of connectivity; in other words, there are a few elements strongly connected in the network, while most of them are weakly connected) [20]. Consequently, viruses (and also pathogenic bacteria) perturb the host network at a global scale (and consequently host physiology) in a guided way, although less significantly than through pure centrality-directed attacks [21]. However, this property reflects an average tendency, as viral proteins also interact with a substantial set of non-central and even peripheral proteins. Previous studies have revealed that some viral proteins bridge different subnetworks within the host interactome that otherwise might appear as disconnected, with an increased ability to spread information [22]. In addition, we now realize that viral targets, at least a subset of them, participate in signaling pathways that are linked to the symptoms of infections and are located in the neighborhood of disease-associated genes within the host proteome, (*e.g.*, genes with differential expression upon infection) [23]. Importantly, those with a marked function

or connectivity in the host tend to be determinants of infection (either as requirements for the virus to replicate or for the host to mount the antiviral immune response) [11-13]. Furthermore, viral proteins showing the largest number of contacts are typically non-structural (*i.e.*, do not form the virion) and have a relevant enzymatic activity [9,13]. From an evolutionary point of view, much controversy exists regarding whether viral targets display faster rates of adaptive evolution, which would indeed be indicative of an evolutionary arms race between the virus and the host [24,25]. Quantitatively, one third of the adaptive mutations in humans appear to be in response to viruses [25].

Because the study of plant viruses has not progressed at the same pace as human virology, there are still multiple unanswered questions regarding their mode of action, from both molecular and holistic perspectives [26]. While various studies have unveiled physical contacts between individual virus-host protein pairs that are relevant in terms of host infection [27-32], an exhaustive analysis of a plant virus-host interactome is still lacking. Moreover, large-scale studies that have been carried out with cellular pathogen effectors in plants reveal a mode of action similar to the animal viral proteins, in network terms. These studies show a similar enrichment in central host proteins that interact with pathogen effector proteins in both plants and animals [33,34]. Therefore, systems level analyses of plant virus-host protein-protein interaction networks are required to advance the field of plant molecular virology and also to uncover general similarities and differences between animal and plant viruses that could contribute to our understanding of viral pathogenesis mechanisms.

In this work, we performed a systematic and stringent identification by high throughput yeast two-hybrid (HT-Y2H) screening [35,36] of the direct protein-protein interactions established between a plant virus and one of its natural hosts, with the aim of providing the first comprehensive view of such a complex interplay in the plant kingdom.

More specifically, we used the turnip mosaic virus (TuMV; species *Turnip mosaic potyvirus*, genus *Potyvirus*, family *Potyviridae*) as a model system [26]. Potyviruses are the largest and most abundant family of plant RNA viruses in nature and are responsible for significant crop losses. As the host plant, we used *Arabidopsis thaliana* (L.) Heynh, the quintessential model in plant biology, with its repertoire of genetic tools, which is naturally infected by TuMV [37]. Here, we present our analyses of the identified virus-targeted host proteins in terms of their biological function, proteomic context, expression level, and evolutionary constraints. Finally, we discuss our results in light of data regarding phenotypic outcomes of infection in wild-type and mutant plants and by comparing them with the results reported for clinically relevant human RNA viruses. Our results represent a valuable resource from which to move forward in the study of plant viruses from a systems biology perspective [38].

Results

Construction of a comprehensive TuMV-*A. thaliana* protein-protein interactome. We performed a HT-Y2H screening for each of the 11 proteins encoded in the TuMV genome against all *A. thaliana* proteins. For that, different clones expressing the virus proteins were constructed and a universal library containing normalized amounts of cDNAs from transcripts isolated from different plant tissues at different developmental stages was used. The screening was done through two mate-and-plate steps (first with a permissive medium to capture putative interactions, then with a more stringent medium to remove false positives), together with a final co-transformation into a single yeast cell to obtain high quality interactors (this was done for those nuclear and cytoplasmic virus proteins). As a result, we obtained 10 unique interactors for HC-Pro, 4 for 6K1, 54 for CI, 33 for VPg, 4 for NIa-Pro, and 245 for NIb. However, we did not obtain positive

interactions in the case of P1 (no colonies at all) and CP (no colonies with the appropriate phenotype). Thus, P1 and CP were removed from the study. We then decided to perform a second HT-Y2H screening capable of identifying interactors of membrane-associated proteins, *i.e.*, with the split-ubiquitin (sUbq) system [39]. This was applied for P3, P3N-PIPO, and 6K2 (indeed, P3 has been shown to attach to the membrane of the endoplasmic reticulum) [26]. This second screen resulted in the identification of 9 unique interactors for P3, 12 for P3N-PIPO, and 10 for 6K2, thereby completing the picture. Collectively, our experiments identified 381 virus-host protein-protein interactions, 378 of them being novel (**Dataset S1**). Intriguingly, the TuMV RNA-dependent RNA polymerase (NIb), perhaps the virus protein with the most clear and specific function, showed the largest number of contacts with the host.

For the specific TuMV-*A. thaliana* pathosystem, we manually curated from the literature 10 additional interactors for VPg, 6 for NIb, 1 for P3N-PIPO, and 1 for 6K2. These 18 interactors escaped our HT-Y2H screening, perhaps the consequence of using a selection medium with increased stringency, but they were considered to generate a final network with 399 virus-host interactions (**Fig. 1**; having also gathered known physical interactions between the different TuMV proteins [40]). In subsequent analyses, the interactions occurring between the *A. thaliana* proteins were also considered (see *e.g.*, **Fig. S1)** [41]. Of note, there are more interactions reported in the literature between other potyviruses and different host plants, such as the interaction between HC-Pro and HUA ENHANCER 1 (HEN1, an RNA methyltransferase) or between VPg and RNA HELICASE-LIKE 8 (RH8; see **Dataset S2** for a complete report) [26], but they were not retrieved by using the TuMV-*A. thaliana* pathosystem and hence were not considered in this work.

To validate that the physical interactions reported here from the HT-Y2H screening

indeed occur *in planta*, we performed bimolecular fluorescence complementation (BiFC) assays [42,43]. For that, we selected a subset of 25 virus targets, some of them with known relevance in infection (*e.g.*, the translation initiation factor eIF(iso)4E) [27] and others that were shown to bind to multiple virus proteins (*e.g.*, the SGS domain-containing protein encoded in locus *AT1G30070*). Notably, we observed reconstituted fluorescence in all cases tested using BiFC in *Nicotiana benthamiana* Domin plants (**Fig. 2**). We concluded that our TuMV-*A. thaliana* protein-protein interactome is of high fidelity and constitutes a solid foundation from which to study this plant pathosystem from a systems biology perspective.

Functional analysis of the *A. thaliana* proteins targeted by TuMV. As a starting point to study the effect of the perturbation introduced by the virus infection on host physiology, we assessed a potential enrichment of certain biological processes within the list of TuMV targets. For that, we took advantage of gene ontology (GO) resources [44]. We found that "response to stress" (and "response to virus" in particular), "post-transcriptional regulation of gene expression", "meristem development", or "photosynthesis" are among the 272 biological processes identified as over-represented (Fisher exact tests for 2×2 contingency tables, adjusted $P < 0.05$; **Fig. 3a**). Arguably, this illustrates the conflict between the host, which needs to mount a defense response at the expense of metabolic and developmental processes, and the virus, which aims to counteract such a defense and create a favorable context for replication [45]. More specifically, we considered a relevant set of GO terms to be mapped against each virus protein (**Fig. 3b**). We observed that some TuMV proteins, despite having contacts with few host proteins (*e.g.*, HC-Pro, P3N-PIPO, or 6K2, all three with only a few interactions), are able to perturb very different processes, from metabolism to regulation

to defense. Certainly, this can be a consequence of targeting *A. thaliana* proteins that participate in multiple functions, such as PLASMA-MEMBRANE ASSOCIATED CATION-BINDING PROTEIN 1 (PCaP1), targeted by P3N-PIPO [31,32], which reflects the highly intricate (mostly hormone-mediated) nature of the signaling and regulatory pathways in plants [46]. Of course, NIb, with the largest number of contacts, is the virus protein that impacts more host functions. The heatmap also revealed functions targeted by all (or almost all) virus proteins, such as "response to stress", as well as functions with a more specific relationship, such as "translation regulation" (only targeted by VPg and NIb).

In addition, a principal component analysis with this data matrix organized the nine virus proteins in a two-dimensional space according to their predicted impact on host physiology (**Fig. 3c**; with 74.2% of explained variance). Interestingly, we observed that virus proteins that physically interact also locate closer in this space, which may be indicative of a strategy evolved by the virus to coordinate the action of its proteins during infection [47]. This is the case, for instance, for HC-Pro and NIa-Pro, VPg and NIb, or CI and P3N-PIPO. This analysis also highlights P3 as the less related protein in terms of host targets as compared to the rest of the virus proteins. Finally, we compared the functions of the different virus proteins [26] with the functions of their host targets (**Fig. 3d**). In essence, the virus requires the exploitation and disruption of multiple biological processes in the host to complete its cycle, involving for that a set of multifunctional proteins.

Topological contextualization of the *A. thaliana* proteins targeted by TuMV. To evaluate the systems-level relevance of each virus target identified by our HT-Y2H screening beyond its function, we performed a computational network analysis. Firstly,

we adopted a virus-centric approach to analyze how the virus proteome is rewired as a consequence of establishing interactions with the host (**Fig. 4a**). We found a significant enrichment in host factors targeted at least by two different virus proteins (*z* test, $P < 0.0001$; **Fig. 4b**), which suggests that such host factors can work as emergent communication channels between the virus proteins [10]. While the length of the shortest paths that connect any two virus proteins is invariant whether or not the host proteome is considered (**Fig. 4c**), the number of paths increases appreciably when it is considered (**Fig. 4d**). Of special note is the emergent relationship that is established between the virus replicase NIb and the virus helicase CI, which may reflect the necessity for coordination in the core replication complex [26,47].

Secondly, we adopted a host-centric approach to contextualize the virus targets into a model of the *A. thaliana* protein-protein interactome (**Fig. 5a**). We focused our analysis on two relevant topological properties: connectivity degree (*i.e.*, the number of direct interactions that a given protein establishes with others) and average shortest path length (*i.e.*, the average length of all shortest paths that connect a given protein with the rest). We found that the connectivity degree distribution for all host proteins has a scale coefficient greater than for the virus targets (as illustrated by the different slopes of the power laws that fit the data in **Fig. 5b**), which is indicative of an enhanced probability of the virus to reach a hub protein. In particular, the scale coefficient dropped from 1.247 to 0.860, a reduction of 31.0% (comparison of slopes in an ANCOVA test, $F_{1,21} = 9.70$, $P = 0.0053$). A similar trend has been quantified in the case of several human viruses [48]. In addition, by comparing the distributions of connectivity degrees and the average shortest path lengths for the virus targets and for different sets of random genes, we found that the distributions associated with the virus targets are significantly shifted towards higher connectivity degrees (Mann-Whitney *U* test, $P = 0.0045$; **Fig. 5c**) and lower path

lengths (Mann-Whitney *U* test, $P = 0.0003$; **Fig. 5d**). For the virus targets, the mean of the degrees was 7.12 and the mean of the path lengths 4.18.

For illustrative purposes, we sketched the virus-host interactome by selecting three virus targets with high degree in the host interactome: the small ubiquitin-like modifier (SUMO) ligase SUMO CONJUGATING ENZYME 1 (SCE1), with 162 interactors (the target with the highest connectivity degree), the PHD finger OBERON 1 (OBE1), with 61 interactors, and the transcription factor TGACG SEQUENCE-SPECIFIC BINDING FACTOR 1 (TGA1), with 38 interactors (**Fig. 5e**). These proteins are targets of NIb and VPg. For completeness (as it interacts with both NIb and TGA1), the salicylic acid (SA)-dependent transcription factor NONEXPRESSER OF PATHOGENESIS-RELATED GENES 1 (NPR1), which controls the expression of genes that exert a response against pathogens, is also shown (note that TGA1 shares interactors with NPR1 and OBE1). These interactions are intended to promote virus replication and within-host movement and block host defenses [28,30,49]. Together, TuMV appears to target *A. thaliana* proteins with increased connectivity degree and ability to spread information, similar to the properties previously described for human proteins targeted by viruses [9,22].

Expression features of the *A. thaliana* proteins targeted by TuMV. Next, we sought to study the expression levels of the virus targets, as well as the relationship between expression and connectivity within the *A. thaliana* protein-protein interactome. For that, we collected the expression values of all *A. thaliana* genes in control conditions and also upon infection with TuMV from previous transcriptomic experiments [50]. By dividing the expression levels into three categories (low, medium and high), we found virus targets in all of them, but with an apparent enrichment in the category of high expression (**Fig. 6a**). Note, for example, the ratio of highly *vs.* lowly expressed proteins in the case of

P3N-PIPO (7 *vs*. 3, out of 13 host interactors), CI (28 *vs*. 6, out of 54 host interactors), or NIb (101 *vs*. 42, out of 251 host interactors). To further explore this issue, we compared the distributions of expression for the virus targets and for different sets of randomly selected genes, revealing that the distribution associated with the virus targets is significantly shifted towards higher expression levels (Mann-Whitney *U* test, $P < 0.0001$; **Fig. 6b**). We repeated the comparison from expression data upon TuMV infection, finding a similar trend (Mann-Whitney *U* test, $P < 0.0001$; **Fig. 6c**). We also noticed that, on average, the expression of the virus targets marginally increases upon infection, as opposed to what occurs in the null case of randomly sampled genes.

