

RESEARCH

Open Access

# The Regulatory Network of *Pseudomonas aeruginosa*

Edgardo Galán-Vásquez, Beatriz Luna and Agustino Martínez-Antonio\*

## Abstract

**Background:** *Pseudomonas aeruginosa* is an important bacterial model due to its metabolic and pathogenic abilities, which allow it to interact and colonize a wide range of hosts, including plants and animals. In this work we compile and analyze the structure and organization of an experimentally supported regulatory network in this bacterium.

**Results:** The regulatory network consists of 690 genes and 1020 regulatory interactions between their products (12% of total genes: 54% sigma and 16% of transcription factors). This complex interplay makes the third largest regulatory network of those reported in bacteria. The entire network is enriched for activating interactions and, peculiarly, self-activation seems to occur more prominent for transcription factors (TFs), which contrasts with other biological networks where self-repression is dominant. The network contains a giant component of 650 genes organized into 11 hierarchies, encompassing important biological processes, such as, biofilms formation, production of exopolysaccharide alginate and several virulence factors, and of the so-called quorum sensing regulons.

**Conclusions:** The study of gene regulation in *P. aeruginosa* is biased towards pathogenesis and virulence processes, all of which are interconnected. The network shows power-law distribution -input degree -, and we identified the top ten global regulators, six two-element cycles, the longest paths have ten steps, six biological modules and the main motifs containing three and four elements. We think this work can provide insights for the design of further studies to cover the many gaps in knowledge of this important bacterial model, and for the design of systems strategies to combat this bacterium.

## Background

*Pseudomonas aeruginosa* is a metabolically versatile Gram-negative bacterium, able to express a wide variety of virulence factors. These allow *P. aeruginosa* to grow in soil and marine habitats, as well as on plant and animal tissues. It is also a significant source of bacteraemia in burn victims, urinary-tract infections, hospital-acquired pneumonia and predominant cause of morbidity and mortality in cystic fibrosis patients [1]. All these makes *P. aeruginosa* the most studied bacterial model regarding the control of pathogenic determinants and the third bacterial model more studied with respect to their molecular biology -after *Escherichia coli* and *Bacillus subtilis*-. The genome sequence of *P. aeruginosa* strain PAO1 was reported in 2000 [1], and since then numerous databases and genomic

resources have been implemented to study their molecular and pathogenic biology [2-4].

It is well known the importance of gene regulation on the organisms' performance as this process defines their metabolic, adaptive and pathogenic capabilities. In this work, we report a collection of known regulatory network interactions connecting transcription factors (TFs), sigma factors ( $\sigma$ ), and anti-sigma factors to their target genes in *P. aeruginosa*. This transcriptional regulatory network (TRN) constitutes the third largest one of any bacteria reported to date. We proceed to analyze the main topological properties of this network and the main functional interactions among their regulatory components. We hope these results will provide insights and guide future studies to increase our knowledge on this important bacterium.

\* Correspondence: amartinez@ira.cinvestav.mx  
Departamento de Ingeniería Genética, Cinvestav, Km. 9.6 Libramiento Norte  
Carr. Irapuato-León 36821 Irapuato Gto. México

## Results and discussion

### The transcriptional regulatory network (TRN) of *Pseudomonas aeruginosa*

With the aim of summarizing all the documented action of the regulatory machinery over the genes encoded in the genome of *P. aeruginosa*, the available published data was searched using a combined strategy: 1) regulatory interactions from dedicated biological databases were extracted [2-4] and, 2) searches in the original literature were performed (see Figure 1 for the general strategy). Both sources of information were verified by analyzing the corresponding papers. Methods frequently used to study transcriptional regulation included microarray analyses and their validation, promoter activity through transcriptional fusions, RT-PCR, EMSA assays and DNA-foot printing [Additional file 1 which contains a complete description of the network interactions along with their experimental evidences and references]. As of May 2010 our curated network consisted of 1020 regulatory interactions among 690 genes products, including 76 transcription factors, 14 sigma factors (nine of these defined with extra-cytoplasmic functions -ECF-), 7 anti-sigma factors, and 593 target genes (Figure 2, a poster version of this figure is available as additional file 2). Given the 5,570 predicted protein coding genes of *P. aeruginosa* PAO1 (the strain on which most of the network reconstruction is based) our network represents roughly 12% of these genes. On the other hand, the regulatory machinery predicted in this bacterium comprises around 500 proteins: 26 sigma factors (one  $\sigma^{54}$ , eight  $\sigma^{70}$ , and 17 of ECF families), the rest corresponding to transcription factors distributed in at least 44 families [5]. This network then, represents roughly 54% of sigma factors and 16% of all the TFs encoded by this

bacterium. In the following sections we will report the structural and the functional properties of this network.

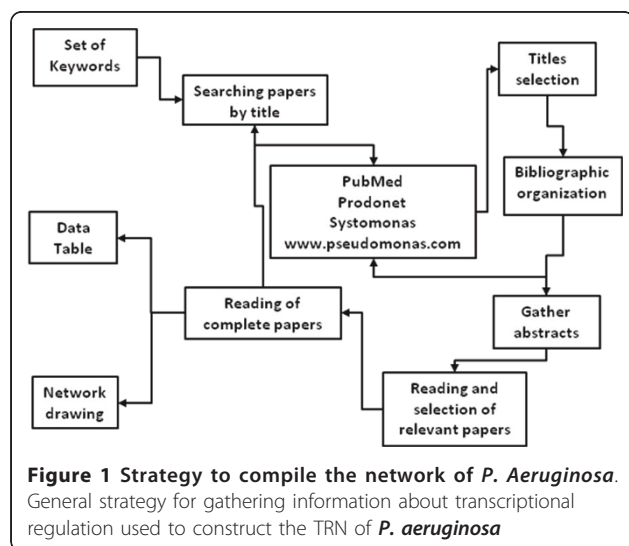
### Topological description of the TRN in *P. aeruginosa*

