# Predicting protein-protein interactions in the *Solanum lycopersicum - Phytophthora infestans* pathosystem

**Daniel Del Hoyo**

**ABSTRACT**
Protein-Protein Interactions (PPIs) play an important role in plant infections, taking part in the detection of the pathogen by the plant and its defence mechanisms, but also in the manipulation of host processes by the pathogen. *Phytophthora infestans* is a plant pathogenic oomycete with a large arsenal of secreted proteins that help the infection process through interactions with the host plant proteins. We constructed a pipeline to predict PPIs between *P. infestans* and one of its most important hosts, *Solanum lycopersicum* (tomato). Two methods based on protein domains are implemented and combined to predict PPIs between these two species. Additionally, various evaluations based on the predicted network architecture, gene expression and targeted processes are performed to validate the predicted interactions and study the influence of different parameters of the pipeline.

**KEYWORDS:** Host-pathogen interactions, pathogen effectors, protein interaction prediction, *Solanum lycopersicum, Phytophthora infestans*, oomycete

**INTRODUCTION**
The oomycete *Phytophthora infestans* is a microscopic and filamentous eukaryote. This pathogen causes the late blight disease on tomato and potato, two of the most important crops in agriculture, leading to billions of dollars in losses and control costs annually [1]. This pathogen was initially categorized as a fungus because of its similar morphology and infection process [2]. However, studies based on genetic structure [3] and phylogeny [4] have revealed that it is evolutionary closer to diatoms and brown algae. Throughout history, *P. infestans* has caused devastating epidemics, one of which led to the well-known Irish potato famine in the 1840s, that caused a population drop of one third in Ireland due to millions of deaths and migrating families [5, 6]. The high genomic mutation rate of *P. infestans* leads to the appearance of new, resistant strains of this pathogen each year, hampering the development of long-term control methods [7, 8].

*Phytophthora* spp. have both asexual and sexual reproductive cycles. The sexual cycle allows the pathogen to diversify into new strains [9]. Plant infection mainly occurs during the asexual cycle that can be classified as hemibiotrophic [10]. The infection cycle has a biotrophic phase, where *P. infestans* grows in close contact with alive host cells through the formation of hyphae and haustoria. This phase is followed by extensive necrosis of the host tissue, followed by fast colonization and sporangium formation (Figure 1) [7, 11, 12].

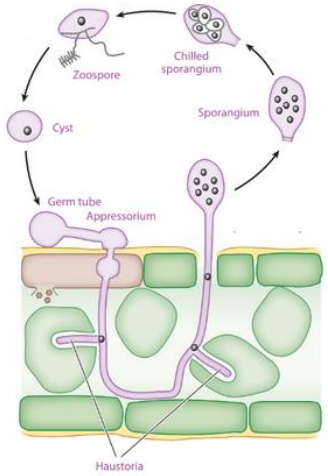


*Figure 1: The Phytophthora asexual life cycle* [12].

During infection, the pathogen adopts different morphological life stages for dispersion, host penetration and growth. Dispersion between hosts is usually governed by sporangia, which are blown from aerial parts of the infected host onto new host plants. When these cells are in contact with a new host, they either directly form a germ tube or release motile zoospores that eventually form a cyst by formation of a cell wall. The cyst germinates and penetrates the host tissues through natural openings in the plant or by forming a germ tube. The germ tube forms an appressorium that pierces the cuticle of a host epidermal cell and once inside the host, a primary vesicle is formed [13]. From this primary vesicle, hyphae start to grow in the extracellular tissue and colonize the plant cells. The hyphae inside the host form a special structure called a haustorium. Here, pathogen proteins are

secreted to disrupt the host defence response and to manipulate its metabolism [13].

Manipulation of host processes is vital for pathogen proliferation. This process is mediated by proteins called effectors. Effectors are secreted by the pathogen from the apoplast and from the haustoria, where the host and the pathogen membranes are in close contact. Genomic analyses revealed that *P. infestans* has an extremely large number of effector genes, many of which code for secreted proteins [14, 15]. Many of these effectors contain an N-terminal signal peptide for secretion that is followed by a C-terminal functional effector module [16]. Some of these effectors are inhibitors of host defence proteases (e.g. Cysteine and Serine protease inhibitors EPI1 and Kazal-like Serine protease inhibitors), proteases to counter attack the host defences and necrosis and ethylene inducing peptides such as the NEP1-Like Family [17]. The rest of the effectors would be translocated inside the host cell through specialized structures and translocation signals. Many of these cytoplasmic effectors are uncategorised, but some others have been included in effector families such as RxLR and CRN, and have been described to modulate and disrupt host cellular signalling and necrosis processes [16]. The RxLR effectors earned this name because they are characterized by an RxLR motif (Arginine, any, Leucine, and Arginine) near the N-terminal domain. This RxLR domain has been shown to act as a translocation signal to penetrate the host cell membrane. RxLR genes are mainly found in genome regions with high transposon activity, which may be the reason for their quick evolution and high variability [18]. Sequence variation is mainly found in the C-terminal regions, as this is likely to be the module with the biochemical effector activity, while the N-terminal regions remain constant with the translocation signal [19]. CRN effectors are known for their crinkling and necrosis-inducing activity and some of them have been described to localize in the host nucleus, modulating the host's gene transcription [20, 21]. These proteins have a highly conserved N-terminal domain containing the translocation motif LxLFLAK (Leucine, any, Leucine, Phenylalanine, Leucine, Alanine, Lysine) and a variable C-terminal region. Although a function for most of these effectors has not been identified yet, one CRN effector of *P. infestans* has been described to exhibit a kinase activity, likely disrupting the host signalling cascade during the infection process [21, 22].

Plants are constantly affected by abiotic and biotic stresses which pose important selection pressures and require them to develop and maintain defences to fight against these factors. This causes pathogens and their host plants to coevolve, in a perpetual fight in which the host tries to destroy the pathogen, while the pathogen tries to avoid the host defences. One of the models that describes this coevolution process of plant-pathogen interactions is the Zig-Zag model proposed by Jones and Dangl in 2006 [23]. Although this model might be too simple for describing all the interactions in the host-pathogen system [24, 25], it still can be used to explain the coevolution of pathogen effectors and host defences.

According to this model, the defences of a plant host are divided into two branches. In the first branch, the plant develops extracellular receptors that recognise conserved molecules among usual pathogens and start a signalling cascade that would activate defences against the colonization. These conserved molecules are called pathogen associated molecular patterns (PAMPs) and the mechanism is called PAMP triggered immunity (PTI). However, some pathogens develop strategies to avoid PTI by hiding their PAMPs or by interfering with the signalling cascade. Among these strategies, we find the secretion of proteins called effectors (Figure 2).

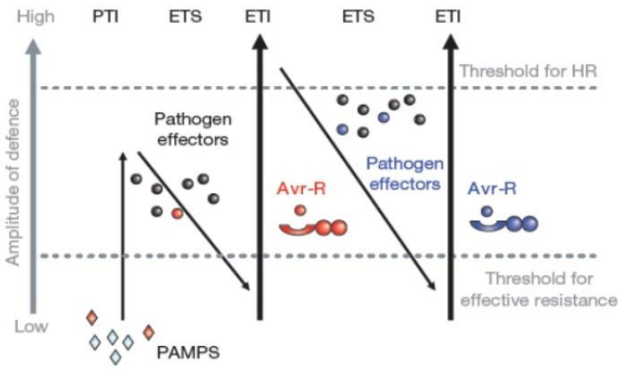


*Figure 2: Different phases of the Zig-Zag model* [23].

In the second branch of the defences, the host plant employs resistance proteins (R) to detect the pathogen effectors and activate the effector triggered immunity (ETI). However, very few direct interactions between effectors and resistance proteins have been described. This led to the formulation of other theories that explain how this recognition happens, for example the guard hypothesis or the decoy hypothesis [26, 27].

In all these models, both the infection processes and the defence responses would be governed through protein-protein interactions (PPI) between host and pathogen proteins. PPIs are key for these pathogen effectors to work. Determining these interactions might reveal new patterns that can be exploited to develop effective controls for this devastating disease. In general, PPIs can be classified into physical and functional interactions. In the first case proteins have physical contact (e.g. two alfa chains and two beta chains form the hemoglobin protein complex). Functional interactions include proteins that work in the same biological processes, but may be physically separated [28]. For experimental determination of

PPIs, several methods have been developed, such as Yeast2Hybrid or co-inmunoprecipitation. However, the experimental study of physical interactions is slow and expensive, and therefore large-scale experiments describing PPIs are rare [29]. Development of *in silico* prediction methods is useful to accelerate this task. Several of these methods have been developed in an attempt to predict whether a pair of proteins could physically interact [30]. These methods are based on several protein features: structure [31], homology [32], sequence [33, 34] or domains [34–36].

Several methods for predicting PPIs that use protein domain information have been developed in the last years, but all of them are based on the same assumption: some protein interactions take place between two or more domains identified in those proteins [37]. Some of these methods even extend this idea to include particular interactions between specific domains and protein motifs [38]. These methods propose that a pair of proteins is likely to interact when they contain domains known to interact. Protein domains are usually defined as compact and stable conserved structural segments of a protein, that are able to fold independently; that is, their structure arises without the intervention of the rest of the protein. On the other hand, protein motifs are conventionally defined as non-independent, short conserved protein sequence segments.

Here we describe the development and application of *in silico* PPI prediction methods to provide insight into potential PPIs between *P. infestans* and tomato (*Solanum lycopersicum*). We examine the application of these computational methods by exploiting protein domain and motif information from various sources.

