In silico Identification of the Indispensable Quorum Sensing Proteins of Multidrug Resistant *Proteus mirabilis*

Shrikant Pawar[1,2‡], Md. Izhar Ashraf[3,4†‡], Shama Mujawar[5], Rohit Mishra[6] and Chandrajit Lahiri[5*]

[1] Department of Computer Science, Georgia State University, Atlanta, GA, United States, [2] Department of Biology, Georgia State University, Atlanta, GA, United States, [3] Department of Computer Applications, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India, [4] Theoretical Physics, The Institute of Mathematical Sciences, Chennai, India, [5] Department of Biological Sciences, Sunway University, Petaling Jaya, Malaysia, [6] Department of Bioinformatics, G.N. Khalsa College, University of Mumbai, Mumbai, India

Catheter-associated urinary tract infections (CAUTI) is an alarming hospital based disease with the increase of multidrug resistance (MDR) strains of *Proteus mirabilis*. Cases of long term hospitalized patients with multiple episodes of antibiotic treatments along with urinary tract obstruction and/or undergoing catheterization have been reported to be associated with CAUTI. The cases are complicated due to the opportunist approach of the pathogen having robust swimming and swarming capability. The latter giving rise to biofilms and probably inducible through autoinducers make the scenario quite complex. High prevalence of long-term hospital based CAUTI for patients along with moderate percentage of morbidity, cropping from ignorance about drug usage and failure to cure due to MDR, necessitates an immediate intervention strategy effective enough to combat the deadly disease. Several reports and reviews focus on revealing the important genes and proteins, essential to tackle CAUTI caused by *P. mirabilis*. Despite longitudinal countrywide studies and methodical strategies to circumvent the issues, effective means of unearthing the most indispensable proteins to target for therapeutic uses have been meager. Here, we report a strategic approach for identifying the most indispensable proteins from the genome of *P. mirabilis* strain HI4320, besides comparing the interactomes comprising the autoinducer-2 (AI-2) biosynthetic pathway along with other proteins involved in biofilm formation and responsible for virulence. Essentially, we have adopted a theoretical network model based approach to construct a set of small protein interaction networks (SPINs) along with the whole genome (GPIN) to computationally identify the crucial proteins involved in the phenomenon of quorum sensing (QS) and biofilm formation and thus, could be therapeutically targeted to fight out the MDR threats to antibiotics of *P. mirabilis*. Our approach utilizes the functional modularity coupled with k-core analysis and centrality scores of eigenvector as a measure to address the pressing issues.

Keywords: *Proteus mirabilis*, urinary tract infection, quorum sensing, eigenvector centrality, k-core analysis

INTRODUCTION

Urinary tract infections (UTI) are the second most common infection prevalent amongst long-term hospital patients, second only to pneumonia. Failure to treat or a delay in treatment can result in systemic inflammatory response syndrome (SIRS), which carries a mortality rate of 20–50% (Jacobsen and Shirtliff, 2011; Schaffer and Pearson, 2015)[1] While *Escherichia coli* remains the most often implicated cause of UTI in previously healthy outpatients, *Proteus mirabilis* take the lead for catheter-associated UTI (CAUTI), causing 10–44% of long-term CAUTIs (Jacobsen and Shirtliff, 2011; Schaffer and Pearson, 2015)[1]. In comparison to normal cases, CAUTI is quite complicated and encountered by patients with multiple prior episodes of UTI, multiple antibiotic treatments, urinary tract obstruction and/or undergoing catheterization as also for those with spinal cord injury or anatomical abnormality (Jacobsen and Shirtliff, 2011; Schaffer and Pearson, 2015)[1]. Such complications of CAUTI caused by *P. mirabilis* arise from the usage of a diverse set of virulence factors by the organism to access and colonize the host urinary tract. These include, but are not limited to, urease and stone formation, fimbriae and other adhesins, iron and zinc acquisition, proteases and toxins and biofilm formation (Schaffer and Pearson, 2015). Despite significant advances made for studying *P. mirabilis* pathogenesis, a meager knowledge of its regulatory mechanism poses an urgent and pressing need to come up with unique health intervention processes for such patients.

In attempts to provide such health interventions, longitudinal, and epidemiological studies on *P. mirabilis* have been reported for extended-spectrum β-lactamase (ESBL) and AmpC β-lactamase (CBL) producers (Luzzaro et al., 2009; Wang J. T. et al., 2014) According to these studies, limited therapeutic options are available for management of such CAUTI which in turn reflects the imminent threats of multi-drug resistance (MDR) *P. mirabilis*. Such MDR phenomenon, exhibited by the gram-negative pathogens like *P. mirabilis*, can be attributed, besides other factors, to the blockade provided by the efflux pumps at the extra-cytoplasmic outer membrane for existing antibiotics entries and remainder drugs expulsion (Eliopoulos et al., 2008; Czerwonka et al., 2016). Besides providing MDR, the cases of CAUTI have been complicated by biofilms formed by the pathogenic *P. mirabilis* (Czerwonka et al., 2016). In fact, different lipopolysaccharide structures of the membrane have been implicated to the adherence of the pathogen on to the surfaces causing CAUTI. Furthermore, along with various other components of the membrane, several cytoplasmic factors interplay among themselves to regulate the cell-density dependent gene regulation. This enables the bacteria for cell-to-cell communication, a phenomenon known as quorum sensing (QS) (Rutherford and Bassler, 2012). Besides other phenotypic traits, QS controls the expression of the virulence factors responsible for pathogenesis of *P. mirabilis* (Stankowska et al., 2012). Again, as per other reports, despite producing two cyclic dipeptides and encoding LuxS-dependent quorum sensing molecule, AI-2, during swarming, *P. mirabilis* has been reported

to have no strong evidence of QS (Holden et al., 1999; Schneider et al., 2002; Campbell et al., 2009; Schaffer and Pearson, 2015). However, a highly ordered swarm cycle suggests an existing mechanism for multicellular coordination (Rauprich et al., 1996). Thus, the fact that *P. mirabilis* are engaged in biofilm formation which is managed, albeit in parts, through quorum sensing brings out the complexity of CAUTI. To deal with such complexity, analyses of the proteins involved in such phenomenon, known as the protein interaction networks (PINs), can reveal important information about key role players of the phenomenon (Lahiri et al., 2014; Pan et al., 2015).

The indispensable role players of phenomenon like QS can be determined by analyzing the PIN involving the proteins in the pathway to produce the QS inducer. The essentiality of such small protein interactome (SPIN) can be brought about by an analysis for the most biologically relevant protein to target for inhibiting that phenomenon, also known as quorum quenching. Ideally, a determination of the number of interacting partners of a particular protein identifies its *degree* centrality (DC) which correlates with its essential nature in the biological scenario (Jeong et al., 2001). However, a much deeper understanding of the essential nature of a particular protein comes upon analyzing its interaction with other partners in the global network of all proteins. In this study, we have discussed the relevance of other centrality measures like *Betweenness centrality* (BC), *Closeness centrality* (CC), and *Eigenvector centrality* (EC) (Jeong et al., 2001) parameters for SPIN having the genes and proteins involved in quorum sensing. Again, analyses of a stipulated sets of QS proteins for a valuable knowledge about the most indispensable virulence proteins to render as drug targets for the QS phenomenon could be uninformative. This led us to conduct a deep probing of the whole genome of *P. mirabilis* (WGPM) for a global analysis of the encoding proteins. This comprises the k-core analysis approach of whole genome protein interactome (GPIN) decomposition to a core of highly interacting proteins (Seidman, 1983). Furthermore, to identify the functional modules in the global network (Guimerà and Nunes Amaral, 2005a), we have performed cartographic analyses and predicted the importance of few proteins sharing similar functional modules. To sum up, the sole objective of this study is to utilize several network based models to analyze and identify crucial role players of QS in *P. mirabilis* and thus, propose their importance as potential drug targets.