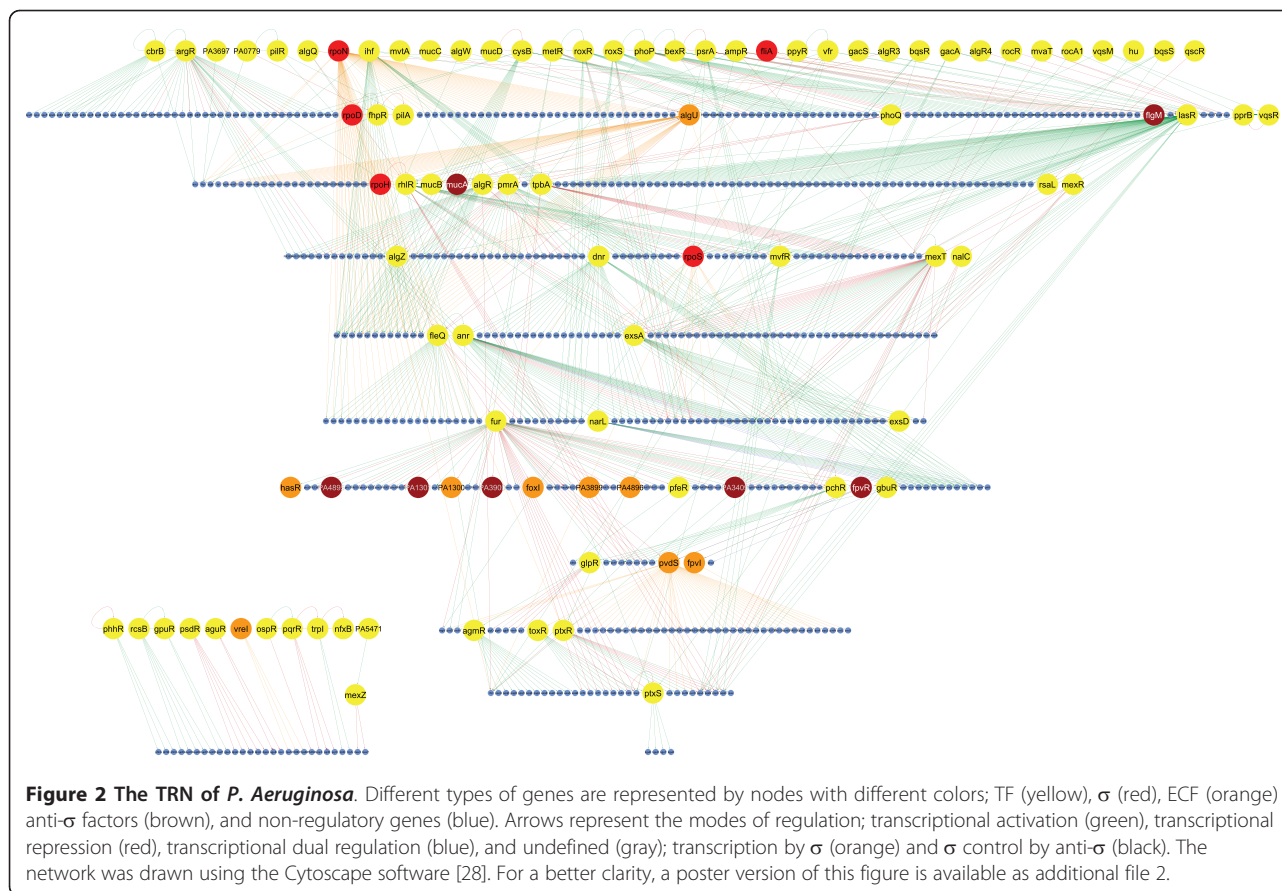
#### Degree distribution

In network and graph theory, the degree ( $k$ ) of a node (gene) is defined as the number of interactions that it has with other nodes. Here we determined the gene's mean degree as the arithmetic average of all the node-degrees ( $k$ ) in each network [Additional file 1]. This result implies that each gene in the TRN of *P. aeruginosa* is connected, on average, with 3 other genes. In directed networks, as in the case of regulatory networks, we can define input ( $k_{in}$ ) and output degree ( $k_{out}$ ) as the number of arrows that enter and leave genes respectively, which corresponds to the number of TFs that regulate a certain gene, and the number of genes that a TF is regulating. The degree distribution gives the probability  $P(k)$ , of finding a node with degree  $k$  [6]. This measure quantifies the diversity of gene-degrees in a network and allows determining which theoretical network is more similar to the type of network we are working with (i.e. classical random networks, scale free, small-world, etc. [7,8]). Some authors claim that biological networks present a well-known distribution closer to a power law  $P(k) = Ak^{-\gamma}$  which indicates that a few genes are highly connected (they are called hubs), while most of them have low connectivity [9]. The constant  $A$  ensures that the  $P(k)$  values are normalized towards 1, and  $\gamma$  is a parameter that provides information about the network structure; for networks with  $\gamma > 3$  a lot of the properties for scale-free networks are not present, which are present for  $2 < \gamma < 3$ , where there is a hierarchy in the degree of nodes, from the most to less connected ones. However, for  $\gamma = 2$ , the highest degree node influences a large fraction of all nodes [9]. In the case of the TRN of *P. aeruginosa*, we find  $A = 0.8856$  and,  $2 < \gamma < 3$  for the input degree distribution (Figure 3A), but without a good trend of this type for the output and overall degree distributions. Because of this, we show instead their corresponding cumulative distributions  $P(k_{out} \leq K)$ ,  $P(k \leq K)$ , (Figures 3B and 3C respectively). Overall the fact that a few genes are highly connected remains valid.

#### Clustering coefficient

The clustering coefficient  $C$ , is a measure that indicates the probability that two genes with a common neighbor in a graph are also interconnected; that is to say, the clustering coefficient quantifies what so much the local neighborhood of a gene is as member of a group of genes. It is common for networks to exhibit a decreasing value of  $C(k)$  with respect to the degree  $k$ , such that in small groups or modules of genes the elements are well





connected, but as the group increases in size the elements are progressively less connected. The regulatory network of *P. aeruginosa* shares this general clustering property in (Figure 3D).

### Connectivity

Connectivity in a network refers to the associations between every pair of genes.