## MATERIALS AND METHODS

### Protein and interactome sets

The set of *P. infestans* proteins used in this project was extracted from Meijer et al. 2014 [15]. In this paper, the secretome of *P. infestans* is obtained by merging an *in silico* predicted secretome [14] with proteins identified extracellularly by mass spectrometry. The resulting secretome consists of 2242 proteins that are potential effectors in the infection process of *S. lycopersicum* (Supplementary File 1). The set of tomato proteins was extracted from the interactions database STRING v10.5 [39]. STRING includes both experimentally determined and predicted PPIs from different species. These predictions are scored on a scale from 150 to 999 using different sources of information: co-expression analysis, ortholog transfer, shared selective signals across genomes and automated text-mining of scientific literature. The database also includes and categorizes the interactions into physical or several kinds of indirect interactions. We maintained a filter for extracting only physical PPIs and applied a score threshold of 900 in order to take only the more reliable interactions. After filtering, the tomato PPI network contained 552 tomato proteins involved in 956 interactions (Supplementary File 1).

### Domain and motif annotation

The sets of proteins of *P. infestans* and *S. lycopersicum* were annotated for domains using the software InterProScan v5 (5.27-66.0) with the Pfam database (v31.0) [37, 40]. Additionally, motifs known to bind domains were extracted from the 3did (v2017_01) [41] and iELM (v1.0) [42] databases in regular expression format. The number of motifs retrieved was 702 and 267 respectively. A motif annotation was performed on the described *P. infestans* secretome and *S. lycopersicum* interactome protein sets. Since some motifs are short peptides that may appear by chance, two different filters were applied: a surface exposure filter and a frequency filter [38]. The surface exposure filter removes motifs of which more than 50% of the residues are predicted to be buried. Surface exposure was calculated with the software NetSurfP v1.0 [43]. Thereafter, we applied a frequency filter to remove motifs that occur in over 5% or 10% of the proteins, on the *S. lycoperosicum* and *P. infestans* sets respectively.

### Domain-based approaches

Two different domain approaches were applied independently to predict PPIs in the *P. infestans* – *S. lycopersicum* pathosystem. These methods are described in Dyer et al. (2007) and Zhang et al. (2016) [35, 36]. Additionally, we added the motif information and worked with these motifs in the same way as domains. Both approaches work in a similar way: they assign a score in the range 0-1 to Domain-Domain Interactions (DDIs) and Domain-Motif Interactions (DMIs); and calculate a score for each possible protein pair from our datasets as follows:

$$\Pr\left(P_{gh}\right) = 1.0 - \prod_{D_{de} \in P_{gh}} \left(1.0 - \Pr\{g, h | d, e\}\right) \quad (1)$$

In Equation 1, we obtain $\Pr(P_{gh})$, the PPI score for proteins *g* and *h,* by substituting $\Pr\{g, h|d, e\}$ for the DDI/DMI scores of domains/motifs *d* and *e*, for all *d* and *e* in *g* and *h* respectively [36]. The two methods differ in how they calculate the score for the DDIs and DMIs. In the Zhang approach, the scores for structurally determined DDIs are extracted from the database 3did (v2017_01), obtaining a total of 11200 DDI scores in the range 0-1. However, this database does not provide scores for DMIs, and neither does the iELM database. We therefore used an arbitrary score of 0.3 for all DMIs, the rounded average of all DDI scores. In the Dyer method, both DDI and DMI scores are calculated based on the

information of the host interactome. A Bayesian formula (Equation 2) is used to score the *a posteriori* probability for a DDI or DMI to appear in interacting proteins [35]. In this way, DDIs and DMIs that appear frequently in interacting proteins obtain a greater score because they are more likely to participate in the interactions.

$$\Pr\{g,h|d,e\} = \frac{\Pr\{d,e|g,h\}\Pr\{I(g,h)\}}{\Pr\{D(g,d),D(h,e)\}} \qquad (2)$$

Pr{g,h|d,e} is the probability of proteins *g* and *h* interacting due to domains *d* and *e*; this is the DDI of *d*, *e*. Pr{d,e|g,h} is the fraction of interactions where one protein contains domain *d* and the other contains domain *e*. Pr{I(g,h)} represents the prior probability that a pair of proteins interact; this is the frequency of interactions in the interactome. Finally, Pr{D(g,d),D(h,e)} is used for normalization as the probability that if we choose two proteins, one will contain domain *d* and the other domain *e*. After all DDIs and DMIs were calculated, both methods were applied to score every possible pair between the *S. lycopersicum* interactome and *P. infestans* secretome proteins.

**Co-localization filter**

To reduce the number of candidate interactions, a filter for subcellular location was used, requiring both proteins of the pair to co-localize in the same subcellular compartment. For this task, the software LOCALIZER (v1.0.4) was used [44]. LOCALIZER predicts the location inside the plant cell of both plant and effector proteins by searching for peptides that contain transit or signal sequences. The algorithm classifies each protein as nucleus, chloroplast, mitochondria or none. If proteins in a pair are predicted to be in different compartments, they are removed. Then, a first subset is built with proteins pairs that were both localized in nucleus, mitochondria or chloroplast. In order to obtain a less restrictive subset, we assumed that proteins classified as none are found in the cytoplasm; a second subset was constructed by adding pairs of proteins found in the cytoplasm to the first.

**Evaluation of predictions**

After these filters, a combined PPI score was calculated as the sum of the Dyer and Zhang scores. We evaluated the interactions using different approaches described below. Approaches 2, 3, 4 and 5 were performed only on the 30 highest scoring interactions, in order to have a better reliability and reduce the complexity of the results.

**1) Comparison to known interactions**

The few known PPI interactions in this pathosystem are reviewed in Whisson et al. 2016 [13]. We manually retrieved the protein sequences of 10 described interactions from the Ensembl and Uniprot databases. In some cases, the exact protein was not found or was not described properly in the paper. For these cases, we manually selected the most similar paralogs and proteins of the same family for *P. infestans* proteins, identified by Uniprot BLAST. Also, some *S. lycopersicum* proteins had to be retrieved as orthologs of the described protein in other organisms such as *Nicotiana benthamiana* and *Solanum tuberosum.* Finally, we obtained a list of 42 interactions (Table S1). In order to test our predictor, the Dyer and Zhang methods were run over the proteins forming these interactions.

**2) Proximity in host PPI network**

Dyer et al. [35] propose that in the host, proteins involved in the same processes usually are part of the same interacting complexes. Therefore, they will be close to each other in the host PPI network. A pathogen effector that targets several host proteins, is likely to target proteins with similar functions. For each pathogen effector with more than one predicted target, the significance of target colocation in the PPI network was estimated by repeatedly randomly selecting the same number of host proteins and computing the sum of distances between all target pairs. A P-value was obtained for each effector with several targets. Since the host PPI graph consists of multiple subgraphs (Figures S1-2), sometimes there was no possible path between the targets. For these cases, a penalization distance was used, calculated as one plus the longest distance between two nodes among all the subgraphs in the PPI network.

**3) Gene correlation on common effectors**

Dyer et al. 2007 [35] proposed that proteins predicted to interact are likely to have similar expression patterns. In addition, the expression of effector genes targeting the same host protein are also likely to be correlated, since they would be playing a similar role on the host target. We used an RNA-Seq transcriptomics dataset of 25 samples of *P. infestans – S. lycoperosicum* infection timepoints (Judelson et al. unpublished), sampled every 4 hours over a period of 2 to 6 days post infection. Read counts were normalized by the DESeq method [45]. Then, in our subset of 30 best interactions, the Pearson correlation between the expression levels was calculated for each target-effector pair. For host proteins targeted by several pathogen effectors, the Pearson correlation was calculated for each effector-effector pair. Each correlation coefficient was tested for significance by calculating the correlations likewise for a set of 1000 random pairs of proteins from the protein.

**4) GO enrichment in host targets**

To investigate whether host targets are infection-related, the predicted *S. lycopersicum* targets were annotated with Gene Ontology (GO) terms. The STRING identifiers were translated to Uniprot identifiers to retrieve the associated GO terms for each host protein. “Cellular component” ontology terms were not considered since we already applied a filter for subcellular location. Then, for each GO

term found for the host predicted targets, a Fisher exact test was performed to test for enrichment compared to the total host interactome. We used an unadjusted P-value threshold of 0.1 to consider a GO term significantly enriched (Supplementary File 3). A list of infection-related GO terms was built using the list of 42 known interactions determined before [13]. The GO terms assigned to the 14 host targets were used as the first infection-related GO terms list. A second and more complete list was built adding the children of every GO term on the first list, since these are also likely to be infection-related (Supplementary File 3).
We counted the number of enriched GO terms that were classified as infection related as True Positives (TP); and those classified as non-infection related as False Positives (FP). As False Negatives (FN) and True Negatives (TN) we used the number of non-enriched GO terms that were infection and non-infection related respectively. Using this data, precision, recall and accuracy were calculated.

**5) Comparison of results with different parameters**
In order to compare the results found when different parameters are used in the pipeline defined in this project, we combine all the P-values from the previous tests using Fisher's omnibus method.

## RESULTS AND DISCUSSION

### The domain annotation is poor among effector proteins

The number of domains per protein is generally low, especially for *Phytophthora* proteins, more than half of them which no domain annotated (Figure 3).

