

Comparative Network Analysis of Two-Component Signal Transducing Protein-Protein Interactions in *Enterococcus faecalis* Sp.

Ranjith N. Kumavath* and Pratap D

Department of Genomic Science, School of Biological Sciences, Central University of Kerala, Riverside Transit Campus, Kasaragod, Kerala, India

Abstract

Background: Protein-Protein interactions network analysis of *Enterococcus faecalis* sp. in a comparative approach helps for the dissection of gene function, potential signal transduction, and virulence pathways. Our research shed light on the comparative network analysis of two component signal transducing protein-protein interactions in ten different *Enterococcus faecalis* strains on the base of protein interactions prediction analysis and Host-Pathogen Interactions analysis using STRING and HPIDB databases. Around 30-40 proteins participate in the Two Component System of *Enterococcus faecalis*.

Results: The network parameters of protein-protein interactions were calculated. The topological coefficient of almost all the strains of *Enterococcus faecalis* network perfectly followed a power law distribution (correlation) 1.000; R-squared) 1.000 and represented the best fit. We have figured out a protein (EF_3329) DNA-binding response regulator, which is not participating in the entire protein interactions network except the X86 strains of *Enterococcus faecalis*. We have detected out one sub network which is not at all having any interactions with the main network except in the *Enterococcus faecalis* D6.

Conclusions: This is a novel comparative study on *Enterococcus faecalis* two component system which provides a useful resource for further analysis of protein interactions of two component signal transductions in other life threatening pathogens. The further analysis spots the light on in silico inhibitor screening which helps in the designing of novel drugs.

Keywords: *Enterococcus faecalis*; Histidine proteins kinases; Host pathogen interactions; Protein-protein interactions; Response regulators; Two component signal transduction

Introduction

Protein-protein interactions are known to be central to most cellular functions. These assemblies govern the dynamics of cellular processes both temporally and spatially. High throughput studies have unveiled an unexpected extended protein networks in various bacteria that challenge our understanding in terms of functionality [1]. A typical two component system consists of a histidine protein kinase containing a conserved kinase core and a response regulator protein containing a conserved regulatory domain. Extracellular stimuli are sensed by and serve to modulate the activities of the histidine kinase. The histidine kinase transfers a phosphoryl group to the regulator protein in a reaction catalyzed by the regulator protein. Phosphotransfer to the regulator protein results in activation of a downstream effector domain that elicits the specific response [2,3]. In an elaborated way we can say that, the ATP-dependent phosphorylation of histidine is generally regulated in response to environmental signals by a family of histidine protein kinases. The phosphorylated aspartate is present within another variety of protein termed as response regulator which undergoes a phosphorylation-induced conformational change that serves to evoke a response [4]. Thus, there are two essential families of proteins that function together the sensory kinases and their associated response regulators.

Histidine Protein Kinases (HPKs) are a large family of signal-transduction enzymes that auto phosphorylate on a conserved histidine residue. HPKs form two-component signaling systems together with their downstream target proteins, the response regulators, which have a conserved aspartate in a so-called 'receiver domain' that is phosphorylated by the Histidine protein kinase [5]. Response regulators of bacterial sensory transduction systems generally consist

of receiver module domains covalently linked to effector domains. The effector domains include DNA binding and/or catalytic units that are regulated by sensor kinase-catalyzed aspartyl phosphorylation within their receiver modules. Most receiver modules are associated with three distinct families of DNA binding domains, but some are associated with other types of DNA binding domains with methylated chemotaxis protein demethylases or with sensor kinases [6].

Enterococci are gram-positive constituents of the normal human micro flora typically colonizing the intestinal tract and skin. However, these organisms are capable of causing disease as opportunistic pathogens, mainly in immune compromised patients. Normally, *Enterococci* are used as probiotics to improve the microbial balance of the intestine and to treat gastroenteritis in humans and animals. These bacteria probably now represent the greatest risk to human health of any bacterial species currently used for these purposes. General features of these organisms are hallmarks of their biology and may also contribute to their pathogenicity. *Enterococci* are distinguished for their ability to grow at temperatures ranging from 10 to 45°C and in 6.5% NaCl or/and to tolerate acidic and alkaline growth conditions. The above observations raise the question of how *Enterococcal* physiology