Connections can be via a direct link or indirectly through a series of intermediate interactions. Connected components are defined for undirected networks, and give us information about how much are connected the elements in a network and their modular structure. Sometimes it is necessary to consider the network as undirected, since it allows us to capture different types of information to perform a better analysis. In the case of the TRN in *P. aeruginosa* there are 12 connected components, with one giant component containing 650 genes, while the rest contain at most six genes. Each connected component in the TRN possesses at least one TF or  $\sigma$ . A skeleton of 65 TFs and 13  $\sigma$  maintain cohesive this giant component (the 12% of its components). We consider that a connected component is composed by  $n$  nodes, and calculate the relative frequency  $P(n)$  for

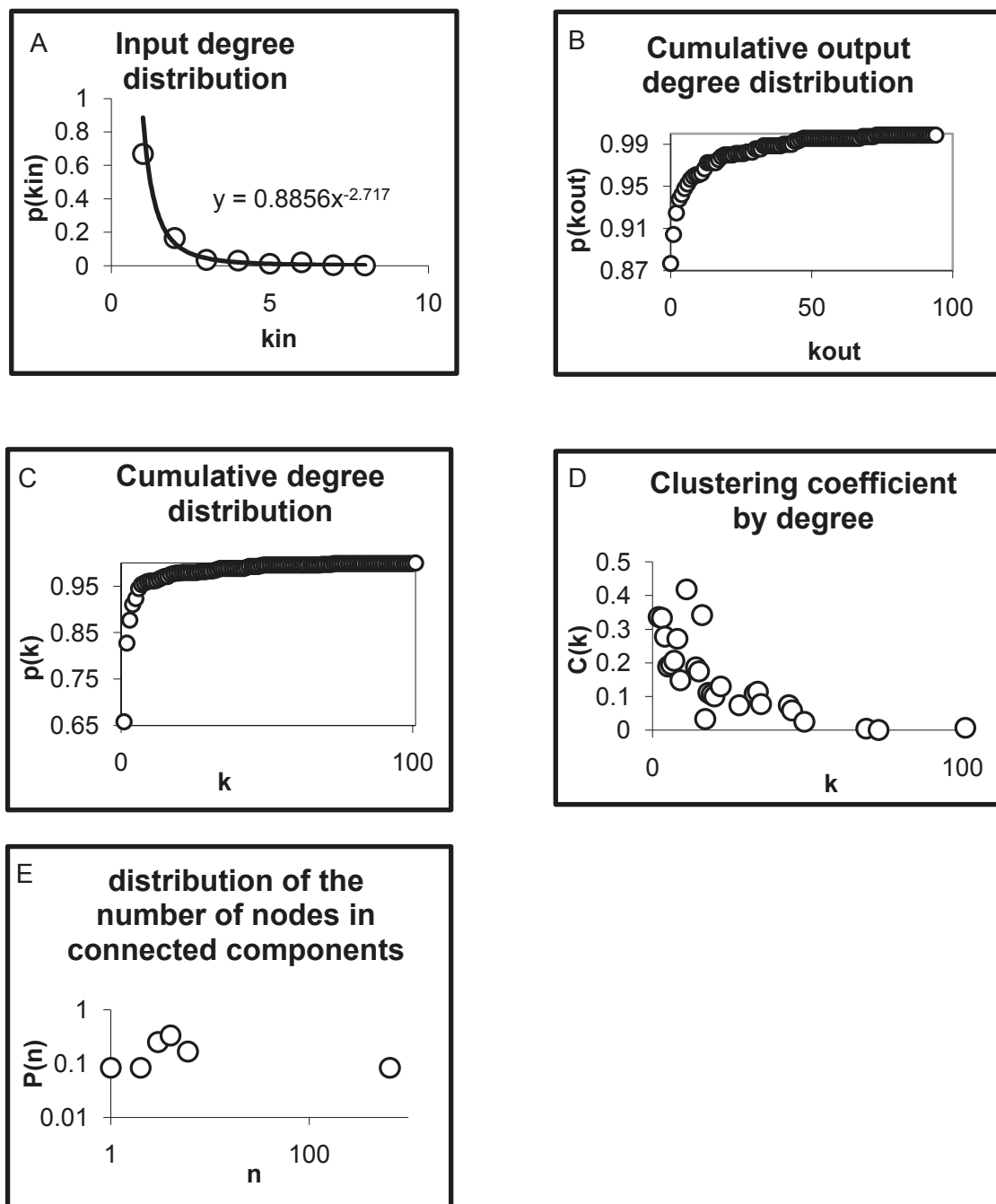
every possible  $n$ , which give us the distribution of the number of nodes in a connected component (Figure 3E), [see also Additional file 1].

### Functional organization of the regulatory network

In order to discern the functional organization of a regulatory network we can study the following aspects of the TRN: i) the regulatory mode and connectivity of each component of the transcriptional machinery and, ii) the manner in which endogenous and exogenous information, relevant for transcriptional regulation, enter and pass through the regulatory machinery until conclude on promoters of target genes. All these computes should be associated with the biological functions of the respective genes. In this sense, some interesting findings in the TRN of *P. aeruginosa* are discussed below.

#### Activation is the dominating activity in the TRN

Analysis of the mode of regulation (excluding interactions by  $\sigma$  and anti- $\sigma$ ) showed, that activation is by far the dominating regulatory activity in *P. aeruginosa* (Table 1). For comparison, the *E. coli* network also shows a higher tendency for activation instead of repression although in *P. aeruginosa* this difference is more



**Figure 3 Topological measurements of the TRN of *P. Aeruginosa*.** Topological measurements of the TRN of *P. aeruginosa* considering the entire network with 690 genes and 1020 regulatory interactions; A) input degree distribution, B) cumulative output degree distribution in the undirected graph, C) cumulative degree distribution, D) clustering coefficient by degree, and E) the distribution of the number of connected components in the network, see details in the text and Methods section.

pronounced. This dominant mode of regulation is also evident in the sub-network consisting of TFs and  $\sigma$  factors in *P. aeruginosa* (Figure 4), and the same was also observed in *E. coli* [10]. Positive regulation in these networks might explain why once a biological process is triggered it can run from the beginning to the end of

defined regulatory pathways, perhaps giving place to conditioned memory as has been observed in *E. coli* and *Saccharomyces cerevisiae* [11,12].