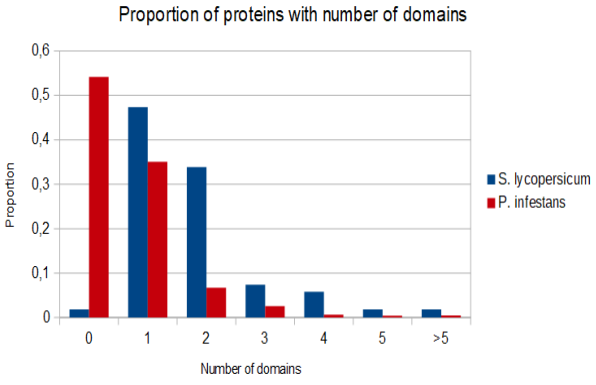


*Figure 3: Domain annotation of P. infestans secretome and S. lycopersicum interactome.*

This is likely due to the high mutation rate of *Phytophthora* effectors, leading to proteins with no significant homology with any other protein used for the domain annotation [16]. The low number of domains unfortunately prevented us to include many of these proteins in further analyses. We therefore expanded the annotation to protein motifs, which were expected to be found in higher number. In fact, the motif annotation yielded a very large number of motifs per protein, even after the use of the exposure and frequency filters (Figure 4).

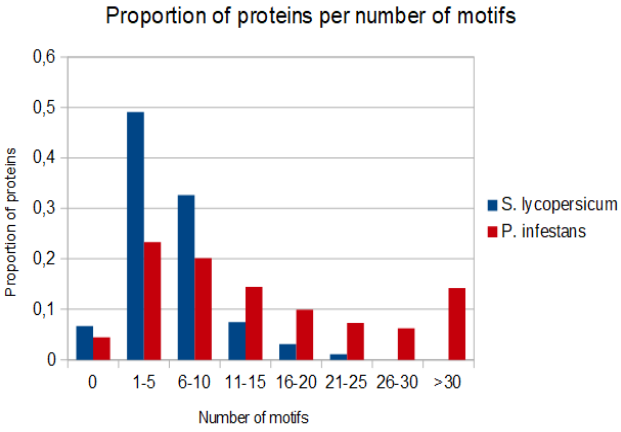


*Figure 4: Motif annotation of S. lycopersicum interactome with 5% of frequency filter; and P. infestans secretome with 10% of frequency filter.*

This use of motifs allowed us to include most of the proteins in our datasets, using domain-motif pairs as a new source of information for predicting PPIs.

### The Dyer and Zhang methods yield different DDI scores

The first step for both the Dyer and Zhang methods is calculating the PPI scores for each DDI. Figure 5 shows the scores for each DDI using only domains in the *Phytophthora and Solanum* pathosystem.

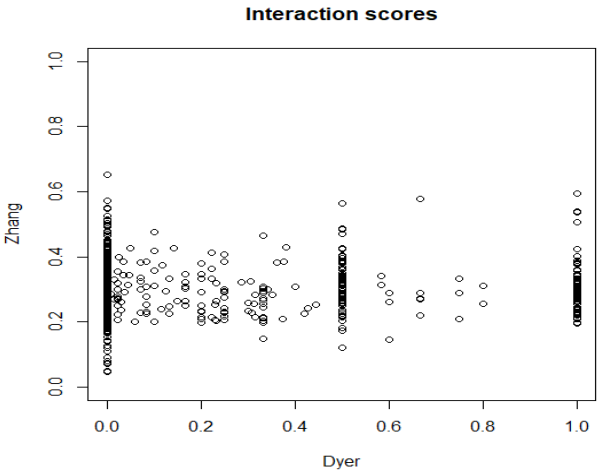


*Figure 5: Dyer and Zhang DDI scores for each pair of domains in our protein sets.*

It is clear that predictions by the two methods are not correlated. The same result is found when including motifs (Figure S3). This lack of correlation is also found when we use any of the DDI scores for the prediction of PPIs in our protein sets. Figure S4 shows that using motifs generally increases the scores of predicted PPIs, yielding scores for protein pairs that could not be found using only domains. It is likely that both methods lack accuracy and are therefore not able to capture and score all the actual DDIs underlying the PPIs. For this reason, we

selected the best predictions based on the sum of both scores, trying to complement one method with the other in order to obtain higher reliability.

**Subcellular location and score filters are set to obtain the most reliable subset of interactions**

The software LOCALIZER allowed us to filter the predicted PPIs based on subcellular location. By filtering for co-localizing proteins, we were able to further narrow down the set of candidate interactions. The predicted PPI datasets were reduced in similar proportions: 37.7% and 8.5%, depending on the in- or exclusion of pairs considered in the cytoplasm. Table 1 shows the number of protein pairs found before and after the co-localization filter.

*Table 1: Number of protein pairs before and after the co-localization filter.*

| *Solanum-Phytophthora* | Original | Localizer | Localizer + cytoplasm |
|---|---|---|---|
| Domains | 555120 | 47087 (8.5%) | 209168 (37.7%) |
| Domains + Motifs | 1193500 | 104150 (8.7%) | 451502 (37.8%) |

Table S2 shows the classification of proteins by LOCALIZER in different compartments and Figure S5 shows the different PPIs predicted after the filtering. After the interactions were filtered for co-location, the best subset of 30 interactions was chosen for each set by filtering on the sum of Dyer and Zhang scores. As a result, 4 sets of 30 interactions are obtained: domain annotation (D), domain annotation with cytoplasm compartment (DC), domain and motif annotation (DM) and domain and motif annotation with cytoplasm compartment (DMC). These 30 interactions and their predicted locations are listed in Table S3. Figures S1-2 show the predicted interactions plotted over the host PPI network. Figure 6 shows that the overlap between the four sets is limited.

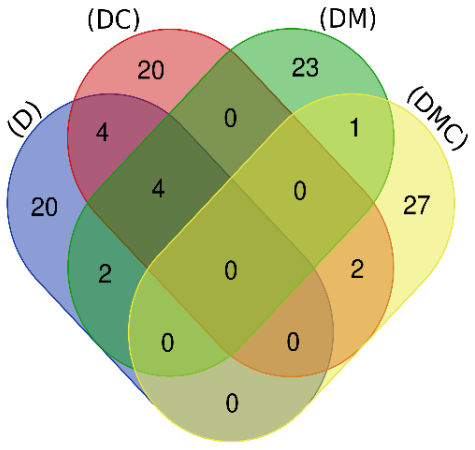


*Figure 6: Overlap of the 30 best interactions, after application of different parameters.*

This implies firstly that the addition of motifs yields PPIs that were not found just with domain information, and secondly, that most of the best PPIs predicted are localized in the cytoplasm (Table S3).

**Scores for known interactions are zero**

A complete and accurate predictor should be able to assign a higher score to known interactions than to most protein pairs. However, all 42 known interactions had a score of 0 for all of the analyses, likely a result of the poor domain annotation for these proteins (Figure 7).

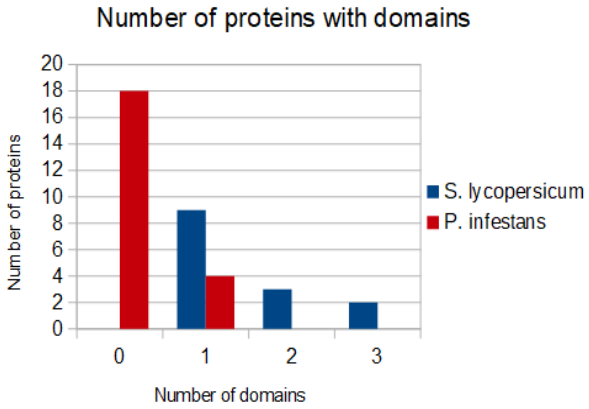


*Figure 7: Domain annotation of P. infestans and S. lycopersicum proteins known to interact.*

The only domain found in some *Phytophthora* effectors was RXLR (PF16810), with no DDI or DMI for neither Dyer nor Zhang. This made it impossible for the predictor, working only with domains, to produce useful information. According to the motif annotation, it seems like the motifs found in these proteins do not yield any DMI that justify the interactions, showing that the additional motif annotation has not been useful in this case (Figure S6).

**Host proteins targeted by the same effector often participate in same processes**

The analysis on the significance of the proximity of targets in the host PPI network was performed over the 4 result sets. For each effector with several host targets, a P-value was calculated for the proximity of these host targets in the host PPI network. Table S4, illustrates these P-values; Table 2 gives the combined P-value for each result.

*Table 2: Fisher combined P-values for proximity of targets evaluation.*

| | D | DC | DM | DMC |
|---|---|---|---|---|
| Fisher P-value | 0.0131 | 0.0107 | 0.025 | 0.854 |

The results show how both sets that use only domains (D, DC) have similar low P-values. The P-value for the set that used motifs (DM) is higher but still significant; for the last set (DCM) the targets are clearly not significantly close. According to these results, running the methods only with protein domains yield more reliable predictions, comparing to also using motifs.

**Gene expression is correlated for predicted interacting proteins and effectors that target the same host proteins**

For each interaction, we evaluated the correlation significance of gene expression associated to the interacting proteins and effectors that target the same host protein (Supplementary File 2). Table 3 shows the Fisher combined P-value for each of the 4 result sets.