***Corresponding author:** Dr. Ranjith N. Kumavath, Department of Genomic Science, School of Biological Sciences, Central University of Kerala, Riverside Transit Campus, Padanakkad (P.O) Kasaragod-671314, Kerala, India, Tel: +91 8547648620; Fax: 0467-2282250; E-mail: nkumavath@gmail.com

Received September 22, 2012; **Accepted** November 14, 2012; **Published** November 16, 2012

Citation: Kumavath RN, Pratap D (2012) Comparative Network Analysis of Two-Component Signal Transducing Protein-Protein Interactions in *Enterococcus faecalis* Sp. J Proteomics Bioinform 5: 270-278. doi:10.4172/jpb.1000249

Copyright: © 2012 Kumavath RN, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

has evolved to allow the organisms to sense environmental changes and respond to the various stimuli with adaptive behavior. Monitoring and adapting to changing environmental conditions is the key function of bacterial signal transduction, which is generally carried out by the so-called two-component systems. Studies of enterococci have identified some general stress proteins; a global view of the *Enterococcal* signal transduction mechanisms has not been gained [7].

For understanding the function of a particular protein, it is usually useful to identify other proteins with which it associates. This can be done by topological identification of specific proteins. To fulfill their biological activities in the cell, most proteins function in association with protein partners or as part of large molecular assemblies. Hence, the knowledge of the interactions context of a protein is crucial to understand its cellular functions. A comprehensive description of the stable and transient protein-protein interactions in a cell would facilitate the functional annotation of all gene products, and provide insight into the higher-order organization of the proteome. Several methodologies have been developed to detect protein-protein interactions, and some have been adapted to chart interactions at a proteome-wide scale [8].

Materials and Methods

The two component signal transducing protein-protein interactions network analysis has been carried out for ten different strains of *Enterococcus faecalis* such as *Enterococcus faecalis* V583, *Enterococcus faecalis* X98, *Enterococcus faecalis* CH188, *Enterococcus faecalis* D6, *Enterococcus faecalis* JH1, *Enterococcus faecalis* HIP11704, *Enterococcus faecalis* TX0102, *Enterococcus faecalis* DAPTO 516, *Enterococcus faecalis* TX0635, *Enterococcus faecalis* 62 and *Enterococcus faecalis* V583 [9-11]. This V583 strain was found to lack the cytolysin gene and a surface adhesin, Esp, that contributes to urinary tract infections. Mobile genetic elements make up one quarter of the genome [12].

Protein tracing

Proteins which are participating in Two Component Signal Transduction (TCST) of specific *Enterococcus faecalis* strains are traced out using NCBI protein database search where we have found number of proteins in a strain participating in TCST were selected and downloaded as fasta format.

Protein interactions, sequence and network visualization

The proteins of specific *Enterococcus faecalis* strains which are participating in the two component signal transduction were identified from NCBI/Protein database. The protein sequences have been downloaded and group of proteins of specific strain are uploaded to the STRING database for the identification of particular protein-protein interactions. STRING does not consider any specific splicing isoforms or posttranslational modifications, but instead represents each protein-coding locus in a genome by a single protein. STRING imports protein association knowledge not only from databases of physical interactions, but also from databases of curated biological pathway knowledge. Functional partnerships between proteins are at the core of complex cellular phenotypes, and the networks formed by interacting proteins provide researchers with crucial scaffolds for modeling, data reduction and annotation. STRING is a database and web resource dedicated to protein-protein interactions, including both physical and functional interactions. It weights and integrates information from numerous sources, including experimental repositories, computational prediction methods and public text collections, thus acting as a Metadatabase

that maps all interactions evidence onto a common set of genomes and proteins. The most important new developments in STRING 8 over previous releases include a URL-based programming interface, which can be used to query STRING from other resources, improved interactions prediction via genomic neighborhood in prokaryotes, and the inclusion of protein structures. Version 8.0 of STRING covers about 2.5 million proteins from 630 organisms, providing the most comprehensive view on protein-protein interactions currently available (Figure 1) [13].