MATERIALS AND METHODS

Dataset Collection

The *P. mirabilis* QS pathways for autoinducer-2 (AI-2) biosynthesis were collected from curated reference databases of genomes and metabolic pathways like KEGG, MetaCyc and BioCyc (Caspi et al., 2015; Kanehisa et al., 2015, 2016). The proteins involved in these pathways were extracted with their annotated names and identification as per UniProt database and submitted as entries for the STRING 10.5 biological meta-database (Szklarczyk et al., 2016) to retrieve protein interaction datasets with at least 10 or 50 interactors having the default medium (0.4) level confidence about the interaction, where the

---

interactor numbers relate to the interacting proteins present in the vicinity of the query [period of access: January to February, 2018]. The interactions of the whole genome proteins of the fully annotated *P. mirabilis* strain HI4320 were retrieved from the detailed protein links file under the accession number 529507 in STRING. The sequenced whole genome of the *P. mirabilis* strain HI4320 contains the profile for the same through its full annotation (Pearson et al., 2008). All proteins data, collected and used for interactome construction hereafter, have been reported in **Supplementary Data 1**.

Interactome Construction

We have taken a stepwise approach to integrate and build the interactomes of the proteins, represented by different sections of **Figure 1**. These are the small protein interactomes (SPIN) comprised of (a) those involved in AI-2 biosynthetic pathway in the organism with small (Holden et al., 1999) and large (Kang et al., 2017) number of interactors retrieved from STRING database (AIPS, AIPL, respectively) (**Figures 1A,B**), (b) only QS genes found (QSPO) (**Figure 1C**), (c) all QS genes reported as homologs (QSPH) present in *P. mirabilis* (**Figure 1D**), (d) all virulence genes reported (QSPV) (**Figure 1E**) and (e) the WGPM (**Figure 1F**). Whereas QSPO contains genes reported to be involved in QS in *P. mirabilis*, QSPH contains additional genes reported to be involved in QS in other organisms and present as homologs in *P. mirabilis*. The virulence genes have been taken from the set reported by Schaffer and Pearson (Schaffer and Pearson, 2015). The number of *P. mirabilis* proteins from the SPIN class of interactomes were 31 for AIPS, 42 for AIPL, 24 for QSPO, 42 for QSPH, 58 for QSPV, and 3548 for GPIN (**Supplementary Data 1**). The medium confidence default values of 0.4 for the individual protein interaction data were obtained from String 10.5. Interactions were 1151 for AIPS, 1571 for AIPL, 30 for QSPO, 129 for QSPH, 2376 for QSPV, and 33462 for GPIN, respectively. These interactions are presented in separate sheets of **Supplementary Data 2**.

All individual interaction data obtained above were imported into Cytoscape version 3.6.0 (Cline et al., 2007) and Gephi 0.9.2 (Bastian et al., 2009) to integrate, build and analyze five SPIN namely AIPS, AIPL, QSPO, QSPH and QSPV and the GPIN (**Figure 1**). Interactomes were considered as undirected graphs represented by $G = (V, E)$ comprising a finite set of V vertices and E edges where an edge $e = (u,v)$ connects two vertices u and v (nodes). In the biological PIN context, a vertex/node represents a protein. The number of physical and functional interactions a protein has with other proteins comprises its degree $d(v)$ (Diestel, 2000).