*Table 3: Fisher combined P-values for evaluation of gene expression correlation.*

| *Fisher P-Value* | *D* | *DC* | *DM* | *DMC* |
|---|---|---|---|---|
| *Target-Effector* | $4.76x10^{-7}$ | $8.79x10^{-4}$ | *0.437* | *0.748* |
| *Effector-Effector* | $6.0x10^{-6}$ | $9.64x10^{-3}$ | *0.0603* | $1.29x10^{-7}$ |

For both sets working only with domains (D, DC) there is a significant correlation of gene expression between both the interacting proteins, and between the effectors that target the same host protein, especially in the case of set D. On the other side, interacting proteins in sets DM and DMC are mostly non-correlated. However, it seems like some correlation is found between the effectors targeting the same host proteins. These results suggest that the sets D and DC contain more promising predictions than DM and DMC, and that the motif annotation introduces noise in the predictions. For the D and DC sets, the gene expression correlation is significant for many predicted interactions, marking them as promising. We have no satisfactory explanation for the low P-value obtained for effector pairs in the DMC set. Preliminary tests show that effectors found in this set are larger than the average secreted protein, and therefore they generally have more annotated domains and motifs. Also, some of these domains are related to glycosidases, so we speculate that these effectors may be apoplastic proteins involved in the process of cell wall degradation, which may explain their correlated gene expression. Nevertheless, further study of these effectors is needed to confirm this hypothesis.

**GO terms related to infection are found enriched in the predicted targets**

For each of the 4 result sets, Supplementary File 3 shows the GO terms enriched in our host targets comparing to the complete set of proteins in the host interactome; and indicates whether these were found in the sets of infection related GO terms. For each set, the precision, recall and accuracy are also shown in Table 4.

*Table 4: Statistics for detection of infection related terms*

| | | D | DC | DM | DMC |
|---|---|---|---|---|---|
| Infection related | Precision | 0.091 | 0.179 | 0.139 | 0 |
| | Recall | 0.154 | 0.269 | 0.192 | 0 |
| | Accuracy | 0.933 | 0.945 | 0.944 | 0.946 |
| Infection related + children | Precision | 0.295 | 0.231 | 0.222 | 0.125 |
| | Recall | 0.07 | 0.049 | 0.043 | 0.016 |
| | Accuracy | 0.78 | 0.777 | 0.778 | 0.78 |

Relatively high accuracies contrast with low precision and recalls which are, at least in part, explained because of the incomplete list of infection related GO terms. This list only contains the GO terms of the 14 known targets of *P. infestans.* Moreover, the GO annotation of these proteins may not be complete and, as the list contains homologs, many GO terms are repeated. The actual number of effector targets is likely to be far higher, so many infection related GO terms are not included in this list. Furthermore, the list is still lacking processes related with the infection, for example GO:0009709 (detection of brassinosteroid stimulus) is classified as being unrelated to infection, yet is highly enriched in D, DC and DM sets. Strikingly, some studies suggest that the brassinosteroid signalling pathway is related with pathogen infections [46], specifically with *Phytophthora* infection [47]. This example together with other GO terms incorrectly classified as not infection-related, like GO:0001578 (microtubule bundle formation) and GO:0060548 (negative regulation of cell death) [48, 49], mean that performance is likely underestimated.

## CONCLUSION

Predicting Protein-Protein interactions is a difficult task in bioinformatics. Several methods have been described and implemented in an attempt to solve this problem and facilitate the experimental determination of these interactions. However, their performance is limited and in general they are based on indirect evidences. Moreover, experimental determination of interactions between proteins is slow, expensive and usually not reliable enough.

These difficulties lead to a lack of reliable protein interaction data. Only model organisms such as *Arabidopsis thaliana* or *Sacharomyces cerevisiae* have a reasonable number of determined interactions. The situation is even worse for inter-species PPIs, e.g. in the *P. infestans* – *S. lycopersicum* pathosystem only 10 interactions have been determined. This is insufficient to train a predictor, as it is obviously just a small subset of the total actual interactions and therefore not representative for the biological interaction.

This lack of data is characteristic for fields where science is still starting to progress. Conversely, the study of PPIs is a promising field where many discoveries and new bioinformatics tools are to be

developed in order to unravel the truth about PPIs. The more we know, the better we will understand how these interactions work and the easier it will be to predict them. It is a slow, but constant process that will eventually improve our knowledge.

In this specific project, it was not possible to apply structural homology and sequence homology approaches since effector proteins of *P. infestans* are small, have a high mutation rate and therefore very little homology with proteins in other (model) organism. For this reason, we worked on a smaller level, not in complete proteins, but looking for conserved fragments, domains or even motifs. With this domain and motif annotation we were able to implement the Dyer and Zhang methods to make PPI predictions. The filters that have been described were applied to improve the results, reducing noise and unlikely interactions in order to get more reliable predictions.

Although our methods were unable to predict the small subset of known interactions, the subsequent evaluations that were carried out suggest that our predictions are not random, and it might therefore be worth to study them in more detail, if possible perhaps even verify them in the lab.

The use of motifs for complementation of the limited domain annotations, gave worse results than when using only domains. Motif annotation allowed these methods to work with proteins without domains, but at the same time introduced noise that could not be corrected for by the exposure and frequency filters. On the other hand, those PPIs that were predicted using only domains yielded promising results for all our evaluations: host targets predicted to interact with the same effector were close in the host PPI network, a result that suggests their role in similar processes; significantly correlated gene expression was found in predicted interacting proteins and also in effectors targeting the same host protein, which indicates that these pairs might be working at the same time and therefore potentially in the same processes; and finally, many of the enriched GO terms found in the predicted targets were likely infection related. Since we could not find an objective way to determine a complete list of infection related GO terms, the results of these evaluations are somewhat subjective. Yet we found indications in literature that suggest the predicted host targets could indeed be related to infection. These evaluations do not prove that the predicted interactions are real, but they make biological sense and are *a priori* promising.

This project is the next step on the research of PPIs between *P. infestans* and its host *S. lycopersicum* and may be helpful for further studies of these pathosystem.

**ACKNOWLEDGEMENTS**

I express my gratitude to my supervisors Dick de Ridder and Sander Rodenburg for their considerable feedback and support during this project; but also, for allowing me to work with independence and helping me to develop my own ideas.

**REFERENCES**

[1] A. J. Haverkort, P. C. Struik, R. G. F. Visser, and E. Jacobsen, “Applied biotechnology to combat late blight in potato caused by phytophthora infestans,” *Potato Res.*, vol. 52, no. 3, pp. 249–264, 2009.

[2] M. Latijnhouwers, P. J. G. M. De Wit, and F. Govers, “Oomycetes and fungi: Similar weaponry to attack plants,” *Trends Microbiol.*, vol. 11, no. 10, pp. 462–469, 2003.

[3] B. M. Tyler *et al.*, “Phytophthora Genome Sequences Uncover Evolutionary Origins and Mechanisms of Pathogenesis,” no. September, pp. 1261–1267, 2006.

[4] S. Baldauf, “The Deep Roots of Eucaryotes,” *Nature*, vol. 300, no. 9, pp. 1703–1706, 2003.

[5] W. E. Fry, “Historical and Recent Migrations of *Phytophthora infestans* : Chronology, Pathways, and Implications,” *Plant Dis.*, vol. 77, no. 7, p. 653, 1993.

[6] W. E. Fry *et al.*, “Five Reasons to Consider *Phytophthora infestans* a Reemerging Pathogen,” *Phytopathology*, vol. 105, no. 7, pp. 966–981, 2015.

[7] S.-K. Oh *et al.*, “In Planta Expression Screens of Phytophthora infestans RXLR Effectors Reveal Diverse Phenotypes, Including Activation of the Solanum bulbocastanum Disease Resistance Protein Rpi-blb2,” *Plant Cell Online*, vol. 21, no. 9, pp. 2928–2947, 2009.

[8] J. Win *et al.*, “Adaptive Evolution Has Targeted the C-Terminal Domain of the RXLR Effectors of Plant Pathogenic Oomycetes © American Society of Plant Biologists Adaptive Evolution Has Targeted the C-Terminal Domain of the RXLR Effectors of Plant Pathogenic Oomycetes,” *Plant Cell*, vol. 19, no. August, pp. 2349–2369, 2007.

[9] a Levin, a Baider, E. Rubin, U. Gisi, and Y. Cohen, “Oospore Formation by Phytophthora infestans in Potato Tubers.,” *Phytopathology*, vol. 91, no. 6, pp. 579–85, 2001.

[10] W. Fry, “Phytophthora infestans: The plant (and R gene) destroyer,” *Mol. Plant Pathol.*, vol. 9, no. 3, pp. 385–402, 2008.

[11] S. Kamoun and C. D. Smart, “Late Blight of Potato and Tomato in the Genomics Era,”

*Plant Heal. Instr.*, vol. 89, no. 7, pp. 692–699, 2005.

[12] H. S. Judelson, “Metabolic Diversity and Novelties in the Oomycetes,” *Annu. Rev. Microbiol.*, vol. 71, no. 1, p. annurev-micro-090816-093609, 2017.

[13] S. C. Whisson, P. C. Boevink, S. Wang, and P. R. Birch, “The cell biology of late blight disease,” *Curr. Opin. Microbiol.*, vol. 34, pp. 127–135, 2016.

[14] S. Raffaele, J. Win, L. M. Cano, and S. Kamoun, “Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of *Phytophthora infestans*.,” *BMC Genomics*, vol. 11, p. 637, 2010.

[15] H. J. G. Meijer *et al.*, “Profiling the Secretome and Extracellular Proteome of the Potato Late Blight Pathogen *Phytophthora infestans*,” *Mol. Cell. Proteomics*, vol. 13, no. 8, pp. 2101–2113, 2014.