In addition, we found a relevant trade-off between the absolute level of differential expression upon infection and the connectivity degree (**Fig. 6d**). That is, the higher the connectivity degree of a given host protein, the lower the absolute differential expression upon infection, indicating that hub proteins in the *A. thaliana* interactome display certain robustness in expression levels in the presence of perturbations (as also observed when certain animal diseases are assessed in terms of networks) [51]. Virus targets widely distribute over this space, even approaching the trade-off front, as it is the case of CALNEXIN 1 (CNX1), targeted by VPg, SCE1, SUMO 3 (SUM3), HEAT SHOCK PROTEIN 70-3 (HSP70-3), the protein encoded in locus *AT1G21440* (a phosphoenolpyruvate carboxylase), and 40S RIBOSOMAL PROTEIN SA (RP40), targeted by NIb, and PLASMA MEMBRANE INTRINSIC PROTEIN 1;3 (PIP1;3), targeted by the membrane-associated P3N-PIPO. Collectively, these results indicate that virus targets display higher expression levels than expected by chance, and also that some virus targets can be at the same time significantly up- or down-regulated upon infection provided they are not highly connected.

Selective constraints upon the *A. thaliana* proteins targeted by TuMV. Much has been discussed about the evolutionary mechanisms that operate on the antiviral defense proteins [52]. Here, we tested the hypothesis of whether the TuMV-interacting proteins come preferentially from a subset of essential genes under strong purifying selection. Firstly, the within-species genome-wide ratio of nonsynonymous to synonymous polymorphisms (p_N/p_S) was computed from the 1001 *A. thaliana* Genomes Project [53], while the genomes of species from a sister clade, *Capsella rubella* and *Boechera retrofracta*, served as references for the analyses (**Fig. 7a**, evolutionary results for each gene in **Dataset S3**). We found that the average p_N/p_S ratio is significantly smaller in the set of TuMV interactors than for randomly selected proteins (Mann-Whitney's *U* test, $P = 0.0016$; **Fig. 7b**); thus, indicating that these interactors are under strong purifying selection and particularly well conserved. Moreover, we estimated the proportion of adaptive nonsynonymous mutations by means of the direction of selection (*DoS*) unbiased statistic in TuMV-interacting proteins and non-interacting ones in the *A. thaliana* lineage [54]. Overall, no significant difference exists between both groups (**Fig. 7c**). However, it is interesting that the NIb interactor SCE1, a highly connected protein in the host, ranks second among the TuMV-interacting proteins with the largest $DoS > 0$ values (**Fig. 7d**). Besides SCE1, none of the other top 5 adapted TuMV-interacting proteins listed in **Fig. 7d** have been previously related to responses to infection. Among the most conserved TuMV-interacting proteins ($DoS < 0$), none have been previously annotated as related to infection but, instead, with different aspects of plant metabolism (**Fig. 7d**). For example, the acylphosphatase encoded in locus *AT5G03370*, the phosphoglycerate mutase encoded in locus *AT5G64460*, and 10-FORMYLTETRAHYDROFOLATE SYNTHETASE (THFS). Of note, two of the most

highly connected proteins, the NIb interactor TGA1 and the VPg interactor OBE1 (see **Fig. 5e**), show $DoS < 0$ values.

Secondly, we evaluated the rates of evolution (d_N and d_S) for each protein in the branch leading to the *Arabidopsis* genus to compute $\omega = d_N/d_S$ (**Fig. 7e**). We found significantly smaller ω values for the TuMV-interacting proteins than for randomly selected proteins (Mann-Whitney's U test, $P = 0.0043$; **Fig. 7f**). This allowed us to conclude, in agreement with the results described above for p_N/p_S, that TuMV-interacting proteins have been subject to purifying selection and are not significantly enriched in fast-evolving genes. In agreement with these observations, several studies have shown that fast-evolving antiviral proteins may not be representative of the many other proteins that physically interact with viruses throughout their infection cycle and that, under normal conditions, play key functions in basic cellular processes that are highjacked by the pathogens. These virus-interacting proteins usually evolve slowly in both animals [12,25,55] and plants [33,56].

Discussion

A valuable, though limited resource in plant systems virology. A main goal of describing protein interactomes is understanding the mechanisms by which the cell maintains its homeostasis under normal conditions and readjusts it in response to stress. Experimental virus-host interactomes are useful because they can be exploited to generate hypothesis regarding the pathogenicity and replication mode of viruses [38]. Indeed, proteins rarely act in isolation within the cell; different protein-protein interactions define major functional pathways crucial for a variety of cellular processes. In this regard, we expect that our data and analyses can guide further studies with plant viruses. In particular, together with known interactions among the different virus proteins and host

proteins (also obtained by Y2H screenings) [40,41], we were able to represent, for the first time, a comprehensive network that characterizes the interaction between a plant virus (TuMV) and one of its natural hosts (*A. thaliana*). In addition, this interaction network may serve as a useful map to guide future biotechnological developments aimed at obtaining plants with increased tolerance to viral infections.

In this direction, it is worth noting that the *A. thaliana* protein with the largest number of interactions targeted by TuMV is SCE1. An emerging topic in systems virology is whether viruses exploit the host SUMOylation pathway [57,58]. SUMO induces proteins to change their stability, activity, and location. Multiple human DNA viruses (*e.g.*, herpesviruses) are known to interfere with the host SUMOylation pathway in diverse ways [57,58]. As previously reported, SUMOylation of NIb at position K172 is a requirement for TuMV to successfully infect the plant, since *sce1* plants are resistant to infection and NIb-K172R viruses have lower infectivity [30].

Furthermore, the 11 interactors that we found for 6K2 (some shared with P3) are significantly enriched in GO terms related to vesicle transport from endoplasmic reticulum to Golgi membranes. This is interesting because all known positive-strand RNA viruses replicate their genomes in membrane-associated compartments dubbed as viral factories [47,59], which facilitate the coordination between the different factors required for RNA replication, while being protected from the RNA silencing machinery [60]. In the case of potyviruses, the small protein 6K2 is responsible for anchoring the replication complex to the endoplasmic reticulum to seed the formation of viral factories [26], so the evaluation of the effect of these host proteins on viral infection would be interesting.

The virus-host protein-protein interaction network reported here is limited in various ways. Firstly, TuMV proteins P1 and CP are not included in the network. We were

unable to detect interactors of P1 and CP in yeast. P1 is an unstable protein only expressed at the beginning of infection. In a previous report, protein complexes that associate with P1 *in planta* were retrieved by following an alternative strategy of affinity purification coupled to mass spectrometry [61]. By contrast, CP is the coat protein of the virus and may display a self-binding ability in yeast that precludes the interaction with other proteins [62]. A new screening with an increased expression level of CP might help. Secondly, some binary interactions already reported between TuMV and *A. thaliana* escaped our HT-Y2H screening [26]. Other interactions that were previously captured by affinity purification and that we did not obtain, such as between HC-Pro and the master RNA silencing factor ARGONAUTE 1 (AGO1) [63], might be rationalized as interactions through third parties. Moreover, the *A. thaliana* interactome is still incomplete because it only covers about 8,000 proteins (about one third of the total proteome) [41]. Consequently, several host proteins identified here to interact with the virus are not included in the computational network analysis. Thirdly, virus proteins usually perform multiple functions at different stages of the infection cycle [26]; thus, the virus-host interaction is presumed to change with time. The interaction can also change if the virus accumulates nonsynonymous mutations that affect the binding interface of the virus protein. In this regard, our results only provide a snapshot of a highly dynamic process [64]. Hence, further experimental and computational work is required to complete the picture for TuMV and also for other potyviruses to assess differences and commonalities in their mode of action.

Linking the TuMV-*A. thaliana* interaction map with disease etiology. Plants and their viruses are engaged in an arms race [52]. On one hand, pathogens deploy the so-called virulence effectors into host cells, wherein they establish highly dynamic physical

interactions with host proteins to redirect the cellular resources into their own benefit [64,65]. On the other hand, plants sense the presence of pathogens at two different levels. Firstly, via recognition of conserved pathogen-associated molecular patterns (PAMPs) by pathogen-recognition receptors exposed on the outer side of the cell membrane. This first level of recognition results in PAMP-triggered immunity [65,66]. Secondly, plants deploy a set of intracellular immune receptors of the nucleotide-binding site leucine-rich repeat protein family, analogous to the animal innate immune NOD-like receptors. This activation results in effector-triggered immunity ETI, which amplifies PAMP-triggered immunity responses, resulting in a burst of reactive oxygen species, changes in ion fluxes, increases in cytosolic Ca^{2+} levels, activation of mitogen- and Ca^{2+}-dependent protein kinases, elevation of phytohormones (most remarkably SA), and transcriptional reprogramming, often leading to host cell death (hypersensitive response) and long-lasting systemic acquired resistance [65].

In this context, we found that NPR1 interacts with the virus replicase NIb. NPR1 acts as a key factor in the plant defense signaling network, mediating the cross-talk between the SA and jasmonic acid/ethylene responses, thus playing a significant role in the establishment of systemic resistance (induced and acquired) [67]. In the nucleus, NPR1 interacts with TGA transcription factors, which also control the expression of pathogenesis-related genes. Furthermore, NPR1 plays a role in histone modification, enforcing the priming of SA-induced defense genes and transgenerational immune memory [68]. By sequestering NPR1 (through the virus replicase NIb), the virus could block the activation of pathogenesis-related genes, thus impeding systemic acquired resistance and ensuring that future plant generations will not be primed against it. Interestingly, NIb also interacts with a TGA transcription factor (TGA1, a hub in the host proteome), which is presumed to enhance the negative effect on SA-mediated resistance

of the aforementioned interaction, especially because this targeting will produce a feedback response in the system by down-regulating the production of SA [69].

Another essential component of the plant immune response affected by viral infection is the production of reactive oxygen species. In particular, H_2O_2 is produced in the apoplast by plasma membrane-associated NADPH oxidases. Then, H_2O_2 is translocated into the cytoplasm by several aquaporins [70]. Once there, it cross-talks with PAMP-triggered immunity and systemic acquired resistance pathways via redox conformational changes of NPR1 and activation of a mitogen-activated protein kinase cascade that upregulates a set of immune responses, including subsequent production of H_2O_2 and callose deposition [70]. Interestingly, we found the membrane-associated P3N-PIPO to interact with the aquaporin PIP1;3, likely interrupting the influx of H_2O_2.

In addition, we took advantage of the virus-host interaction map to gain mechanistic insight about one of the most remarkable symptoms of TuMV infection, *i.e.*, the sterilization of *A. thaliana* plants. This results from the abortion of the apical meristem growth [37]. The interaction between VPg and OBE1 (an evolutionarily conserved hub in the host proteome) is suggestive, as this protein is involved in the establishment and maintenance of the shoot and root apical meristems through the regulation of the expression of the stem cell factor WUSCHEL (WUS). Notice that VPg also interacts with the PDH fingers OBERON 2 (OBE2), an element that functions like OBE1 [71], and OBERON 3 (OBE3), one of the transcriptional activators of WUS and the CLAVATA (CLV) pathway [72]. By looking at the interactors of OBE1 in the network, we found OBE3, STOMATAL CLOSURE-RELATED ACTIN BINDING 1 (SCAB1), which stabilizes actin filaments and controls stomatal movement [73], and various WRKY transcription factors, which activate the expression of defense genes against pathogens [74]. But as recently shown, WUS is also responsible for triggering innate antiviral

immunity in the meristems [75]. Consequently, we may argue that OBE1 and OBE3 are sequestered by TuMV in order to facilitate its within-host (mostly systemic) movement [28] by subverting the regulation of both the stomatal mechanics and the antiviral response in stem cells, with the side effect of sterilization.

Comparing the mode of action of TuMV and human viruses. As discussed above, TuMV interferes with PAMP- and effector-triggered immunity. Mammalian cells also trigger their innate immune systems by PAMP recognition through two groups of receptors, the Toll-like and RIG-I-like receptors [76]. Similar to what occurs in plants, these receptors initiate a signaling cascade that converge on transcription factors that induce type I interferon (IFN-α/β) expression. Not surprisingly, mammalian viruses also block the IFN-α/β pathway; the V protein of measles virus providing a well-studied example [77].