Network Analyses

SPIN

The constructed five SPIN were subsequently analyzed individually through the common four measures of centrality applied to biological networks, namely, eigenvector centrality (EC), betweenness centrality (BC), degree centrality (DC) and closeness centrality (CC) (Koschützki and Schreiber, 2004; Özgür et al., 2008; Pavlopoulos et al., 2011; **Supplementary Data 3**). This was done either via Gephi or the Cytoscape integrated
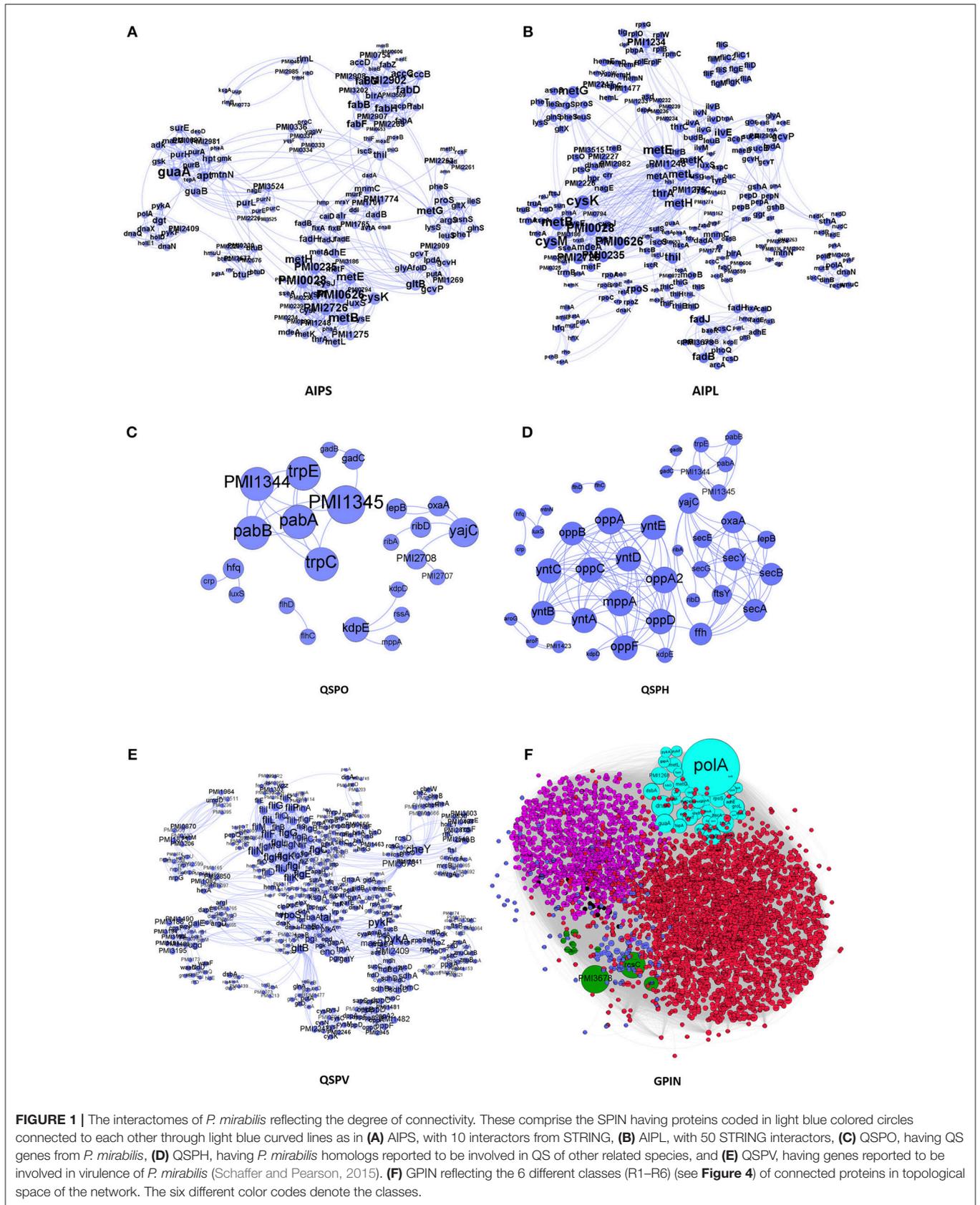
java plugin CytoNCA (Tang et al., 2015). For computing CytoNCA scores, the combined scores obtained from different parameters in STRING were taken as edge weights. The combined scores ranging from 0 to 1, considered in STRING for reporting interactions, generally indicate the confidence of the interaction among the proteins with the level of evidence from the parameters like gene neighborhood, gene fusion, gene co-occurrence, gene co-expression, experiments, annotated pathways and text mining. To find common proteins from each centrality measures, the top 5 proteins were taken for drawing Venn diagrams through online tool Venny 2.1 (Oliveros, 2007–2015) to (**Figure 2**).
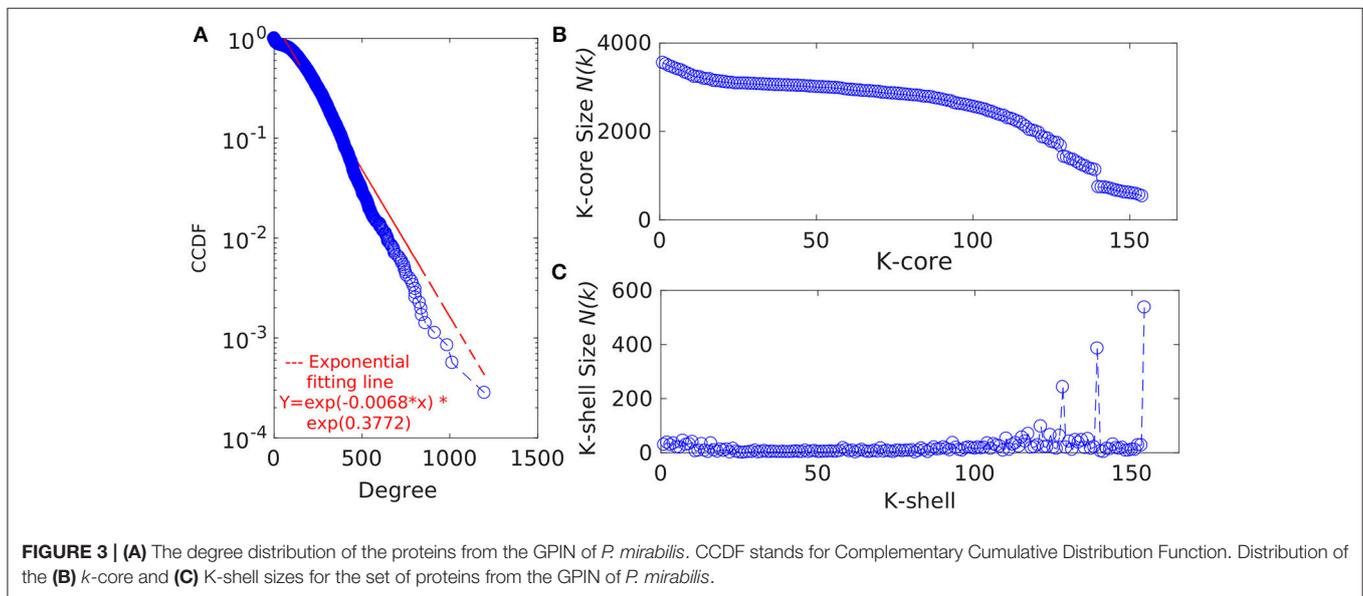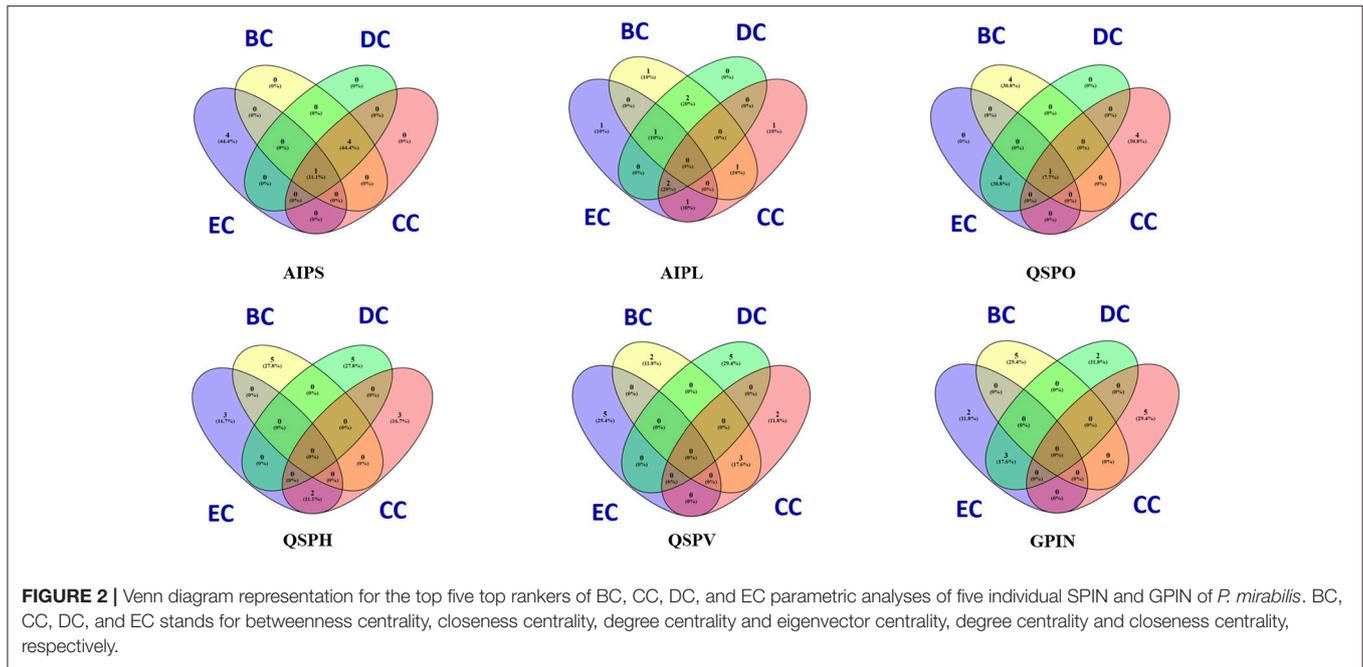
GPIN

