Decision making in host-pathogen interactions: Exploring *Phytophthora infestans* regulatory network.

Castro[1], J.C., Valdés[1], I., Gonzalez[1], L.N., Cañas[1], S., Núster[2], C.E., Restrepo[1], S., and D.M. Riaño-Pachón[3].

1. Department of Biological Sciences, Universidad de los Andes, Colombia. 2. School of Agricultural Sciences, Universidad Nacional de Colombia. 3. Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE), Campinas, Brazil.

Abstract

With the advent of novel methods for inference and evaluation of regulatory functions, and the generation of massive transcriptomic and proteomic datasets, it is necessary to find a common framework to integrate both data and theoretical background to gain significative insight into biological problems. Transcriptional regulation is a major participant in complex cellular systems, where it plays a central role in decision making. We thus attempted to link expression data with mathematical models in order to infer gene regulatory networks assembled by transcription factor interactions. Using yeast expression data and transcription sub-networks we tested four different models for network inference. Among these we found "directed partial correlation" to be the one with the best performance. We then applied this method to rebuild the transcription network of the oomycete *Phytophthora infestans* using publicly available data. We used this network to analyze the changes in regulation during the life cycle of the pathogen and to assess how the topology of the network is related to the evolutionary history of the oomycetes. Additionally, using qRT-PCR we evaluated, the expression levels of 14 *P. infestans* transcription factors during its interaction with *Solanum tuberosum* group *phureja.* With these data we also reconstructed the regulatory network of *P. infestans* in two different varieties of the host, one resistant and one susceptible. Preliminary results of this network reconstruction showed rewiring according to the resistance level of the host, which indicates that different regulatory processes are activated in response to different environmental cues. This type of study can be used to assess issues in pathogens' biology like understanding features of complex life cycles.

Introduction

Novel approaches in the area of systems biology have made possible to study a myriad of features in systems from which available information is inherently incomplete or noisy [1]. Applying tools from control theory, simulation strategies and based always on experimental data, biological properties that were elusive and difficult to study in the past century (e.g. synergy, modularity, robustness), can be effectively addressed following a network approach [2–5]. The understanding of complex systems involves the study of many of these tools [6].

Pathogenesis has been considered as a complex system [7–9]. Therefore, we will consider a plant-oomycete interaction and through the use of networks, we will evaluate how regulatory events can mold pathogenic behavior. Among Oomycetes, *Phytophthora infestans* is considered a model, and has been considered as the most destructive pathogen of potato (*Solanum tuberosum*) crops, as the causing agent of the late blight disease [10–12]. The response of the host, *S. tuberosum*, can be modulated by several genes which in turn control large traits of resistance and susceptibility [13, 14]. However its inner dynamics are not fully understood. [11, 15]. The *P. infestans* – *S. tuberosum* group *phureja* interaction does not rely on a single and simple deterministic mechanism; on the contrary, the pathogen makes use of a complex web of elements used to overcome quantitative plant responses. Another characteristic of this oomycete's life cycle is the hemibiotrophy it displays. During the firsts days of infection it suppresses host defenses and feeds on the host behaving as a biotroph. Then, a destructive necrotrophic phase follows [16, 17], it is likely that the concerted action of different genes is then modulating this two phase phenomenon, however this is not yet understood.

To better understand these characteristics we will assess the effect of different environments, a resistant and a susceptible host, on the Gene Regulatory Network (GRN) of *P. infestans*, this network refers to the set of interactions among the different transcription

factors of *P. infestans*, which in turn are regulating gene expression levels. Transcriptome studies have shown that only a small fraction of genes of the *S. tuberosum* behave in a constant fashion in compatible and incompatible interactions, which supports the idea that different genes are expressed in each case [18]. Likewise, different processes should be modulated in the pathogen by different sets of genes. The study of GRNs on the other hand, cannot only enable the better understanding of the different types of gene regulation of this organism during its life cycle but in the evolutive time scale as well. Analyzing the gene expression levels and their change through time can tell us about the physiological characteristics of the organism, if we take it a step further, translating expression profiles into interaction networks, could link genetic information with phenotypic traits[19]. Also, understanding the relative time of appearance of the nodes in the network and how this affect network topology can serve as a tool for the formulation of evolutionary hypothesis [20–22]. Indeed, the nodes of the network, the genes, appeared at different time points in the evolutive history of the organism, and thus, if we study the GRN of any species, our representation of the network is a frame of a constantly changing process [22, 23].

Genome scale models of metabolism and protein-protein interaction networks have provided insights into some of these complex biological questions in an attempt to bridge the genotype-phenotype gap [20, 24–26]. GRNs determine the structure of transcriptional changes in a particular physiological state of an organism [27–30]. Therefore, the study of the transcription regulation plays an important role into addressing that bridge as well. Even more, transcription factors play key functions in the regulation of gene expression [31, 32]. Thus transcription regulation could be the key to bind genotypic information and phenotypic responses to environmental cues. As part of a changing life cycle, the regulatory events will control which proteins are expressed, thus determining cellular response to different situations, this process is in a sense analogous to decision making.

However, inference of these networks often requires large amounts of expression data as well as mutant and knock-out collections, which are usually sparse [27, 29, 33–35].

The strengths and weaknesses of these methods vary considerably with the amount of data, especially in data-poor scenarios [36]. The availability of longitudinal data (i.e. time series), has enabled the arise of methods that relate variables over time, particularly expression time series data has received focus over the last years [33, 37–39], and community efforts, such as the DREAM challenge, promote the research on this field. However, a master strategy is still lacking. Methods for network reconstruction over time series data have been proposed, and autoregressive models serve as a base for their construction [39–41]. The key assumption is the fact that the effect of a variable over another might have a shift in time, and this is the basis for the inference of regulatory events as causal relationships between genes. However these may not suffice for datasets with small sampling frequency, as the ability to effectively determine gene interactions and topological measures tend to be affected in short time series inference [42, 43].

We were interested in evaluating the effect of different network inference methods with time series data of different lengths. Advances in the study of GRNs of model organisms such as *Saccharomyces cerevisiae*, have created a benchmark for the evaluation of theoretical methods. We used this model as reference to evaluate different methods for GRN inference [36, 44]. The performance of a method depends on the size of the data and in the time series analysis, the frequency of the sampling has particular importance. Inference methods should balance computation efficiency with big data and accuracy with small datasets. After evaluating the methods on *S. cerevisae* we used the better in performance to determine the possible GRN of *P. infestans* in order to explore evolutive and physiological decision processes by the pathogen. The evaluation in a controlled scenario, allowed us to choose the best method to be applied to real datasets from *P. infestans*, first to publicly available microarray data and then to our own qPCR data. Our main objective was to approach the reconstruction of GRN that is active in the pathogen during the compatible interaction with two different varieties *S. tuberosum* group *phureja*, and associate this

network with changes in the life style as well with events on the evolutive history of the pathogen.

Methods

In silico expression data of Yeast

To address the prediction of regulatory relationships based on expression data, we used the software GeneNetWeaver [44] to sample 100 gene regulatory sub-networks from *Saccharomyces cerevisiae*, and generate time series expression data, this will be the **gold-standard** sub-networks that will be used later to evaluate the network inference methods. Each sub-network consisted of 200 nodes and an average of 350 edges (data not shown), expression data were generated using an Ordinary Differential Equation Model of transcriptional regulation [44, 45]. Expression data for each gene (node in the network) was *in silico* generated. Time series of 21 time points for each gene were generated, with a 50min step between time points. Time series data was evaluated by several methods to determine the capacity of these to accurately infer the correct sub-networks. Inference was carried out with the full time series as well with a sub-set of 6 time points with a step of 200 minutes.

The nodes of the network are the genes, and the regulation events are represented as edges. The number of connections a node has is the degree of the node, the number of routes that pass through a node is the betweenness and maximal sub-graph in which each node has degree K is the coreness [46].

Network inference from expression data

Transcription regulation is a process that inherently conveys a shift in time. For this reason we incorporated Granger causality [47] as a viable strategy to determine gene relationships. In this model a variable is said to Granger cause another, i.e., the expression of TF X Granger causes the expression of gene Y, if given for any two variables, it is possible to determine a significant correlation when one variable is shifted respect to the other [47,

48]. A multivariate extension has been proposed for systems controlled by more than one variable as follows [40, 49, 50]:

$$h(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n + \phi \quad (1)$$

Or in a vectorized form:

$$h(x) = \theta' X + \phi \quad (2)$$

We used equation 2 to estimate the possible relationship between any two variables. Where $h(x)$ represents a variable that is Granger caused by X, i.e., Y, θ represents the fit parameter vector and ϕ is the lag parameter. With this same framework, we used Mutual Information as the measure of relationship between the two variables, taking into account a shift in time. Mutual information has been proposed to capture non-linear interactions [41]. We defined shifted mutual information (SMI) as:

$$SMI(X;Y) = \sum_{y+\phi \in Y} \sum_{x \in X} p(x,y) \log_2 \left(\frac{p(x,y)}{p(x)p(y)} \right) \quad (3)$$

where $p(x,y)$ is the joint probability function of X and Y, and $p(x)$ and $p(y)$ are the marginal probability functions of X and Y respectively. In order to evaluate the significance of this measure, we took the time series data and carried out a randomization procedure, after this we recalculated the mutual information, randomized data served as a basis to determine a null distribution of the mutual information. Using this information we calculated the significance based on the number of appearances of the initial mutual information value in the null distribution, thus providing an empirical p-value.

These methods however, do not account for correlations due to a third partner which can result in a high number of false positives. For this reason, we assessed directed partial

correlation (DPC) [39] as a way to estimate if nodes interact based on their correlation, as well as considering the residual correlation of a third variable. This is particularly easy to compute in comparison with the previously mentioned methods, and has showed reliable results in large scale datasets. Finally we evaluated inner composition alignment (IOTA) [42, 43]. This method relies on the finding that related variables in time series data have a monotonically ascendant behavior one to another. We implemented IOTA as:

$$1 - \frac{\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\omega_{ij}\theta\left[\left(g_{j+l}^{(k,l)} - g_{i}^{(k,l)}\right)\left(g_{i}^{(k,l)} - g_{j}^{(k,l)}\right)\right]}{\Delta} \quad (4)$$

Where n is the length od the time series, $g^{(k,l)}$ is the rearranged permutation of the l time series respect to k, ω_{ij} represents a weighting function and θ is the Heaviside step function This model served us as a comparison method as it does not relies on an autoregressive procedure to estimate variable relationship.

Evaluation of predicted sub-networks from yeast.