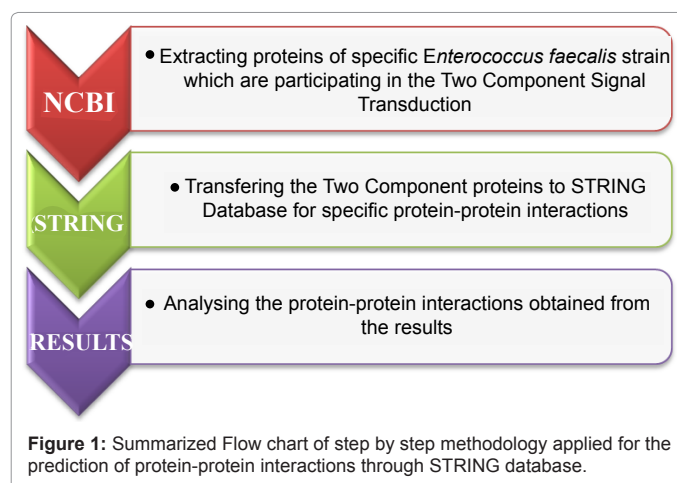
STRING generates the protein interactions based on the Neighborhood, Gene fusion, Co-occurrence, Experiments, Databases, Text mining and Homology. Comparatively STRING gives score between 0.001 to 0.999. The STRING score is divided into Low Confidence (0.150), Medium Confidence (0.400), High Confidence (0.700) and Higher Confidence (0.900) which shows a perfect interactions network.

Analysis of host-pathogen protein-protein interactions

A pathogen causing an infectious disease generally exhibits extensive interactions with the host [14]. These complex crosstalks between a host and a pathogen may assist the pathogen in successfully invading the host organism, breaching its immune defense, as well as replicating and persisting within the organism. Systematic determination and analysis of Host-Pathogen Interactions (HPIs) is a challenging task from both experimental and computational approaches, and is critically dependent on the previously obtained knowledge about these interactions.

The molecular mechanisms of Host-Pathogen Interactions (HPIs) include interactions between proteins, nucleotide sequences, and small ligands [15-18]. "Intra-species PPI", where two proteins from the same species interact with each other and "Inter-species PPI" where two proteins from two different species interact. Host-pathogen protein-protein interactions play a vital role in initiating infection are a subset of inter-species interactions. Identification and study of HPIs is critical for understanding molecular mechanisms of infection and subsequent development of drug targets (Figure 2).

The Host-Pathogen Interactions data of 10 strains of *Enterococcus faecalis* was generated using HPIDB (Host Pathogen Interactions Database). The two component signal transducing proteins are given for BLASTP in HPIDB full database BLAST with parameters of Blossom 62



Algorithm Matrix and E-Value of 10.0 [19]. The obtained PPI data was loaded in to Cytoscape 2.8.2 for the identification of protein-protein interactions of specific two component signal transducing proteins of different strains of *Enterococcus faecalis*.

Network topology analysis

The combined BLAST data has been generated from Host Pathogen Interactions Data Base 1.0. (Figure 3) [20] which was imported into the Cytoscape 2.8.2. When the blast data was imported taking E-value of the proteins and predicted partners as source interactions, bit score was taken as a target interactions and interactions type was given as default. The Cytoscape 2.8.2 generates a grid view of protein-protein interactions which are arranged according to their individual blast scores. The topological parameters of *Enterococcus faecalis* strains PPI networks were analyzed using Network Analysis plugin [21] of Cytoscape.2.8.2. The edges in all PPI networks were treated as undirected. The definition of network topological measure can be found in Network Analyzer Online Help (<http://med.bioinf.mpi-inf.mpg.de/netanalyzer/help/2.6.1/index.html>).

Fitting a line

Network Analyzer provides another useful feature - fitting a line on the data points of some complex parameters. The method applied is the least squares method for linear regression [22]. Network Analyzer gives the correlation between the given data points and the corresponding points on the fitted line. In addition, the R-squared value (also known as coefficient of determination) is reported. Fitting a line can be used to identify linear dependencies between the values of the *x* and *y* coordinates in a complex parameter shows the fitted line on a neighborhood connectivity distribution. The correlation between the data points and corresponding points on the line is approximately 0.969. The R-squared value is 0.939 giving a relatively high confidence that the underlying model is indeed linear.