MATLAB version 7.11, a programming language developed by MathWorks (MATLAB Statistics Toolbox Release, 2010), was used for further analyses of the GPIN. To gain an overview of the technical aspects of the GPIN, the distributions of network degree (k) was plotted against the Complementary Cumulative Distribution Function (CCDF) (**Figure 3A**). Further concepts about the core group, comprising very specific proteins, was obtained from a *k*-core analysis of the proteins in the whole genome context. This essentially prunes the network to a *k*-core with proteins having degree at least equal to *k* and classifying in K-shell based on their classes of interacting partners (**Figures 3B,C**). A network decomposition (pruning) technique was adopted to produce gradually increasing cohesive sequence of subgraphs (Seidman, 1983). Further, a significant knowledge of the functional connectivity and participation of each protein was obtained from the network topological representation of the within-module degree z-score of the protein vs. its participation coefficient, P, cartographically represented first by Guimerà and Nunes Amaral (2005b) (**Figure 4**). The intra-connectivity of a node "i" to other nodes in the same module is measured by the z-score while the positioning of the node "i" in its own module with respect to other modules measures the participation coefficient, P. Participation of each protein reflected its intra- and inter-modular positioning, where functional modules were calculated based on Rosvall method (Rosvall and Bergstrom, 2011). A modular network has high intra-connectivity and sparse inter-connectivity due to which each module has relatively high density and high separability. Each group of nodes in these type of networks share a common biological function as mentioned by Vella et al. (2018). This analysis divided the proteins into mainly two major categories namely the non-hub nodes and hub nodes, where the latter is the connecting point of many nodes. The category of the former has been assigned the roles of ultra-peripheral nodes (R1), peripheral nodes (R2), non-hub connector nodes (R3), and the non-hub kinless nodes (R4). Likewise, the hub nodes have been designated as provincial hubs (R5), connector hubs (R6), and kinless hubs (R7) (Guimerà and Nunes Amaral, 2005b) (**Figure 4**, **Supplementary Data 4**).

RESULTS

To have an understanding of the important protein(s) of QS in *P. mirabilis*, we have taken a stepwise approach of

FIGURE 1 | The interactomes of *P. mirabilis* reflecting the degree of connectivity. These comprise the SPIN having proteins coded in light blue colored circles connected to each other through light blue curved lines as in **(A)** AIPS, with 10 interactors from STRING, **(B)** AIPL, with 50 STRING interactors, **(C)** QSPO, having QS genes from *P. mirabilis*, **(D)** QSPH, having *P. mirabilis* homologs reported to be involved in QS of other related species, and **(E)** QSPV, having genes reported to be involved in virulence of *P. mirabilis* (Schaffer and Pearson, 2015). **(F)** GPIN reflecting the 6 different classes (R1–R6) (see **Figure 4**) of connected proteins in topological space of the network. The six different color codes denote the classes.

FIGURE 2 | Venn diagram representation for the top five top rankers of BC, CC, DC, and EC parametric analyses of five individual SPIN and GPIN of *P. mirabilis*. BC, CC, DC, and EC stands for betweenness centrality, closeness centrality, degree centrality and eigenvector centrality, degree centrality and closeness centrality, respectively.

FIGURE 3 | (A) The degree distribution of the proteins from the GPIN of *P. mirabilis*. CCDF stands for Complementary Cumulative Distribution Function. Distribution of the **(B)** *k*-core and **(C)** K-shell sizes for the set of proteins from the GPIN of *P. mirabilis*.

building five SPIN, with an ultimate goal to identify the key role playing proteins in the phenomenon of QS to serve as potential candidates for therapeutic targets. **Table 1** represents the comparative picture of the most common topmost proteins, as per centrality measures, in their descending order. In most of the cases, at least three or two of the centrality measures brought out the same protein. These proteins are the ones reflected to be important through each SPIN analysis. For instance, AIPS has MetG and MtnN as the top rankers while LuxS, MnmC, and PMI3678 turns out to be important for AIPL (**Table 1**). Others like QSPO, QSPH, and QSPV have YajC, PMI1345, OppA, RpoS, flagellar proteins of the *flh and fli* operon and some other two-component systems proteins like CheY and KdpE as

important rankers. The functions of these proteins are mentioned in **Table 2**. The top ranking proteins for each of these five SPINs have been reflected in **Figure 2** with Venn diagrams. The common topmost rankers across all the five SPINs are reflected in **Supplementary Figure 1**.

An overview of the important proteins, from individual SPIN as well as across all SPIN, is obtained upon such aforementioned analyses. However, to tackle the MDR *P. mirabilis*, in a global perspective for a drug to be effective, the proteins need to be essentially indispensable. Thus, the whole genome proteins interactome (GPIN) of *P. mirabilis* was then analyzed to understand the global scenario.
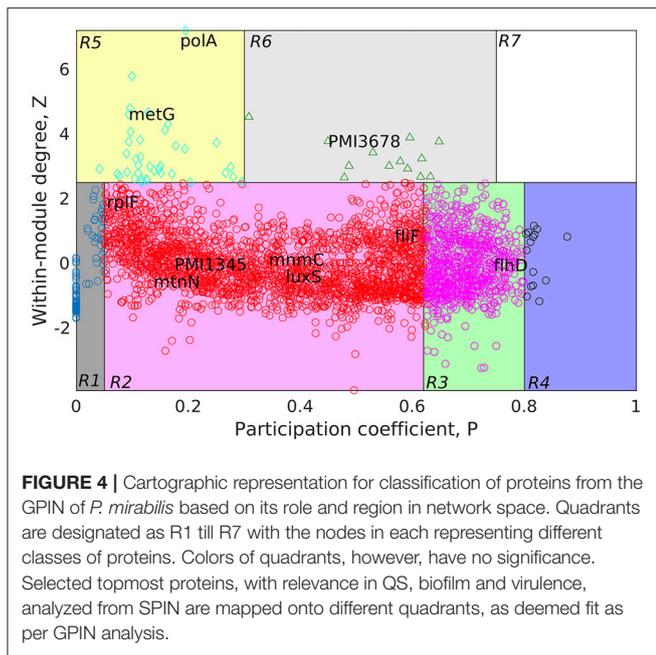
FIGURE 4 | Cartographic representation for classification of proteins from the GPIN of *P. mirabilis* based on its role and region in network space. Quadrants are designated as R1 till R7 with the nodes in each representing different classes of proteins. Colors of quadrants, however, have no significance. Selected topmost proteins, with relevance in QS, biofilm and virulence, analyzed from SPIN are mapped onto different quadrants, as deemed fit as per GPIN analysis.

TABLE 1 | The most common topmost proteins of *P. mirabilis* SPIN and GPIN.

Network	EC	BC	DC	CC
AIPS	**MetG**, LuxS, GcvP, **Hpt**, PMI3524	MtnN, **CysK**, LuxS, MetB, MnmC	MtnN, MnmC, **CysK**, LuxS, MetB	MtnN, **CysK**, LuxS, MetB, MnmC
AIPL	LuxS, **ThrA**, MetH, **MetL**, PMI0028	MnmC, MtnN, LuxS, **PMI3678**, **TrmA**	MnmC, LuxS, MtnN, **ThrA**, MetH	**PMI3678**, **ThrA**, MetH, PMI0028, PMI0626
QSPO	PMI1345, PMI1344, **TrpE**, PabA, **PabB**	YajC, PMI1345, GadC, **RibD**, PMI2708	PMI1345, PMI1344, **TrpE**, PabA, **PabB**	KdpE, **Hfq**, FlhD, FlhC, PMI1345
QSPH	PMI1345, GadC, **TrpE**, FlhD, FlhC	KdpE, **Ffh**, KdpD, **LepB**, **FtsY**	OppA, MppA, OppA2, OppD, OppC	FlhD, FlhC, PMI1423, AroF, AroG
QSPV	FliF, FliK, FlgG, FlgC, FlgI	**RpoS**, **Eno**, Irp, Pgm, **PMI3678**	CheY, PykA, **PykF**, **Tal**, FliN	**RpoS**, **Eno**, **PMI3678**, FliC, CsrA
GPIN	**PolA**, **GuaA**, **DnaK**, **MetG**, **RecA**	**RplP**, PMI1727, PMI1033, PMI2007, **RpoS**	**PolA**, **PMI3678**, **RcsC**, **DnaK**, **GuaA**	PMI2375, PMI2723, PMI0739, PMI3495, PMI2629

The bold cased proteins are present in the innermost 154th k- core. EC, BC, DC, and CC stands for eigenvector centrality, betweenness centrality, degree centrality, and closeness centrality, respectively.