We used the above mentioned methods to infer the *S. cerevisae* sub-networks and for each edge a *p*-value was provided values for each edge were adjusted using the Benjamini-Hochberg correction [51]. These values were used as a confidence score on the existence of an edge. We evaluated the performance of the predicted sub-networks estimating the Receiving Operating Characteristic (ROC) and Precision-Recall (PR) curves. To do this we used a variable cutoff threshold to determine the number of positive interactions and the number of false interactions. In the first iteration, if the adjusted/corrected confidence value between any pair of genes was 1, then we considered that there was an edge in the network between these genes. If the confidence was smaller than 1, then we considered that there was no edge. In subsequent iterations, the threshold was reduced, the step used was 0.01, in the ROC curve when the threshold reached zero all interactions were classified as positive, i.e., an edge, yielding the maximum true positive rate. All negative interactions were as well classified as positive yielding the maximum false

positive rate. With this information we calculated the area under the ROC (AUROC) and area under PR (AUPR) in each method for the 100 sub-networks, in the short and long time series. As the AUROC is equivalent to a Mann-Withney test, as so it is used to determine the aceptance threshold of positive and negative values of a predictor [52, 53], as these measurements showed variation over different estimations, we determined the mean AUROC to be the acceptance threshold of an edge. Additionally we evaluated the correlation of our prediction with the gold-standard network for two centrality measures K-coreness and betweenness, to determine the impact of the inference over the sub-network topology. The k-core of a graph is a maximal sub-graph in which each node has at least degree k. The coreness of a node is k if it belongs to the k-core but not to the (k+1)-core, the betweenness centrality is the number of short paths from all vertices that pass trough a given node. Thus we determined these measurements for each node in both networks, and compared the values.

In silico reconstruction of *Phytophthora infestans* regulatory network

We first used the expression data publicly available to reconstruct the *P. infestans* regulatory network. Data was derived from a custom GPL809 chip of the Roche Nimblegen technologies (Roche Nimblegen), to measure the expression levels of 18155 genes of the *P. infestans* genome. RNA samples were taken from 2 media (rye sucrose and V8 agar) and infected potato leaves from a time series infection (2, 3 ,4 and 5 days post inoculation) [54]. We considered the media samples to be the time 0 thus we builded a time series of 5 points.

We also reconstructed the GRN, using SMI and DPC, from the time series expression data of *P. infestans* in its compatible interaction with *S. tuberosum* group *phureja.* We estimated the topology measures of this networks and compared these to a random assembled network using the Erdos-Reyni model [2].

Network-Phylostratigraphic analysis of *Phytophthora* transcription factors

Using OrthoMCL [55, 56], with an inflation value of 2.5, we compared the *P. infestans* annotated genome to 19 genomes representative of the evolutionary relationships of the species, The genomes were assigned to a phylostratum depending on their phylogenetic distance to *P. infestans*, the more basal, older, groups have the lowest number of phylostratum and the closer, more recent, groups, have higher number [57, 58]. Phylostratum I contain the genomes of Bacteria and Archea groups, phylostratum II represents the branching from Eukarya of the other cellular organisms, III contains the organisms as defined in the SARP group [59], this group contains the groups Stramenophyla, Alveolata, Rhizaria and Plantae. The IV phylostratum represents the branching of the oomycetes from the other members of the SARA group. Finally V and VI are the *Phytophthora* genus and the species of interest: *P. infestans*. The phylostratum gives an idea of the age of the genes, and thus of the nodes in the GRN, and this could be related to network topology measurements.

Infection assays, RNA extraction and cDNA preparation

Col2 and Col3 cultivars of *S. tuberosum* group *phureja* were grown under greenhouse conditions (temperature 18ºC, 12 light hours and 60% relative humidity), Col2 is a susceptible variety whereas Col3 is highly resistant to late blight. Leaflets from 6-week old plants were collected and infected with *P. infestans* strain Z3-2 [60]. The strain was grown on Potato Dextrose Agar (PDA) at room temperature (21C on average), and an inoculum was prepared at a concentration of 4.0 $*10^5$ sporangia per ml as previously described [61]. Infection assays were performed in moist chambers at room temperature. Samples were collected every 12 hours and up to 72 hours post infection and flash frozen in liquid nitrogen, for a total of 6 time points, an important difference with the microarray data which used a shorter time series (5 time points). Total RNA was extracted using the Qiagen RNeasy extraction kit (Qiagen, Valencia, CA, USA) according to manufacturer's protocol and resuspended in 50µl of RNAse-free water. Treatment with DNAse (Thermo Scientific,

Suwanee, GA, USA) was performed to avoid contamination with genomic DNA. Reverse transcription was performed with the DyNAmo 2step synthesis kit (Thermo Scientific, Suwanee, GA, USA), using 1µl of RNA in a 50µl final volume and using oligo-dT as primer. cDNA quantification was performed using Nanodrop 1000 (Thermo Scientific, Suwanee, GA, USA), and cDNA was then diluted to a final concentration of 800ng/µl of total cDNA.

Real Time quantitative PCR

Using the computational annotation of the TFs and transcription regulators (TRs) of *P. infestans* [62] and based on microarray and transcriptome studies [18, 54, 63], we determined a set of TFs that were significantly overexpressed in the interaction *P. infestans* - *S. tuberosum*. Initially, we performed a significance microarray analysis (SAM) to determine the set of differentially expressed genes in the available microarray experiment from Cooke *et al* (GEO accession: GSE33240) [54]. We selected the genes with a *p*-value≤0.05, log2 fold-change (logfc) > 1 and false discovery rate (FDR) ≤ 0.01. We then cross-validated our results with the transcriptome analysis, chose the TFs that were differentially expressed on both sets of data, according to the criteria shown above.

We selected 49 genes to analyze a sub-network of the transcription regulation in *P. infestans*. These correspond to the Myb, Myb. Related and Top families. The Myb family has also been suggested to play an important role in the infection process in previous studies [64]. We designed primers for RT-qPCR using the QuantPrime software [65]. Primers were selected with an annealing temperature of 60ºC ± 0.5ºC and to span an exon-exon border to avoid genomic DNA amplification, these were against a cDNA pull of all the sampling points, primers with unspecific amplification or no amplification at all were discarded, the amplicon expected size was confirmed in an 1.2% agarose gel (data not shown). Primers for Actin A, GADPH and Tubulin A, Tubulin B, and WS21 were designed, these genes were used as reference (normalizing) genes for PCR experiments. Real time quantitative PCR (RT-qPCR) experiments were performed using SyBRGreen technology using Dynamo

SyBRGreen RT-qPCR kit (Thermo Scientific, Suwanee, Georgia, USA) according to manufacturer instructions, in 96-well reaction plates using 1µl of cDNA to a total volume of 10µl for 40 cycles. Expression values were calculated as the relative ratio of expression compared to the reference gene according to Pfaffl *et al* [66, 67]:

$$\text{ratio} = \frac{E^{\Delta\text{Ct Target}}}{E^{\Delta\text{Ct Reference}}} \quad (5)$$

Where ΔCt correspond to the difference in Ct values from the control condition (0 hours post infection) and the treatment (sampling points after infection).

Results

Inference of Yeast sub-networks

First, we used and compared for accuracy, four network inference methods using *in silico*-generated time series expression data from the yeast *Saccharomyces cerevisiae*. The analysis showed that SMI and DPC had a higher accuracy than the other two methods (Table 1, Figure 1a & 2). However, the variance of the AUROC was higher for SMI that for DPC. Additionally when SMI had a high AUROC, it is often under the selection of very small *p*-value cutoff, which causes a rapid diminishing in the precision as the recall increases (Figure 1b). DPC on the other hand, showed a consistent behavior between the ROC and PR curves as it recover a high number of true positives at a high cutoff, and precision remained stable over the PR space.

We then assessed the effect of the predictions over the network topology. We evaluated two topological measurements of network centrality, i.e., K-coreness and betweenness, both in the gold-standard and the predicted network. To assess this we selected DPC inferred sub-networks, as it is the method with the highest performance at a high threshold. K-coreness was generally overestimated, while betweenness centrality tended to be underestimated, compared to the metrics computed on the whole real network (Figure 3). This can be due to the fact that several bona fide edges were omitted from the

prediction, i.e., they are false negatives, which can decrease the betweenes centrality for highly connected nodes. The number of false positives although low (Figure 4), can cause the K-core to increase for certain nodes as its centrality was increased. Thus, these metrics hint that the inferred network is topologically different from the original network. It was already presented that in the case of scale-free networks, sub-networks could have a different topology [68].

Network inference/reconstruction from short time series

To assess the effect of the time series length over our reconstruction we attempted the network reconstruction using time series of length 6, instead of 21. This had a large impact on the performance of the different methods (Figure 5). All the models showed an overall decrease in the AUROC measurement. Although for DPC and IOTA the variance was slightly reduced, the predictive power of the model was severely affected by the length of the time series, increasing the error in the inferred networks.

Reconstruction of *P. infestans* GRN from publicly available expression data.

Using SMI and DPC we estimated the GNR of *P. infestans*, from the 589 TFs annotated in the genome. We were able to establish an interaction with a partner for 525, yielding a network with a total of 1750 edges. The network showed an average degree of 6.62 and an average path length of 3.35 with a diameter of 7. Newman-Girvam modularity had a value of 2.59 relatively large for a net with small diameter [69, 70]. Using a Poisson dispersion [71] test we showed that the networks degree distribution to be in accordance with a Poisson distribution (Figure 6).

We found 9 TFs to be differentially expressed in the biotrophy stage (0 to 3 days post infection), and 94 TFs to be differentially expressed in the necrotrophy phase (4 to 6 days post infection). Among these PITG_11750 and PITG_10768 showed different behavior in both phases, being up regulated in the biotrophy stage (p-value= 0.0449 log2 fold-change=

1.683 and p-value= 0.0479 log2 fold-change= 1.781 respectively) and down regulated in the necrotrohy stage (p-value= 0.0112 log2 fold-change= -2.121 and p-value= 0.0116 log2 fold-change= -2.037 respectively). Notably all of the TFs that were differentially expressed in the necreotrophy stage, were down regulated.

Phylostratigraphic analysis of network properties.

To determine the age of the nodes of the *P. infestans* regulatory network and the role of age in the topology, we assessed 6 phylostrata for the *P. infestans* genome. Phylostrata I and II contained one TF each, phylostratum III contained 360 TFs, IV contained 52 TFs and V and VI contained 105 and 6 genes respectively. Figure 7 shows the variation in K-coreness and degree according to each phylostratum. We found that coreness had a high value in the old phylostrata and then suddenly decreases and remains stable through the evolutionary history of *Phytophthora*. Degree on the other hand showed a fluctuating behavior, in the branching of Eukarya it increases and shows a decrease in the latter phylostratum. However as the divergence of *Phytophthora* takes place, a sudden increase in degree is shown, with a major decrease in the genes belonging to *P. infestans*. Network coreness has a different behavior however, as has a higher values in the first two phyostrata and the decreases in the branching of SARA group. The latter strata does not show any changes in the coreness values.