#### **Most of the TFs are positively auto-regulated**

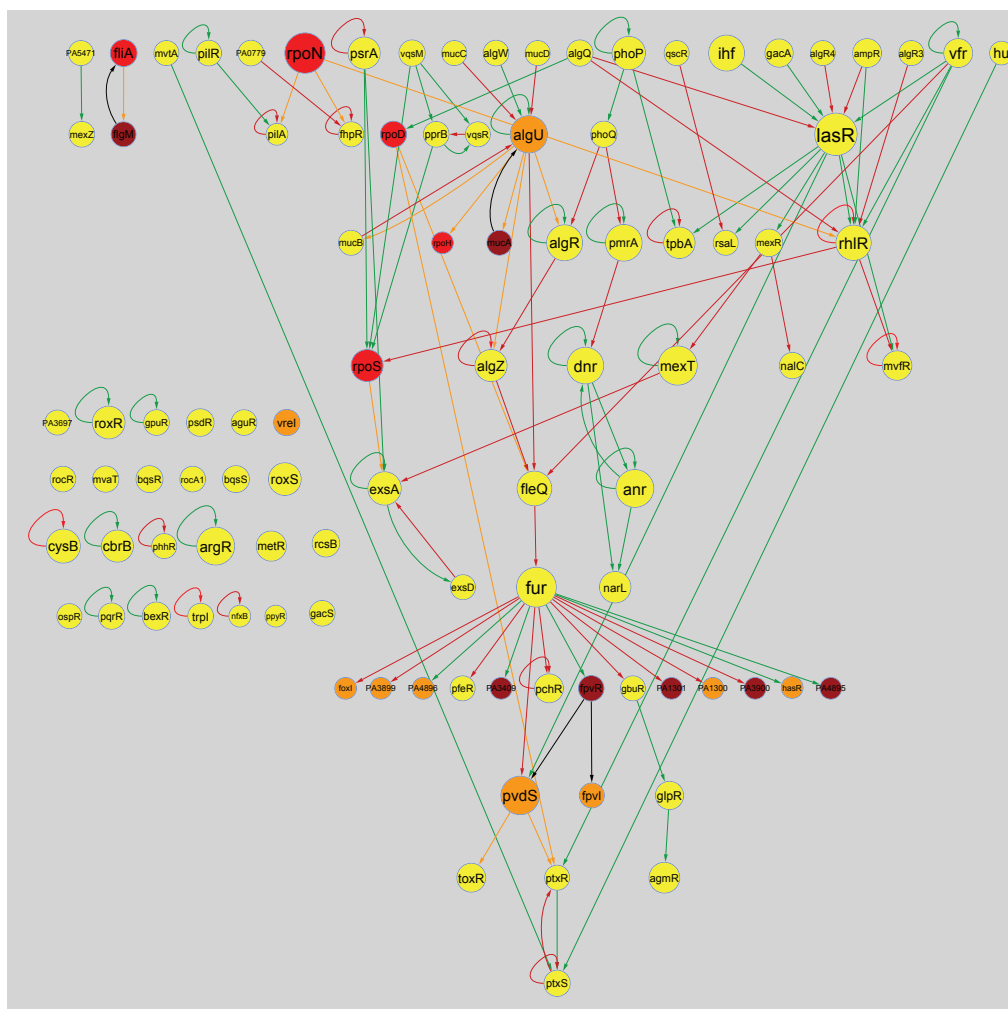
The mode in which regulatory genes are auto-regulated is important for network dynamics. It is known that

**Table 1 Network statistics**

	Whole network	Sub network
Number of TFs	76	76
Auto-regulations	29 (38.15%)	29
Positive auto-regulations	16 (55.17%)	16
Negative auto-regulations	13 (44.82%)	13
Regulatory arrows	1020	122
Positive arrows	779 (76.37%)	70 (57.37%)
Negative arrows	218 (21.35%)	52 (42.62%)
Dual arrows	11 (1.07%)	0
Unknown arrows	12 (1.17%)	0
Average path length	4.08	2.13
Maximum out degree	95 (lasR)	14 (fur)
Maximum in degree	8 (rhII)	7 (rhIR)

The number and sign of interactions and other topological measures for the whole network and the sub-network, which includes only the regulatory machinery of *P. aeruginosa*, are shown (Figure 4).

auto-repressions are important controllers to keep homeostatic levels of biological functions, while auto-activation is a condition to reach multiple steady states and differentiation [13]. Operatively, the TFs are important points of genetic control since they are the master switches whose regulatory activity extends over many genes. Consequently, it is not surprising to realize that nature has developed self-regulation on these genes as a quick and effective mode of control over a wide range of physiological processes. In contrast to what happens in *E. coli* and *B. subtilis*, auto-regulation is mostly positive for the TFs of *P. aeruginosa* (Table 2 and Figure 4). Given that negative self-regulation maintains homeostasis, is normal to observe this mode as dominant in regulatory networks as in *E. coli* [10,14], and *B. subtilis* (Galán-Vásquez *et al*, unpublished observations), with the exception of developmental processes for biofilm



**Figure 4 Transcriptional machinery sub-network in *P. aeruginosa* (TF and sigma factors).** Nodes and arrows are the same as in Figure 2 with the exception that it lacks of the non-regulatory genes. The network is presented in a hierarchical structure, in order to better appreciate the hierarchical organization among TFs and  $\sigma$ .



**Table 2 Top ten most influencing regulators in the TRN of *P. aeruginosa***

Transcription factor	TFs and $\sigma$ regulated (excluding self-regulation) TFR	Total of genes regulated GR	Type of $\sigma$ used by the regulated promoters SF	Number of TFs used as co-regulators CR	G coefficient
lasR	6	89	3	20	0.155527032
Fur	14	54	3	6	0.114938218
mexT	1	46	3	15	0.10771137
Vfr	4	3	4	12	0.09930047
algR	1	29	3	14	0.097211089
Anr	2	42	2	12	0.086724434
lhf	1	29	2	14	0.085306327
ptxR	1	11	3	12	0.082955889
rhIR	2	25	2	13	0.082890819
algW	1	1	3	7	0.062073371

The highest values of G coefficient obtained from the TFs of whole network are shown.

and flagella formation in *E. coli* where self-activation is enriched [10]. This observation is in agreement with the postulation that auto-activation is normally found at the core of differentiation and development processes. Dynamic analysis shows that auto-activation causes a slow and delayed response compared to auto-repression and simple regulation [15]. It is supposed that this delayed response gives time enough for regulation pass through different check points, before the differentiation processes goes ahead. Given the bias of knowledge toward the study of only a minor part of the network in *P. aeruginosa* (virulence and pathogenic processes), it is premature to conclude whether that distinction should be a property of the whole network, or if, it mainly represents an evolutionary design for the execution of pathogenesis and virulence actions by this bacterium. Using the database of Reciprocal Best Hits (RBH) orthologs genes in bacteria [16], we found orthologs for seven regulatory genes from *P. aeruginosa* in *E. coli*, in both organisms with experimental evidence about their mode of self-regulation [Additional file 1]. From these, the three negatively auto-regulated genes in *P. aeruginosa* conserve this mode of self-regulation in *E. coli*. For the remaining four positively auto-regulated in *P. aeruginosa*, one conserved their positive mode of auto-regulation in *E. coli* and the rest three changes; two are dual and one self-regulate negatively in *E. coli*. Although this information is scarce, it might be possible that positive auto-regulation in TFs might be effectively selected in *P. aeruginosa*.