From a holistic perspective, some principles have been put forward to characterize virus-host interactions in humans. Broadly, viruses tend to interact with proteins that are (*i*) hubs and bottlenecks in the host interactome [9,14], (*ii*) evolutionarily conserved and under positive selection [24,25], (*iii*) involved in key biological processes instrumental for their infection and replication [14], and (*iv*) close to other proteins involved in disease symptoms [23]. Our results demonstrate that these principles generally hold true in the case of a plant virus, while adding some nuances.

Firstly, we show that the set of TuMV targets is significantly enriched in hub proteins, which suggests that the perturbation introduced by the virus is deliberate and not merely random. This pattern has been described for several human viruses, including both RNA viruses (*e.g.*, *Hepatitis C virus* or *Dengue virus*) and DNA viruses (*e.g.*, *Epstein-Barr virus*) [48]. This way, the virus has the potential of dismantling the entire

system through a selective attack [21], which aligns with the so-called centrality-lethality rule [19]. Moreover, by leaning on hub proteins of the host, a small set of viral proteins can lead to the simultaneous rewiring of many cellular processes. However, hubs only represent a small fraction of all nodes in a scale-free network, so this proportion is also transmitted to the virus targets. Our results also reveal that TuMV selectively targets proteins that, while not being highly connected, bridge different parts of the *A. thaliana* interactome, thereby with increased ability to spread information (*i.e.*, proteins that have high neighborhood connectivity) [22,34]. In addition, the virus protein by itself may work as a bridge to coordinate the perturbation among different host factors, as may be the case of HC-Pro, P3N-PIPO, or NIb, by targeting proteins with very different functions in the cell.

Secondly, by analyzing both the p_N/p_S and ω ratios, we found that the TuMV targets are significantly enriched in proteins whose genes have evolved slowly (*i.e.*, conserved genes). This is in tune with previous analyses of human genes targeted by viruses [12,25,55] and suggests that viruses look for constrained elements with the aim of ensuring broad host ranges. We then evaluated if even at slow rates these virus targets have been subject to positive selection, finding non-significant results. This contrasts with the concept of an evolutionary arms race, as well as with previous analyses indicating that virus targets in humans have been adapted in response to infection [25]. Perhaps, a new evolutionary study restricted to subsequences encoding viral protein binding surfaces rather than whole gene sequences would offer more enlightening results [24]. We also found that the TuMV targets display higher expression levels, which indeed favors the interaction with multiple proteins (*e.g.*, viral proteins) in multiple tissues. Interestingly, it is well accepted that highly expressed genes tend to evolve at slower rates [78], arguably because the exploration of new variants is more costly [79]. Thus,

conservation and expression are two sides of the same coin that the virus conveniently exploits.

Thirdly, our functional analysis showed that the host proteins targeted by the virus are significantly enriched in GO terms associated with defense and regulation, but also with general metabolic processes. Although it is difficult to establish a frontier, the study of the mode of action of multiple human viruses revealed that proteostasis, signaling (*e.g.*, JAK-STAT pathway, which acts downstream of IFN-α/β), transport, and RNA metabolism are host functions preferentially targeted by RNA viruses, while transcription, proteostasis, macromolecular assembly, DNA and RNA metabolisms, and cell cycle (*e.g.*, cancer pathways) are the focus of DNA viruses [14]. TuMV, a plant RNA virus, seems to be closer to human RNA viruses, especially because it impairs hormone signaling pathways, host protein expression at the level of translation, and the RNA silencing machinery.

Fourthly, it has been argued that proteins involved in disease susceptibility and symptomatology in humans (*e.g.*, cancer-related proteins altered by viral infections, like *Epstein-Barr virus*-associated lymphoma) should reside in the network neighborhood of the corresponding viral targets [16,23]. In this regard, the link established between VPg, OBE1 (virus target), WUS (host protein regulated by the virus target), and sterilization (disease) in our phytopathosystem is a good example. All in all, we anticipate gaining a deeper understanding about the mode of action of TuMV and other plant viruses if the interrelation interface is enlarged with more interactions (*e.g.*, by curating and integrating different sources of information) [38] and also if the *A. thaliana* interactome is upgraded, which now contains about 22,000 interactions. In addition, further work should be aimed at combining protein-protein interactions, transcriptional regulation, and metabolic

pathways in order to generate a mechanistic picture of viral infection in plants as comprehensive as possible [7].

Methods

Plasmid construction. The plasmid p35STunos contains an infectious cDNA clone (GeneBank accession AF530055.2) corresponding to the TuMV isolate YC5 obtained from infected calla lily (*Zantedeschia* sp.) [80]. The 11 TuMV coding regions (corresponding to the virus proteins P1, HC-Pro, P3, P3N-PIPO, 6K1, CI, 6K2, VPg, NIa-Pro, NIb, and CP) were amplified by polymerase chain reaction (PCR) from p35STunos with the Phusion High-Fidelity DNA polymerase (Thermo) by using the corresponding pairs of primers, including Gateway adapters, listed in **Dataset S4**.

For the Y2H system based on the *GAL4* promoter (interactions occurring in the cytoplasm) [35], the PCR products were cloned by recombination with the In-Fusion enzyme (Clontech) into the yeast bait vector pGBKT7 (Clontech), which was digested with *Eco*RI and *Bam*HI. This generated a translational fusion of the virus protein (bait protein) with the GAL4 DNA-binding domain. The construction for P3N-PIPO was done in two steps. First, part of the P3 cistron was amplified by PCR and cloned into the plasmid pGBKT7. Second, one adenine was inserted in the putative frameshift site (GGAAAAAA) by site-directed mutagenesis to express the virus protein without the need of frameshifting [81].

For the screening based on the sUbq (interactions occurring in the membrane) [39], the PCR products were cloned by recombination *in vivo* to obtain the CubPLV translational fusion (in the case of P3, P3N-PIPO, and 6K2). For that, the bait vector pMetYC-gate was digested with *Pst*I and *Hin*dIII and then was co-transformed together

with the PCR product into the yeast strain THY.AP4. Transformants were selected on SD/-Leu medium after incubation at 30 °C for 5 days.

For the BiFC constructs, PCR products from all TuMV genes were recombined into the plasmid pDONR207 by using the BP Clonase II Enzyme mix (Invitrogen). For cloning the different *A. thaliana* genes, total RNA was extracted from plant tissues (*A. thaliana* Col-0) by using the Trizol reagent (Invitrogen) following the manufacturer's recommendations and further purified by lithium chloride precipitation. The corresponding cDNAs were synthesized by using the RevertAid H Minus First Strand cDNA synthesis kit (Fermentas) with a polyT+N-primer. Full-length ORFs were amplified by PCR from those cDNAs with the Phusion High-Fidelity DNA polymerase (Thermo) by using suitable primers, including Gateway adapters, also listed in **Dataset S4**. The constructs were then recombined into the plasmid pDONR207 by using the LR Clonase II Enzyme mix (Invitrogen). All constructed plasmids were amplified in *Escherichia coli* strain DH5α, purified, and verified by sequencing.

HT-Y2H screening. To identify the host proteins that interact with the eleven TuMV proteins, an *A. thaliana* Col-0 cDNA library (Clontech) was screened by using the Matchmaker Gold Yeast Two-Hybrid System (Clontech). For this, the Y187 haploid yeast strain (with library proteins) and the Y2HGold haploid reporter strain (with the plasmid pGBKT7 that expresses each of the 11 TuMV proteins) were mated and plated on a double dropout medium (SD/-Leu/-Trp) containing 40 µg/mL X-α-Gal and 200 ng/mL aureobasidin A, and then incubated at 30 °C for 5 days. Co-transformants that were phenotypically positive for α-galactosidase activity were subjected to a further, more stringent phenotypic assay on a quadruple dropout medium (SD/-Leu/-Trp/-Ade/-His) containing X-α-Gal and aureobasidin A. Plasmids pGBKT7-T-antigen, pGADT7-

laminin C, and pGADT7-murine p53 (Clontech) were used as negative controls. Colony PCRs were performed from the yeast colonies that displayed a positive interaction to eliminate duplicate clones, and prey plasmids were rescued with an isolation kit to be subsequently transformed into *E. coli* DH5α for amplification and sequencing, as described by the manufacturer. DNA and protein sequence analyses were performed with the WU-BLAST algorithm as formerly implemented in the TAIR website [82]. For each novel interaction, the prey and bait plasmids were co-transformed into the Y2HGold strain to verify genuine positive interactions from false positives.

An *A. thaliana* Col-0 cDNA library based on the sUbq system (Dualsystems) was also screened to identify membrane-associated interactions. For this, the yeast THY.AP4 (already with the plasmid pMetYC-gate that expresses each of the four membrane-associated TuMV proteins) was transformed with the library proteins fused to the NubG domain, plated on a triple dropout medium (SD/-Leu/-Trp/-Ade), and incubated at 30 °C for 10 days. Methionine at 10 µg/mL was added to the medium to optimize the CubPLV fusion expression. Co-transformants were subjected to further, more stringent phenotypic assay on a quadruple dropout medium (SD/-Leu/-Trp/-Ade/-His) containing 80 µg/mL X-Gal and 10 µg/mL methionine. Colony PCRs were performed from the yeast colonies that displayed a positive phenotype for β-galactosidase activity to eliminate duplicate clones, and prey plasmids pDSL-Nx were rescued in *E. coli* DH5α for sequencing. The pMetYC-gate Ost3 plasmid (Dualsystems) served as a negative control.

BiFC assay with confocal microscopy. The Gateway destination vectors pYFN43 and pYFC43 (kindly provided by Prof. Pablo Vera, IBMCP) were used to obtain the coding sequences of the two moieties of a yellow fluorescent protein (YFP) [83]. The primers used to amplify by PCR the N-terminal sequence of YFP (corresponding to residues 1 to

154) and the C-terminal sequence of YFP (corresponding to residues 155 to 240) are also listed in **Dataset S4**. The Phusion High-Fidelity DNA polymerase (Thermo) was used. Both PCR products were cloned into the plasmid pEarlyGate101 with the restriction enzymes *Avr*II and *Spe*I, replacing the native YFP cDNA. The Gateway destination vectors p101-YFN and p101-YFC were created in this work to generate the translational fusions with the virus and host proteins.

Cultures of *Agrobacterium tumefaciens* strain C58 harboring appropriate binary plasmids were grown overnight and then centrifuged, and OD_{600} was adjusted to 0.5 with 10 mM MES pH 5.6, 10 mM $MgCl_2$, and 150 mM acetosyringone. Individual bacterial cultures were mixed and used to agroinfiltrate the young leaves of 2-3 weeks old *N. benthamiana* plants (for the simultaneous transient expression of the two YFP moieties fused to the proteins of interest). After 48 h, the yellow fluorescence of agroinfiltrated leaves was analyzed by using an inverted confocal microscopy (Zeiss LSM780) with a CApo 40X/1.2 objective (Zeiss). To detect yellow fluorescence (from reconstituted YFP), excitation was done with a 488 nm argon laser and the resulting emission signal was collected in the 520-550 nm window; while to detect red fluorescence (from chloroplasts), excitation was done at 488 nm and emission collected at 680-750 nm. Image processing was performed with ImageJ v1.8 [84].

Computational functional analysis. With the whole list of TuMV-targeted host proteins (all virus proteins), a functional analysis was performed by using the agriGO webserver [85] to identify which gene ontology categories (related to biological processes) are over-represented within. The statistical significance, with respect to the complete plant genome (TAIR release 10), was evaluated by a Fisher exact test (2×2 contingency table) with a correction for multiple testing using the Benjamini-Hochberg false discovery rate

procedure (to adjust the *P* value) [86], only considering GO terms with five or more mapping entries. With those identified biological processes, a functional network in a semantic space was constructed by using the REVIGO tool [87]. This analysis also served to identify the particular biological processes in which the targets of a given virus protein participate. A map between the different virus proteins and 112 relevant biological processes (corresponding to metabolism, development, organization, signaling, regulation, and defense) was generated with the fraction of virus targets implicated in each process. Note that a given host protein can participate in multiple processes.

Autogenous host and virus interactomes. The *A. thaliana* protein-protein interactome was used to contextualize the host proteins identified as targets for the different virus proteins. This interactome was constructed by accounting for all known physical interactions with experimental evidence (mainly by Y2H screens) [41,88]. This network covers about 8,000 plant proteins and has about 22,000 non-redundant interactions between them. A pre-analysis of the global topological properties was performed with Cytoscape [89]. This quantitative evaluation included the computation of the connectivity degree and average shortest path length for each node in the network. The whole *A. thaliana* interactome as well as the TuMV-*A. thaliana* interactome here identified were also represented with Cytoscape. In addition, a general protein-protein interactome of potyviruses was taken from previous work that collected experimental data from multiple Y2H and BiFC assays [40]. The recently described interaction between P3 and P3N-PIPO was also included [90].