[16] J. Mcgowan and D. A. Fitzpatrick, “Genomic, Network, and Phylogenetic Analysis of the Oomycete Effector Arsenal,” vol. 2, no. 6, pp. 1–22, 2017.

[17] T. Kanneganti, E. Huitema, C. Cakir, and S. Kamoun, “Synergistic Interactions of the Plant Cell Death Pathways Induced by Phytophthora infestans Nep1-Like Protein PiNPP1 . 1 and INF1 Elicitin,” *Mol. Plant-Microbe Interact.*, vol. 19, no. 8, pp. 854–863, 2006.

[18] B. J. Haas *et al.*, “Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans,” *Nature*, vol. 461, no. 7262, pp. 393–398, 2009.

[19] S. C. Whisson, R. R. Vetukuri, A. O. Avrova, and C. Dixelius, “Can silencing of transposons contribute to variation in effector gene expression in Phytophthora infestans ?,” *Mob. Genet. Elements*, vol. 2, no. 2, pp. 110–114, 2012.

[20] T. Song *et al.*, “An Oomycete CRN Effector Reprograms Expression of Plant HSP Genes by Targeting their Promoters,” *PLoS Pathog.*, vol. 11, no. 12, pp. 1–30, 2015.

[21] R. Stam *et al.*, “Identification and Characterisation CRN Effectors in Phytophthora capsici Shows Modularity and Functional Diversity,” *PLoS One*, vol. 8, no. 3, pp. 1–13, 2013.

[22] M. van Damme *et al.*, “The Irish Potato Famine Pathogen Phytophthora infestans Translocates the CRN8 Kinase into Host Plant Cells,” *PLoS Pathog.*, vol. 8, no. 8, 2012.

[23] J. D. G. Jones and J. L. Dangl, “The plant immune system,” *Nature*, vol. 444, no. 7117, pp. 323–329, 2006.

[24] L. Pritchard and P. R. J. Birch, “The zigzag model of plant-microbe interactions: Is it time to move on?,” *Mol. Plant Pathol.*, vol. 15, no. 9, pp. 865–870, 2014.

[25] D. E. Cook, C. H. Mesarich, and B. P. H. J. Thomma, “Understanding Plant Immunity as a Surveillance System to Detect Invasion,” *Annu. Rev. Phytopathol.*, vol. 53, no. 1, pp. 541–563, 2015.

[26] E. A. Van Der Biezen and J. D. G. Jones, “Plant disease-resistance proteins and the gene-for-gene concept,” *Trends Biochem. Sci.*, vol. 23, no. 12, pp. 454–456, 1998.

[27] R. A. L. van der Hoorn and S. Kamoun, “From Guard to Decoy: A New Model for Perception of Plant Pathogen Effectors,” *Plant Cell Online*, vol. 20, no. 8, pp. 2009–2017, 2008.

[28] M. F. Seidl, A. Schneider, F. Govers, and B. Snel, “A predicted functional gene network for the plant pathogen *Phytophthora infestans* as a framework for genomic biology,” *BMC Genomics*, vol. 14, p. 483, 2013.

[29] V. S. Rao, K. Srinivas, G. N. Sujini, and G. N. S. Kumar, “Protein-Protein Interaction Detection: Methods and Analysis,” *Int. J. Proteomics*, vol. 2014, no. ii, pp. 1–12, 2014.

[30] E. Nourani, F. Khunjush, and S. Durmus, “Computational approaches for prediction of pathogen-host protein-protein interactions,” *Front. Microbiol.*, vol. 6, no. FEB, pp. 1–10, 2015.

[31] T. Cui, W. Li, L. Liu, Q. Huang, and Z. G. He, “Uncovering new pathogenhost proteinprotein interactions by pairwise structure similarity,” *PLoS One*, vol. 11, no. 1, pp. 1–18, 2016.

[32] S. Wuchty, “Computational prediction of Host-Parasite protein interactions between P. falciparum and H. sapiens,” *PLoS One*, vol. 6, no. 11, 2011.

[33] J. Shen *et al.*, “Predicting protein-protein interactions based only on sequences information.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 11, pp. 4337–41, 2007.

[34] M. Dyer, T. M. Murali, and B. W. Sobral, “Supervised learning and prediction of physical interactions between human and HIV proteins,” *Infect. Genet. Evol.*, vol. 11, no. 5, pp. 917–923, 2011.

[35] M. D. Dyer, T. M. Murali, and B. W. Sobral, “Computational prediction of host-pathogen protein-protein interactions,” *Bioinformatics*, vol. 23, no. 13, pp. 159–166, 2007.

[36] X. Zhang, X. Jiao, J. Song, and S. Chang, “Prediction of human protein-protein interaction

by a domain-based approach," *J. Theor. Biol.*, vol. 396, pp. 144–153, 2016.

[37] R. D. Finn, B. L. Miller, J. Clements, and A. Bateman, "IPfam: A database of protein family and domain interactions found in the Protein Data Bank," *Nucleic Acids Res.*, vol. 42, no. D1, pp. 364–373, 2014.

[38] A. Zhang, L. He, and Y. Wang, "Prediction of GCRV virus-host protein interactome based on structural motif-domain interactions," *BMC Bioinformatics*, vol. 18, no. 1, pp. 1–13, 2017.

[39] D. Szklarczyk *et al.*, "The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D362–D368, 2017.

[40] P. Jones *et al.*, "InterProScan 5: Genome-scale protein function classification," *Bioinformatics*, vol. 30, no. 9, pp. 1236–1240, 2014.

[41] R. Mosca, A. Céol, A. Stein, R. Olivella, and P. Aloy, "3did: A catalog of domain-based interactions of known three-dimensional structure," *Nucleic Acids Res.*, vol. 42, no. D1, pp. 374–379, 2014.

[42] R. J. Weatheritt, P. Jehl, H. Dinkel, and T. J. Gibson, "iELM-a web server to explore short linear motif-mediated interactions," *Nucleic Acids Res.*, vol. 40, no. W1, pp. 364–369, 2012.

[43] B. Petersen, T. N. Petersen, P. Andersen, M. Nielsen, and C. Lundegaard, "A generic method for assignment of reliability scores applied to solvent accessibility predictions," *BMC Struct. Biol.*, vol. 9, pp. 1–10, 2009.

[44] J. Sperschneider *et al.*, "LOCALIZER: Subcellular localization prediction of both plant and effector proteins in the plant cell," *Sci. Rep.*, vol. 7, no. December 2016, pp. 1–14, 2017.

[45] S. Anders and W. Huber, "Differential expression analysis for sequence count data," *Genome Biol.*, vol. 11, no. 10, 2010.

[46] L. De Bruyne, M. Höfte, and D. De Vleesschauwer, "Connecting growth and defense: The emerging roles of brassinosteroids and gibberellins in plant innate immunity," *Mol. Plant*, vol. 7, no. 6, pp. 943–959, 2014.

[47] A. Chaparro-Garcia *et al.*, "The receptor-like kinase serk3/bak1 is required for basal resistance against the late blight pathogen Phytophthora infestans in Nicotiana benthamiana," *PLoS One*, vol. 6, no. 1, 2011.

[48] D. Cahill, J. Rookes, A. Michalczyk, K. McDonald, and A. Drake, "Microtubule dynamics in compatible and incompatible interactions of soybean hypocotyl cells with Phytophthora sojae," *Plant Pathol.*, vol. 51, no. 5, pp. 629–640, 2002.

[49] D. Takemoto, A. R. Hardham, and D. A. Jones, "Differences in Cell Death Induction by Phytophthora Elicitins Are Determined by Signal Components Downstream of MAP Kinase Kinase in Different Species of Nicotiana and Cultivars of Brassica rapa and Raphanus sativus," *Plant Physiol.*, vol. 138, no. 3, pp. 1491–1504, 2005.