Fitting a power law

The degree distribution of many biological networks approximates a power law: $DD(k) \sim k^{-\alpha}$ for some negative constant α . Several studies have reported similar properties of the average clustering coefficient distribution [23] and the topological co-efficient [24]. Network Analyzer can fit a power law to some topological parameters. Please note that Network Analyzer uses the least squares method [22] and only points with positive coordinate values are considered for the fit. This

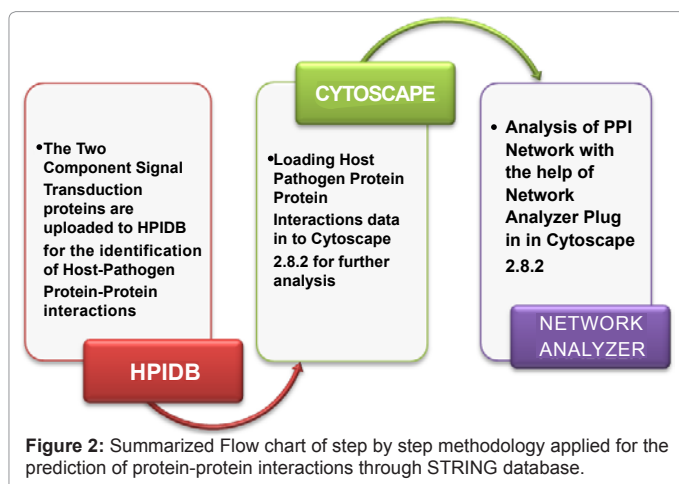


Figure 2: Summarized Flow chart of step by step methodology applied for the prediction of protein-protein interactions through STRING database.

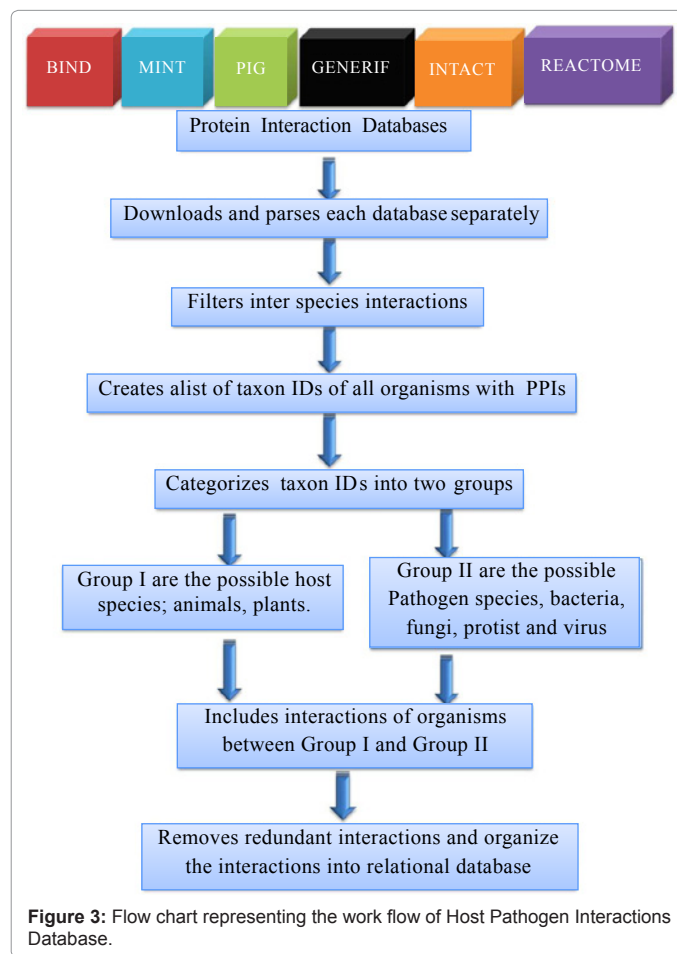


Figure 3: Flow chart representing the work flow of Host Pathogen Interactions Database.

approach fits a line on logarithmized data and may be inappropriate for supporting certain hypotheses. Network Analyzer gives the correlation between the given data points and the corresponding points on the fitted curve. In addition, the R-squared value (also known as coefficient of determination) is reported. This coefficient gives the proportion of variability in a data set, which is explained by a fitted linear model [25,26]. Therefore, the R-squared value is computed on logarithmized data, where the power-law curve: $y = \beta x^{\alpha}$ is transformed into linear model: $\ln y = \ln \beta + \alpha \ln x$.