The Complete GPIN

In an attempt to analyze the type of network being built from the functional and physical interactions empirically found and theoretically predicted among the whole genome proteins retrieved from STRING, we have observed the degree distribution of GPIN to be exponential showing a non-linear

TABLE 2 | Functions of centrality based topmost proteins of individual *P. mirabilis* networks.

Protein name	Description of function
MetG	Is required not only for elongation of protein synthesis but also for the initiation of all mRNA translation through initiator tRNA(fMet) aminoacylation
MtnN	Catalyzes the irreversible cleavage of the glycosidic bond in both 5′-methylthioadenosine (MTA) and S-adenosylhomocysteine (SAH/AdoHcy) to adenine and the corresponding thioribose, 5′-methylthioribose and S-ribosylhomocysteine, respectively
LuxS	Involved in the synthesis of autoinducer 2 (AI-2) which is secreted by bacteria and is used to communicate both the cell density and the metabolic potential of the environment. The regulation of gene expression in response to changes in cell density is called quorum sensing. Catalyzes the transformation of S-ribosylhomocysteine (RHC) to homocysteine (HC) and 4,5-dihydroxy-2,3-pentadione (DPD)
MnmC	Catalyzes the last two steps in the biosynthesis of 5-methylaminomethyl-2-thiouridine (mnm_5s_2U) at the wobble position (U34) in tRNA. Catalyzes the FAD-dependent demodification of $cmnm_5s_2U34$ to nm_5s_2U34, followed by the transfer of a methyl group from S-adenosyl-L-methionine to nm_5s_2U34, to form mnm_5s_2U34
PMI3678	Catalyzes the Phosphorelay through sensor kinase activity of two-component Regulatory system
PMI1345	Catalyzes the transfer of the phosphoribosyl group of 5-phosphorylribose-1-pyrophosphate (PRPP) to anthranilate to yield N-(5′-phosphoribosyl)-anthranilate (PRA)
FlhD	Functions in complex with FlhC as a master transcriptional regulator that regulates transcription of several flagellar and non-flagellar operons by binding to their promoter region. Activates expression of class 2 flagellar genes, including fliA, which is a flagellum-specific sigma factor that turns on the class 3 genes. Also regulates genes whose products function in a variety of physiological pathways
FliF	Flagellar protein whose M ring may be actively involved in energy transduction
PolA	In addition to polymerase activity, this DNA polymerase exhibits 5′-3′ exonuclease activity
RplP	Binds 23S rRNA and is also seen to make contacts with the A and possibly P site tRNAs

The functions of the selected proteins are derived from UniProt database.

preferential attachment nature (**Figure 3A**; Vázquez, 2003). Hereafter, we have framed an idea of the important proteins from an array of proteins involved in the five individual SPIN, upon performing a k-core analysis for them (**Figures 3B,C**). Notably, the innermost core was 154th shell and had genes like *thrA, cysK, metG, metL, trpE, rpoS, eno*, etc. which have already been reflected from the four network centrality analyses of the SPINs (**Table 1**, **Supplementary Data 3**: **Sheet 1–5**). Additionally, it is to be noted that top 5 EC and DC measures of the GPIN also had their position in the innermost 154th core, thereby indicating their importance in the global scenario. Other important genes e.g., *luxS, PMI1345* from the k-core analyses were found in the 139th shell. The latter category was found to have direct involvement in QS.

Furthermore, to classify the proteins based on their regional positioning and functional role in the network topological space of *P. mirabilis*, we have analyzed the GPIN represented

cartographically (**Figure 4**, **Supplementary Data 4**). Essentially, such representation would classify the complete set of proteins in the genome with respect to their connectivity within similar classes of proteins performing similar biological function (functional module) along with their participation with other related and/or non-related functional module (also see materials and methods and discussion section). Noticeably, the R6 quadrant had the top 5 proteins belonging to either the innermost 154th core or almost close to the 139th core containing most of the proteins related to QS (**Supplementary Data 4**). These are GltB and PMI3678 for the former and PMI3348, PMI0587, and PMI3517 for the latter. Moreover, upon looking deep into EC classification of R6 quadrants, all top 5 proteins, namely PolA, GuaA, DnaK, MetG, and RecA were from the innermost 154th core. Furthermore, analysis after sorting of module followed by R quadrant, k-core followed by either module or EC measures, all revealed the proteins to be mostly belonging to the R6 or R5 categories, besides their 154th or 139th core classification (**Supplementary Data 4**). It is worthwhile to mention here that a similar sorting analyses of BC with respect to Quadrant and k-core had revealed proteins mostly from R2 or R3, none of them occupying the innermost 154th core, except RplP, and RpoS.

DISCUSSION

We have started with the proteins involved in *P. mirabilis* AI-2 biosynthesis pathway (**Supplementary Figure 2**) and derived the AIPS besides AIPL (**Figures 1A,B**). While the former connects the proteins of the pathway as reported by default in STRING with only 10 interactors, supposedly directly involved in the phenomenon of AI biosynthesis, the latter has been formed upon extending those to 50 interactors per protein query. The idea was to incorporate other related proteins having connectivity to the AI-2 whose analysis might give more insight about QS in *P. mirabilis*. Moreover, it was necessary to have an idea of the robustness of the proteins involved in QS pathways and thus, QSPO was constructed to have an idea of the proteins directly involved in the phenomenon of QS in *P. mirabilis* only (**Figure 1C**). Again, with the homologous proteins reported to be involved in QS in other species from KEGG database, it was necessary to look into their association with acknowledged QS proteins of *P. mirabilis* (**Supplementary Figure 3**). Thus, QSPH was constructed to take into consideration of this fact and analyze further (**Figure 1D**). Furthermore, with multiple genes and proteins reviewed for the virulence of *P. mirabilis* (Schaffer and Pearson, 2015), including those involved for QS phenomenon, it was necessary to have an interactome QSPV constructed to analyse their interactions and involvement (**Figure 1E**). All these SPIN were constructed to have an understanding of the indispensable proteins responsible for QS in *P. mirabilis*. Finally, a complete whole genome analyses for other plausible indispensable proteins connecting biofilm formation, AI-2 biosynthesis, quorum sensing and even MDR was necessary to have a bird's eye view of the global scenario. This was done with the construction and analyses of the GPIN (**Figure 1F**).