Computational network analysis. The degree distribution (for the whole *A. thaliana* interactome) was represented for the TuMV-targeted proteins and for all plant proteins, fitted in both cases to a probability power law: $P(k) \sim k^{-\gamma}$, where k is the connectivity degree and γ the scale coefficient [20], done with MATLAB (MathWorks). In addition, for the list of TuMV-interacting proteins, the observed distributions of connectivity degree, average shortest path length (topological properties), expression levels (in healthy and infected states), rates of gene evolution, and genome-wide ratios of nonsynonymous to synonymous polymorphisms (see below) were represented as violin plots. One thousand random lists of proteins were also generated to compute null distributions of these variables. The statistical significance was assessed by means of Mann-Whitney *U* tests with MATLAB.

Gene expression data. Transcriptomic data of healthy and TuMV-infected *A. thaliana* plants (corresponding to Affymetrix data from microarray experiments) were retrieved from previous work [50], which were subsequently normalized by state-of-the-art procedures in a meta-analysis that studied multiple plant viruses [91] to obtain absolute and relative (infected *vs*. healthy) expression values. In addition, a scatter plot between differential expression and connectivity degree was generated for all host proteins, highlighting the virus targets.

Comparative genomics and quantifying adaptation. To determine the evolutionary rate leading to the *Arabidopsis* genus, we selected three *Arabidopsis* species (*Arabidopsis halleri*, *Arabidopsis lyrata* and *A. thaliana*), three species from a sister clade (*B. retrofracta*, *C. rubella* and *Crucihimalaya himalaica*) and three representations of an outgroup (*Cardamine hirsuta*, *Erysimum cheiranthoides* and *Barbarea vulgaris*; **Table**

S1). Only long splicing variants were kept to minimize the complexity of the dataset. Furthermore, proteins shorter than 50 amino acids were filtered out. Orthologous gene groups were detected with OrthoFinder [92], which was run with default setting on the nine genomes. The orthologous groups were pruned for duplicates by only keeping the best BLAST hit against the *A. thaliana* gene copy. Further increment of the dataset was done by allowing a certain amount of absence in some of the genomes, with the absence being allowed if the closest relative was present (**Fig. S2**). Amino acid alignments were generated with MAFFT [93] using the accurate option (L-INS-i). Regions contains a large number of gaps were removed from the back-translated codon alignments using trimAL [94] with 0.85 gap-score cut-off. The alignment dataset contains a variable number of taxa, meaning that CODEML from the PAML package [95] could not be run directly. GWideCodeML [96] overcomes this problem by pruning the given species tree to fit each of the alignments. The $\omega = d_N/d_S$ heterogeneous branch model of CODEML was run with default setting on GWideCodeML providing the unrooted species tree generated by OrthoFinder.

The genome set was reduced for the population study to containing *A. thaliana* and two close relatives (*B. retrofracta* and *C. rubella*). The information from the 1001 *A. thaliana* Genomes Project [53] was used for generating artificial *A. thaliana* genomes contains all SNP variants observed. Orthologous groups obtained in the previous analysis was used again here. Amino acid alignments were done as previously specified with MAFFT and trimAL. Determining the proportion of adaptive amino acid substitutions within *the A. thaliana* population was done using the direction of selection unbiased statistic (*DoS*) [54]. *DoS* calculates the difference between the proportion of substitutions and polymorphisms that are nonsynonymous as $DoS = D_N/(D_N + D_S) - p_N/(p_N + p_S)$, where D_S and D_N are the numbers of fixed and p_S and p_N are the numbers of polymorphic

mutations per gene within the *A. thaliana* population (subindexes *S* and *N* refer, respectively, to synonymous and nonsynonymous mutations). *DoS* takes values in the interval [−1, 1] and under the null hypothesis of neutral evolution $DoS = 0$; adaptive evolution would result in $DoS > 0$, while $DoS < 0$ values are expected for purifying selection. The R library PopGenome [97] was used for the calculation of D_N, D_S, p_N, p_S. The results of all the above analyses are presented in **Dataset S3**.

Data availability

All relevant data are included in the files of the publication (main figures and supplementary material). Any other data that support the findings of this study are available from the corresponding authors upon request.

Acknowledgements

We thank Francisca de la Iglesia and Paula Agudo for excellent technical assistance and the rest of the EvolSysVir lab, especially José L. Carrasco, for fruitful discussions. This work was supported by grants PID2019-103998GB-I00 and PGC2018-101410-B-I00 (Agencia Estatal de Investigación - FEDER) to S.F.E. and G.R., respectively, and PROMETEO2019/012 (Generalitat Valenciana) to S.F.E.

Authors' contributions

S.F.E. conceived the research with input of G.R.; F.M. performed the experimental work (HT-Y2H screens and BiFC assays), supervised by L.Y. (partly) and S.F.E.; G.R. performed the computational work (functional, network, and evolutionary analyses) and prepared the figures, with input of C.T. (evolutionary analysis) and S.G-S. (literature curation); J.H. worked in the intravirus interaction network; L.Y. contributed with

materials (sUbq system); G.R. and S.F.E. analyzed the results; G.R. and S.F.E. wrote the paper.

Competing interests

The authors declare no competing interests.

References

1. Fauci, A.S., Morens, D.M. The perpetual challenge of infectious diseases. *N. Engl. J. Med.* 366, 454-461 (2012).
2. Navas-Castillo, J., López-Moya, J.J., Aranda, M.A. Whitefly-transmitted RNA viruses that affect intensive vegetable production. *Ann. Appl. Biol.* 165, 155-171 (2014).
3. Dyurgerov, M.B., Meier, M.F. Twentieth century climate change: evidence from small glaciers. *Proc. Natl. Acad. Sci. USA* 97, 1406-1411 (2000).
4. Culver, J.N., Padmanabhan, M.S. Virus-induced disease: altering host physiology one interaction at a time. *Annu. Rev. Phytopathol.* 45, 221-243 (2007).
5. Hasin, Y., Seldin, M., Lusis, A. Multi-omics approaches to disease. *Genome Biol.* 18, 83 (2017).
6. Law, G.L., Korth, M.J., Benecke, A.G., Katze, M.G. Systems virology: host-directed approaches to viral pathogenesis and drug targeting. *Nat. Rev. Microbiol.* 11, 455-466 (2013).
7. Vidal, M., Cusick, M.E., Barabási, A.L. Interactome networks and human disease. *Cell* 144, 986-998 (2011).
8. Mahmoudabadi, G., Phillips, R. A comprehensive and quantitative exploration of thousands of viral genomes. *eLife* 7, e31955 (2018).
9. De Chassey, B., Navratil, V., Tafforeau, L., Hiet, M.S., Aublin-Gex, A., Agaugué, S., *et al*. Hepatitis C virus infection protein network. *Mol. Syst. Biol.* 4, 230 (2008).
10. Uetz, P., Dong, Y.A., Zeretzke, C., Atzler, C., Baiker, A., Berger, B., *et al*. Herpesviral protein networks and their interaction with the human proteome. *Science* 311, 239-242 (2006).
11. Calderwood, M.A., Venkatesan, K., Xing, L., Chase, M.R., Vazquez, A., Holthaus,

A.M., *et al*. Epstein-Barr virus and virus human protein interaction maps. *Proc. Natl. Acad. Sci. USA* 104, 7606-7611 (2007).

12. Jäger, S., Cimermancic, P., Gulbahce, N., Johnson, J.R., McGovern, K.E., Clarke, S.C., *et al*. Global landscape of HIV-human protein complexes. *Nature* 481, 365-370 (2012).
13. Khadka, S., Vangeloff, A.D., Zhang, C., Siddavatam, P., Heaton, N.S., Wang, L., *et al*. A physical interaction network of dengue virus and human proteins. *Mol. Cell. Proteomics* 10, M111.012187 (2011).
14. Pichlmair, A., Kandasamy, K., Alvisi, G., Mulhern, O., Sacco, R., Habjan, M., *et al*. Viral immune modulators perturb the human molecular network by common and unique strategies. *Nature* 487, 486-490 (2012).
15. Dyer, M.D., Murali, T.M., Sobral, B.W. The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathog.* 4, e32 (2008).
16. Tisoncik-Go, J., Gasper, D.J., Kyle, J.E., Eisfeld, A.J., Selinger, C., Hatta, M., *et al*. Integrated omics analysis of pathogenic host responses during pandemic H1N1 influenza virus infection: the crucial role of lipid metabolism. *Cell Host Microbe* 19, 254-266 (2016).
17. Rolland, T., Taşan, M., Charloteaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., *et al.* A proteome-scale map of the human interactome network. *Cell* 159, 1212-1226 (2014).
18. Menche, J., Sharma, A., Kitsak, M., Ghiassian, S.D., Vidal, M., Loscalzo, J., *et al*. Uncovering disease-disease relationships through the incomplete interactome. *Science* 347, 1257601 (2015).
19. Albert, R., Jeong, H., Barabási, A.L. Error and attack tolerance of complex networks. *Nature* 406, 378-382 (2000).
20. Barabási, A.L., Oltvai, Z.N. Network biology: understanding the cell's functional

organization. *Nat. Rev. Genet*. 5, 101-113 (2004).

21. Crua Asensio, N., Muñoz Giner, E., de Groot, N.S., Torrent Burgas, M. Centrality in the host-pathogen interactome is associated with pathogen fitness during infection. *Nat. Commun.* 8, 1409 (2017).
22. Rachita, H.R., Nagarajaram, H.A. Viral proteins that bridge unconnected proteins and components in the human PPI network. *Mol. BioSyst*. 10, 2448-2458 (2014).
23. Gulbahce, N., Yan, H., Dricot, A., Padi, M., Byrdsong, D., Franchi, R., *et al.* Viral perturbations of host networks reflect disease etiology. *PLoS Comput. Biol.* 8, e1002531 (2012).
24. Franzosa, E.A., Xia, Y. Structural principles within the human-virus protein-protein interaction network. *Proc. Natl. Acad. Sci. USA* 108, 10538-10543 (2011).
25. Enard, D., Cai, L., Gwennap, C., Petrov, D.A. Viruses are a dominant driver of protein adaptation in mammals. *eLife* 5, e12469 (2016).
26. Revers, F., García, J.A. Molecular biology of potyviruses. *Adv. Virus Res.* 92, 101-199 (2015).
27. Schaad, M.C., Anderberg, R.J., Carrington, J.C. Strain-specific interaction of the tobacco etch virus NIa protein with the translation initiation factor eIF4E in the yeast two-hybrid system. *Virology* 273, 300-306 (2000).
28. Dunoyer, P., Thomas, C., Harrison, S., Revers, F., Maule, A. A cysteine-rich plant protein potentiates Potyvirus movement through an interaction with the virus genome-linked protein VPg. *J. Virol.* 78, 2301-2309 (2004).
29. Endres, M.W., Gregory, B.D., Gao, Z., Foreman, A.W., Mlotshwa, S., Ge, X., *et al*. Two plant viral suppressors of silencing require the ethylene-inducible host transcription factor RAV2 to block RNA silencing. *PLoS Pathog.* 6, e1000729 (2010).

30. Xiong, R., Wang, A. SCE1, the SUMO-conjugating enzyme in plants that interacts with NIb, the RNA-dependent RNA polymerase of turnip mosaic virus, is required for viral infection. *J. Virol.* 87, 4704-4715 (2013).
31. Vijayapalani, P., Maeshima, M., Nagasaki-Takekuchi, N., Miller, W.A. Interaction of the trans-frame potyvirus protein P3N-PIPO with host protein PCaP1 facilitates potyvirus movement. *PLoS Pathog.* 8, e1002639 (2012).
32. Poque, S., Wu, H.W., Huang, C.H., Cheng, H.W., Hu, W.C., Yang, J.Y., *et al*. Potyviral gene-silencing suppressor HCPro interacts with salicylic acid (SA)-binding protein 3 to weaken SA-mediated defense responses. *Mol. Plant Microbe Interact.* 31, 86-100 (2018).
33. Mukhtar, M.S., Carvunis, A.R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., *et al*. Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science* 333, 596-601 (2011).
34. Ahmed, H., Howton, T.C., Sun, Y., Weinberger, N., Belkhadir, Y., Mukhtar, M.S. Network biology discovers pathogen contact points in host protein-protein interactomes. *Nat. Commun.* 9, 2312 (2018).
35. Fields, S., Song, O. A novel genetic system to detect protein-protein interactions. *Nature* 340, 245-246 (1989).
36. Nagy, P.D. Yeast as a model host to explore plant virus-host interactions. *Annu. Rev. Phytopathol.* 46, 217-242 (2008).
37. Pagán, I., Alonso-Blanco, C., García-Arenal, F. Host responses in life-history traits and tolerance to virus infection in *Arabidopsis thaliana*. *PLoS Pathog*. 4, e1000124 (2008).
38. Elena, S.F., Rodrigo, G. Towards an integrated molecular model of plant-virus interactions. *Curr. Opin. Virol.* 2, 719-724 (2012).