Results and Discussion

We have generated protein-protein interactions networks in different strains of *Enterococcus faecalis*. We have observed several proteins which have not been participating in the interactions network. All the interactions are predicted under the parameters adjusted to highest confidence (0.900) and the numbers of predicted interactions are adjusted to not more than 10. The color schema of bands in the network denotes the specific parameters at which the proteins are interacting with one another.

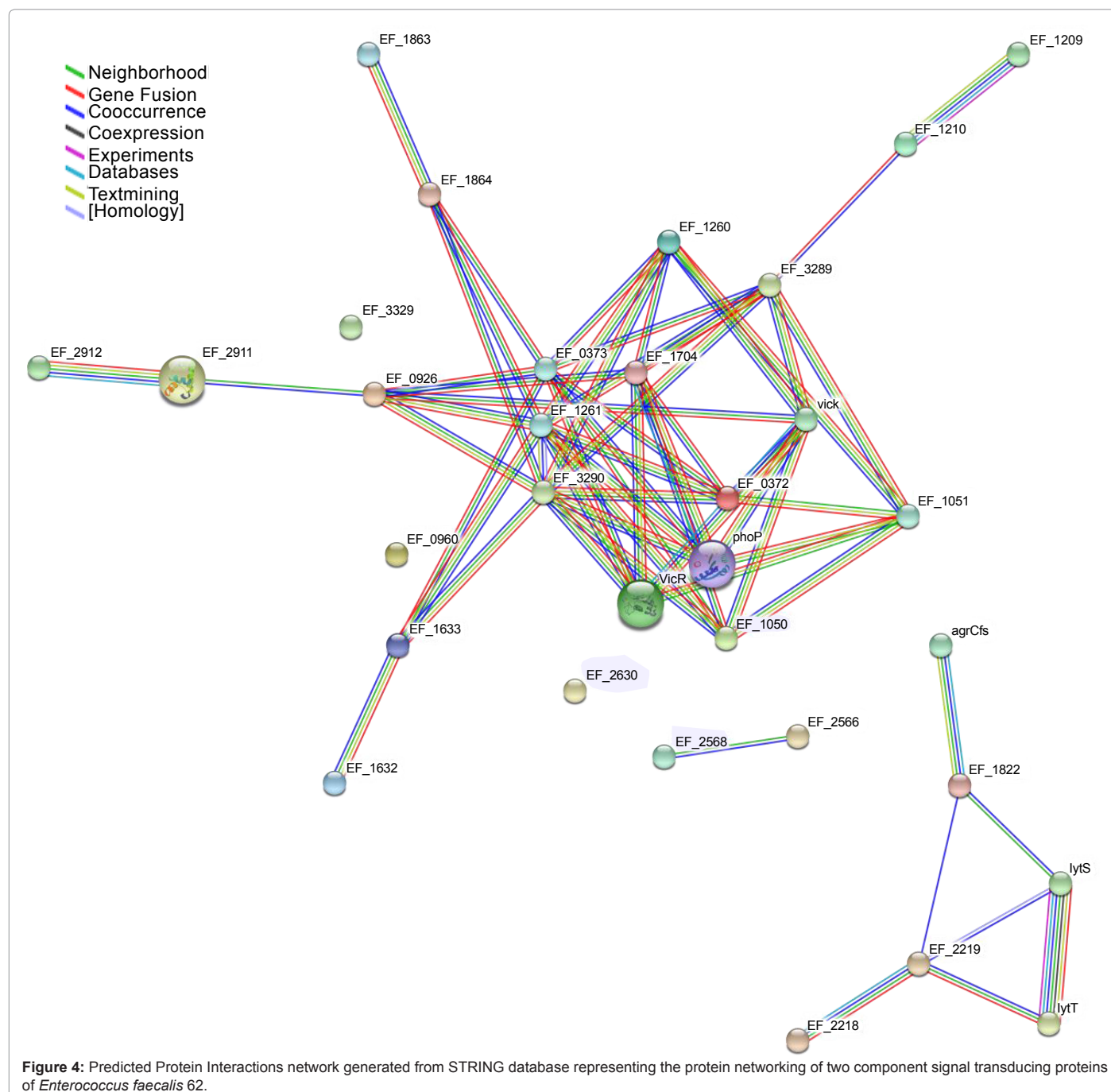
Protein Interactions of *Enterococcus faecalis* 62