The five SPIN were then analyzed individually by utilizing the four important centrality measures of DC, CC, BC, and EC. Of these, DC is the most basic, informing the connectivity of any protein in the network. CC might reflect the proximity of a protein in terms of its communication with others to render a functionally virulent phenotype. Being a comparatively better measure in terms of bridging different functionally important groups of virulent proteins, BC might bring out the importance of a protein to be targeted for therapeutic purposes. However, EC might show the most important proteins having their impact on other important proteins in a virulent network and thus, turn out to be indispensable protein to target. We have found a varying range of proteins ranging from the locomotive flagellar proteins of the *flh* and *flg* operon (Claret and Hughes, 2000), LuxS (Schneider et al., 2002) and MtnN directly involved in AI-2 biosynthetic pathway, MetG and MnmC involved in the protein translation machinery along with the proteins PMI1345 (Wang M. C. et al., 2014) and PMI3678 with catalytic activities/domains, chaperone protein, Hfq (Wang M. C. et al., 2014), signal transduction protein, KdpE (Rhoads et al., 1978), and a pre-protein translocase subunit, YajC (Pearson et al., 2008). Among the proteins PMI1345 and PMI3678, as per UniProt database, the former is having an activity as anthranilate phosphoribosyltransferase catalyzing the transfer of the phosphoribosyl group of 5-phosphorylribose-1-pyrophosphate (PRPP) to anthranilate to yield N-(5′-phosphoribosyl)-anthranilate (PRA). Essentially, PMI1345 is involved in the 2nd step of the subpathway synthesizing L-tryptophan from chorismate. Again, PMI3678 has the histidine kinase domain and displays activities of kinase through ATP binding and in-turn regulates transcription via a two-component regulatory system. Thus, as analyzed above, with the different proteins, pertaining to the biofilm formation, flagellar locomotion, translation and signal transduction, a level of complexity of the *P. mirabilis* QS machinery could be perceived.

To gain more insight into the global scenario of the whole genome, we have constructed the GPIN (**Figure 1F**) and analyzed it through several network topological and centrality parametric measures (**Supplementary Datas 3, 4**). For this GPIN, we have observed that the connectivity distribution, P(k), of a particular node gets connected to k other nodes, for large values of k. This confirms that the GPIN is indeed a large network and neither a random, Erdos and Renyi type (Erdos and Rényi, 1960) nor a small-world, Watts and Strogatz type (Watts and Strogatz, 1998). Our GPIN is free of a characteristic scale and roughly followed the power-law (Albert et al., 2000) with an exponential decay of the degree distribution (**Figure 3A**). Initially, we have analyzed the constructed GPIN with k-core/K-shell topological parameters (**Figures 3B,C**). Technically speaking, a k-core is a subnetwork with a minimum number of k-links. A K-shell is a set of nodes having exactly k-links. In another words, K-shell is the part of k-core but not of (k+1)-core. Thus, proteins belonging to the outer shell have lower k value thereby reflecting the limited number of interacting partner proteins. On the contrary, proteins from the inner k-core/K-shell are very specific ones having high interaction with each other and are considered to be the most

important ones. It has been observed that the inner core member proteins are highly interactive due their robust and central character (Alvarez-Hamelin et al., 2006). In this light, a complete decomposition of the network, achieved by decomposing the core, would reveal the innermost important part of the network. We have found the 154th core as the innermost one for our GPIN having many proteins involved in the biosynthesis of amino acids, including cysteine and methionine, the amino acid precursor of the components of AI-2 biosynthetic pathway. These proteins rank top for most of the EC measures across the other five SPIN as well. Furthermore, the 139th core was on focus due to its nearby proximity to the innermost core and comprising most of the proteins directly involved in QS. Our analyses till this far revealed LuxS and PMI1345 to be the prominent EC proteins in the 139th core of the genome. Interestingly, only PolA and RplP, top rankers of BC measures, made it to the innermost 154th core compared to the other topmost EC proteins in that core. This probably reflects the importance of EC measure to reveal the prominent stakeholders of the machinery responsible for the very survival and probably virulence of the organism. Any effective drug target should, thus, be selected from this core group with high EC rank.

A further delving deep into the functional connectivity of the modules formed in network topological space reinforced our findings this far. The topological orientation of the nodes in space are being represented cartographically where *P*-values have been put in the x-axis and z-score values in the y-axis. In this context, R1 has low *P*-values and low z-scores while R7 has the highest for both of them. Following this representation, the non-hubs and the hubs are classified into the protein groups of R1-4 and R5-7, respectively. Among them, the kinless hubs proteins (R7), having high connection within module (z) as well as between modules (P) scores, becomes important in terms of functionality. Similarly, the ultra-peripheral proteins (R1), with least P and z measures, are the least connected across the network followed by the peripheral proteins (R2). Such proteins can be detached easily and thus, are perceived, not much to affect the whole network when attempted to reach the core upon decomposition. This is nothing but the outermost shells of the k-core measures (refer previous section) which has proteins not grossly affecting the survival of the organism. Likewise, proteins belonging to the non-hub connectors (R3) group might be involved in only a small but fundamental sets of interactions. On the contrary, proteins of the provincial hubs class (R5) have many connections which are within-module. Again, the non-hub kinless proteins (R4) link other proteins which are evenly distributed across all the modules. However, the connector hub proteins (R6) link most of the other modules and are expected to be the most conserved in terms of decomposition as well as evolution. This could be the very set of proteins which the organism would maintain as the essential ones for their very survival. We have observed mostly R5 and R6 classes of proteins occupying the innermost 154th and the QS-involved 139th cores. Furthermore, the EC measures brings out the importance when compared to other measures of centralities.

In order to bring out the biological implication of the cartographic analyses, we now discuss the relevance of the

proteins identified as essential in the context of virulence, biofilm formation and QS phenomenon. In this context, it is important to note that, we have observed many of the already known genes and proteins, viz LuxS, FlhDC to be reflected from our *in silico* cartographic analyses as well. For example, with the highest number (17) of fimbrial operons reported in any sequenced bacterial species, four *P. mirabilis* fimbriae, namely, MR/P, UCA, ATF, and PMF have shown prominent roles in biofilm formation (Scavone et al., 2016). The thickness, structure, and the amount of exopolysaccharides produced by some biofilms formed by *P. mirabilis* are influenced by important acylated homoserine lactones (Stankowska et al., 2012). Moreover, some virulence factors are regulated by QS molecules like acylated homoserine lactones (acyl-HSLs) (Henke and Bassler, 2004). Of the two QS types, LuxS is an essential enzyme for AI-2 type which is coded by luxS gene having S-ribosylhomocysteine lyase activity (Schneider et al., 2002). Acetylated homoserine lactone derivatives modifies the expression of virulence factors of *P. mirabilis* strains (Stankowska et al., 2008). The flhDC master operon is a key regulator in swarmer cell differentiation in *P. mirabilis*, it is known to cause an increased viscosity and intracellular signals (Fraser and Hughes, 1999). Furthermore, the extracellular signals can be sensed by two-component regulators such as RcsC–RcsB (Fraser and Hughes, 1999).