39. Johnsson, N., Varshavsky, A. Split ubiquitin as a sensor of protein interactions in vivo. *Proc. Natl. Acad. Sci. USA* 91, 340-344 (1994).
40. Bosque, G., Folch-Fortuny, A., Picó, J., Ferrer, A., Elena, S.F. Topology analysis and visualization of *Potyvirus* protein-protein interaction network. *BMC Syst. Biol.* 8, 129 (2014).
41. *Arabidopsis* Interactome Mapping Consortium. Evidence for network evolution in an Arabidopsis interactome map. *Science* 333, 601-607 (2011).
42. Hu, C.D., Chinenov, Y., Kerppola, T.K. Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol. Cell* 9, 789-798 (2002).
43. Walter, M., Chaban, C., Schutze, K., Batistic, O., Weckermann, K., Nake, C., *et al*. Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. *Plant J.* 40, 428-438 (2004).
44. Gene Ontology Consortium. Creating the gene ontology resource: design and implementation. *Genome Res.* 11, 1425-1433 (2001).
45. Pumplin, N., Voinnet, O. RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. *Nat. Rev. Microbiol.* 11, 745-760 (2013).
46. Pieterse, C.M.J., Van der Does, D., Zamioudis, C., Leon-Reyes, A., Van Wees, S.C.M. Hormonal modulation of plant immunity. *Annu. Rev. Cell Dev. Biol*. 28, 489-521 (2012).
47. Mäkinen, K., Hafrén, A. Intracellular coordination of potyviral RNA functions in infection. *Front. Plant Sci.* 5, 110 (2014).
48. Meyniel-Schicklin, L., de Chassey, B., André, P., Lotteau, V. Viruses and interactomes in translation. *Mol. Cell. Proteomics* 11, M111.014738 (2012).

49. Cheng, X., Xiong, R., Li, Y., Li, F., Zhou, X., Wang, A. SUMOylation of turnip mosaic virus RNA polymerase promotes viral infection by counteracting the host NPR1-mediated immune response. *Plant Cell* 29, 508-525 (2017).
50. Yang, C., Guo, R., Jie, F., Nettleton, D., Peng, J., Carr, T., *et al*. Spatial analysis of *Arabidopsis thaliana* gene expression in response to turnip mosaic virus infection. *Mol. Plant Microbe Interact.* 20, 358-370 (2007).
51. Lu, X., Jain, V.V., Finn, P.W., Perkins, D.L. Hubs in biological interaction networks exhibit low changes in expression in experimental asthma. *Mol. Syst. Biol.* 3, 98 (2007).
52. Daugherty, M., Malik, H.S. Rules of engagement: molecular insights from host-virus arms races. *Annu. Rev. Genet.* 46, 677-700 (2012).
53. 1001 Genomes Consortium. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* 166, 481-491 (2016).
54. Stolezki, N., Eyre-Walker, A. Estimation of the neutrality index. *Mol. Biol. Evol*. 28, 63-70 (2011).
55. Halehalli, R.R., Nagarajaram, H.A. Molecular principles of human virus protein-protein interactions. *Bioinformatics* 31, 1025-1033 (2015).
56. Weβling R., Epple P., Altmann S., He Y., Yang L., Henz S.R., *et al*. Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell Host Microbe* 16, 364-375 (2014).
57. Boggio, R., Chiocca, S. Viruses and SUMOylation: recent highlights. *Curr. Opin. Microbiol*. 9, 430-436 (2008).
58. Wimmer, P., Schreiner, S., Dobner, T. Human pathogens and the host cell SUMOylation system. *J. Virol.* 86, 642-654 (2012).
59. Fernández de Castro, I., Volonté, L., Risco, C. Virus factories: biogenesis and

structural design. *Cell Microbiol*. 15, 24-34 (2013).

60. Baulcombe, D. RNA silencing in plants. *Nature* 431, 356-363 (2004).
61. Martínez, F., Daròs, J.A. Tobacco etch virus protein P1 traffics to the nucleolus and associates with the host 60S ribosomal subunits during infection. *J. Virol.* 88, 10725-10737 (2014).
62. Perlmutter, J.D., Qiao, C., Hagan, M. F. Viral genome structures are optimal for capsid assembly. *eLife* 2, e00632 (2013).
63. Ivanov, K.I., Eskelin, K., Bašić, M., De, S., Lõhmus, A., Varjosalo, M., Mäkinen, K. Molecular insights into the function of the viral RNA silencing suppressor HCPro. *Plant J*. 85, 30-45 (2016).
64. Rodrigo, G., Daròs, J.A., Elena, S.F. Virus-host interactome: Putting the accent on how it changes. *J. Proteomics* 156, 1-4 (2017).
65. Dodds, P., Rathjen, J. Plant immunity: towards and integrated view of plant-pathogen interactions. *Nat. Rev. Genet*. 11, 539-548 (2010).
66. Teixeira, R.M., Ferreira, M.A., Raimundo, G.A.S., Loriato, V.A.P., Reis, P.A.B., Fontes, E.P.B. Virus perception at the cell surface: revisiting the roles of receptor-like kinases as viral pattern recognition receptors. *Mol. Plant Pathol*. 20, 1196-1202 (2019).
67. Backer, R., Naidoo, S., van den Berg, N. The *NONEXPRESSOR OF PATHOGENESIS-RELATED GENES 1* (*NPR1*) and related family: mechanistic insights in plant disease resistance. *Front. Plant Sci*. 10, 102 (2019).
68. Choi, S.M., Song, H.R., Han, S.K., Han, M., Kim, C.Y., Park, J., *et al*. HDA19 is required for the repression of salicylic acid biosynthesis and salicylic acid-mediated defense responses in *Arabidopsis*. *Plant J*. 71, 135-146 (2012).
69. Sun, T., Busta, L., Zhang, Q., Ding, P., Jetter, R., Zhang, Y. TGACG-BINDING

FACTOR 1 (TGA1) and TGA4 regulate salicylic acid and pipecolic acid biosynthesis by modulating the expression of *SYSTEMIC ACQUIRED RESISTANCE DEFICIENT 1* (*SARD1*) and *CALMODULIN-BINDING PROTEIN 60g* (*CBP60g*). *New Phytol*. 217, 344-354 (2017).

70. Tian, S., Wang, X., Li, P., Wang, H., Ji, H., Xie, J., *et al*. Plant aquaporin AtPIP1;4 links apoplastic H_2O_2 induction to disease immunity pathways. *Plant Physiol*. 171, 1635-1650 (2016).
71. Saiga, S., Furumizu, C., Yokoyama, R., Kurata, T., Sato, S., Kat, T., *et al*. The *Arabidopsis OBERON1* and *OBERON2* genes encode plant homeodomain finger proteins and are required for apical meristem maintenance. *Development* 135, 1751-1759 (2008).
72. Lin, T.F., Saiga, S., Abe, M., Laux, T. OBE3 and WUS interaction in shoot meristem stem cell regulation. *PLoS ONE* 11, e0155657 (2016).
73. Zhao, Y., Zhao, S., Mao, T., Qu, X., Cao, W., Zhang, L., *et al*. The plant-specific actin binding protein SCAB1 stabilizes actin filaments and regulates stomatal movement in *Arabidopsis*. *Plant Cell* 23, 2314-2330 (2011).
74. Eulgem, T., Rushton, P.J., Robatzek, S., Somssich, I.E. The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* 5, 199-206 (2000).
75. Wu, H., Qu, X., Dong, Z., Luo, L., Shao, C., Forner, J., *et al.* WUSCHEL triggers innate antiviral immunity in plant stem cells. *Science* 370, 227-231 (2020).
76. Vidalain, P.O., Tangy, F. Virus-host protein interactions in RNA viruses. *Microbes Infect*. 12, 1134-1143 (2010).
77. Ramachandran, A., Parisien, J.P., Horvath, C.M. STAT2 is a primary target for measles virus V protein-mediated alpha/beta interferon signaling inhibition. *J. Virol.* 82, 8330-8338 (2008).

78. Pál, C., Papp, B., Hurst, L.D. Highly expressed genes in yeast evolve slowly. *Genetics* 158, 927-931 (2001).
79. Drummond, D.A., Bloom, J.D., Adami, C., Wilke, C.O., Arnold, F.H. Why highly expressed proteins evolve slowly. *Proc. Natl. Acad. Sci. USA* 102, 14338-14343 (2005).
80. Chen, C.C., Chao, C.H., Chen, C.C., Yeh, S.D., Tsai, H.T., Chang, C.A. Identification of turnip mosaic virus isolates causing yellow stripe and spot on calla lily. *Plant Dis.* 87, 901-905 (2003).
81. Chung, B.Y., Miller, W.A., Atkins, J.F., Firth, A.E. An overlapping essential gene in the *Potyviridae*. *Proc. Natl. Acad. Sci. USA* 105, 5897-5902 (2008).
82. Poole, R.L. The TAIR database. *Methods Mol. Biol*. 406, 179-212 (2005).
83. Belda-Palazón, B., Ruiz, L., Martí, E., Tárraga, S., Tiburcio, A.F., Culiáñez, F., *et al*. Aminopropyltransferases involved in polyamine biosynthesis localize preferentially in the nucleus of plant cells. *PLoS ONE* 7, e46907 (2012).
84. Schneider, C.A., Rasband, W.S., Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671-675 (2012).
85. Du, Z., Zhou, X., Ling, Y., Zhang, Z., Su, Z. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res*. 38, W64-W70 (2010).
86. Benjamini, Y., Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* 57, 289-300 (1995).
87. Supek, F., Bosnjak, M., Skunca, N., Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* 6, e21800 (2011).
88. Chatr-Aryamontri, A., Breitkreutz, B.J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., *et al*. The BioGRID interaction database: 2015 update. *Nucleic Acids Res*. 43, D470-D478 (2015).

89. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498-2504 (2003).
90. Chai, M., Wu, X., Liu, J., Fang, Y., Luan, Y., Cui, *et al*. (2020) P3N-PIPO interacts with P3 via the shared N-terminal domain to recruit viral replication vesicles for cell-to-cell movement. *J. Virol.* 94, e01898-19.
91. Rodrigo, G., Carrera, J., Ruiz-Ferrer, V., del Toro, F.J., Llave, C., Voinnet, O., *et al*. A meta-analysis reveals the commonalities and differences in *Arabidopsis thaliana* response to different viral pathogens. *PLoS ONE* 7, e40526 (2012).
92. Emms, D.M., Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16, 157 (2015).
93. Nakamura, T., Yamada, K.D., Tomii, K., Katoh, K. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* 34, 2490-2492 (2018).
94. Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972-1973 (2009).
95. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol*. 24, 1586-1591 (2007).
96. Macías, L., Barrio, L., Toft, C. GWideCodeML: A Python package for testing evolutionary hypotheses at the genome-wide level. *G3 - Genes Genom. Genet.* doi.org/10.1534/g3.120.401874 (2020).
97. Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S.E., Lercher, M.J. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol*. 31, 1929-1936 (2014).

Figures

Figure 1: Virus-host protein-protein interactome constructed in this work by yeast two-hybrid screening. This corresponds to the interaction of the plant virus TuMV and the host plant *A. thaliana* (blue edges). Virus proteins highlighted. The interactors of P3, P3N-PIPO, and 6K2 were retrieved with a library of *A. thaliana* proteins expressed in the membrane, whilst the interactors of HC-Pro, 6K1, CI, VPg, NIa-Pro, and NIb with a library of proteins expressed in the cytoplasm. Virus-virus protein-protein interactions (red edges) already reported in previous work complement this network view.

Figure 2: Validation of some virus-host protein-protein interactions by bimolecular fluorescence complementation *in planta* with a split YFP system. The TuMV proteins were cloned with the YFP N-terminus and the host proteins with the YFP C-terminus. Confocal images of plant tissues (from *N. benthamiana*) to reveal the interaction by fluorescence (GFP filter), together with merged images (GFP and RFP filters) to localize the chloroplasts due to their autofluorescence. Scale bar, 20 μm.

Figure 3: Functional analysis of the host proteins targeted by the plant virus TuMV. a) Representation in a semantic space of the functional categories (related to biological processes) enriched within the virus targets. Bubble size scales with the total number of proteins per category. Statistical significance assessed by Fisher exact tests, 2×2 tables, adjusted $P < 0.05$. b) Heat map for the main functions (related to metabolism, development, organization, signaling, regulation, and defense) targeted by each TuMV protein. Color scale indicates fraction relative to the total number of targets of that virus protein. c) Principal component analysis showing the spatial arrangement of the different virus proteins regarding the functional overlap of their targets (from data shown in panel c). d) Associative map between the function of each virus protein (relative to the virus)

and the main functions of each virus target (relative to the host plant).

Figure 4: Virus-host interactions allow establishing new communication channels between the virus proteins. a) Partial virus-host interactome showing those host proteins that are targeted by two or more TuMV proteins (19 host proteins in total). b) Null probability distribution of the number of host proteins targeted by two or more virus proteins after 10^4 random realizations. Arrow marks the actual value; *statistical significance (z test, $P < 0.0001$). c) Length and d) number of the different shortest paths connecting the virus protein pairs in this partial interactome (blue and red edges). Numbers in the upper hemi-matrixes indicate how these values change when only virus-virus interactions are considered (red edges).