**SUPPLEMENTARY DATA**

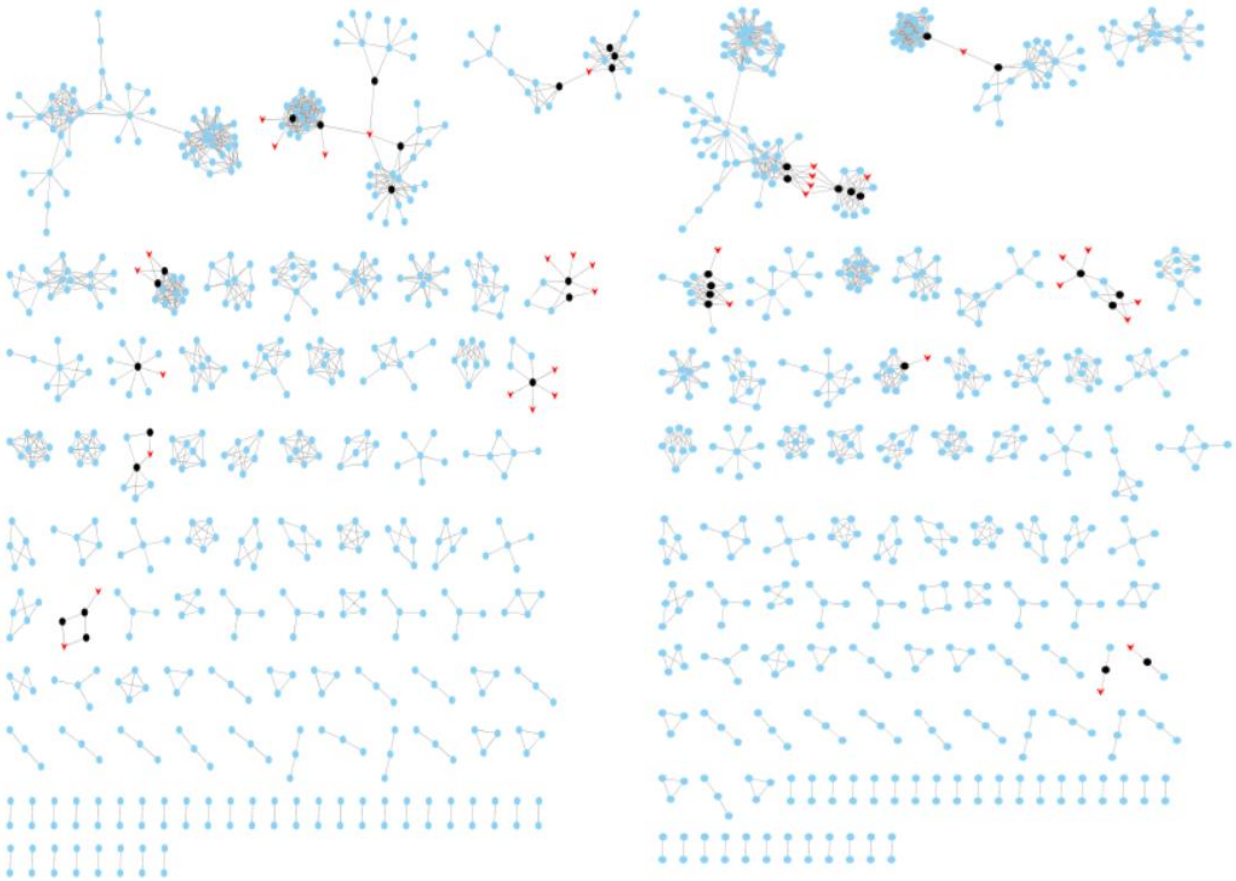

*Figure S1: Best 30 predicted interactions for the S. lycopersicum – P. infestans pathosystem using only domains in the pipeline. Left: D set (not including cytoplasm). Right: DC set (including cytoplasm). Pathogen effectors are plotted as red triangles, host targets as black circles and the rest of host proteins as blue circles.*

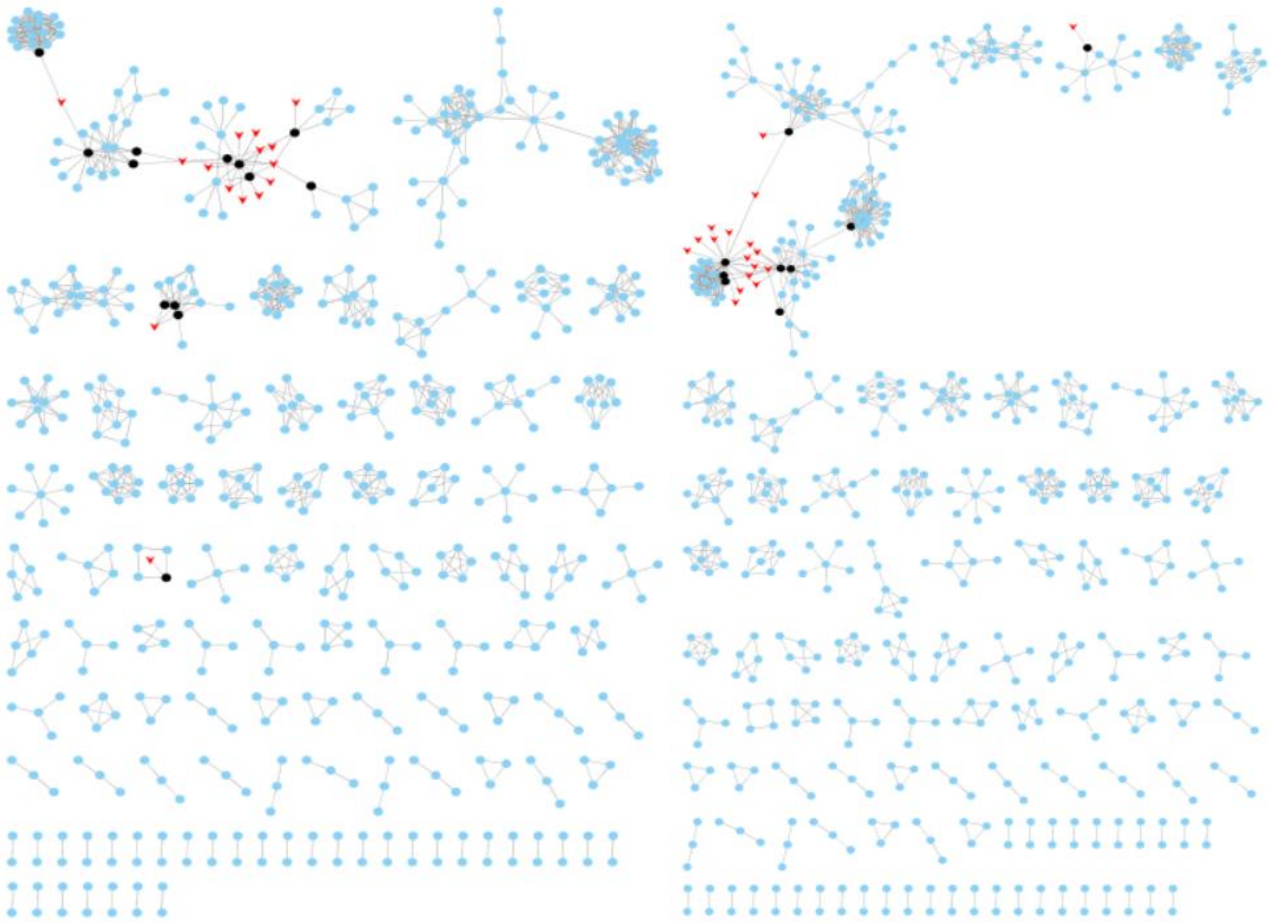

*Figure S2: Best 30 predicted interactions for the S. lycopersicum – P. infestans pathosystem using also motifs in the pipeline. Left: DM set (not including cytoplasm). Right: DMC set (including cytoplasm). Pathogen effectors are plotted as red triangles, host targets as black circles and the rest of host proteins as blue circles.*

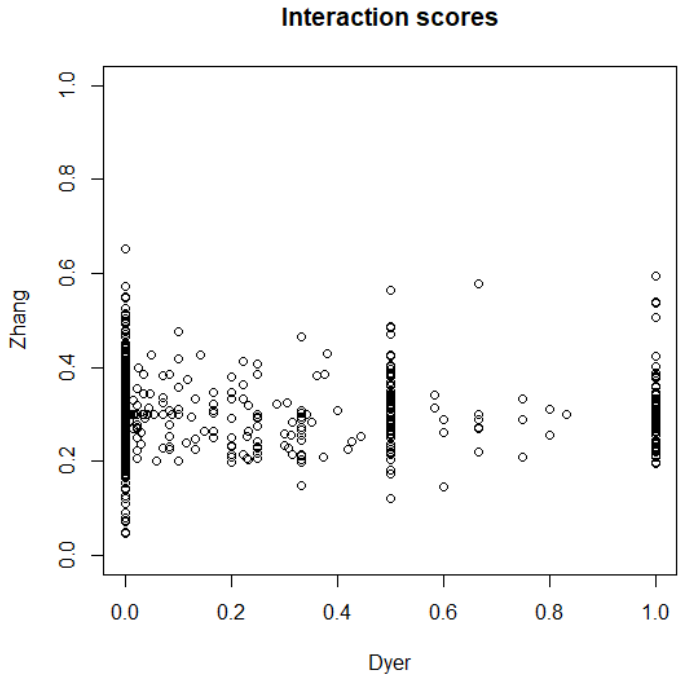


*Figure S3: Dyer and Zhang DDI scores for each pair of domains, including motifs, in our protein sets.*

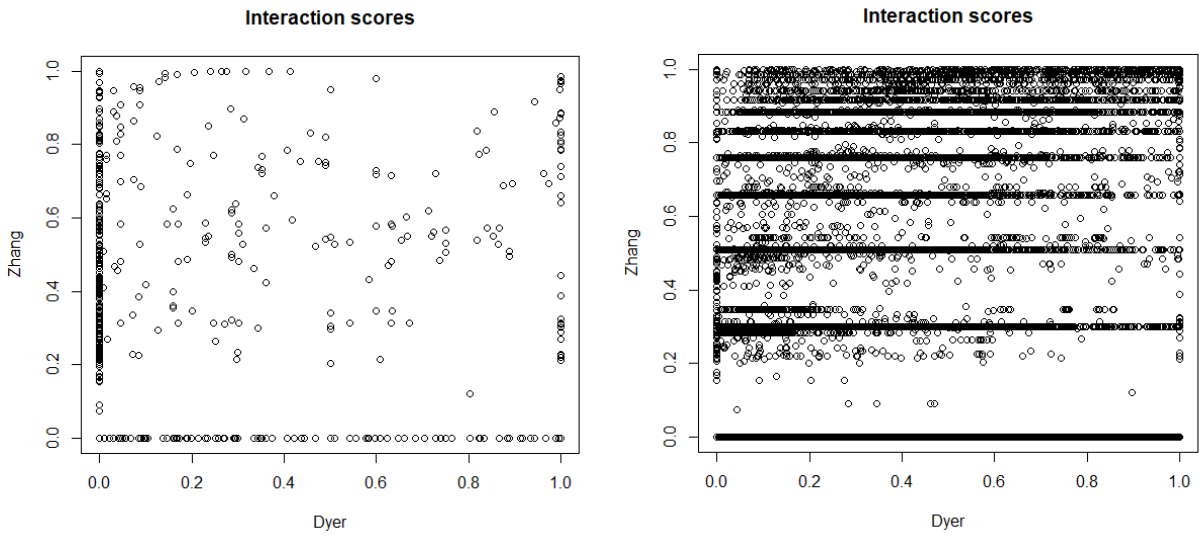


*Figure S4: Dyer and Zhang PPI scores for each pair of proteins in our protein sets.*
*Left: Using only domains. Right: Including motifs.*