#### Path lengths

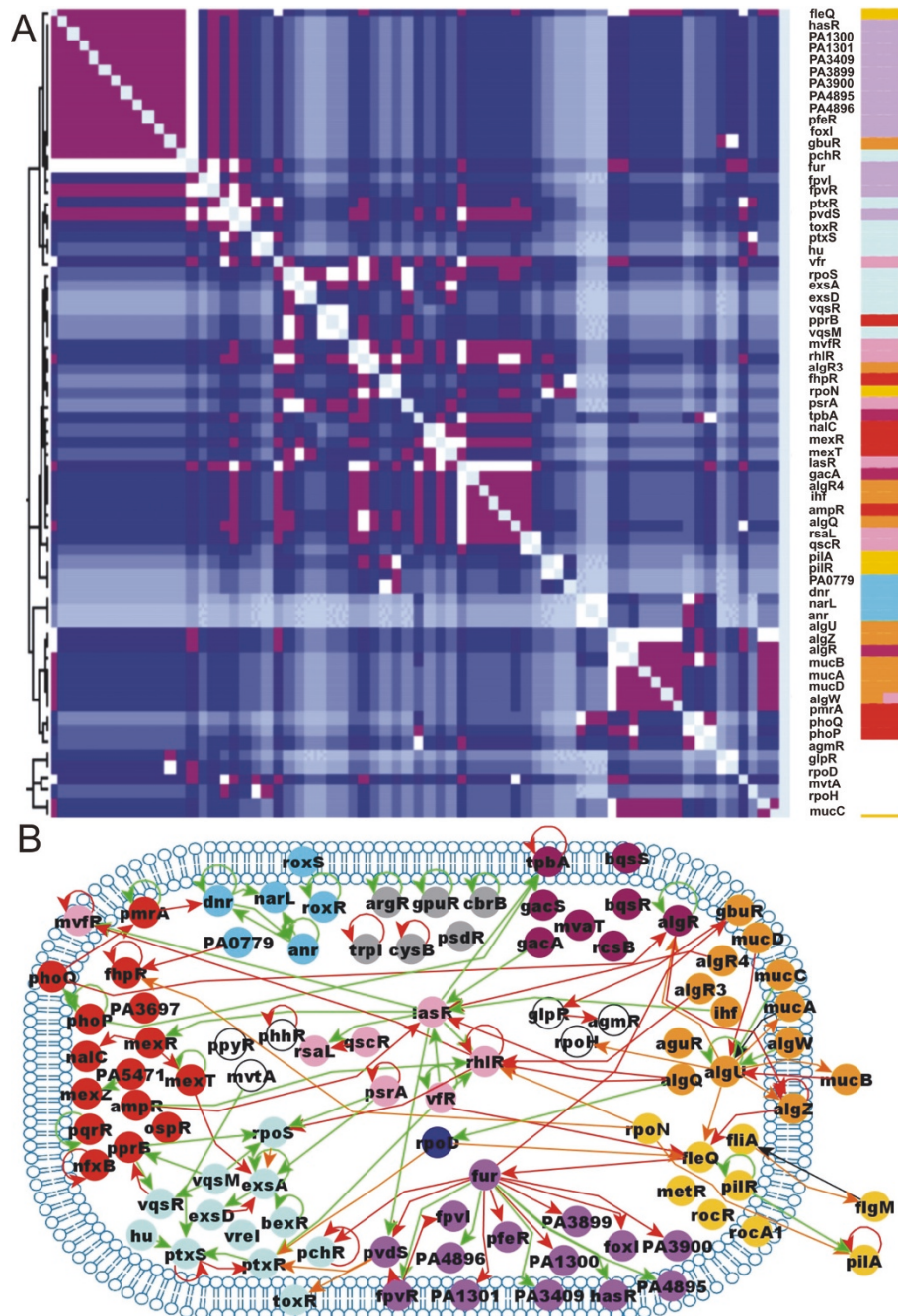
A path in a TRN refers to a chain of regulatory interactions between the genes constituting it. The longest path in the TRN of *P. aeruginosa* consists of 11 steps. This is almost the same size than in their *E. coli* counterpart where the longest paths include 14 steps. The

size of paths in *P. aeruginosa* is interesting if we consider that the TRN is far from complete, therefore it is reasonable to expect that longer paths can be found as the network will be further characterized. The longer regulatory path in *P. aeruginosa* goes along biological processes that include alginate biosynthesis, iron metabolism and pioverdine synthesis, implying that these processes could be physiologically connected (Figures 2 and 5). The most frequent path-size (1102 times) found in the *P. aeruginosa* TRN consists of five steps [Additional file 1].

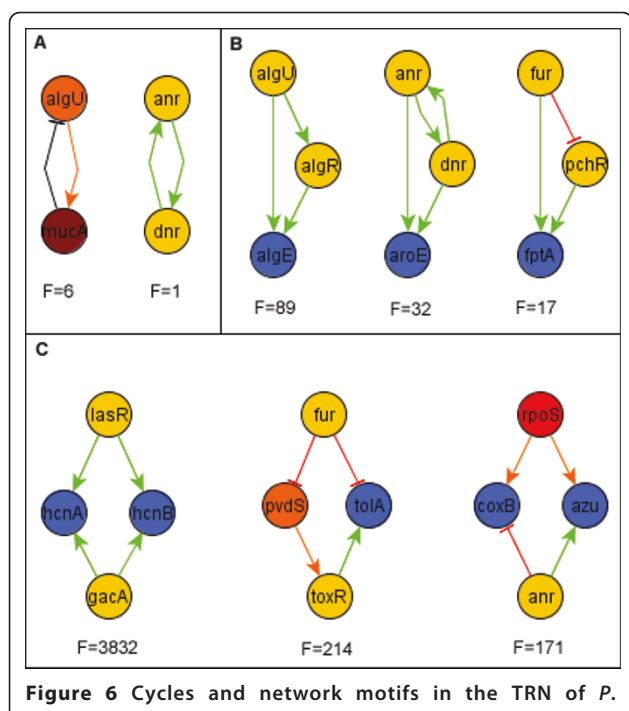
Short paths are common in metabolic and signal transduction networks, since this arrangement ensures fast and efficient response to changes in food use and to environmental perturbations [6]. The longest paths in the TRN of *E. coli* include regulatory process for biofilm formation and flagella assembly [10], both of these are considered development processes and are also amongst the longest paths in *P. aeruginosa*.