Figure 5: The plant virus TuMV targets host proteins with higher connectivity. a) Whole protein-protein interactome of the host plant *A. thaliana* contextualizing those proteins that interact with the plant virus TuMV (red nodes). b) Probability distribution of the degree for only the virus targets (red) or all proteins in the host interactome (black). Points correspond to the data, whilst lines to the power law probability ~ degree$^{-\gamma}$ (γ = 0.860 for virus targets, γ = 1.247 for all proteins). c) Comparison between the actual degree distribution (from virus targets) and a representative null distribution (from randomly picked genes). *Statistical significance (Mann-Whitney U test, $P < 0.05$). The inset shows the distribution of P values after 10^3 random realizations, with geometric mean 0.0045. d) Comparison between the actual shortest path length distribution (from virus targets) and a representative null distribution (from randomly picked genes). *Statistical significance (Mann-Whitney U test, $P < 0.05$). The inset shows the distribution of P values after 10^3 random realizations, with geometric mean 0.0003. e) Network-function detail of four virus targets with markedly high degree, which interact

with TuMV proteins NIb and VPg. The area of the shadow regions is log-proportional to the degree (indicated in number). TGA1 shares interactors with NPR1 and OBE1.

Figure 6: The plant virus TuMV targets host proteins with higher expression levels. a) Virus-host protein-protein interactome contextualizing gene expression data (from healthy *A. thaliana*). Three expression levels are categorized. Host proteins whose expression significantly changes upon TuMV infection are represented by bigger nodes (+ indicates up-regulation, − down-regulation). b, c) Comparison between the actual Affymetrix expression distribution (from virus targets) and a representative null distribution (from randomly picked genes). *Statistical significance (Mann-Whitney *U* test, $P < 0.05$). The inset shows the distribution of *P* values after 10^3 random realizations, with geometric mean $< 10^{-10}$ (in both cases). Expression levels corresponding to b) a healthy or c) an infected plant. d) Scatter plot between the degree and differential expression (virus targets in red). Virus targets that are more in the trade-off front are highlighted, indicating in parenthesis the corresponding virus protein.

Figure 7: TuMV-interacting host proteins are evolutionarily conserved hubs. a) Phylogeny of the three species used for the population study. b) Comparison between the actual p_N/p_S distribution [computed as $p_N/(p_S + 1)$ to account for genes with $p_S = 0$] and a representative null distribution (from randomly picked genes). *Statistical significance (Mann-Whitney *U* test, $P < 0.05$). The inset shows the distribution of *P* values after 10^3 random realizations, with geometric mean 0.0016. c) Comparison between the actual direction of selection (*DoS*) distribution and a representative null distribution (from randomly picked genes). ns means statistically non-significant (Mann-Whitney *U* test, $P > 0.05$). The inset shows the distribution of *P* values after 10^3 random realizations, with geometric mean 0.0654. d) Top five host factors with $DoS > 0$ (positive selection) and

DoS < 0 (negative selection), together with their interacting virus proteins. e) Species tree generated by OrthoFinder. The branch of interest, used in the CODEML analysis, is the green branch leading to the *Arabidopsis* genus. f) Comparison between the actual ω distribution and a representative null distribution (from randomly picked genes). *Statistical significance (Mann-Whitney *U* test, $P < 0.05$). The inset shows the distribution of *P* values after 10^3 random realizations, with geometric mean 0.0043.

Supplementary Information

Dataset S1 (Excel file): The list of all host interactors identified in this work by HT-Y2H screening.

Dataset S2 (Excel file): A literature-curated list of all physical potyvirus-plant interactions already described.

Dataset S3 (Excel file): Results of the different selection analyses (p_N/p_S and ω) for each *A. thaliana* gene.

Dataset S4 (Excel file): All primers used to generate the different genetic constructions of this work.

Table S1: List of plant species used for the evolutionary analyses.

Species	Reference to the genome data
Arabidopsis halleri L.	Briskine, R.V., Paape, T., Shimizu-Inatsugi, R., Nishiyama, T., Akama, S., Sese, J., *et al*. Genome assembly and annotation of *Arabidopsis halleri*, a model for heavy metal hyperaccumulation and evolutionary ecology. Mol. Ecol. Resour. 17, 1025-1036 (2017).
Arabidopsis lyrata L.	Hu, T.T., Pattyn, P., Bakker, E.G., Cao, J., Cheng, J.F., Clark, R.M., *et al*. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat. Genet. 43, 476-481 (2011).
Arabidopsis thaliana (L.) Heyhn	Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796-815 (2000).
Barbarea vulgaris W.T. Aiton	Byrne, S.L., Erthmann, P.Ø., Agerbirk, N., Bak, S., Hauser, T.P., Nagy, I., *et al*. The genome sequence of *Barbarea vulgaris* facilitates the study of ecological biochemistry. *Sci. Rep*. 7, 40728 (2017).
Boechera retrofracta (Graham) Á. Love & D. Love	Kliver, S., Rayko, M., Komissarov, A., Bakin, E., Zhernakova, D., Prasad, K., *et al*. Assembly of the *Boechera retrofracta* genome and evolutionary analysis of apomixis-associated genes. *Genes* 9, 185 (2018).
Capsella rubella Reut.	Slotte, T., Hazzouri, K.M., Ågren, J.A., Koenig, D., Maumus, F., Guo, Y.L., *et al*. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat. Genet.* 45, 831-835(2013).
Cardamine hirsuta L.	Gan, X., Hay, A., Kwantes, M., Haberer, G., Hallab, A., Dello, R., *et al*. The *Cardamine hirsuta* genome offers insight into the evolution of morphological diversity. *Nat. Plants* 2, 16167 (2016).
Crucihimalaya himalaica (Edgew) Al-Shehbaz, O'Kane & R.A. Price	Zhang, T., Qiao, Q., Novikova, P.Y., Wang, Q., Yue, J., Guan, Y., *et al*. Genome of *Crucihimalaya himalaica*, a close relative of *Arabidopsis*, shows ecological adaptation to high altitude. *Proc. Natl. Acad. Sci. USA* 116, 7137-7146 (2019).
Erysimum cheiranthoides L.	Züst, T., Strickler, S.R., Powell, A.F., Mabry, M.E., An, H., Mirzaei, M., *et al*. Independent evolution of ancestral and novel defenses in a genus of toxic plants (*Erysimum*, *Brassicaceae*). *eLife* 9, e51712 (2020).

1

Figures

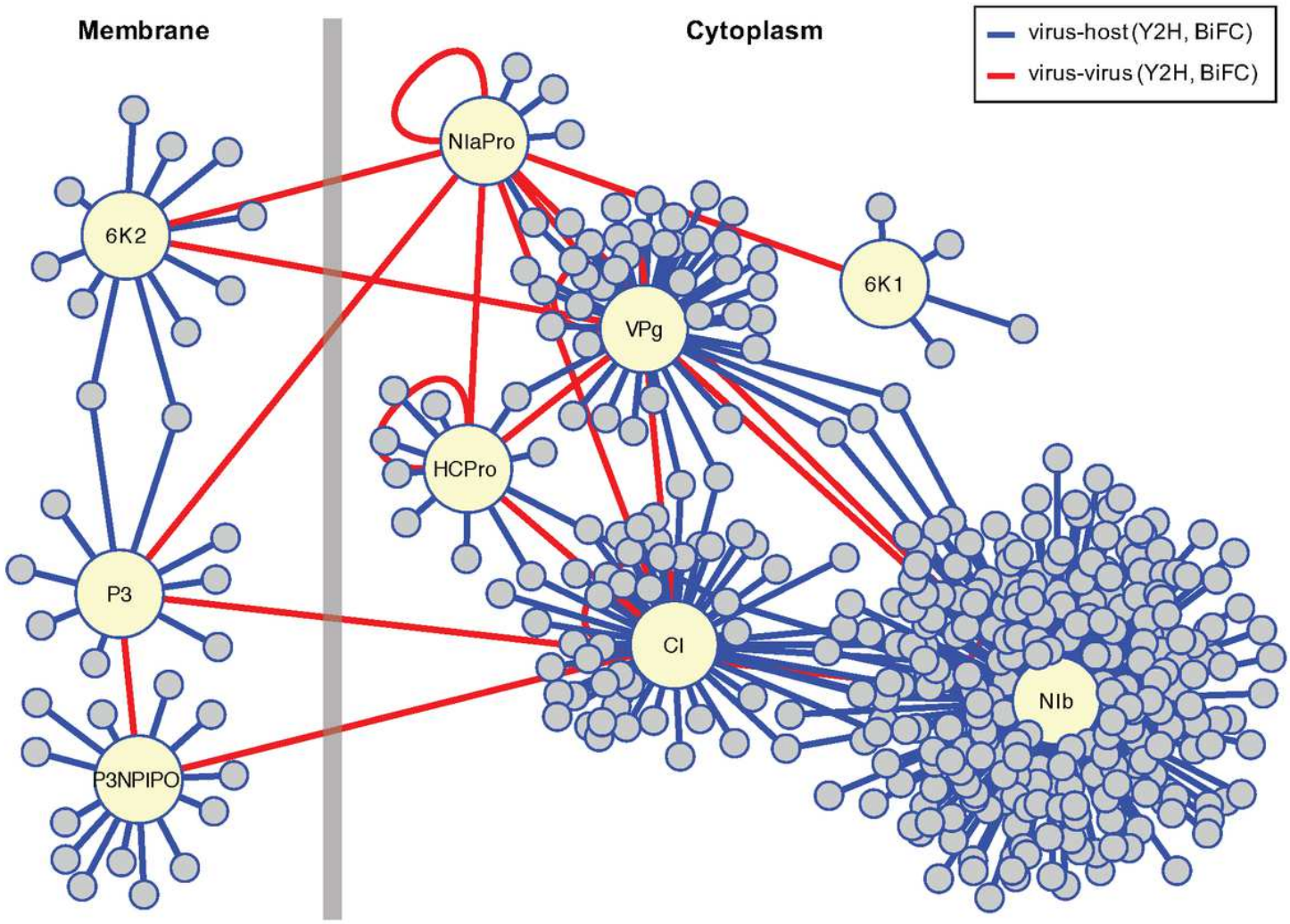

Figure 1

Virus-host protein-protein interactome constructed in this work by yeast two-hybrid screening. This corresponds to the interaction of the plant virus TuMV and the host plant A. thaliana (blue edges). Virus proteins highlighted. The interactors of P3, P3N-PIPO, and 6K2 were retrieved with a library of A. thaliana proteins expressed in the membrane, whilst the interactors of HC-Pro, 6K1, CI, VPg, NIa-Pro, and NIb with a library of proteins expressed in the cytoplasm. Virus-virus protein-protein interactions (red edges) already reported in previous work complement this network view.