P. infestans regulatory sub-network inference from q-RTPCR data

The expression profile of 14 genes of *P. infestans* (Figure 8), in its interaction with Col2 and Col3 varieties, was used to reconstruct the regulatory sub-network in both cases. Our reconstruction shows that under infection over a susceptible cultivar (Col2) several interactions are identified, however the average degree (1.875) increased in the Col3 reconstruction, these sub-network depicted no interaction for genes PITG_14634 and PITG_10311 with the other elements (Figure 9), in this network PITG_19362 and PITG_16114 are depicted as an isolated component whereas in the Col2 network show a

353 link with PITG_11647 (*p*-value = 0.0467). Genes PITG_01528 and PITG_03670 show no
354 interaction with other components of the network in both reconstructions.
355
356 ## Discussion
357 In the study of GRNs it is important to assess inference methods that can behave
358 consistently with versatile sets of data. We have shown that inference methods currently
359 used, are sensible to the length of the time series used, and this should be taken into
360 account, as this can have several effects on the analysis and the predictions derived from it.
361 These issues could be addressed by the use of hybrid methods that increase accuracy of
362 the predictions without an excess of data [36], as well as the combination of low cost raising
363 strategies for detection of regulatory elements such as FAIRE-seq [72]. Conclusions of our
364 analysis should be heavily scrutinized. However under careful inspection, inferences derived
365 from network reconstruction can be faithful related to several biological contexts. We have
366 formulated several hypotheses of the phenotypic qualities and evolutive history of *P.*
367 *infestans* based on its GRN. These can be seen when crossing the topological information
368 of the networks with the divergence events of the pathogen evolutive course. It is, however,
369 still necessary to keep developing research on this field and develop new methods that can
370 adapt to different types of data. Especially in the study of plant-pathogen interactions as
371 several organisms are sampled at once. The analysis of the transcription regulation in an
372 evolutive context shows that topology of the network is a feature closely related to lifestyle
373 and that through its study, several clues of an organism life history can be extracted.
374 We showed that the ROC and PR spaces were consistent for the DPC inference
375 method, but far from satisfactory. However, in some sub-networks, AUROC can be as low
376 as 0.85 and even lower in small time series, e.g., 0.75. Low performance on GNW
377 benchmarks has been reported for perturbation datasets, due to small effect of perturbed
378 nodes and highly noisy data [39]. In large scale network reconstructions, the multiple effects
379 of several genes cannot be accounted for by DPC. Biological datasets often provide a limited

set of time points that are highly spaced in time, and poor performance of inference methodologies could severely affect the analysis derived from these reconstructions [27, 30, 68]. Major changes in the sub-network topology due to poor inference, can bias studies towards false conclusions [68]. However, predicted interactions have high accuracy as shown by PR curves, this fact can be used to detailed study of particular interactions.

Time series expression data vary over time and within the biological samples as well [73]. This property of longitudinal data can be considered as a bi-dimensional distribution. Neither DPC nor SMI consider this feature on their estimation, probably causing high variance of the AUROC values especially for SMI. As mutual information considers the conditional entropy of the variables in the dataset it is important to consider how this variable changes over the time domain [74]. This indicates that the distribution of variables over multidimensional spaces can also be an important feature to consider when modeling this type of interactions, and should also be included in methodologies such as DPC.

Evaluation of the inference methods is achieved by taking into account the ROC and PR curves [36, 75]. However, importance of changes in network topology, due to sampling effects and errors in edge inference, are to be considered as well in the evaluation of methods for network reconstruction [68]. Some topology measurements are more prone to error than others. This was seen in the case of betweenness centrality which has a higher rate of overestimation compared to K-coreness (Figure 4). This suggests that some measurements have more robustness to the inference method used, and that analysis should be carried out taking these into consideration.

Our inference of the GRN of *P. infestans* showed a topology consistent with a small network [2, 3, 35]. This was confirmed by comparison with a random network according to Erdos-Reyni model [2] (degree = 6.5 diameter 16 average path length=7.5). Although the network only included a total of 527 nodes, it has been predicted that *P. infestans* can have up to 589 predicted transcription factors [62]. This difference can be caused by both failure to infer the edges of this nodes, and a misclassification of these as transcription factors.

However the lack of connection of these components can also be due to the fact that these genes did not show significative changes in the time series, this could be due to the fact that microarray experiments have a poor dynamic range to evaluate the gene expression in most cases [76, 77], therefore relationships of these with the other regulatory elements of the network, could not be addressed by our method. To evaluate this issue it would be necessary to test the model under a different scenario, such as other stages of the life cycle of the organism, although for other stages of the life cycle (e.g. sporulation, zoosporogenesis, mycelium) data x available [78] this comes from cross-sectional studies which cannot be analyzed by the methods shown in this study. It is also possible that the regulatory functions of these genes are regulated in a context that are not exclusively dependent of other TFs.

Animal decisions are modulated by changes in hormone level and preferential activation of neurological components [79]. The changes in life style in the cycle of *P. infestans* such as the activation of different sets of genes during the change of biotrophy to necrotrophy, could be understood as GRN rewiring under environmental context. The fact that TFs were down regulated during necrotrophy phase could be indicating, that this stage does not rely on the activation of a network component, but rather it could be allowing the expression of necrotrophic proteins that were negatively regulated by transcriptional control, an example of this can be seen with the SNE1 protein which acts as a suppressor of cell death inducing proteins in *P. infestans* [80]. The behavior of PITG_17750 and PITG 10768, showed an antagonistic feature of both cycles, This antagonism of expression between the two phases has been proposed to be key in the switch between them [17, 80].

In a more general sense, natural selection process favors decisions made by species over long periods of time, over events of divergence organisms build their own history and differentiate from other by “evolutive decisions”. The GRN of *P. infestans* describes a very interesting behavior along the evolutive history of this organism. Analysis over degree and K-coreness showed a variable behavior over the different phylostrata: the genes orthologous to basal Eukaryotes have a higher degree than the ones originated in the first phylostratum

(i.e orthologous to Archea & Bacteria). Conservation of regulatory interactions changes in a step-wise manner over time [81], novel TFs of eukaryotes are likely to regulate eukaryote genes as the regulation mechanism are different from those of Bactria and Archea, [82, 83], therefore genes on the first phylostratum are unlikely to gain ore interactions rather than loose [81]. However K-coreness showed a high value in the genes related to this clade (Phylostratum I) and this is maintained over the second phylostratum. The latter decrease in coreness of the nodes, could describe a preferential attachment model: in this model the nodes that are incorporated to the networks bind to a highly connected node [2, 5]. Thus, the more recent genes (Phylostrata III-VI) are expected to decrease the average number of connections, consistent with the reduction observed in the branching to SARA from the other eukaryotes and as well for Oomycetes. K-core had also a lower value for the *Phytophthora* genus. However the expected degree distribution for a preferential attachment model is a power law distribution [2, 84, 85] and our reconstruction showed a Poisson distribution. This could be due to the size of the network, as the model predicts nodes with infinite degree [1-4, 35]. This distribution could be partially power law as highly connected nodes are sparse and low connected nodes are abundant. The changes of degree observed in the appearance of the *Phytophthora* clade could be related to a particular selective pressure over regulatory functions associated to the life style cues [86]. The rapid diversification of *P. infestans* shows that population sizes are highly variable and several bottleneck have occurred over the life history of this organism [86, 87]. This could also indicate that a non-adaptive process such as genetic drift could be influencing network topology [23], explaining why in other scenarios such as the mitochondrial old genes tend to have higher centrality that novel genes [20].

Our sub-network of the TFs of *P.infestans* in its interaction with *S. tuberosum* group *phureja* yielded a network that showed several differences among the different varieties studied. Col 2 sub-network had more connected elements than Col3 sub-network. This could be suggesting that regulatory functions have host-characteristics dependent variation, thus a larger rearrangement of the network could be taking place. However, this large

rearrangement was not seen due to the low sampling of the regulatory elements. The shared elements of both networks indicate that particular elements of the regulation are responding independently of the host, but under certain environments, this regulation can change. In comparison with our microarray-based reconstruction, these elements could be showing that a component of transcription regulation is necessary for the infection process to occur. Indeed, in the microarray analysis a large fraction of the TFs (83%), did not show major changes in the expression levels. qRT-PCR is considered to be the gold-standard of gene expression analysis, and can measure a more dynamic range of changes in gene expression [67, 88], additionally as the length of the time series heavily influences the sensitivity and specificity of the network inference, a collection of 6 time point presents several advantages in the inferences when compared to the microarray data, we highlight the importance of extending the number of samples genes ti better assess the topology of the network and associate with biological phenomena if different physiological states such as different hosts.

Transcription regulation can give insights of an organisms physiological responses as well as the long time process of evolution. Phenotypic responses can be seen as decisions taken by the organism in a particular context, in animals the brain and particularly the *Gyrus Cinguli* are in charge of these functions [79]. In oomycetes this function could be addressed by transcription regulation. We have shown that a negative regulation may be responsible for the control of the activation of genes in charge of necrotrophic processes, and a large fraction of the GRN is devoted to this function (16%). Antagonistic function can also be responsible for maintaining balance of certain modules of genes [17]. However as seen in the interaction of *P. infestans* with Col2 and Col3 varieties, there are possibly some components that need constant regulation in order to carry on with a complex process such as infection. Therefore decisions would not only be accounted by network rewiring but also to conserved components as well. In an evolutive perspective, this type of study can provide hypotheses of the organization and assembling of these GRNs as well as the selective pressures that lead to them. Gene duplication is the most common form of novel gene

appearance [57], new genes inherit the connections from old genes, and over selective events of the regulatory region, old genes tend to gain more connections that new genes which only have the inheritance passed to them [21]. It is important to point out that other variables can affect network topology [23], and that coupling network inference with population and genetic studies could show how networks reshape upon different conditions.

References

1. Barabási A-L: **The network takeover**. *Nat Phys* 2011, **8**:14–16.

2. Albert R, Barabási A-L: **Statistical mechanics of complex networks**. *Rev Mod Phys* 2002, **74**:47–97.

3. Barabási A-L, Oltvai ZN: **Network biology: understanding the cell's functional organization.** *Nat Rev Genet* 2004, **5**:101–13.