Enterococcus faecalis 62 contains the gelatinase *gel E* and serine proteinase *sprE* genes but displays a gelatinase-negative phenotype [27]. In this Protein-Protein Interactions data of *Enterococcus faecalis* 62 we have figured out three signal transducing proteins which are not participating in the interactions. In figure 3 you can observe EF_3329

DNA-binding response regulator, EF_0960 endonuclease/exonuclease/phosphatase family protein and EF_2630 transcriptional regulator which are not participating in the protein interactions network, whereas EF_2568 aminotransferase, class V and EF_2566 hypothetical protein are interacting with each other without interacting with the other network proteins. We have identified a sub network of EF_2218 AraC family DNA-binding response regulator, EF_2219 sensor histidine kinase, EF_1822 response regulator, *lytS* sensor histidine kinase; member of the two-component regulatory system, *agrCfs* histidine kinase, putative, and *lytT* response regulator; Member of the two-component regulatory system *lytS/lytT* that probably regulates genes involved in cell wall metabolism (Figure 4).

Comparative network analysis of predicted protein interactions

The protein-protein interactions built for the 9 strains of *Enterococcus faecalis* are on the basis of protein-protein interactions of *Enterococcus faecalis* V583 which are available in the STRING 9.0 database. According to the protein interactions networks predicted in 10 different strains of *Enterococcus faecalis*, we have figured out a protein which is not participating in any protein interactions network except in protein interactions network of X98 where we have identified the absence of the protein EF_3329 DNA-binding response regulator. Sub networks which are not at all having any interactions with the



main network have been identified in 9/10 networks except in the *Enterococcus faecalis* D6 where no sub networks are formed. Most of the sub networks consists of following proteins EF_2218 AraC family DNA-binding response regulator, EF_2219 sensor histidine kinase, EF_1822 response regulator, *lytS* sensor histidine kinase; member of the two-component regulatory system, *agrCfs* histidine kinase; putative, and *lytT* response regulator; Member of the two-component regulatory system *lytS/lytT* that probably regulates genes involved in cell wall metabolism which did not have any interactions with the main protein interactions network in all the 10 *Enterococcus faecalis* strains.

Topological analysis of *Enterococcus faecalis* host-pathogen PPI networks

We have calculated the topological parameters of PPI networks of *Enterococcus faecalis* using Network Analysis [21] plug in present in the Cytoscape.2.8.2. The obtained results consists of connected components, Network diameter, Network radius, Network Centralization, Shortest paths, Characteristic Path length, Average numbers of neighbors, Number of Nodes, Network Density and Network Heterogeneity. During the power law fit of Network analysis some data points has identified containing non-positive coordinates so, only points with positive coordinates are induced in the fit. Here we have comparatively analyzed the results of Network Analysis of ten *Enterococcus faecalis* strains.

Host-Pathogen protein interactions analysis of *Enterococcus faecalis* 62

The Network analysis data of *Enterococcus faecalis* 62 contains connected components 20, Network diameter is 6, Network radius is 1, Network Centralization is 0.118, Shortest paths are 1796 (10%), Characteristic Path length is 2.606, Average numbers of neighbors are 1.712, Number of Nodes are 132, Network Density is 0.013 and Network Heterogeneity is 1.311. In the graph of Node Degree Distribution (Figure 5A) the number of proteins with a given link (k) in the protein interactions network of *Enterococcus faecalis* 62 follows a fitted line $y = a+bx$ (a) 23.053, (b) -1.849. Correlation is 0.377, $R^2 = 0.142$ and power law $y = ax^b$ (a) 23.639, (b) -1.410. Correlation is 0.944, $R^2 = 0.685$ for the power law fit. The Topological Co-efficient (Figure 5B) was plotted against the number of links. The topological coefficient with a given link (k) in the *Enterococcus faecalis* 62 protein interactions network follows a same fitted line (a) 0.528, (b) -0.036. Correlation is 0.884.