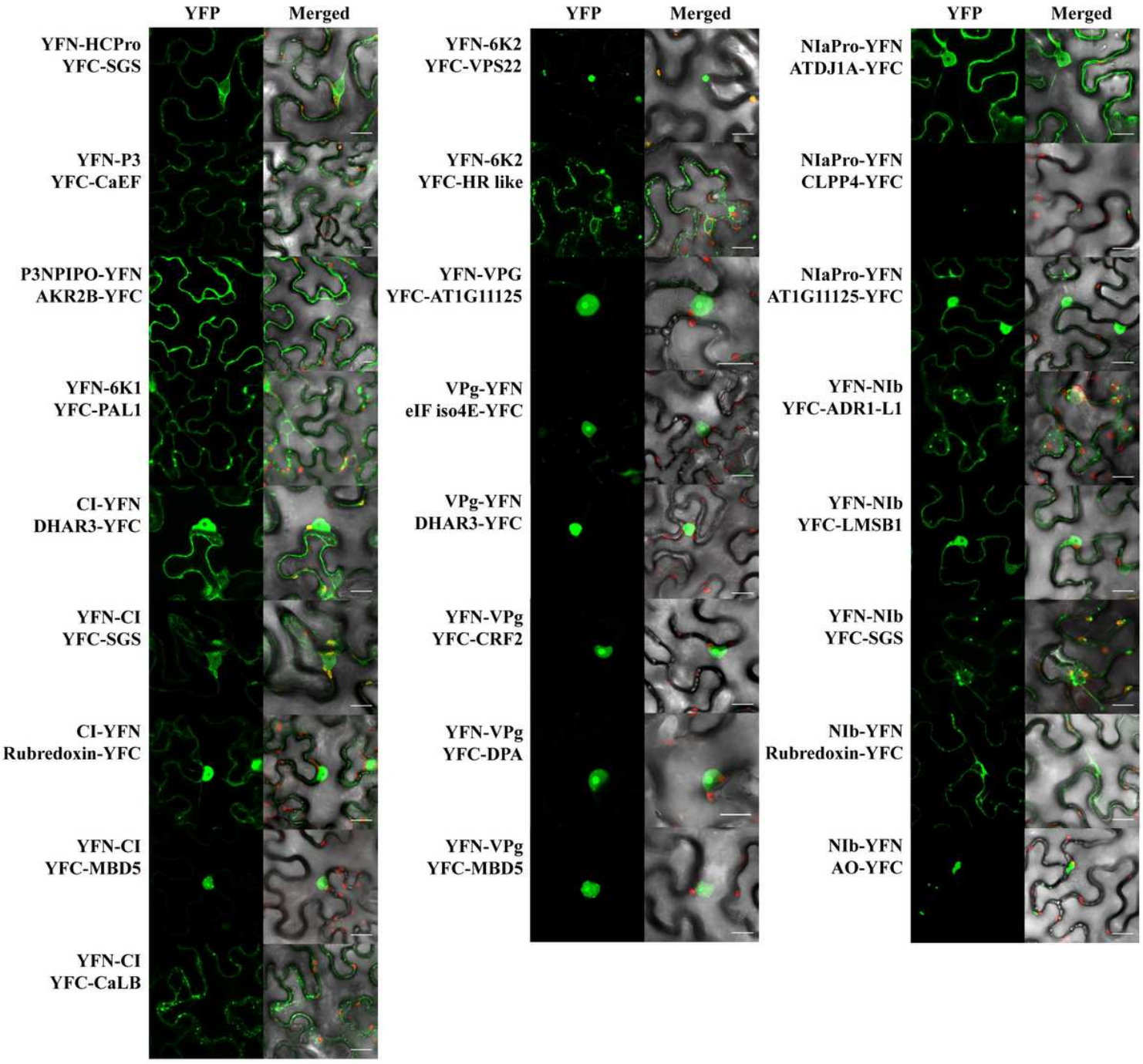

Figure 2

Validation of some virus-host protein-protein interactions by bimolecular fluorescence complementation in planta with a split YFP system. The TuMV proteins were cloned with the YFP N-terminus and the host proteins with the YFP C-terminus. Confocal images of plant tissues (from N. benthamiana) to reveal the interaction by fluorescence (GFP filter), together with merged images (GFP and RFP filters) to localize the chloroplasts due to their autofluorescence. Scale bar, 20 µm.

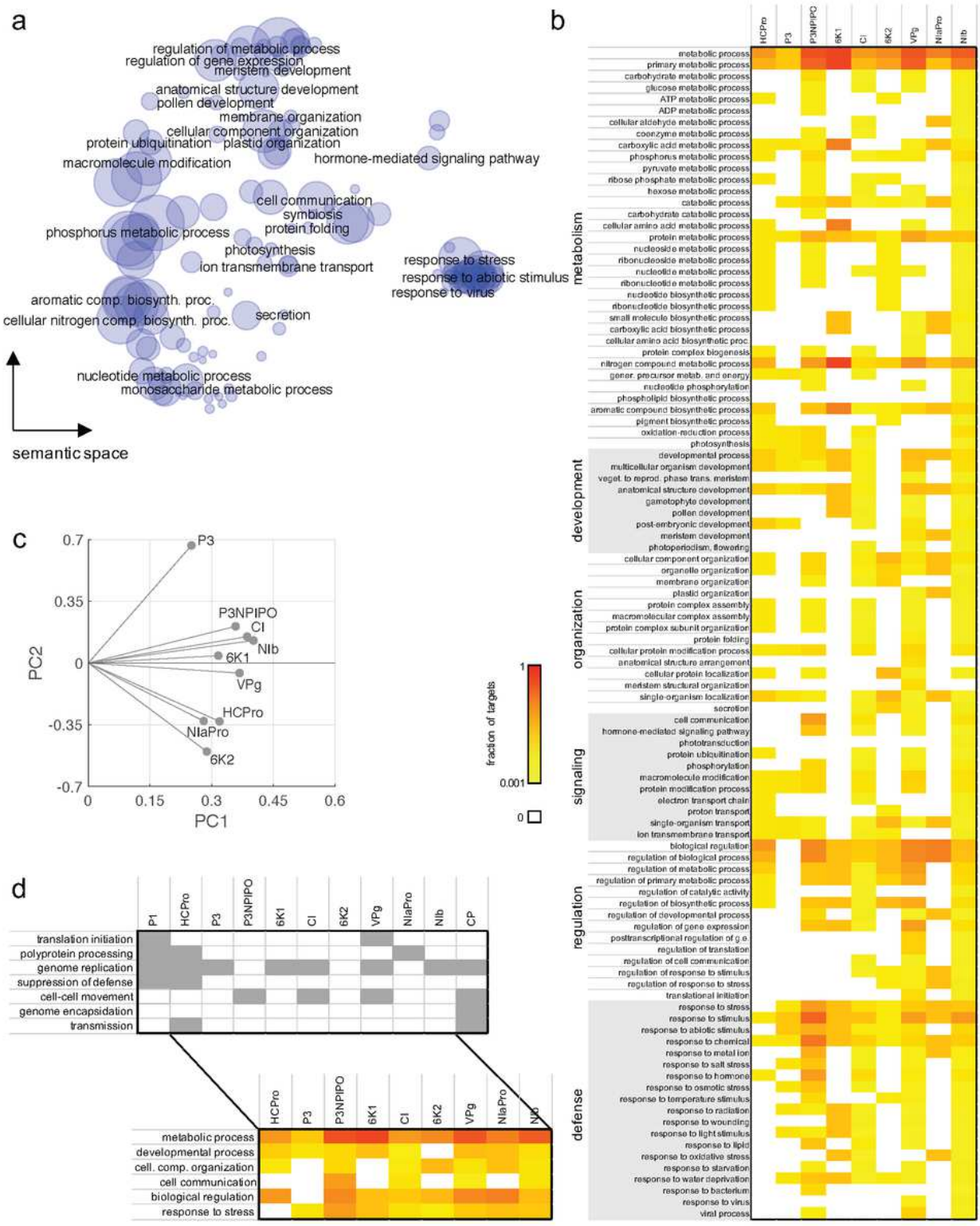

Figure 3

Functional analysis of the host proteins targeted by the plant virus TuMV. a) Representation in a semantic space of the functional categories (related to biological processes) enriched within the virus targets. Bubble size scales with the total number of proteins per category. Statistical significance assessed by Fisher exact tests, 2´2 tables, adjusted $P < 0.05$. b) Heat map for the main functions (related to metabolism, development, organization, signaling, regulation, and defense) targeted by each TuMV

protein. Color scale indicates fraction relative to the total number of targets of that virus protein. c) Principal component analysis showing the spatial arrangement of the different virus proteins regarding the functional overlap of their targets (from data shown in panel c). d) Associative map between the function of each virus protein 1 (relative to the virus) and the main functions of each virus target (relative to the host plant).

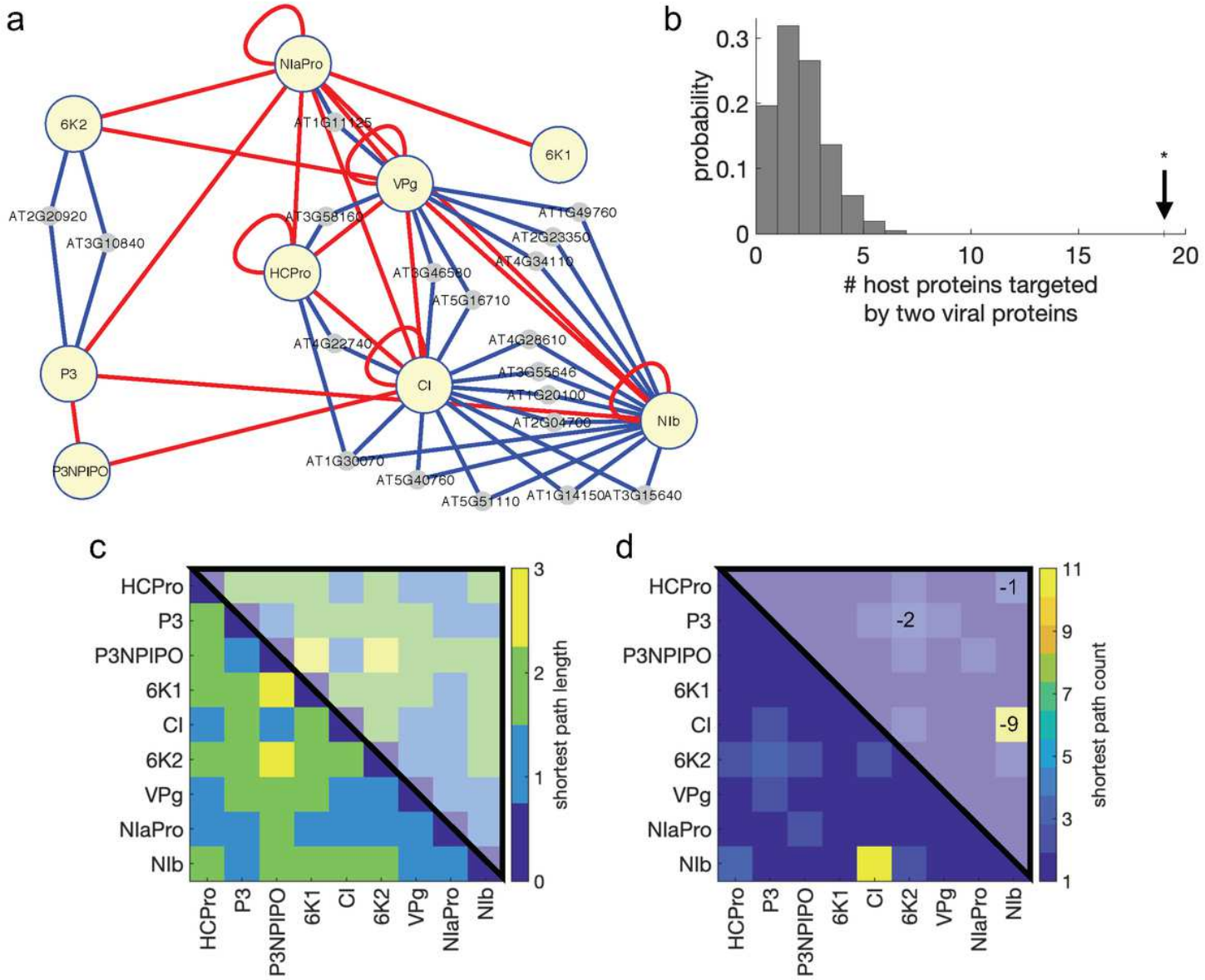

Figure 4

Virus-host interactions allow establishing new communication channels between the virus proteins. a) Partial virus-host interactome showing those host proteins that are targeted by two or more TuMV proteins (19 host proteins in total). b) Null probability distribution of the number of host proteins targeted by two or more virus proteins after 104 random realizations. Arrow marks the actual value; *statistical significance (z test, $P < 0.0001$). c) Length and d) number of the different shortest paths connecting the virus protein pairs in this partial interactome (blue and red edges). Numbers in the upper hemi-matrixes indicate how these values change when only virus-virus interactions are considered (red edges).