4. Strogatz SH: **Exploring complex networks.** *Nature* 2001, **410**:268–76.

5. Kitano H: **Computational systems biology.** *Nature* 2002, **420**:206–10.

6. Lansing JS: **Complex Adaptative Systems**. *Annu Rev Anthropol* 2003, **32**:183–204.

7. Barrat A, Barthelemy M, Vespignani A: *Dynamical Processes on Complex Networks*. 1st edition. Cambridge: Cambridge University Press; 2008.

8. Kumar H, Kawai T, Akira S: **Pathogen recognition by the innate immune system.** *Int Rev Immunol* 2011, **30**:16–34.

9. Hethcote H: **The mathematics of infectious diseases**. *SIAM Rev* 2000, **42**:599–653.

10. Kamoun S: **A catalogue of the effector secretome of plant pathogenic oomycetes.** *Annu Rev Phytopathol* 2006, **44**:41–60.

11. Fry W: **Phytophthora infestans: the plant (and R gene) destroyer.** *Mol Plant Pathol* 2008, **9**:385–402.

12. Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, Bozkurt TO, Ah-Fong AM V, Alvarado L, Anderson VL, Armstrong MR, Avrova A, Baxter L, Beynon J, Boevink PC, Bollmann SR, Bos JIB,

515 Bulone V, Cai G, Cakir C, Carrington JC, Chawner M, Conti L, Costanzo S, Ewan R,
516 Fahlgren N, et al.: **Genome sequence and analysis of the Irish potato famine pathogen**
517 **Phytophthora infestans.** *Nature* 2009, **461**:393–8.
518 13. Van der Vossen E, Sikkema A, Hekkert B te L, Gros J, Stevens P, Muskens M,
519 Wouters D, Pereira A, Stiekema W, Allefs S: **An ancient R gene from the wild potato**
520 **species Solanum bulbocastanum confers broad-spectrum resistance to Phytophthora**
521 **infestans in cultivated potato and tomato**. *Plant J* 2003, **36**:867–882.
522 14. Gebhardt C, Bellin D, Henselewski H, Lehmann W, Schwarzfischer J, Valkonen JPT:
523 **Marker-assisted combination of major genes for pathogen resistance in potato.** *Theor Appl*
524 *Genet* 2006, **112**:1458–64.
525 15. Ewig EE, Šimko I, Smart CD, Bonierbale MW, Mizubuti ESG, May GD, Fry WE:
526 **Genetic mapping from field tests of qualitative and quantitative resistance to Phytophthora**
527 **infestans in a population derived from Solanum tuberosum and Solanum berthaultii**. *Mol*
528 *Breed* 2000, **6**:25–36.
529 16. Judelson HS: **Dynamics and innovations within oomycete genomes: insights into**
530 **biology, pathology, and evolution.** *Eukaryot Cell* 2012, **11**:1304–12.
531 17. Lee S-J, Rose JKC: **Mediation of the transition from biotrophy to necrotrophy in**
532 **hemibiotrophic plant pathogens by secreted effector proteins**. *Plant Signal Behav* 2010,
533 **5**:769–772.
534 18. Gyetvai G, Sønderkær M, Göbel U, Basekow R, Ballvora A, Imhoff M, Kersten B,
535 Nielsen K-L, Gebhardt C: **The transcriptome of compatible and incompatible interactions of**
536 **potato (Solanum tuberosum) with Phytophthora infestans revealed by DeepSAGE**
537 **analysis.** *PLoS One* 2012:e31526.
538 19. Shen-Orr SS, Milo R, Mangan S, Alon U: **Network motifs in the transcriptional**
539 **regulation network of Escherichia coli.** *Nat Genet* 2002, **31**:64–8.
540 20. Yang J-S, Kim J, Park S, Jeon J, Shin Y-E, Kim S: **Spatial and functional organization**
541 **of mitochondrial protein network.** *Sci Rep* 2013, **3**:1403.

542 21. Arabidopsis Interactome Mapping Consortium: **Evidence for network evolution in an**
543 **Arabidopsis interactome map.** *Science* 2011, **333**:601–7.
544 22. Abouheif E, Wray GA: **Evolution of the gene network underlying wing polyphenism in**
545 **ants.** *Science* 2002, **297**:249–52.
546 23. Lynch M: **The evolution of genetic networks by non-adaptive processes.** *Nat Rev*
547 *Genet* 2007, **8**:803–13.
548 24. O'Brien EJ, Lerman J a, Chang RL, Hyduke DR, Palsson BØ: **Genome-scale models of**
549 **metabolism and gene expression extend and refine growth phenotype prediction.** *Mol Syst*
550 *Biol* 2013, **9**:693.
551 25. Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, Feist AM,
552 Palsson BØ: **Genome-scale metabolic reconstructions of multiple Escherichia coli strains**
553 **highlight strain-specific adaptations to nutritional environments.** *Proc Natl Acad Sci U S A*
554 2013, **110**:20338–43.
555 26. Karr JR, Sanghvi JC, Macklin DN, Gutschow M V, Jacobs JM, Bolival B, Assad-Garcia
556 N, Glass JI, Covert MW: **A whole-cell computational model predicts phenotype from**
557 **genotype.** *Cell* 2012, **150**:389–401.
558 27. Kim HD, Shay T, O'Shea EK, Regev A: **Transcriptional regulatory circuits: predicting**
559 **numbers from alphabets.** *Science* 2009, **325**:429–32.
560 28. Davidson E, Levin M: **Gene regulatory networks.** *Proc Natl Acad Sci U S A* 2005,
561 **102**:4935.
562 29. Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M: **Genomic**
563 **analysis of regulatory network dynamics reveals large topological changes.** *Nature* 2004,
564 **431**:308–12.
565 30. Jothi R, Balaji S, Wuster A, Grochow J a, Gsponer J, Przytycka TM, Aravind L, Babu
566 MM: **Genomic analysis reveals a tight link between transcription factor dynamics and**
567 **regulatory network architecture.** *Mol Syst Biol* 2009:294.

568 31. Spitz F, Furlong EEM: **Transcription factors: from enhancer binding to developmental**
569 **control.** *Nat Rev Genet* 2012, **13**:613–26.
570 32. Chu D, Zabet NR, Mitavskiy B: **Models of transcription factor binding: sensitivity of**
571 **activation functions to model assumptions.** *J Theor Biol* 2009:419–29.
572 33. Zhang X, Liu K, Liu Z-P, Duval B, Richer J-M, Zhao X-M, Hao J-K, Chen L: **NARROMI:**
573 **a noise and redundancy reduction technique improves accuracy of gene regulatory**
574 **network inference.** *Bioinformatics* 2013, **29**:106–13.
575 34. Balaji S, Babu MM, Aravind L: **Interplay between network structures, regulatory modes**
576 **and sensing mechanisms of transcription factors in the transcriptional regulatory network**
577 **of E. coli.** *J Mol Biol* 2007:1108–22.
578 35. Balázsi G, Barabási a-L, Oltvai ZN: **Topological units of environmental signal**
579 **processing in the transcriptional regulatory network of Escherichia coli.** *Proc Natl Acad Sci*
580 *U S A* 2005:7841–6.
581 36. Marbach D, Prill RJ, Schaffter T, Mattiussi C, Floreano D, Stolovitzky G: **Revealing**
582 **strengths and weaknesses of methods for gene network inference.** *Proc Natl Acad Sci U S*
583 *A* 2010, **107**:6286–91.
584 37. Hsiao Y-T, Lee W-P: **Inferring robust gene networks from expression data by a**
585 **sensitivity-based incremental evolution method.** *BMC Bioinformatics* 2012, **13 Suppl**
586 **7**(Suppl 7):S8.
587 38. Chemmangattuvalappil N, Task K, Banerjee I: **An integer optimization algorithm for**
588 **robust identification of non-linear gene regulatory networks.** *BMC Syst Biol* 2012, **6**:119.
589 39. Yuan Y, Li C-T, Windram O: **Directed partial correlation: inferring large-scale gene**
590 **regulatory network through induced topology disruptions.** *PLoS One* 2011, **6**:e16835.
591 40. Lütkepohl H: *New Introduction to Multiple Time Series Analysis*. Springer; 2007:786.
592 41. Kantz H, Schreiber T: *Nonlinear Time Series Analysis [Paperback]*. Cambridge
593 University Press; 2 edition; 2004:388.

42. Hempel S, Koseska A, Nikoloski Z: **Data-driven reconstruction of directed networks**. *Eur Phys J B* 2013, **86**:250.

43. Hempel S, Koseska A, Kurths J, Nikoloski Z: **Inner Composition Alignment for Inferring Directed Networks from Short Time Series**. *Phys Rev Lett* 2011, **107**:054101.

44. Schaffter T, Marbach D, Floreano D: **GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods.** *Bioinformatics* 2011, **27**:2263–70.

45. Marbach D, Schaffter T, Mattiussi C, Floreano D: **Generating realistic in silico gene networks for performance assessment of reverse engineering methods.** *J Comput Biol* 2009, **16**:229–39.

46. Seidman S: **Network structure and minimum degree**. *Soc Networks* 1983, **5**:269–287.

47. Granger CWJ: **Investigating Causal Relations by Econometric Models and Cross-spectral Methods**. *Econometrica* 1969, **37**:424.

48. Behmard H, Faridani A: **Sampling of Bandlimited Functions on Unions of Shifted Lattices**. *J Fourier Anal Appl* 2002, **8**:43–58.

49. Stern DI: **A multivariate cointegration analysis of the role of energy in the US macroeconomy**. *Energy Econ* 2000, **22**:267–283.

50. Hatemi-J A: **Asymmetric causality tests with an application**. *Empir Econ* 2011, **43**:447–456.

51. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing**. *J R Stat Soc Ser B …* 1995, **57**:289–300.

52. Davis J, Goadrich M: **The relationship between Precision-Recall and ROC curves**. In *Proc 23rd Int Conf Mach Learn - ICML '06*. New York, New York, USA: ACM Press; 2006:233–240.