#### Cycles

Biological systems frequently contain positive or negative feedback loops. Multi-element biological cycles (with two or more components) can also be positive or negative (depending on the product of the signs of the constitutive interactions). The existence of positive cycles is a necessary condition to have multiple steady states or attractors. Negative cycles are important to keep homeostasis, since they maintain the system functioning through periodic orbits. Besides self-regulation, multi-element circuits in a TRN can be defined as self-enclosed paths. Until now, excluding self-regulation, there are only seven cycles with two regulatory factors in *P. aeruginosa* (Figure 6A); six negatives (*algU-mucA* and *algU-mucB*, for the control of alginate synthesis; *exsA-exsD*, for the control of secretion systems; *flgM-flhA*, for motility control; *pprB-vqsR* which enhance



**Figure 5 Functional modules.** Determination of biological modules more represented in the regulatory machinery of *P. aeruginosa*. A) Using a short path metric distance  $1/D^2$  where  $D$  is the distance between two nodes. White color represents interactions with  $D = 1$ , purple color shows the interactions with  $D = 2$ , interactions with  $D = 3$  are displayed in blue, with diminishing intensity as distance increases until a light blue color, representing interactions that do not exist. B) The major clusters of biological functions were also identified by analyzing the scientific literature: quorum sensing (pink), alginate biosynthesis (orange), iron metabolic (violet), nitrogen metabolic (cyan), motility (yellow), antibiotics resistance (red), expression of virulence factors (sky blue), biofilm formation (purple) and amino acid metabolism (gray). For the sake of comparison between both approaches, the same colors were used for the regulatory machinery in both figures.



**Figure 6 Cycles and network motifs in the TRN of *P. aeruginosa*.** A) The most abundant two-node motif are the negative feedback with a frequency (F) of six in the network. The three most abundant motifs of three and four nodes are shown, B) the most abundant motif of three nodes is the coherent feed forward loop with a frequency of 89, C) the most abundant four-node motif is the bi-fan with a frequency of 3832. The frequency of each cycle and motif was compared with the average of those found in 1000 random networks.

exotoxin A production and; *ptxR-ptx S* that control protease and pyocyanine synthesis) and one positive (*anr-dnr*, for the control of aerobic/anaerobic respiration), and no cycles with more TF-elements are present in the TRN. The dynamics of positive and negative two-node cycles has been widely studied. They represent important core components for network dynamics acting as robust switches to respond to signals from environmental conditions [17].

#### Motifs

A motif in a TRN is a topological structure that is more frequented than expected [18]. The most represented motifs in the *P. aeruginosa* network are those formed by three and four genes (Figure 6), [Additional file 1]. Previous research suggests that motifs represent elements for optimal network design given their relationship with the network dynamics and structural stability. The prevalence of certain types of motifs has been considered a product of the evolution acting on the organization of biological networks [19-21]. In particular, motifs such as feed-forward loops (FFL; networks with three vertex, composed of two input transcription factors, one of which regulates the other, both jointly regulating a

target gene) have a higher abundance in TRN than expected from random networks with the same number of nodes and arrows [18,22]. The dynamic behavior of feed-forward loops has been extensively analyzed [23]; these studies revealed that FFL have two main functions: a) to speed up the response time of the target gene (incoherent FFL, when the signs of the direct and indirect regulation are opposites) and, b) to act as sign-sensitive delays for one of the two TFs (coherent FFL, with the same sign for both the direct and indirect regulation). Considering all the biological process where they participate, FFL are also implicated in pulse generation and cooperativity. In *P. aeruginosa* the most common motifs are those of three nodes known as coherent feed-forward loops [23], where the sign of the interactions is the same, positive in this case (Figure 6B). This type of motif is present 89 times in the *P. aeruginosa* regulatory network. Additionally, we found that the most common motifs of four nodes, which occurred 3832 in the network, are those known as bi-fan, where two TFs each positively co-regulate to two target genes (Figure 6C). This motif is also frequent in other organisms such as *S. cerevisiae* and *E. coli* [24].

#### Hierarchical organization of the TRN of *P. aeruginosa*

A hierarchical organization is given by a directed informational flux, beginning from the most influencing regulators. In this way, the TFs constitute the skeleton and the non-regulatory genes are the leaves in a hierarchical network (Figure 2 and Additional file 1). The first level is populated by 33 TFs and 2 sigma factors. The origins [25], which are the points of informational inputs into the network, are set at this level. The second level is the most populated level but includes a high proportion of non-regulatory genes. Most of  $\sigma$  are set at higher levels, except for those involved in iron metabolism, which, as also observed for *E. coli*, are in the lowest levels as dedicated sigma factors for specific functions.

#### Co-regulation

Promoter and regulatory regions are zones in transcription units where regulatory information is integrated. This regulatory integration is evidenced by the presence of DNA-binding sites for multiple regulators or promoters for different  $\sigma$  respectively. In the TRN of *P. aeruginosa*, regulators of level one can co-regulate with regulators of any other level and this co-regulatory activity diminishes as TFs are set lower in the network hierarchy. Regulators that frequently co-regulate with other regulators (in parenthesis the number of co-regulations) are: *lasR*(20), *mexT*(15), *algR*(14), *ihf*(14), *rhlR*(13), *anr*(12), *ptxR*(12). On the other hand, the most regulated genes are (in parenthesis the number of TFs regulating it): *rhlI*(8), *rhlAR*(7), *algD*(7), *alg44*(6), *alg8*(6), *algA*(6), *algE*(6), *algF*(6), *algG*(6), *algI*(6), *algJ*(6), *algK*(6), *algL*(6), *algX*(6), *hcnA*(6), *lasR*(6), *algU*(6). It is interesting that the most regulated genes in



both *E. coli* and *P. aeruginosa* encode for TFs, FhlCD (flagella synthesis) and RhlI (quorum sensing) respectively.

#### **Most influencing regulators in the TRN of *P. aeruginosa***

The most influencing regulators in a regulatory network are called “global regulators” and are defined by a series of operative properties, including: i) they should regulate a large number of genes; ii) should regulate other sigmas and regulators; iii) should co-regulate together with many TFs and, iv) their target genes should have promoters using more than one kind of  $\sigma$  [26]. All these properties were computed for regulators found in *P. aeruginosa* (see Methods section) and the top ten are shown in Table 2. A coefficient  $G$  was introduced here, which indicates if a regulator is more or less global taking into account the regulatory criteria mentioned above. The most influencing regulators in *Pseudomonas* have a lower qualification than the corresponding seven global TFs in *E. coli*. This might be due to the limited knowledge of the transcriptional regulation in *P. aeruginosa* compared to *E. coli*.