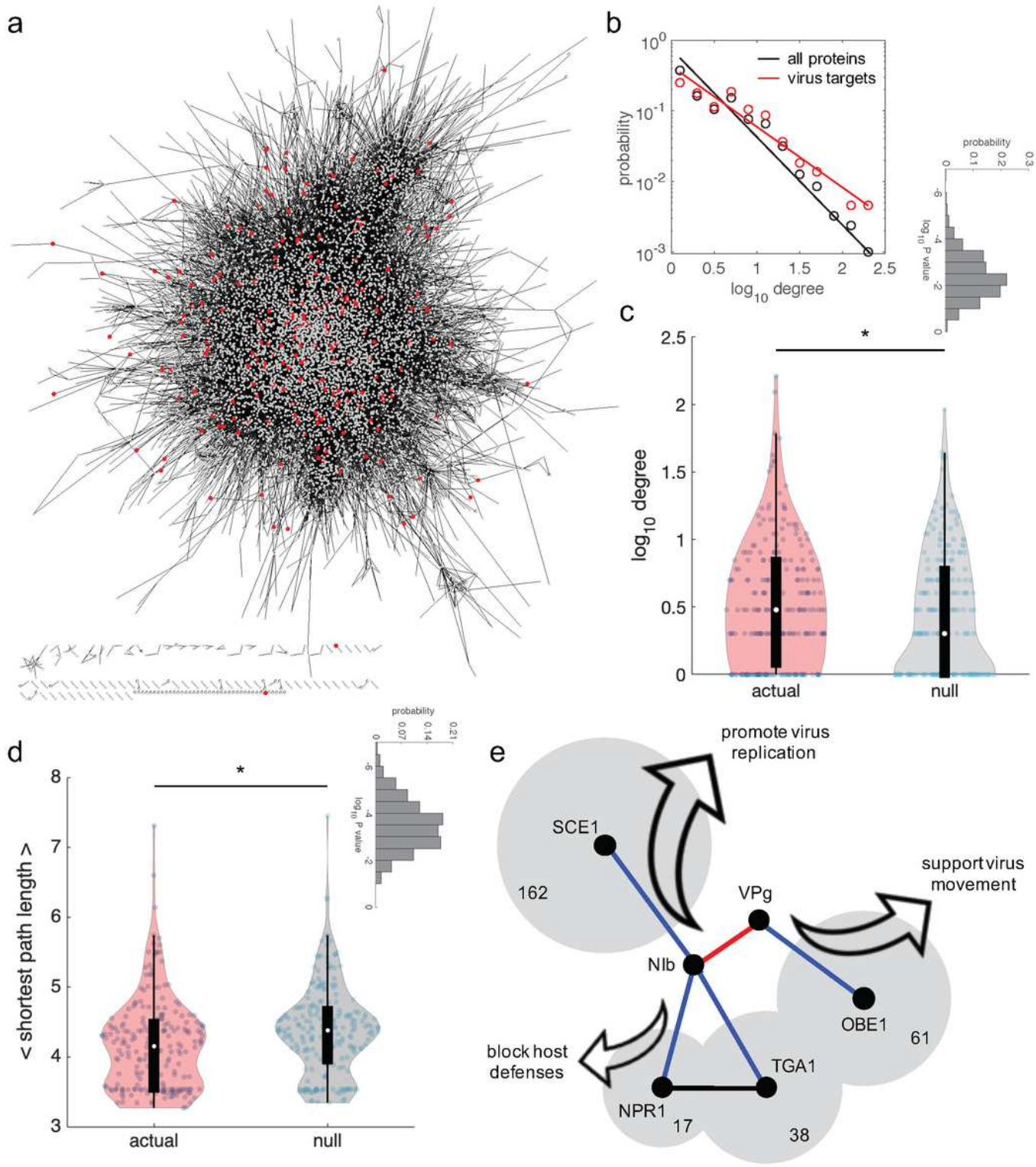

Figure 5

The plant virus TuMV targets host proteins with higher connectivity. a) Whole protein-protein interactome of the host plant A. thaliana contextualizing those proteins that interact with the plant virus TuMV (red nodes). b) Probability distribution of the degree for only the virus targets (red) or all proteins in the host interactome (black). Points correspond to the data, whilst lines to the power law probability ~ degree−g (g = 0.860 for virus targets, g = 1.247 for all proteins). c) Comparison between the actual degree distribution

(from virus targets) and a representative null distribution (from randomly picked genes). *Statistical significance (Mann-Whitney U test, P < 0.05). The inset shows the distribution of P values after 103 random realizations, with geometric mean 0.0045. d) Comparison between the actual shortest path length distribution (from virus targets) and a representative null distribution (from randomly picked genes). *Statistical significance (Mann-Whitney U test, P < 0.05). The inset shows the distribution of P values after 103 random realizations, 1 with geometric mean 0.0003. e) Network-function detail of four virus targets with markedly high degree, which interact with TuMV proteins NIb and VPg. The area of the shadow regions is log-proportional to the degree (indicated in number). TGA1 shares interactors with NPR1 and OBE1.

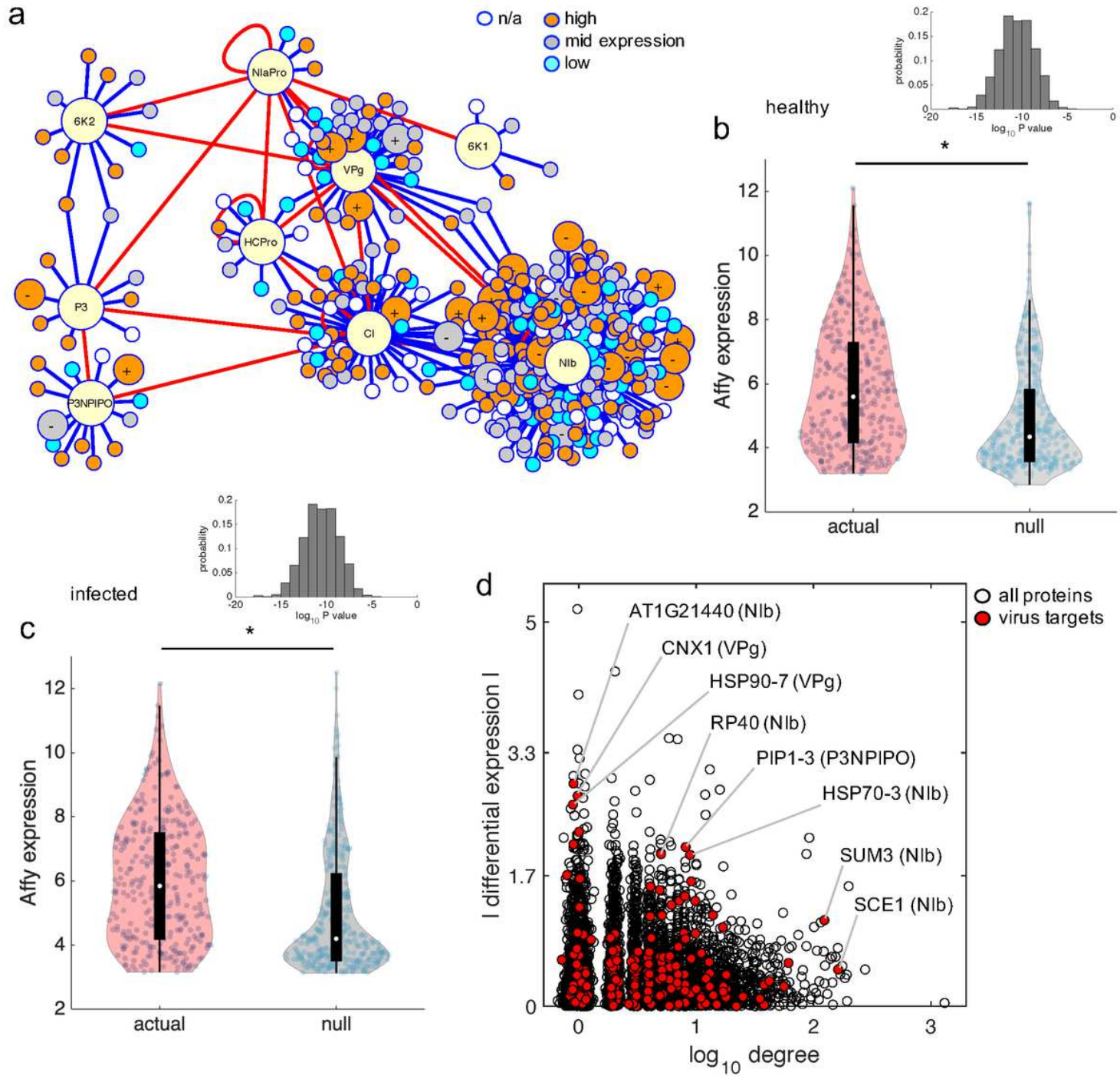

Figure 6

The plant virus TuMV targets host proteins with higher expression levels. a) Virus-host protein-protein interactome contextualizing gene expression data (from healthy A. thaliana). Three expression levels are categorized. Host proteins whose expression significantly changes upon TuMV infection are represented by bigger nodes (+ indicates up-regulation, - down-regulation). b, c) Comparison between the actual Affymetrix expression distribution (from virus targets) and a representative null distribution (from randomly picked genes). *Statistical significance (Mann-Whitney U test, P < 0.05). The inset shows the distribution of P values after 103 random realizations, with geometric mean < 10-10 (in both cases). Expression levels corresponding to b) a healthy or c) an infected plant. d) Scatter plot between the degree and differential expression (virus targets in red). Virus targets that are more in the trade-off front are highlighted, indicating in parenthesis the corresponding virus protein.

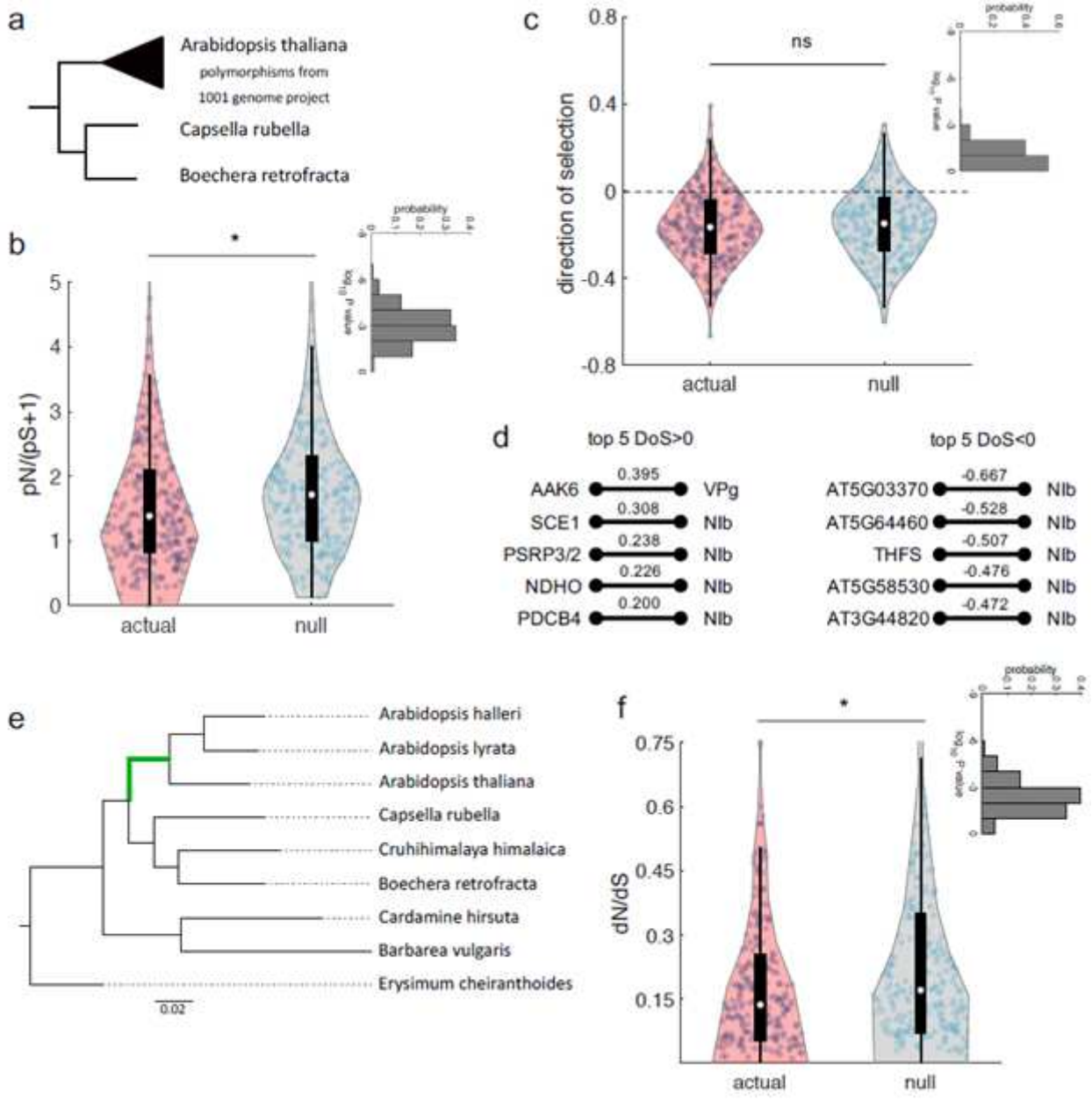

Figure 7

TuMV-interacting host proteins are evolutionarily conserved hubs. a) Phylogeny of the three species used for the population study. b) Comparison between the actual pN/pS distribution [computed as pN/(pS + 1) to account for genes with pS = 0] and a representative null distribution (from randomly picked genes). *Statistical significance (Mann-Whitney U test, P < 0.05). The inset shows the distribution of P values after 103 random realizations, with geometric mean 0.0016. c) Comparison between the actual direction of selection (DoS) distribution and a representative null distribution (from randomly picked genes). ns means statistically non-significant 1 (Mann-Whitney U test, P > 0.05). The inset shows the distribution of P values after 103 random realizations, with geometric mean 0.0654. d) Top five host factors with DoS > 0 (positive selection) and DoS < 0 (negative selection), together with their interacting virus proteins. e) Species tree generated by OrthoFinder. The branch of interest, used in the CODEML analysis, is the green branch leading to the Arabidopsis genus. f) Comparison between the actual w distribution and a representative null distribution (from randomly picked genes). *Statistical significance (Mann-Whitney U test, P < 0.05). The inset shows the distribution of P values after 103 random realizations, with geometric mean 0.0043.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- figureS1.pdf
- figureS2.pdf
- datasetS1.xlsx
- datasetS2.xlsx
- datasetS3.xlsx
- datasetS4.xlsx
- tableS1.docx