53. Cortes C, Mohri M: **AUC Optimization vs. Error Rate Minimization.** *Neural Inf Process Syst 15* 2003.

54. Cooke DEL, Cano LM, Raffaele S, Bain R a, Cooke LR, Etherington GJ, Deahl KL, Farrer R a, Gilroy EM, Goss EM, Grünwald NJ, Hein I, MacLean D, McNicol JW, Randall

621 E, Oliva RF, Pel M a, Shaw DS, Squires JN, Taylor MC, Vleeshouwers VG a a, Birch PRJ,
622 Lees AK, Kamoun S: **Genome analyses of an aggressive and invasive lineage of the Irish**
623 **potato famine pathogen.** *PLoS Pathog* 2012, **8**:e1002940.
624 55. Li L, Stoeckert CJ, Roos DS: **OrthoMCL: identification of ortholog groups for eukaryotic**
625 **genomes.** *Genome Res* 2003, **13**:2178–89.
626 56. Chen F, Mackey AJ, Stoeckert CJ, Roos DS: **OrthoMCL-DB: querying a**
627 **comprehensive multi-species collection of ortholog groups.** *Nucleic Acids Res* 2006,
628 **34**(Database issue):D363–8.
629 57. Tautz D, Domazet-Lošo T: **The evolutionary origin of orphan genes.** *Nat Rev Genet*
630 2011, **12**:692–702.
631 58. Domazet-Loso T, Brajković J, Tautz D: **A phylostratigraphy approach to uncover the**
632 **genomic history of major adaptations in metazoan lineages.** *Trends Genet* 2007, **23**:533–9.
633 59. He D, Fiz-Palacios O, Fu C-J, Fehling J, Tsai C-C, Baldauf SL: **An alternative root for**
634 **the eukaryote tree of life.** *Curr Biol* 2014, **24**:465–70.
635 60. Céspedes MC, Cárdenas ME, Vargas AM, Rojas A, Morales JG, Jiménez P, Bernal AJ,
636 Restrepo S: **Physiological and molecular characterization of Phytophthora infestans**
637 **isolates from the Central Colombian Andean Region.** *Rev Iberoam Micol* , **30**:81–7.
638 61. Vargas AM, Quesada Ocampo LM, Céspedes MC, Carreño N, González A, Rojas A,
639 Zuluaga a P, Myers K, Fry WE, Jiménez P, Bernal AJ, Restrepo S: **Characterization of**
640 **Phytophthora infestans populations in Colombia: first report of the A2 mating type.**
641 *Phytopathology* 2009, **99**:82–8.
642 62. Buitrago-Floréz F, Restrepo S, Riaño-Pachón DM: **Identification of transcription factor**
643 **genes and their correlation with the high diversity of Stramenopiles**. 2014. In publication
644 process.
645 63. Avrova AO, Venter E, Birch PR., Whisson SC: **Profiling and quantifying differential**
646 **gene transcription in Phytophthora infestans prior to and during the early stages of potato**
647 **infection**. *Fungal Genet Biol* 2003, **40**:4–14.

64. Seidl MF, Wang R-P, Van den Ackerveken G, Govers F, Snel B: **Bioinformatic inference of specific and general transcription factor binding sites in the plant pathogen Phytophthora infestans.** *PLoS One* 2012, **7**:e51295.

65. Arvidsson S, Kwasniewski M, Riaño-Pachón DM, Mueller-Roeber B: **QuantPrime--a flexible tool for reliable high-throughput primer design for quantitative PCR.** *BMC Bioinformatics* 2008, **9**:465.

66. Pfaffl MW: **A new mathematical model for relative quantification in real-time RT-PCR.** *Nucleic Acids Res* 2001, **29**:e45.

67. Bustin SA, Benes V, Nolan T, Pfaffl MW: **Quantitative real-time RT-PCR--a perspective.** *J Mol Endocrinol* 2005, **34**:597–601.

68. Stumpf MPH, Wiuf C, May RM: **Subnets of scale-free networks are not scale-free: sampling properties of networks.** *Proc Natl Acad Sci U S A* 2005, **102**:4221–4.

69. Lancichinetti A, Fortunato S: **Community detection algorithms: A comparative analysis**. *Phys Rev E* 2009, **80**:056117.

70. Newman MEJ: **Communities, modules and large-scale structure in networks**. *Nat Phys* 2011, **8**:25–31.

71. Cox DR, Lewis PAW: *The Statistical Analysis of Series of Events*. Methuen; 1966:285.

72. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD: **FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin.** *Genome Res* 2007, **17**:877–85.

73. Diggle P, Heagerty P, Liang K-Y, Zeger S: *Analysis of Longitudinal Data*. Oxford University Press, USA; 2002:396.

74. Balagani KS, Phoha V V: **On the feature selection criterion based on an approximation of multidimensional mutual information.** *IEEE Trans Pattern Anal Mach Intell* 2010, **32**:1342–3.

75. Prill RJ, Iglesias P a, Levchenko A: **Dynamic properties of network motifs contribute to biological network organization.** *PLoS Biol* 2005, **3**:e343.

675 76. Allison DB, Cui X, Page GP, Sabripour M: **Microarray data analysis: from disarray to**
676 **consolidation and consensus.** *Nat Rev Genet* 2006, **7**:55–65.
677 77. Cantu D, Govindarajulu M, Kozik A, Wang M, Chen X, Kojima KK, Jurka J, Michelmore
678 RW, Dubcovsky J: **Next generation sequencing provides rapid access to the genome of**
679 **Puccinia striiformis f. sp. tritici, the causal agent of wheat stripe rust.** *PLoS One* 2011,
680 **6**:e24230.
681 78. Judelson HS, Narayan RD, Ah-Fong AM V, Kim KS: **Gene expression changes during**
682 **asexual sporulation by the late blight agent Phytophthora infestans occur in discrete**
683 **temporal stages.** *Mol Genet Genomics* 2009, **281**:193–206.
684 79. Doya K: **Modulators of decision making.** *Nat Neurosci* 2008, **11**:410–6.
685 80. Kelley BS, Lee S-J, Damasceno CMB, Chakravarthy S, Kim B-D, Martin GB, Rose
686 JKC: **A secreted effector protein (SNE1) from Phytophthora infestans is a broadly acting**
687 **suppressor of programmed cell death.** *Plant J* 2010, **62**:357–66.
688 81. Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann S a: **Structure and**
689 **evolution of transcriptional regulatory networks.** *Curr Opin Struct Biol* 2004, **14**:283–91.
690 82. Sonenberg N, Hinnebusch AG: **Regulation of translation initiation in eukaryotes:**
691 **mechanisms and biological targets.** *Cell* 2009, **136**:731–45.
692 83. Jackson RJ, Hellen CUT, Pestova T V: **The mechanism of eukaryotic translation**
693 **initiation and principles of its regulation.** *Nat Rev Mol Cell Biol* 2010, **11**:113–27.
694 84. Stumpf MPH, Ingram PJ: **Probability models for degree distributions of protein**
695 **interaction networks**. *Europhys Lett* 2005, **71**:152–158.
696 85. Li L, Alderson D, Tanaka R, Doyle J, Willinger W: **Towards a theory of scale-free**
697 **graphs: Definition, properties, and implications (extended version)(2005)**. *Internet Math*
698 2005, **2**:4.
699 86. Li Y, van der Lee T a J, Evenhuis a, van den Bosch GBM, van Bekkum PJ, Förch MG,
700 van Gent-Pelzer MPE, van Raaij HMG, Jacobsen E, Huang SW, Govers F, Vleeshouwers
701 VG a a, Kessel GJT: **Population dynamics of Phytophthora infestans in the Netherlands**

reveals expansion and spread of dominant clonal lineages and virulence in sexual offspring. *G3 (Bethesda)* 2012:1529-40.

87. Raffaele S, Kamoun S: **Genome evolution in filamentous plant pathogens: why bigger can be better.** *Nat Rev Microbiol* 2012, **10**:417-30.

88. Schefe JH, Lehmann KE, Buschmann IR, Unger T, Funke-Kaiser H: **Quantitative real-time RT-PCR data analysis: current concepts and the novel "gene expression's CT difference" formula.** *J Mol Med (Berl)* 2006, **84**:901-10.

Figures and tables

Method	Lag step	p-Value Threshold	p-Value Correction	AUROC
IOTA	NA	0.15	FDR	0.85
Directed Partial Correlation	1	0.34	FDR	0.66
Directed Partial Correlation	2	0.26	FDR	0.74
Directed Partial Correlation	3	0.09	FDR	0.91
Shifted Mutual Information	1	0.42	FDR	0.58
Shifted Mutual Information	2	0.35	FDR	0.65
Shifted Mutual Information	3	0.32	FDR	0.68
Granger Causality	1	0.39	FDR	0.61
Granger Causality	2	0.38	FDR	0.62
Granger Causality	3	0.35	FDR	0.65

Table 1. Network inference methods over long time series. In order to determine the Yeast network from time series expression data, several methods were implemented, some methods rely of shift over time the lag-step is shown for several of these, the significance values are estimated and set under a threshold for each method based on the area under the Receiving Operating Characteristic (AUROC) values, correction of these values was applied using Benjamini-Hochberg correction (FDR), the relative performance of these as measured in the AUROC is shown.