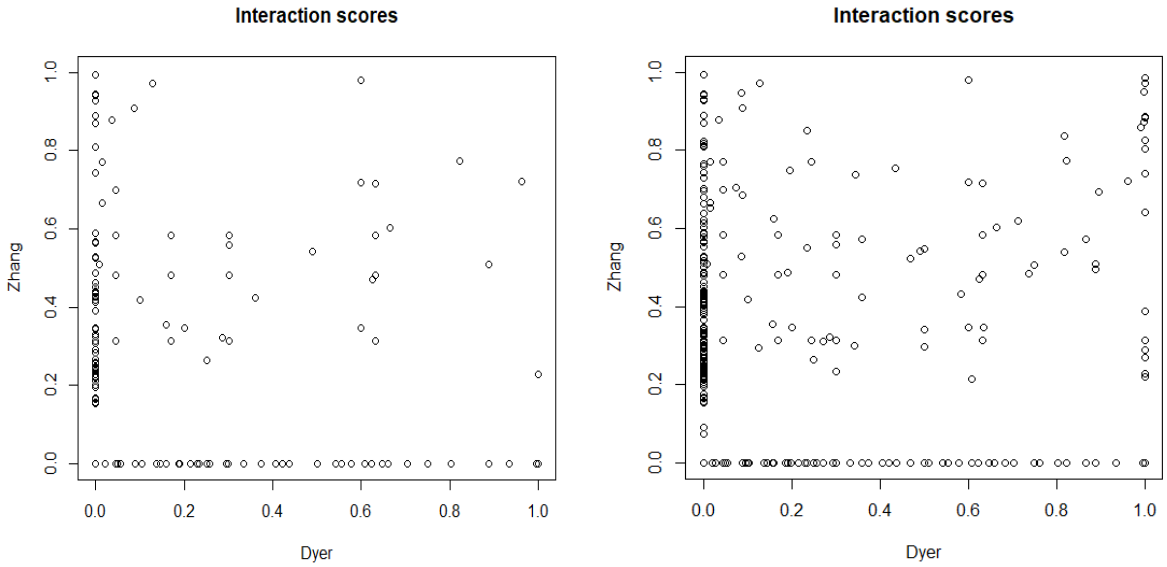


*Figure S5: Dyer and Zhang PPI scores for each pair of proteins in our protein set using only domains (Figure S4) after being filtered using LOCALIZER. Left: Without cytoplasm compartment; Right: With cytoplasm compartment.*

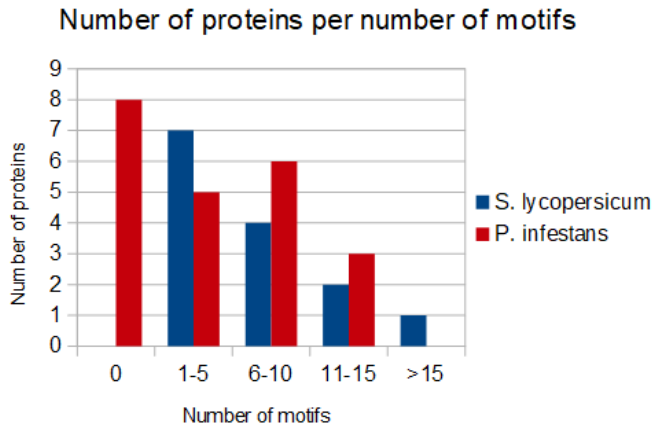


*Figure S6: Motif annotation of proteins known to interact from P. infestans with 10% of frequency filter; and S. lycopersicum with 10% frequency filter.*

*Table S 1: Known interactions of S. lycopersicum – P. infestans proteins [15]*

| *Solanum* | *Phytophthora* | *Solanum* | *Phytophthora* |
|---|---|---|---|
| 4081.Solyc01g005160.2.1 | PITG_14368 | 4081.Solyc05g007350.1.1 | PITG_06432 |
| 4081.Solyc01g005160.2.1 | PITG_14371 | 4081.Solyc05g007350.1.1 | PITG_16663 |
| 4081.Solyc01g005160.2.1 | PITG_14374 | 4081.Solyc06g005170.2.1 | PITG_11350 |
| 4081.Solyc01g005160.2.1 | PITG_23016 | 4081.Solyc06g005170.2.1 | PITG_11383 |
| 4081.Solyc01g090480.2.1 | PITG_04089 | 4081.Solyc06g005170.2.1 | PITG_11384 |
| 4081.Solyc01g105310.2.1 | PITG_04085 | 4081.Solyc06g005170.2.1 | PITG_21422 |
| 4081.Solyc01g105310.2.1 | PITG_04086 | 4081.Solyc06g005170.2.1 | PITG_22935 |
| 4081.Solyc01g105310.2.1 | PITG_18683 | 4081.Solyc07g063420.2 | PITG_03192 |
| 4081.Solyc01g105310.2.1 | PITG_20300 | 4081.Solyc07g063420.2 | PITG_21238 |
| 4081.Solyc01g105310.2.1 | PITG_20301 | 4081.Solyc11g071430.1.1 | PITG_14368 |
| 4081.Solyc01g105310.2.1 | PITG_20303 | 4081.Solyc11g071430.1.1 | PITG_14371 |
| 4081.Solyc02g064720.2 | PITG_00366 | 4081.Solyc11g071430.1.1 | PITG_14374 |
| 4081.Solyc02g064720.2 | PITG_02860 | 4081.Solyc11g071430.1.1 | PITG_23016 |
| 4081.Solyc02g092560.2 | PITG_00366 | 4081.Solyc12g056520.1.1 | PITG_06432 |
| 4081.Solyc02g092560.2 | PITG_02860 | 4081.Solyc12g056520.1.1 | PITG_16663 |
| 4081.Solyc03g080090.2.1 | PITG_03192 | 4081.Solyc12g088670.1.1 | PITG_04085 |
| 4081.Solyc03g080090.2.1 | PITG_21238 | 4081.Solyc12g088670.1.1 | PITG_04086 |
| 4081.Solyc04g071350.2.1 | PITG_06432 | 4081.Solyc12g088670.1.1 | PITG_18683 |
| 4081.Solyc04g071350.2.1 | PITG_16663 | 4081.Solyc12g088670.1.1 | PITG_20300 |
| 4081.Solyc04g072220.2.1 | PITG_03192 | 4081.Solyc12g088670.1.1 | PITG_20301 |
| 4081.Solyc04g072220.2.1 | PITG_21238 | 4081.Solyc12g088670.1.1 | PITG_20303 |

*Table S2: Classification of proteins in different compartments by LOCALIZER.*

| LOCALIZER | Mitochondria | Chloroplast | Nucleus | Cytoplasm (none) |
|---|---|---|---|---|
| *Phytophthora* | 155 | 140 | 484 | 1463 |
| *Solanum* | 50 | 68 | 187 | 247 |

*Table S3: Best 30 interactions for each of the 4 sets of S. lycopersicum – P. infestans pathosystem.*

| Domains (D) | Domains (cytoplasm) (DC) | Domains+Motifs (DM) | Domains+Motifs (cytoplasm) (DMC) |
|---|---|---|---|
| Solanum - Phytophthora - Location | Solanum - Phytophthora - Location | Solanum - Phytophthora - Location | Solanum - Phytophthora - Location |
| 4081.Solyc01g088020.2.1 - PITG_02427 - nucleus | 4081.Solyc01g067490.2.1 - PITG_00983 - cytoplasm | 4081.Solyc01g088020.2.1 - PITG_02427 - nucleus | 4081.Solyc02g083940.2.1 - PITG_04980 - cytoplasm |
| 4081.Solyc01g088700.2.1 - PITG_02664 - nucleus | 4081.Solyc01g067490.2.1 - PITG_08048 - cytoplasm | 4081.Solyc01g088700.2.1 - PITG_02664 - nucleus | 4081.Solyc02g087300.1.1 - PITG_04980 - cytoplasm |
| 4081.Solyc02g070890.2.1 - PITG_05338 - nucleus | 4081.Solyc01g067490.2.1 - PITG_09585 - cytoplasm | 4081.Solyc03g007100.2.1 - PITG_02427 - nucleus | 4081.Solyc02g087300.1.1 - PITG_05502 - cytoplasm |
| 4081.Solyc03g007100.2.1 - PITG_02427 - nucleus | 4081.Solyc01g067490.2.1 - PITG_17982 - cytoplasm | 4081.Solyc03g007610.2.1 - PITG_19099 - chloroplast | 4081.Solyc02g087300.1.1 - PITG_09457 - cytoplasm |
| 4081.Solyc03g007100.2.1 - PITG_14123 - nucleus | 4081.Solyc01g088700.2.1 - PITG_02664 - nucleus | 4081.Solyc04g012170.2.1 - PITG_03335 - nucleus | 4081.Solyc02g087300.1.1 - PITG_16950 - cytoplasm |
| 4081.Solyc03g083440.2.1 - PITG_03694 - nucleus | 4081.Solyc02g070890.2.1 - PITG_05338 - nucleus | 4081.Solyc04g012170.2.1 - PITG_06804 - nucleus | 4081.Solyc02g087300.1.1 - PITG_19889 - cytoplasm |
| 4081.Solyc03g117120.2.1 - PITG_02427 - nucleus | 4081.Solyc02g070910.1.1 - PITG_11236 - cytoplasm | 4081.Solyc04g012170.2.1 - PITG_07302 - nucleus | 4081.Solyc02g087300.1.1 - PITG_20529 - cytoplasm |
| 4081.Solyc04g009950.2.1 - PITG_14108 - nucleus | 4081.Solyc02g070910.1.1 - PITG_11728 - cytoplasm | 4081.Solyc04g012170.2.1 - PITG_14108 - nucleus | 4081.Solyc02g087300.1.1 - PITG_23035 - cytoplasm |
| 4081.Solyc04g049070.2.1 - PITG_14939 - nucleus | 4081.Solyc02g070910.1.1 - PITG_17832 - cytoplasm | 4081.Solyc04g051510.1.1 - PITG_03335 - nucleus | 4081.Solyc03g120900.1.1 - PITG_04980 - cytoplasm |