#### **Biological processes in *P. aeruginosa* TRN**

Defining functional modules in a formal computational way is a difficult task. However, it has been shown that employing a simple metric of shorter distances among TFs in the *E. coli* network, it might be possible to recover modules with a good approximation to those that are manually defined, on the basis of the knowledge of biological functions of their products [27]. In this work we used this metric for the TFs and  $\sigma$ -anti- $\sigma$  sub-network of *P. aeruginosa* (Figure 4), and get the following biological modules: alginate biosynthesis, quorum sensing, iron capture and metabolism, production of virulence factors, antibiotic resistance and motility (Figure 5A). This finding was corroborated by manual inspection of TFs participating in the same biological processes (Figure 5B). It is clear that the processes that are more thoroughly studied in *P. aeruginosa* correspond to those related to pathogenesis and virulence properties while little attention has been given to biological processes, such as, central metabolism, membrane biogenesis, cell-division, etc. Most of the best-studied biological processes are connected, beginning from alginate biosynthesis to quorum sensing, and from there to those involved in the production of virulence factors. Additionally, there is a directed regulatory connection from alginate biosynthesis to iron metabolism and to some mechanisms of antibiotic resistance (Figure 5B). Since these processes act cooperatively during infection and pathogenesis, it is very important to give a detailed characterization of *P. aeruginosa* regulatory network. The latter may lead to the development of strategies to disrupt its connectivity, thus, possibly decreasing the pathogenicity of this bacterium.

## **Conclusions**

Here we report the topological and functional organization of the third largest regulatory network in bacteria. From our analysis, it is evident that the study of regulation in *P. aeruginosa* is biased towards particular biological processes, involved in pathogenesis and virulence. These processes include alginate and biofilm formation, production of virulence factors and antibiotic resistance, many of which are coordinated by quorum sensing in the bacterial population. Current data suggests, that motility, iron metabolism and anaerobic respiration might be less connected to these processes by now. All these processes are connected in the network via a hierarchical organization with 11 levels, and the connected parts of the network form a giant component with 650 genes, the 10% of them corresponded to TFs. Overall, the network has degree distribution and structural organizations as other biological networks known to date. A peculiar property of this network is the fact that its TFs are mainly auto-activated. This is the first time this mode of self-regulation is reported as dominant in a bacterial TRN. It remains to be revealed whether this property is really a characteristic of the entire network in this bacterium, or is just a property of this part of the network, which clearly controls adaptive, pathogenic and virulence processes. As it can be observed, regulatory information related to several important biological processes of *P. aeruginosa* is lacking; for instance, the regulation on the uptake of carbon sources and their metabolism, amino acid biosynthesis or cell-division. This bias makes difficult a complete analysis on the regulatory network of this bacterium and better compare it with regulatory networks of bacteria most characterized such as *E. coli* or *B. subtilis*. It might be that studying basic biological functions on these organisms we can understand the basis of their versatile metabolism, adaptiveness and pathogenicity. In special it is lacking the knowledge of the activity of the housekeeping sigma and transcription factors controlling activities of central metabolism. Because of this it will be very important for the community working on the biology of *P. aeruginosa* to study additional biological processes in order to have a more complete picture of the regulatory network in this bacterium. We hope this analysis will give insights in this direction to guide future work, with the aim of covering the many gaps of knowledge on this important bacterial model.

## **Methods**

### **Biological data and representation**

The general strategy for the curation of regulatory interactions is shown in Figure 1. Briefly, we searched PubMed with relevant key words, such as: *P. aeruginosa*,

sigma or transcription factor, transcriptional regulation, etc. Data on regulatory networks were obtained from the literature and compiled in an Excel table including experimental evidence and references. The Additional file 1 shows the complete information for the interactions of the entire network. The regulatory interactions were drawn in a form of network using the Cytoscape software [28].

#### Transcriptional machinery sub-network

With the aim of analyzing the regulatory behavior of the transcriptional machinery of *P. aeruginosa*, the regulatory interactions present only among TFs, sigmas and anti-sigmas from the whole network were extracted (Figure 4).

#### Computational analysis of the regulatory network

All the computational analyses on the network were made using the Octave free software <http://www.octave.org>. Analyses of degree, centrality, clustering coefficient, connectivity, cycles, paths and hierarchical levels were made according to previous definitions and following the approach as in [10]. Motif determination was made following the work by Uri Alon and coworkers calculating the probability of finding the same motif in a random network as the average of the motifs found in 1000 randomized networks, maintaining the same number of nodes, edges and the proportion of the type of regulatory interactions (positive, negative, dual) [24].

#### The G coefficient for global regulators

We computed the coefficient *G*, which indicates the global activity of a TF in a TRN as follows:

$$G = \frac{1}{4} \left( \frac{TFR}{N_{TF} + N_{SF} - 1} + \frac{GR}{N_G} + \frac{SF}{N_{SF}} + \frac{CR}{N_{TF} - 1} \right)$$

Where,  $N_{TF}$  indicates the total number of TFs (in the known network in each case),  $N_G$  is the number of non-regulatory genes, and  $N_{SF}$  is the number of sigma factors in the whole network. Additionally, *TFR* and *GR*, represents the number of TFs and non-regulatory genes regulated by each TF, respectively; *SF* represents the distinct sigma factors used by the promoters of genes regulated by each TF; and *CR* represents the number of TFs each TF co-regulates with.

#### Determination of biological modules/processes in the regulatory machinery network

With the aim of determining biological modules in the TRN we used a shortest path metric criteria among TFs and sigmas (we used the relation  $1/D^2$  where *D* is the distance between two nodes), as reported for *E. coli*

[27]. Additionally, we manually grouped TFs and sigma factors in agreement to the functional classification of their regulated genes [4].

#### Additional material

**Additional file 1:** This excel document contains additional information about: sheet 1) the type of regulatory interactions on the network, strain, experimental evidences and references; sheet 2) the components and length of directed paths in the whole network; sheet 3) the components and length of directed paths in the regulatory sub-network; sheet 4) representative motifs in the entire network with two, three and four genes; sheet 5) components and measures of the whole network and the sub-network; sheet 6) orthologs TFs with auto-regulation identified in *E. coli*.

**Additional file 2:** This is a poster version of Figure 2 to better appreciate the name of genes which are difficult to distinguish in the corresponding figure in the ms.

#### List of abbreviations

TF: transcription factor;  $\sigma$ : sigma factor; ECF: extra-cytoplasmic factors; TRN: transcriptional regulatory network.