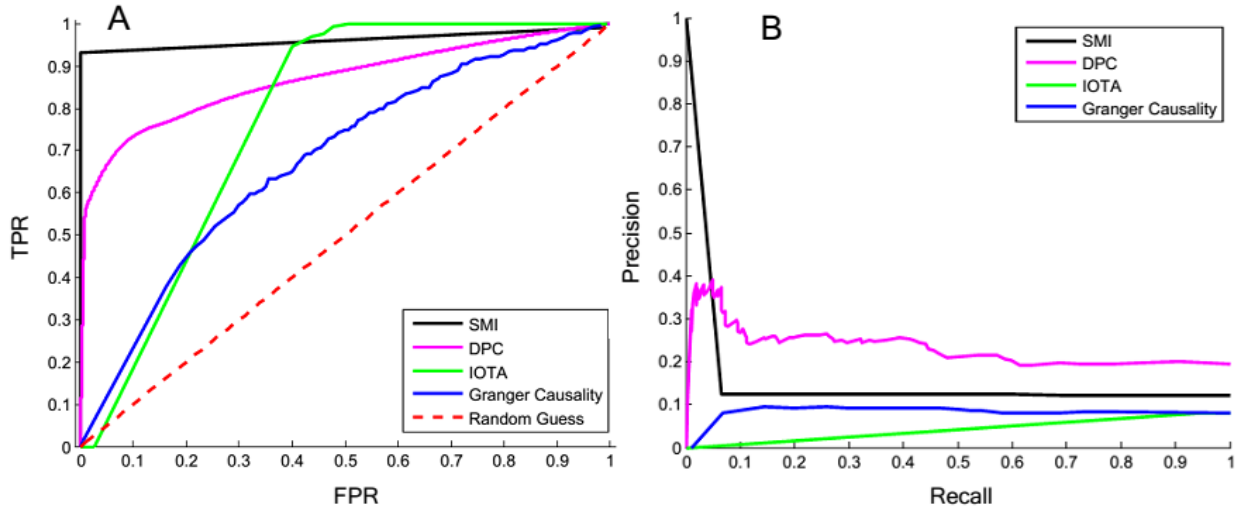
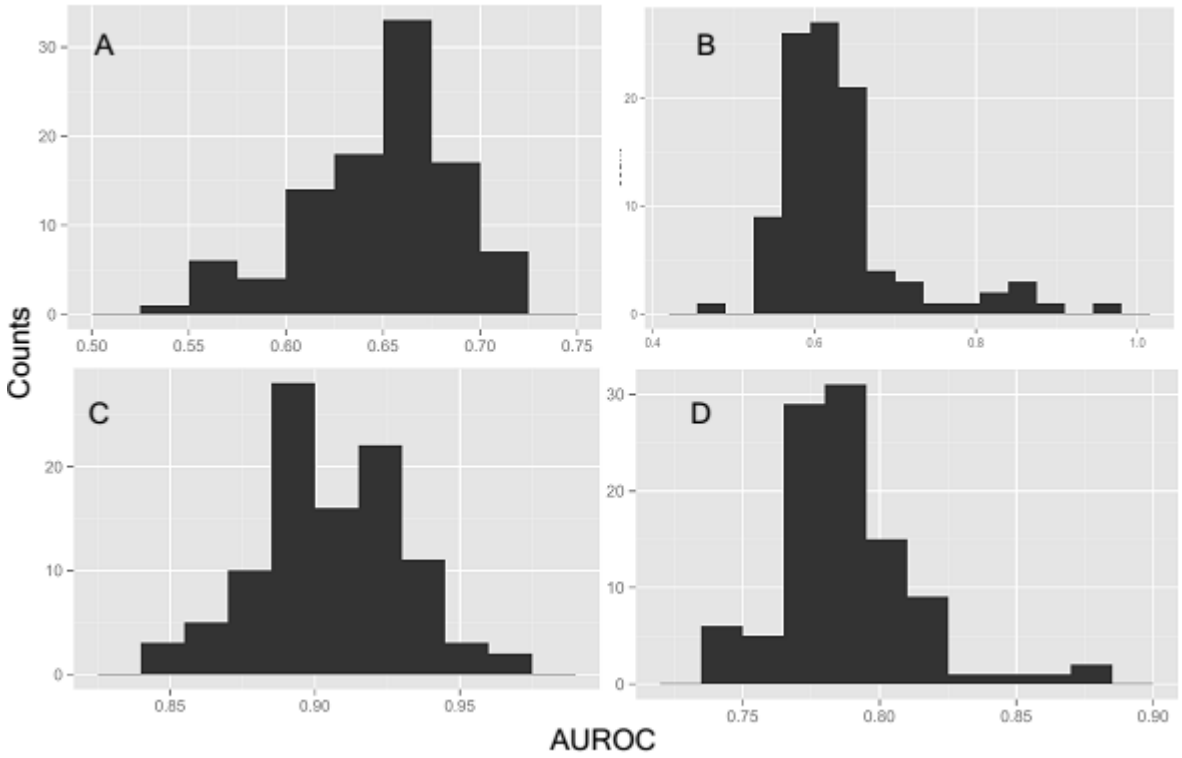
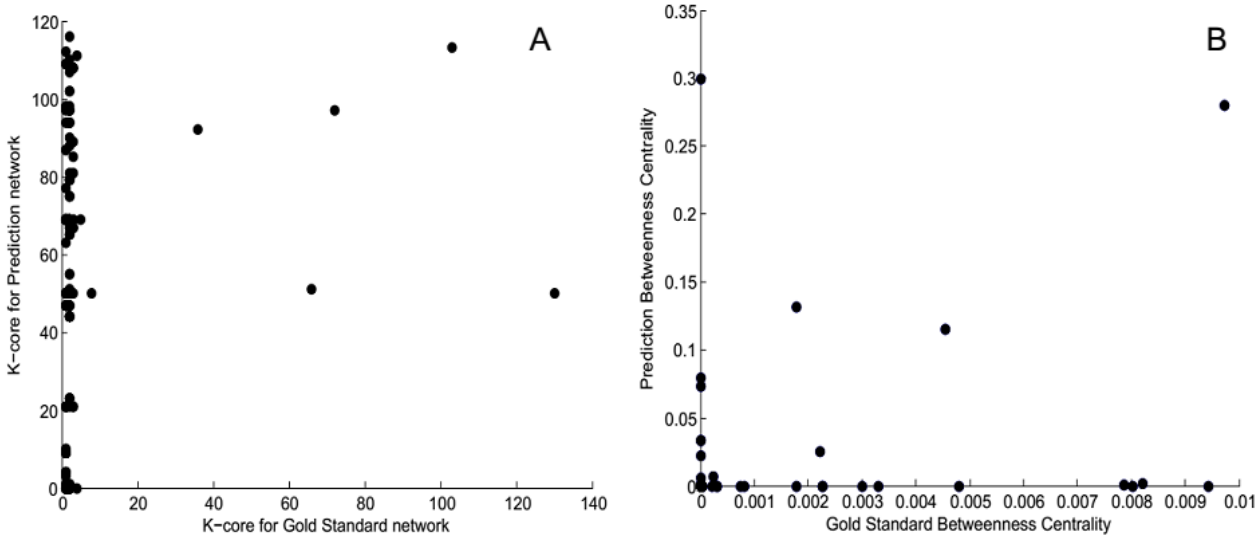


Figure 1. Receiving Operating Characteristic and Precision-Recall curves for the different estimation methods using 21 time points time seires. A) The performance of the different network reconstruction methods is evaluated based on the ROC curve, true positive rate (TPR) plotted against false positive rate (FPR) using a variable cutoff of 0.01, shifted mutual information (SMI) and directed partial correlation (DPC) show better performance compared to Granger causality and inner composition alingment (IOTA), this is confirmed by the PR curves B) PR curve shows prescion and recal over a variable cutoff of 0.01. DPC and SMI show the high level of confidence to recover the positive edges of the network.



725

726 **Figure 2. Distribution of AUROC values for the different inference methods used in *in silico* generated long**
727 **time series (21 time points).** A) Granger causality mean = 0.649 variance = 0.001 B) SMI distribution mean =
728 0.629 variance = 0.005 C) DPC mean = 0.914 variance = $6.3*10^{-4}$ D) IOTA.



729

730 **Figure 3. Differences in topology for the prediction and gold standard networks of *S cerevisae* using DPC over**
731 **a 21 time points time series**. A) K-core measurements for the gold standard and the prediction network, the
732 predicted network over estimates the coreness of several vertices. The omission of edges leads to higher
733 fragmentation of the network and over decreases the robustness of it. B) Betweenness centrality for the gold
734 standard and the prediction network, over representation of node betweenness for several nodes is shown due
735 to the presence of false positive interactions and the misclassification of edges as false negatives

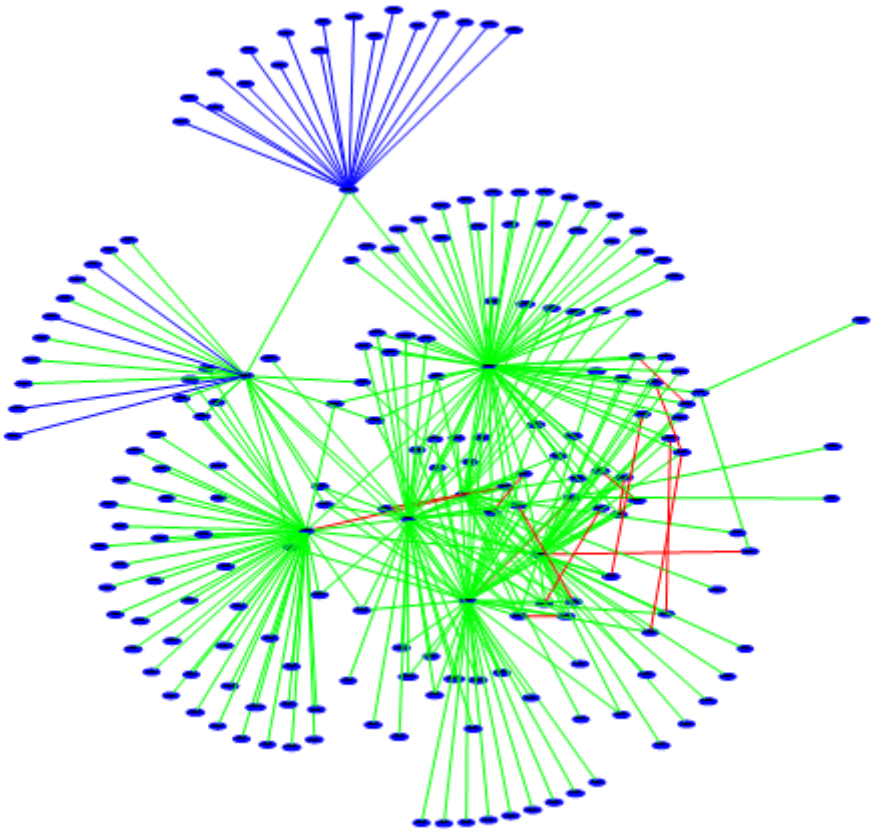


Figure 4. Similarities between predicted and gold standard networks for *S. cerevisae* using DPC over a 21 time points time series. 22 false negatives (blue) appear only in the gold standard but are ignored by the prediction and 299 true positives (green) and 15 false positives (red).