Having said the above, we observe that, many such genes and proteins, not reported to have connections with QS and virulence, have also been unearthed from our study. Thus, it is imperative to have an in-depth analysis to bring out the importance of the proteins unearthed through the process. In order to achieve the same, we rely on the fact that the innermost 154th core could harbor the genes/proteins essential for the very survival of the organism. Moreover, our cartographic analysis shows that R6 classes of proteins having high intra- and inter-connectivity, within and between the functional modules might play a crucial role in the maintenance of the organismal structure. This adds up to another level of indispensable nature. Furthermore, the very concept of Eigenvector centrality, which reflects the important proteins' connectivity with other such important proteins in terms of their function, finalize the indispensable factor. This method of utilizing the *k*-core, functional module and centrality measure, like that of Eigenvector, has been used to analyze large networks to reveal the important proteins, albeit, in a complete different scenario (Ashraf et al., 2018). Utilizing this method, referred to as KFC, we found the three topmost indispensable factors for *P. mirabilis* are *gltB*, *PMI3678*, and *rcsC* (**Supplementary Data 5**). It is important to note that the glutamate synthase encoding gene *gltB*, has been shown to be involved in a quorum sensing-dependent glutamate metabolism which affects the homeostatic osmolality and outer membrane vesiculation in *Burkholderia glumae* (Kang et al., 2017). Expression level of *gltB* has been shown to affected in *E. coli* by the stationary phase QS signals (Ren et al., 2004). Again, *rcsC* encodes an sensor histidine kinase protein which is known to be involved in swarming migration and capsular polysaccharide synthesis along with *yojN* (Belas et al., 1998; Fraser and Hughes, 1999). The sensor kinase activity for *PMI3678* encoding an aerobic respiration control protein, however, has not

been reported earlier for *P. mirabilis*, and thereby could serve as one of the important therapeutic targets. All these proteins are quite different to those reported to be quite important in a recent study to unearth the fitness factors in a single-species and polymicrobial CAUTI setting, performed with a genome wide transposon mutagenesis of *P. mirabilis* (Armbruster et al., 2017). In this study, Armbuster *et al.* has observed the polyamine uptake and biosynthesis to the fitness factor for single species CAUTI while branched chain amino acid (BCAA) synthesis turned out to be important for polymicrobial infection along with *Providencia stuartii* (Armbruster et al., 2017). None of these fitness factors, found to be helpful in colonizing either the catheterized bladder (referred to as *F*actors for *B*ladder *C*olonization, FBC) or the kidney (*F*actors for *K*idney *C*olonization, FKC), were observed in our analysis to be belonging to the R6 quadrant despite some falling within the innermost 154th core (**Supplementary Data 5**). While the reports by Armbuster et al. is in a live and dynamic setting, ours is, a static and theoretical network analysis. However, given the fact that this theoretical analysis reflects only a few indispensable ones, they might have some relevance in therapeutic intervention strategies to tackle CAUTI caused by MDR *P. mirabilis*.

CONCLUSION

This study takes a stepwise approach to identify the crucial role players from different sets of interacting proteins of *P. mirabilis* involved primarily in QS phenomenon. Essentially, this delineates the building of theoretical interactomes comprising the five individual SPIN which are analyzed through network parametric measures to reveal the most important proteins for such phenotype of QS and biofilm formation. All these lead to the identification of LuxS and PMI1345 to be important proteins

of this organism. Furthermore, the results are supplemented through a decomposition of the *P. mirabilis* genome interactome, GPIN, followed by analysis of centrality measurements to reach the innermost core of the proteins essential for virulence and survival. Such in-depth analysis of the GPIN revealed other classes of important conserved proteins like GltB, PMI3678, and RcsC having the potential for being the most important ones and thus, indispensable among the set of whole genome proteins of *P. mirabilis*.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2018.00269/full#supplementary-material

REFERENCES

Albert, R., Jeong, H., and Barabási, A. L. (2000). Error and attack tolerance of complex networks. *Nature* 406, 378–382. doi: 10.1038/35019019

Alvarez-Hamelin, I., DallAsta, L., Barrat, A., and Vespignani, A. (2006). *Advances in Neural Information Processing Systems 18*. Cambridge, MA: MIT press.

Armbruster, C. E., Forsyth-DeOrnellas, V., Johnson, A. O., Smith, S. N., Zhao, L., Wu, W., et al. (2017). Genome-wide transposon mutagenesis of *Proteus mirabilis*: essential genes, fitness factors for catheter-associated urinary tract infection, and the impact of polymicrobial infection on fitness requirements. *PLoS Pathog.* 13:e1006434. doi: 10.1371/journal.ppat.1006434

Ashraf, M. I., Ong, S. K., Mujawar, S., Pawar, S., More, P., Paul, S., et al. (2018). A side-effect free method for identifying cancer drug targets. *Sci. Rep.* 8:6669. doi: 10.1038/s41598-018-25042-2

Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *Icwsm* 8, 361–362.

Belas, R., Schneider, R., and Melch, M. (1998). Characterization of *Proteus mirabilis* precocious swarming mutants: identification of rsbA, encoding a regulator of swarming behavior. *J. Bacteriol.* 180, 6126–6139.

Campbell, J., Lin, Q., Geske, G. D., and Blackwell, H. E. (2009). New and unexpected insights into the modulation of LuxR-type quorum sensing by cyclic dipeptides. *ACS Chem. Biol.* 4, 1051–1059. doi: 10.1021/cb900165y

Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., et al. (2015). The MetaCyc database of metabolic pathways and enzymes and

the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 44, D471–D480. doi: 10.1093/nar/gkv1164

Claret, L., and Hughes, C. (2000). Rapid turnover of FlhD and FlhC, the flagellar regulon transcriptional activator proteins, during proteus swarming. *J. Bacteriol.* 182, 833–836. doi: 10.1128/JB.182.3.833-836.2000

Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., et al. (2007). Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.* 2, 2366– 2382. doi: 10.1038/nprot.2007.324

Czerwonka, G., Guzy, A., Kałuza, K., Grosicka, M., Danczuk, M., Lechowicz, Ł., et al. (2016). The role of *Proteus mirabilis* cell wall features in biofilm formation. *Arch. Microbiol.* 198, 877–884. doi: 10.1007/s00203-016-1249-x

Diestel, R. (2000). *Graph Theory*. Berlin; Heidelberg: Springer-Verlag.

Eliopoulos, G. M., Maragakis, L. L., and Perl, T. M. (2008). *Acinetobacter baumannii*: epidemiology, antimicrobial resistance, and treatment options. *Clin. Infect. Dis.* 46, 1254–1263. doi: 10.1086/529198

Erdos, P., and Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* 5, 17–61.

Fraser, G. M., and Hughes, C. (1999). Swarming motility. *Curr. Opin. Microbiol.* 2, 630–635. doi: 10.1016/S1369-5274(99)00033-8

Guimerà, R., and Nunes Amaral, L. A. (2005a). Cartography of complex networks: modules and universal roles. *J. Statist. Mech.* 2005:P02001. doi: 10.1088/1742-5468/2005/02/P02001

Guimerà, R., and Nunes Amaral, L. A. (2005b). Functional cartography of complex metabolic networks. *Nature* 433, 895–900. doi: 10.1038/nature03288

Henke, J. M., and Bassler, B. L. (2004). Bacterial social engagements. *Trends Cell Biol.* 14, 648–656. doi: 10.1016/j.tcb.2004.09.012

Holden, M. T., Ram Chhabra, S., De Nys, R., Stead, P., Bainton, N. J., Hill, P. J., et al. (1999). Quorum-sensing crosstalk: isolation and chemical characterization of cyclic dipeptides from *Pseudomonas aeruginosa* and other gram-negative bacteria. *Mol. Microbiol.* 33, 1254–1266. doi: 10.1046/j.1365-2958.1999.01577.x

Jacobsen, S. M., and Shirtliff, M. E. (2011). *Proteus mirabilis* biofilms and catheter-associated urinary tract infections. *Virulence* 2, 460–465. doi: 10.4161/viru.2.5.17783

Jeong, H., Mason, S. P., Barabási, A. L., and Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature* 411, 41–42. doi: 10.1038/35075138

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2016). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2015). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070

Kang, Y., Goo, E., Kim, J., and Hwang, I. (2017). Critical role of quorum sensing-dependent glutamate metabolism in homeostatic osmolality and outer membrane vesiculation in Burkholderia glumae. *Sci. Rep.* 7:44195. doi: 10.1038/srep44195

Koschützki, D., and Schreiber, F. (2004). "Comparison of centralities for biological networks," in *German Conference on Bioinformatics* (Bielefield), 199–206.