| | | | |
|---|---|---|---|
| 4081.Solyc04g049770.2.1 - PITG_02664 - nucleus | 4081.Solyc03g007100.2.1 - PITG_02427 - nucleus | 4081.Solyc04g051510.1.1 - PITG_06804 - nucleus | 4081.Solyc07g066060.2.1 - PITG_06804 - nucleus |
| 4081.Solyc04g051510.1.1 - PITG_05338 - nucleus | 4081.Solyc03g007670.2.1 - PITG_00983 - cytoplasm | 4081.Solyc04g051510.1.1 - PITG_19041 - nucleus | 4081.Solyc08g067040.2.1 - PITG_00983 - cytoplasm |
| 4081.Solyc04g051510.1.1 - PITG_09665 - nucleus | 4081.Solyc03g007670.2.1 - PITG_05498 - cytoplasm | 4081.Solyc05g055690.1.1 - PITG_14108 - nucleus | 4081.Solyc08g067040.2.1 - PITG_02728 - cytoplasm |
| 4081.Solyc04g051510.1.1 - PITG_11752 - nucleus | 4081.Solyc03g007670.2.1 - PITG_08048 - cytoplasm | 4081.Solyc06g084210.2.1 - PITG_14108 - nucleus | 4081.Solyc08g067040.2.1 - PITG_04980 - cytoplasm |
| 4081.Solyc04g051510.1.1 - PITG_19041 - nucleus | 4081.Solyc03g007670.2.1 - PITG_09585 - cytoplasm | 4081.Solyc07g040790.2.1 - PITG_03335 - nucleus | 4081.Solyc08g067040.2.1 - PITG_05502 - cytoplasm |
| 4081.Solyc05g056130.2.1 - PITG_16366 - nucleus | 4081.Solyc03g007670.2.1 - PITG_17982 - cytoplasm | 4081.Solyc07g040790.2.1 - PITG_06804 - nucleus | 4081.Solyc08g067040.2.1 - PITG_06741 - cytoplasm |
| 4081.Solyc05g056130.2.1 - PITG_21027 - nucleus | 4081.Solyc03g115050.2.1 - PITG_06222 - cytoplasm | 4081.Solyc07g040790.2.1 - PITG_07302 - nucleus | 4081.Solyc08g067040.2.1 - PITG_08536 - cytoplasm |
| 4081.Solyc07g049480.2.1 - PITG_04891 - chloroplast | 4081.Solyc03g117120.2.1 - PITG_02427 - nucleus | 4081.Solyc07g066060.2.1 - PITG_00353 - nucleus | 4081.Solyc08g067040.2.1 - PITG_09457 - cytoplasm |
| 4081.Solyc07g049480.2.1 - PITG_18396 - chloroplast | 4081.Solyc04g051510.1.1 - PITG_05338 - nucleus | 4081.Solyc07g066060.2.1 - PITG_03335 - nucleus | 4081.Solyc08g067040.2.1 - PITG_10517 - cytoplasm |
| 4081.Solyc07g066060.2.1 - PITG_02427 - nucleus | 4081.Solyc04g051510.1.1 - PITG_09665 - nucleus | 4081.Solyc07g066060.2.1 - PITG_06804 - nucleus | 4081.Solyc08g067040.2.1 - PITG_16950 - cytoplasm |
| 4081.Solyc08g077700.2.1 - PITG_02664 - nucleus | 4081.Solyc06g036420.1.1 - PITG_05498 - cytoplasm | 4081.Solyc07g066060.2.1 - PITG_07302 - nucleus | 4081.Solyc08g067040.2.1 - PITG_17575 - cytoplasm |
| 4081.Solyc09g008470.2.1 - PITG_14108 - nucleus | 4081.Solyc06g075340.2.1 - PITG_03462 - cytoplasm | 4081.Solyc07g066060.2.1 - PITG_12108 - nucleus | 4081.Solyc08g067040.2.1 - PITG_18117 - cytoplasm |
| 4081.Solyc10g008100.2.1 - PITG_10846 - nucleus | 4081.Solyc07g005810.2.1 - PITG_09792 - cytoplasm | 4081.Solyc07g066060.2.1 - PITG_14108 - nucleus | 4081.Solyc08g067040.2.1 - PITG_18397 - cytoplasm |
| 4081.Solyc11g065690.1.1 - PITG_00594 - nucleus | 4081.Solyc08g077700.2.1 - PITG_02664 - nucleus | 4081.Solyc07g066060.2.1 - PITG_17603 - nucleus | 4081.Solyc08g067040.2.1 - PITG_19889 - cytoplasm |
| 4081.Solyc11g065690.1.1 - PITG_01349 - nucleus | 4081.Solyc11g068400.1.1 - PITG_18776 - cytoplasm | 4081.Solyc07g066060.2.1 - PITG_17661 - nucleus | 4081.Solyc08g067040.2.1 - PITG_20529 - cytoplasm |
| 4081.Solyc11g065690.1.1 - PITG_02948 - nucleus | 4081.Solyc11g072290.1.1 - PITG_05498 - cytoplasm | 4081.Solyc07g066060.2.1 - PITG_18833 - nucleus | 4081.Solyc08g067040.2.1 - PITG_23035 - cytoplasm |
| 4081.Solyc11g065690.1.1 - PITG_03842 - nucleus | 4081.Solyc12g009960.1.1 - PITG_02664 - nucleus | 4081.Solyc07g066060.2.1 - PITG_18834 - nucleus | 4081.Solyc10g047000.1.1 - PITG_04980 - cytoplasm |
| 4081.Solyc12g008360.1.1 - PITG_16366 - nucleus | 4081.Solyc12g014320.1.1 - PITG_00983 - cytoplasm | 4081.Solyc07g066060.2.1 - PITG_22856 - nucleus | 4081.Solyc11g017070.1.1 - PITG_04980 - cytoplasm |
| 4081.Solyc12g008360.1.1 - PITG_21027 - nucleus | 4081.Solyc12g014320.1.1 - PITG_08048 - cytoplasm | 4081.Solyc08g077700.2.1 - PITG_02664 - nucleus | 4081.Solyc12g014320.1.1 - PITG_00983 - cytoplasm |
| 4081.Solyc12g009960.1.1 - PITG_02664 - nucleus | 4081.Solyc12g014320.1.1 - PITG_09585 - cytoplasm | 4081.Solyc11g005110.1.1 - PITG_06804 - nucleus | 4081.Solyc12g014320.1.1 - PITG_08048 - cytoplasm |
| 4081.Solyc12g099660.1.1 - PITG_10846 - nucleus | 4081.Solyc12g014320.1.1 - PITG_17982 - cytoplasm | 4081.Solyc12g009960.1.1 - PITG_02664 - nucleus | 4081.Solyc12g035360.1.1 - PITG_04980 - cytoplasm |

*Table S4: P-values for each effector with several targets in proximity of targets evaluation.*

| (D) | | (DC) | | (DM) | | (DMC) | |
|---|---|---|---|---|---|---|---|
| PITG_02427 | 0.186 | PITG_00983 | 0.075 | PITG_02427 | 1 | PITG_00983 | 1 |
| PITG_02664 | 0.008 | PITG_02427 | 1 | PITG_02664 | 0.005 | PITG_04980 | 0.006 |
| PITG_05338 | 0.031 | PITG_02664 | 0.005 | PITG_03335 | 0.213 | PITG_05502 | 1 |
| PITG_10846 | 0.031 | PITG_05338 | 0.031 | PITG_06804 | 0.328 | PITG_09457 | 1 |
| PITG_14108 | 0.035 | PITG_05498 | 0.001 | PITG_07302 | 0.122 | PITG_16950 | 1 |
| PITG_16366 | 0.022 | PITG_08048 | 0.07 | PITG_14108 | 0.206 | PITG_19889 | 1 |
| PITG_21027 | 0.034 | PITG_09585 | 0.087 | | | PITG_20529 | 1 |
| | | PITG_17982 | 0.07 | | | PITG_23035 | 1 |