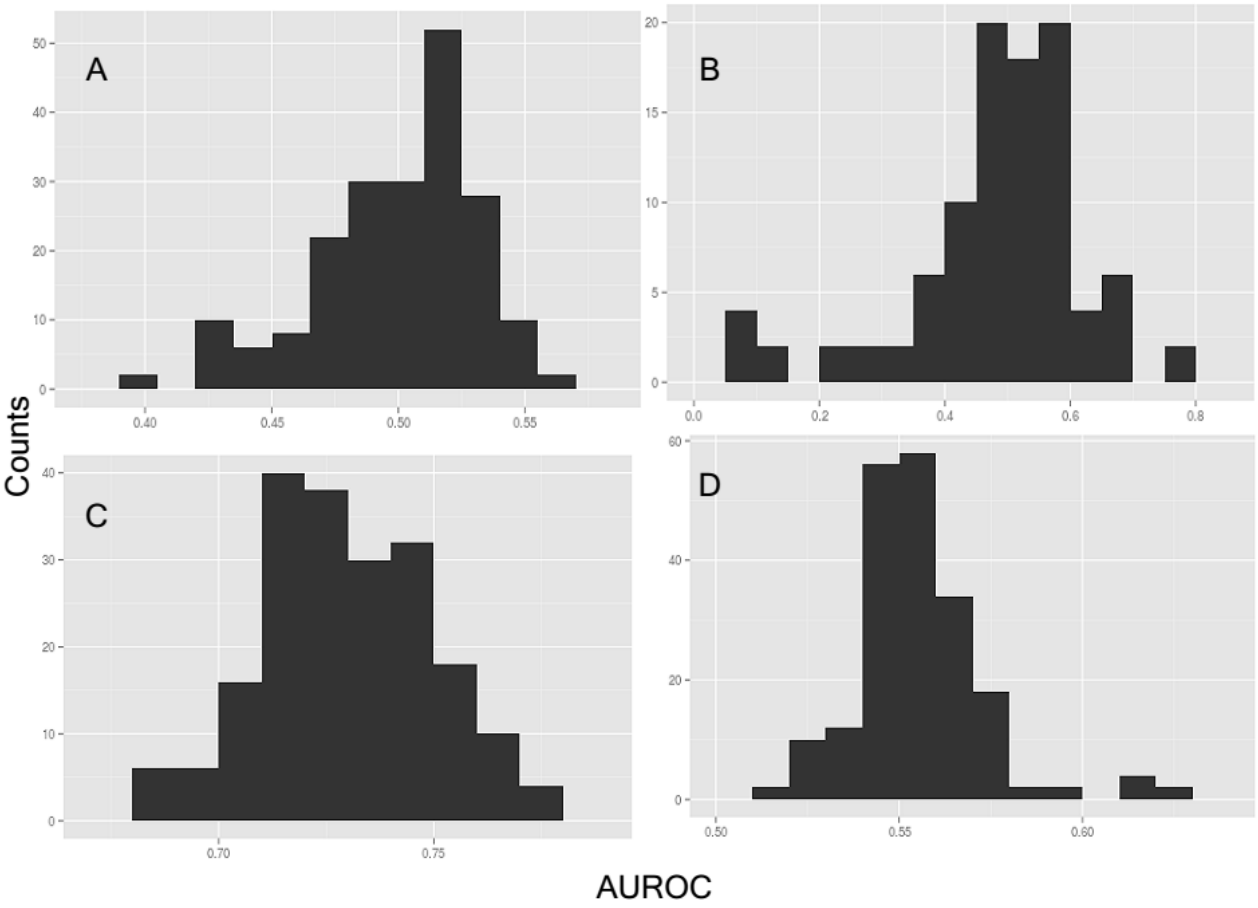
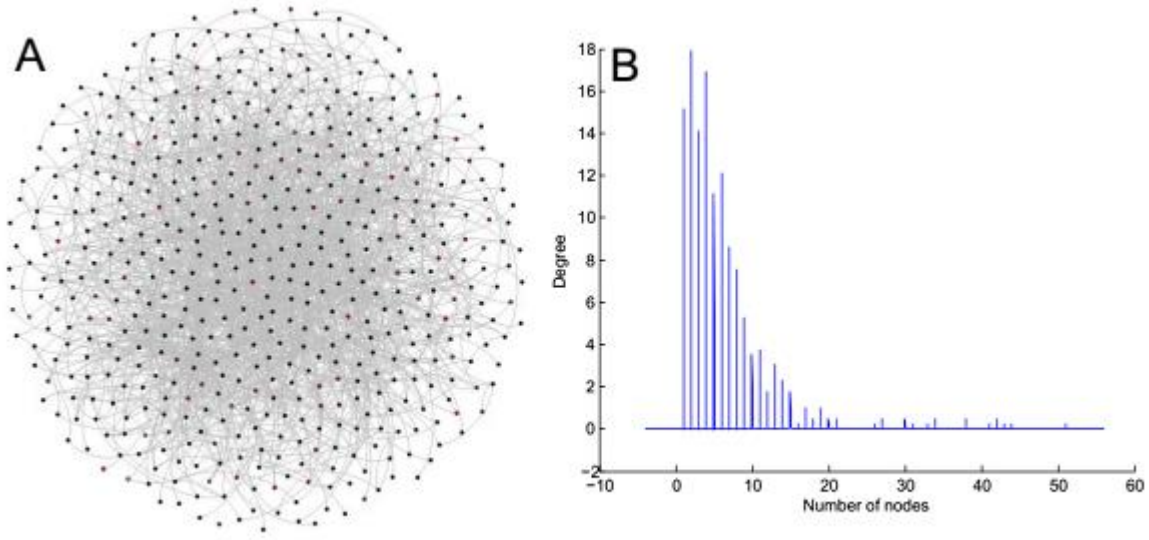
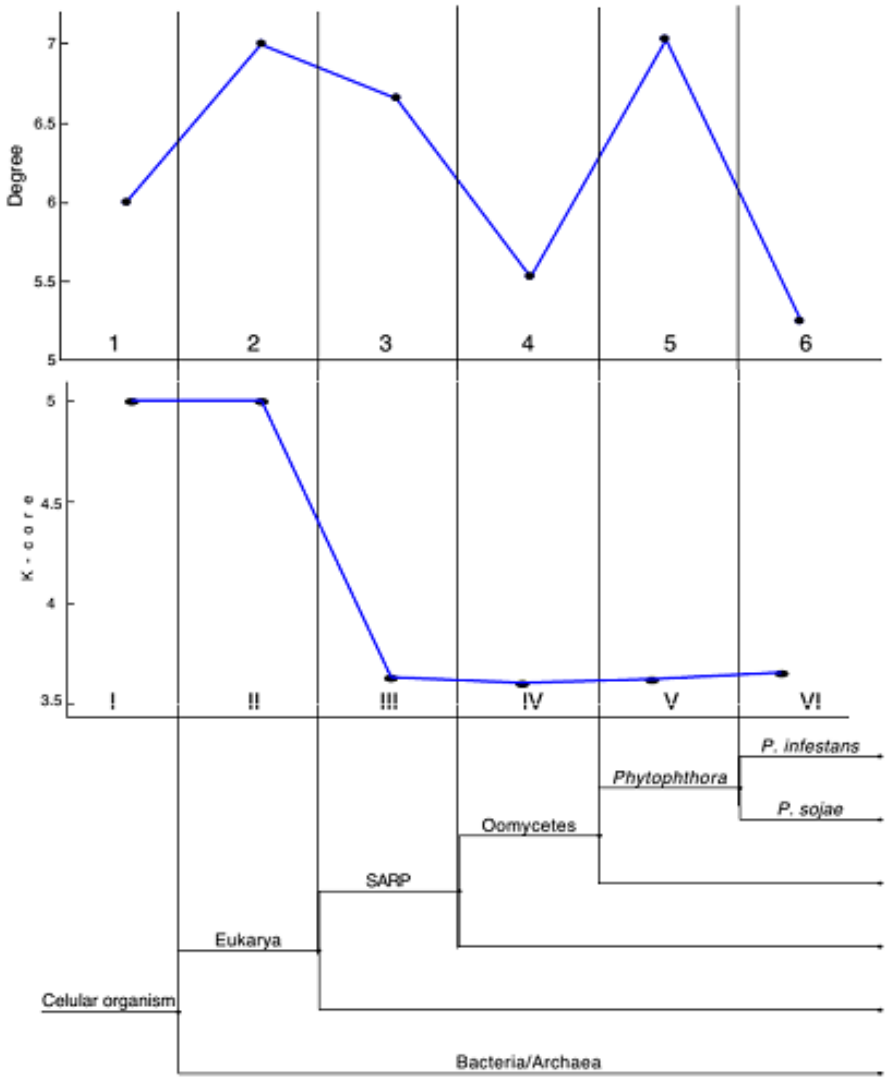


Figure 5. Distribution of AUROC values for the different inference methods used in *in silico* generated short time series (6 time points). A) Granger causality mean = 0.499 variance = $9.8*10^{-4}$ B) SMI distribution mean = 0.530 variance = 0.024 C) DPC mean = 0.730 variance = $3.9*10^{-4}$ D) IOTA.



744

745 **Figure 6.** ***Phytophthora infestans*** **regulatory network.** The regulatory network (A) composed of 525 TFs and
746 1750 interactions (grey lines), estimated from microarray data time series using DPC with a seven points time
747 series, green dots represent genes differentially expressed in biotrophy phase, red dots represent genes
748 differentially expressed in the necrotrophic phase, black dots represent genes with not significative change on
749 the microarray analysis using SAM. The degree distribution is in accordance with a Poisson distribution (B).



750

751 **Figure 7. *Phytophthora* phylostratigraphic map for the networks topology using DPC reconstruction.** 20
752 genomes were used to compare the genome of *P. infestans* each genome is associated to a phylostratum.
753 Orthologous relationships are determined using OrthoMCL, gene with common ancestry are considered to
754 have arisen before the divergence of the groups in which the genomes are been compared. 6 genomic
755 phylostrata corresponding to phylogenetic internodes (lower panel) are bordered by vertical grids that denote
756 sets of *Phythopthora* founder genes. In each phylostratum the average degree and K-core centrality are shown.
757 Fluctuations in the degree of the network are observed in the evolutive history as very the first genes to appear
758 are low connected, but the in the branching of eukarya the connectivity rises, to decrease again in the
759 oomycetes taxa. Genes exclusive to *Phythopthora* genus are also highly connected. K-core shows a different
760 behavior as the ancestral genes have high coreness while recently derived genes tend to have low coreness
761 with low variation along evolutive history.

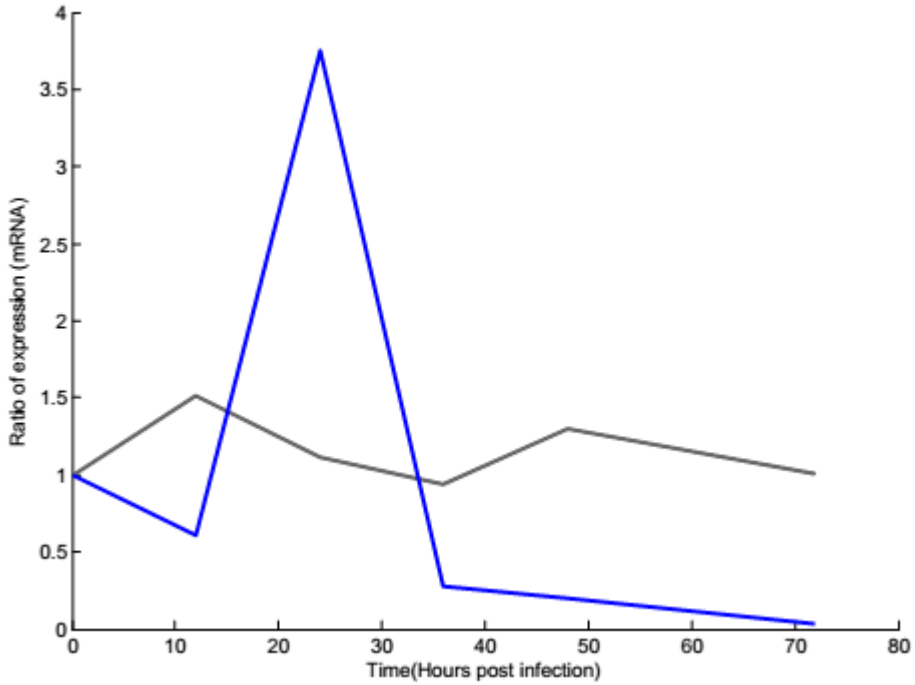


Figure 8. Expression profile of PITG_02143 in comparison to Actin as a reference gene. Expression levels are measured as relative expression to the pathogen growing in culture medium. Six points of time are depicted. Not significative changes are observed in the expression levels of actin.