Lahiri, C., Pawar, S., Sabarinathan, R., Ashraf, M. I., Chand, Y., and Chakravortty, D. (2014). Interactome analyses of Salmonella pathogenicity islands reveal SicA indispensable for virulence. *J. Theor. Biol.* 363, 188–197. doi: 10.1016/j.jtbi.2014.08.013

Luzzaro, F., Brigante, G., D'Andrea, M. M., Pini, B., Giani, T., Mantengoli, E., et al. (2009). Spread of multidrug-resistant *Proteus mirabilis* isolates producing an AmpC-type β-lactamase: epidemiology and clinical management. *Int. J. Antimicrob. Agents* 33, 328–333. doi: 10.1016/j.ijantimicag.2008.09.007

MATLAB and Statistics Toolbox Release (2010). Natick, MA: The MathWorks, Inc.

Oliveros, J. C. (2007–2015). *Venny. An Interactive Tool for Comparing Lists with Venn's Diagrams*. Available online at: http://bioinfogp.cnb.csic.es/tools/venny/index.html

Özgür, A., Vu, T., Erkan, G., and Radev, D. R. (2008). Identifying gene-disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics* 24, i277–i285. doi: 10.1093/bioinformatics/btn182

Pan, A., Lahiri, C., Rajendiran, A., and Shanmugham, B. (2015). Computational analysis of protein interaction networks for infectious diseases. *Brief. Bioinform.* 17, 517–526. doi: 10.1093/bib/bbv059

Pavlopoulos, G. A., Secrier, M., Moschopoulos, C. N., Soldatos, T. G., Kossida, S., Aerts, J., et al. (2011). Using graph theory to analyze biological networks. *BioData Min.* 4:10. doi: 10.1186/1756-0381-4-10

Pearson, M. M., Sebaihia, M., Churcher, C., Quail, M. A., Seshasayee, A. S., Luscombe, N. M., et al. (2008). Complete genome sequence of uropathogenic *Proteus mirabilis*, a master of both adherence and motility. *J. Bacteriol.* 190, 4027–4037. doi: 10.1128/JB.01981-07

Rauprich, O., Matsushita, M., Weijer, C. J., Siegert, F., Esipov, S. E., and Shapiro, J. A. (1996). Periodic phenomena in *Proteus mirabilis* swarm colony development. *J. Bacteriol.* 178, 6525–6538. doi: 10.1128/jb.178.22.6525-6538.1996

Ren, D., Bedzyk, L. A., Ye, R. W., Thomas, S. M., and Wood, T. K. (2004). Stationary-phase quorum-sensing signals affect autoinducer-2 and gene expression in *Escherichia coli*. *Appl. Environ. Microbiol.* 70, 2038–2043. doi: 10.1128/AEM.70.4.2038-2043.2004

Rhoads, D. B., Laimins, L., and Epstein, W. (1978). Functional organization of the kdp genes of *Escherichia coli* K-12. *J. Bacteriol.* 135, 445–452.

Rosvall, M., and Bergstrom, C. T. (2011). Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PLoS ONE* 6:e18209. doi: 10.1371/journal.pone.0018209

Rutherford, S. T., and Bassler, B. L. (2012). Bacterial quorum sensing: its role in virulence and possibilities for its control. *Cold Spring Harb. Perspect. Med.* 2:a012427. doi: 10.1101/cshperspect.a012427

Scavone, P., Iribarnegaray, V., Caetano, A. L., Schlapp, G., Härtel, S., and Zunino, P. (2016). Fimbriae have distinguishable roles in *Proteus mirabilis* biofilm formation. *FEMS Pathog. Dis.* 74:ftw033. doi: 10.1093/femspd/ftw033

Schaffer, J. N., and Pearson, M. M. (2015). *Proteus mirabilis* and urinary tract infections. *Microbiol. Spectrum* 3:UTI-0017-2013. doi: 10.1128/microbiolspec.UTI-0017-2013

Schneider, R., Lockatell, C. V., Johnson, D., and Belas, R. (2002). Detection and mutation of a luxS-encoded autoinducer in *Proteus mirabilis*. *Microbiology* 148, 773–782. doi: 10.1099/00221287-148-3-773

Seidman, S. B. (1983). Network structure and minimum degree. *Soc. Netw.* 5, 269–287. doi: 10.1016/0378-8733(83)90028-X

Stankowska, D., Czerwonka, G., Rozalska, S., Grosicka, M., Dziadek, J., and Kaca, W. (2012). Influence of quorum sensing signal molecules on biofilm formation in *Proteus mirabilis* O18. *Folia Microbiol.* 57, 53–60. doi: 10.1007/s12223-011-0091-4

Stankowska, D., Kwinkowski, M., and Kaca, W. (2008). Quantification of *Proteus mirabilis* virulence factors and modulation by acylated homoserine lactones. *J. Microbiol. Immunol. Infect.* 41, 243–253.

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2016). The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, D362–D368. doi: 10.1093/nar/gkw937

Tang, Y., Li, M., Wang, J., Pan, Y., and Wu, F. X. (2015). CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *BioSystems* 127, 67–72. doi: 10.1016/j.biosystems.2014.11.005

Vázquez, A. (2003). Growing network with local rules: preferential attachment, clustering hierarchy, and degree correlations. *Phys. Rev. E* 67:056104. doi: 10.1103/PhysRevE.67.056104

Vella, D., Marini, S., Vitali, F., Di Silvestre, D., Mauri, G., and Bellazzi, R. (2018). MTGO: PPI network analysis via topological and functional module identification. *Sci. Rep.* 8:5499. doi: 10.1038/s41598-018-23672-0

Wang, J. T., Chen, P. C., Chang, S. C., Shiau, Y. R., Wang, H. Y., Lai, J. F., et al. (2014). Antimicrobial susceptibilities of *Proteus mirabilis*: a longitudinal nationwide study from the Taiwan surveillance of antimicrobial resistance (TSAR) program. *BMC Infect. Dis.* 14:486. doi: 10.1186/1471-2334-14-486

Wang, M. C., Chien, H. F., Tsai, Y. L., Liu, M. C., and Liaw, S. J. (2014). The RNA chaperone Hfq is involved in stress tolerance and virulence in uropathogenic *Proteus mirabilis*. *PLoS ONE* 9:e85626. doi: 10.1371/journal.pone.0085626

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature* 393, 440–442.