#### Acknowledgements

Authors thank Angela Porras Dorantes and Viridiana Sanchez Paez for their help on compiling original literature, Johannes Klein for supplying some regulatory data from the ProDoric database and to Heladia Salgado for data management. We also thank Cei Abreu-Goodger, Gloria Soberón-Chávez, Carlos Cervantes, Luis Delaye and Jesús Campos for their useful comments on of the manuscript. EG-V received a CONACYT fellowship (234827) for a master's degree. This work was funded by young-researcher grants from CONCYTEG and CONACYT (102854) given to AM-A.

#### Authors' contributions

EG-V carried out the acquisition of data, analysis and interpretation of data, and drafted the manuscript. BL performed the structural analysis and interpretation of data, and drafted the manuscript. AM-A conceived the study, and participated in its design and coordination, the interpretation of data, and drafted the manuscript. All authors read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

Received: 15 December 2010 Accepted: 14 June 2011

Published: 14 June 2011

#### References

1. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrenner P, Hickey MJ, Brinkman FS, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E, Westbrook-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim R, Smith K, Spencer D, Wong GK, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock RE, Lory S, Olson MV: **Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen.** *Nature* 2000, **406**:959-964.
2. Klein J, Leupold S, Münch R, Pommerenke C, Johl T, Kärst U, Jänsch L, Jahn D, Retter I: **ProdoNet: identification and visualization of prokaryotic gene regulatory and metabolic networks.** *Nucleic Acids Res* 2008, **36**:W460-W464.
3. Choi C, Münch R, Leupold S, Klein J, Siegel I, Thielen B, Benkert B, Kucklick M, Schobert M, Barthelmes J, Ebeling C, Haddad I, Scheer M, Grote A, Hiller K, Bunk B, Schreiber K, Retter I, Schomburg D, Jahn D: **SYSTEMONAS- an integrated database for systems biology analysis of *Pseudomonas*.** *Nucleic Acids Res* 2007, **35**:D533-D537.
4. Winsor G, Van T, Lo R, Bhavjinder K, Whiteside M, Hancock R, Brinkman S: ***Pseudomonas* Genome Database: facilitating user-friendly,**

- comprehensive comparisons of microbial genomes. *Nucleic Acids Res* 2009, **37**:D483-D488.
5. Perez-Rueda E, Janga SC, Martínez-Antonio A: **Scaling relationship on the gene content of transcriptional machinery in bacteria.** *Mol. BioSyst* 2009, **12**:494-501.
  6. Réka A: **Scale-free networks in cell biology.** *J. Cell Sci* 2005, **118**:4947-4957.
  7. Watts J: **The "new" science of networks.** *Annu. Rev. Sociol* 2004, **30**:243-270.
  8. Erdős P, Rényi A: **On random graphs I.** *Publ. Math* 1959, **6**:290-297.
  9. Barabási A, Oltvai Z: **Networks biology: understanding the cell's functional organization.** *Nat. Rev. Genet* 2004, **5**:101-113.
  10. Martínez-Antonio A, Janga SC, Thieffry D: **Functional organisation of *Escherichia coli* transcriptional regulatory network.** *J. Mol. Biol* 2008, **381**:238-247.
  11. Tagkopoulos I, Liu YC, Tavazoie S: **Predictive behavior within microbial genetic networks.** *Science* 2008, **320**:1313-7.
  12. Amir M, Romano G, Groisman B, Yona A, Dekel E, Kupiec M, Dahan O, Pikpel Y: **Adaptive prediction of environmental changes by microorganisms.** *Nature* 2009, **460**:220-225.
  13. Thomas R: **Boolean formalization of genetic control circuits.** *J. Theor. Biol* 1973, **42**:563-585.
  14. Lagomarsino M, Jona P, Bassetti B, Isambert H: **Hierarchy and feedback in the evolution of the *Escherichia coli* transcription network.** *Proc. Natl. Acad. Sci. USA* 2007, **104**:5516-5520.
  15. Alon U: **Network motifs: theory and experimental approaches.** *Nat. Rev. Genet* 2007, **8**:450-461.
  16. Moreno-Hagelsieb G, Latimer K: **Choosing BLAST options for better detection of orthologs as reciprocal best hits.** *Bioinformatics* 2008, **24**: 319-24.
  17. Watts D, Strogatz S: **Collective dynamics of 'small-world' networks.** *Nature* 1998, **393**:440-442.
  18. Shen-Orr S, Milo R, Mangan S, Alon U: **Network motifs in the transcriptional regulation network of *Escherichia coli*.** *Nature Genet* 2002, **31**:64-68.
  19. Doyle J, Csete M: **Motifs, control, and stability.** *PLoS Biol* 2005, **3**(11):e392.
  20. Prill RJ, Iglesias PA, Levchenko A: **Dynamic properties of network motifs contribute to biological network organization.** *PLoS Biol* 2005, **3**(11):e343.
  21. Mangan S, Zaslaver A, Alon U: **The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks.** *J. Mol. Biol* 2003, **334**:197-204.
  22. Balázsi G, Barabási AL, Oltvai ZN: **Topological units of environmental signal processing in the transcriptional regulatory network of *Escherichia coli*.** *Proc. Natl. Acad. Sci. USA* 2005, **102**:7841-7846.
  23. Mangan S, Alon U: **Structure and function of the feed-forward loop network motif.** *Proc. Natl. Acad. Sci. USA* 2003, **100**:11980-11985.
  24. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U: **Network Motifs: Simple building blocks of complex networks.** *Science* 2002, **298**:824-827.
  25. Balazsa G, Babarasi A L, Oltvai ZN: **Topological units of environmental signal processing in the transcriptional regulatory network of *Escherichia coli*.** *Proc. Natl. Acad. Sci. USA* 2005, **102**:7841-7846.
  26. Martínez-Antonio A, Collado-Vides J: **Identifying global regulators in transcriptional regulatory networks in bacteria.** *Curr. Opin. Microbiol* 2003, **6**:482-489.
  27. Resendis-Antonio O, Freyre-González J, Menchaca-Méndez R, Gutiérrez-Ríos R, Martínez Antonio A, Ávila-Sánchez C, Collado-Vides J: **Molecular analysis of the transcriptional regulatory network of *Escherichia coli*.** *Trends Genet* 2005, **21**(1):16-20.
  28. Shannon P, Markiel A, Ozier O, Baliga N, Wang J, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res* 2003, **13**:2498-2504.

doi:10.1186/2042-5783-1-3

Cite this article as: Galán-Vásquez *et al.*: The Regulatory Network of *Pseudomonas aeruginosa*. *Microbial Informatics and Experimentation* 2011 **1**:3.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