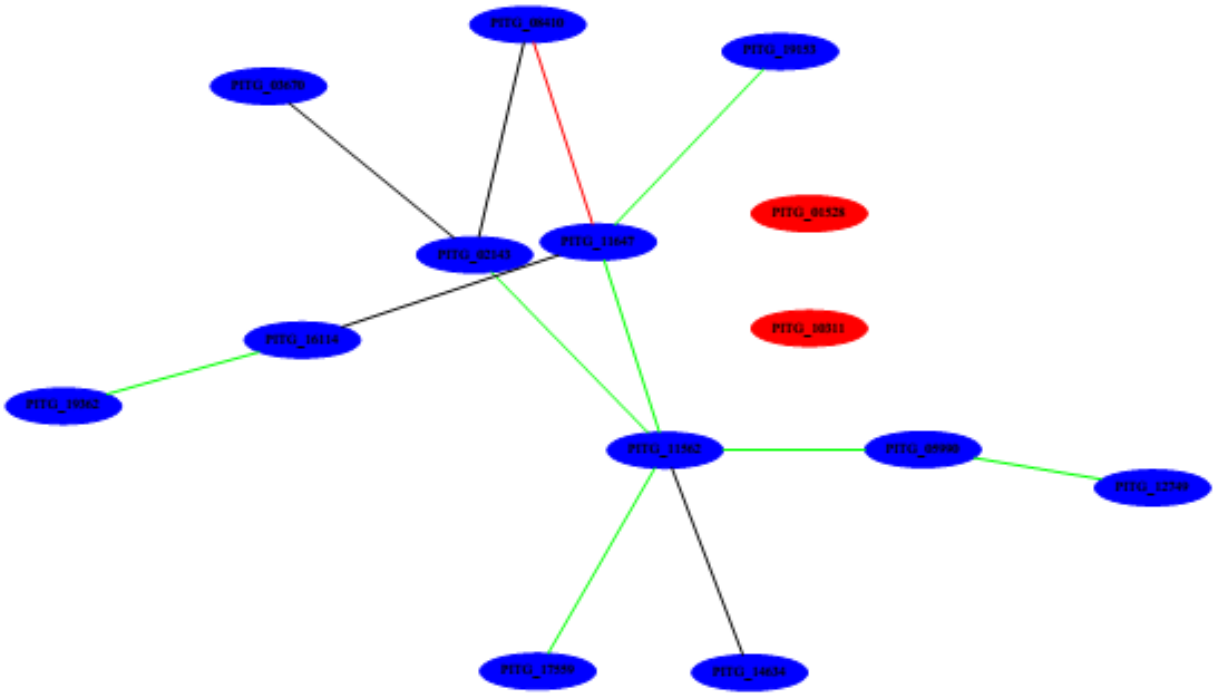


Figure 9. Regulatory network of *P. infestans* for the cultivar Col2 and Col3. The GRN was obtained from time series expression data acquired by RT-qPCR experiments over 6 time points during the interaction of *P. infestans* with *S. tuberosum* group *phureja*, using DPC and a *p*-value threshold of 0.09 Interactions for 14 nodes are shown, Shared edges between Col and Col3 sub-networks are shown in green, edges exclusive of Col2 are shown in black and edges exclusive of Col3 are shown in red, 2 nodes are seen as isolated in both networks (red).

Supplementary data

Transcript identifier	Forward sequence	Reverse sequence
PITG_14401T0	CCGTAACAACAGCAACAGGTTCG	ACTTCGCTCGCTTCCACTCATC
PITG_05989T0	ACCCGAGCATCAACAAATCACC	TCCGTACGTCCAGGCAACAATTC
PITG_19851T0	GATAGCCGATAACCAGCAGCAAGC	TCTTGCAACAACGCCTGCAGACTC
PITG_04464T0	AAAGAACGCACCCGTTGTTCAC	GAACTCTCCGAGTTTCCAGCAC
PITG_08960T0	TGAAAGGCAAAGGGACGTGGAC	CAGGTAAGAACGCAGCAACCTC
PITG_13032T0	TGTCACCAGGTCCATGCCAAAC	TCTCTTCAGAGCGGGTCGAATG
PITG_14420T0	CCAACACCAACGTTGCACACAC	GAGGAGATTGTCATGGTGTTTGGG
PITG_00988T0	CACTAAACTGCCTGGCCGTAATGG	TCAAGCTGGTTGTGCCACCTTTC
PITG_00513T0	AGGCCAAGTCCATTTCGCCAAC	AAGCGGTCATGCTCCTCCAAAG
PITG_16114T0	ACAGCAGGCTGCGCATATCTTG	AGAATCCTTTCCGTTTGCCGTAGC
PITG_01056T0	GCAAGAGGAACGGAGTCAATGG	AAGAACTCAGCCACTTGCTGCTC
PITG_13133T0	CGTCTATCTCAGACTTCGGCAAGC	TCTGTGAACACACGCTTGTTGG
PITG_21561T0	AAGGGCCCTGGAAAGCCATTAC	CTGTGCGTGGGTTTGAACTTGG
PITG_08755T0	ACAACGCGATCAAGAACCGCTAC	CGCTTGTAGAGGTCTCGGACATTG
PITG_05990T0	AACAGCATTCCCGGACGAACTG	ACTTGGGTCCAATCGGTTTCGC
PITG_11731T0	CATGACCCATGCGCAGAAATATCG	TTCTTCTGGTTGCGCAGAACGC
PITG_12749T0	TCACCTGTGAGCCAAACGAGTG	AAGCGAGCGTGTTCGTCTTCTG
PITG_17559T0	CTTGCTCGCTGCATGCAAAGTC	TGTTCCACTGTAATGCGGTTGGG
PITG_06748T0	ACAGCCAGAACGAGACCTTTGAC	TGTAGTTGCCAGGCTGTTGACG
PITG_01528T0	ACGTGCTCGAGTCGTTAATGGTG	TGATCACGCAACAGCTCCAAGC
PITG_08697T0	ATTGCGTCGTCCTGAGCTTTCC	CCGATGAAACCAGTTGGAAGCG
PITG_00038T0	TTATCGCTCAGCAGCTCAAGGG	TCGTCCGCGCATTAATGATTTCC
PITG_19361T0	TACGCAAGCTCCCAGTGTCTAC	TCAACACGCTTGGGTCGATCTTG
PITG_17567T0	AAGCACGGCATTTGGTCCGAAG	ACCGCGAGGAAACTCCTTGATG
PITG_03670T0	TTCACGGCCACAAGTGGAAACG	TGCGTGTGTGCGTACTTGAGTC

PITG_20209T0	TTGACTTCGCGGAGGACAAAGC	TAGAAGCGAGGGTCTGGTCAAC
PITG_19153T0	AAGATGACGTCAAGTGCGCAGAG	AGCCATCTGAGCGGATGTGTTG
PITG_19362T0	GAAAGGACCGAGGGAGAGAAGAAG	AGTTGCATTTGCAGCCCTCCTC
PITG_02143T0	TCCAAGCACTCGCTTCACCATC	TTGGTCGGAAATGGCCATCGTG
PITG_11562T0	ACATCGCTGCAACAGGAACAGC	AACTCCTGCTGGACGCTTTGTG
PITG_11710T0	AGCACGAACTCGACATGCTTGG	TGCATTAGCATTTGCATGGTTGGG
PITG_11223T0	ACTGCAAGTGATCCGCATCGAG	AGACGCTGGTGATGCCTTTGTC
PITG_08410T0	ATCGCCCAGTTTACGTCTGTGC	TCCACTTCACCCGTTGCCAAAC
PITG_14400T0	ACCCGCAGATGACCTTAATCCC	AGGCATGAAGCACCCGTGAAAG
PITG_17552T0	CACAACCTCGACAGACAGTGAC	ACTGAACGACAGTGAAGAATGCG
PITG_05317T0	AAGCTCAAGTTGCTGCCAAAGC	GATAACACTCGCGAACCAACGC
PITG_14364T0	TCGGAGTTCAGCCAATTGAGCAC	TCCATGAGGAAGTCCATGGAGCAG
PITG_00514T0	ATTGCAGCCCTTTGGAACCG	TTCCTGCACGTCTTGTCGGTAG
PITG_05595T0	TGACAAGTCTTCGCCCGAGTTC	AGGTGAAGCGTCCGACAAGATG
PITG_11647T0	TCCGACGCTGATTCACTTCCATC	TCAACTTGTCGTGGCGGTGTTG
PITG_10311T0	CATGCCTTAACGCACGTTTGCC	AACGCATGGCGTGACCTTTACG
PITG_10310T0	TGGGATCAGCAAAGACGACCTC	TGGGAATTCGCCCGAATCACTG
PITG_19607T0	ACGAAAGCGTTCAGTGTCATTCC	ATGGCATCTCGGTTTGCAGTGG
PITG_10768T0	GCAAGCGATTTAGCACGTCTGG	TACGCATCGCATGGAGCTTTCG
PITG_00982T0	ACGGAATGGTTCTGCAACGAGTG	ATCAAGCTTGGCGCATTTCAGG
PITG_02520T0	TGGCTGTCATCCACTCGAACTC	GTTGCCACGTATTCACCTCCAG
PITG_07109T0	TCGCACGAGTTGCAGTTGAATG	ATTCCAGCGAAACGTCCTCCTC
PITG_01850T0	ACCCTGAGCTGGTGTTTAGCAG	TCCACTCATCGCAGTTGTGGAC

Supplementary Table 1. Primers used for the RT-qPCR experiments. 49 primers were designed for real time PCR experiments, 28 of these span in exon- exon border to prevent amplification of genomic DNA.