$R^2 = 0.781$ and power law (a) 0.973, (b) -0.918, Correlation = 0.987, $R^2 = 0.929$ and for the Neighboring Connectivity Distribution (Figure 5C) (a) 4.002, (b) -0.252, Correlation = 0.495, $R^2 = 0.245$ for fitted line and for power law fit (a) 4.420, (b) -0.582, Correlation = 0.831, $R^2 = 0.398$. Stress Centrality Distribution (Figure 5D) (a) 42.846, (b) -0.355, Correlation = 0.325, $R^2 = 0.106$ for fitted line and for power law fit (a) 2.286, (b) 0.370, Correlation = 0.958, $R^2 = 0.931$. The stress of a node n is the number of shortest paths passing through n . The stress distribution gives the number of nodes with stress s for different values of s . The values for the stress are grouped into bins whose size grows exponentially by a factor of 10. The Betweenness centrality distribution (Figure 5E) was built as (a) 0.282, (b) 0.489, Correlation = 0.346, $R^2 = 0.236$ for the power law fit, closeness centrality distribution (Figure 5F) (a) 0.486, (b) 0.154, Correlation = 0.218, $R^2 = 0.051$ for power law fit.

The correlation and R-Squared values of power law fit showed a high fit when compare to the standard (Power law perfect fit Correlation = 1.000 and R-Squared = 1.000) in Topological Co-efficient (Correlation = 0.987, $R^2 = 0.929$) and Node Degree Distribution (Correlation =

0.944, $R^2 = 0.685$). The interactions analysis also has showed shortest paths 1796 (10%). The above all parameters prove that the protein interactions network of *Enterococcus faecalis* 62 is perfectly fit.

Comparative network analysis of host-pathogen protein interactions

The two component signal transduction host-pathogen protein interactions network analysis of 10 different *Enterococcus faecalis* strains have yielded us antithetic statistical data containing Node degree distribution, Topological coefficients, Neighborhood connectivity distribution, Stress centrality distribution, Betweenness centrality and Closeness centrality. Here we have given the compared statistical data analysis.

Network parameters

The network parameters have been generated by undirected network analysis. Network parameters of 10 variant *Enterococcus faecalis* strains have been comparatively analyzed (Table 1). The network parameters contain Connected Components, Network Diameter, Network Radius, Network Centralization, Shortest Paths, Characteristic Path Length, Average Numbers of Neighbors, Number of Nodes, Network Density and Network Heterogeneity. *Enterococcus faecalis* TX0635 shows high number of connected components (Table 1)(Figure 6) compared to other strains of *Enterococcus faecalis*. Network diameter of *Enterococcus faecalis* CH188 is high (Figure 7C), Network radius of all the strains is same. Network Centralization of *Enterococcus faecalis* HIP11704 is higher (Figure 7A).

Shortest paths

Shortest paths are high in *Enterococcus faecalis* CH188, Characteristic Path Length is high in *Enterococcus faecalis* CH188 (Table 1)(Figure 7E), Average Numbers of Neighbors are more in *Enterococcus faecalis* X98 (Table 1)(Figure 7G), Number of Nodes are high in *Enterococcus faecalis* X98 (Table 1)(Figure 7B), Network Density is high in *Enterococcus faecalis* D6 and *Enterococcus faecalis* V583 (Table 1)(Figure 7D) and Network Heterogeneity is high in *Enterococcus faecalis* TX0102 (Table 1)(Figure 7F).

Node degree distribution

The correlation value of *Enterococcus faecalis* X86 is comparatively high (Supplementary Table 1) whereas *Enterococcus faecalis* TX0635 gave a peak value in the comparison of R-Squared values under the fitted line algorithm ($y = a+bx$). In the Fitted Power Law $Y = ax^b$ correlation value of *Enterococcus faecalis* X86 is high whereas *Enterococcus faecalis* CH188 is high in the comparison of R-Squared values (Supplementary Table 2).

Topological coefficients

The correlation and R-Squared values of *Enterococcus faecalis* TX0102 (Supplementary Table 3) are high compared to the other strains of *Enterococcus faecalis* in the fitted line algorithm ($y = a+bx$). Compared to the values of all other *Enterococcus faecalis* strains Correlation and R-Squared values of *Enterococcus faecalis* HIP11704 showed a highest fit of network (Supplementary Table 4). The topological coefficient fitness is considerably high only which the protein interactions network is highly linked giving a highest fit value in proteins interacting in a network.

Neighborhood connectivity distribution

The correlation and R-Squared values of *Enterococcus faecalis*

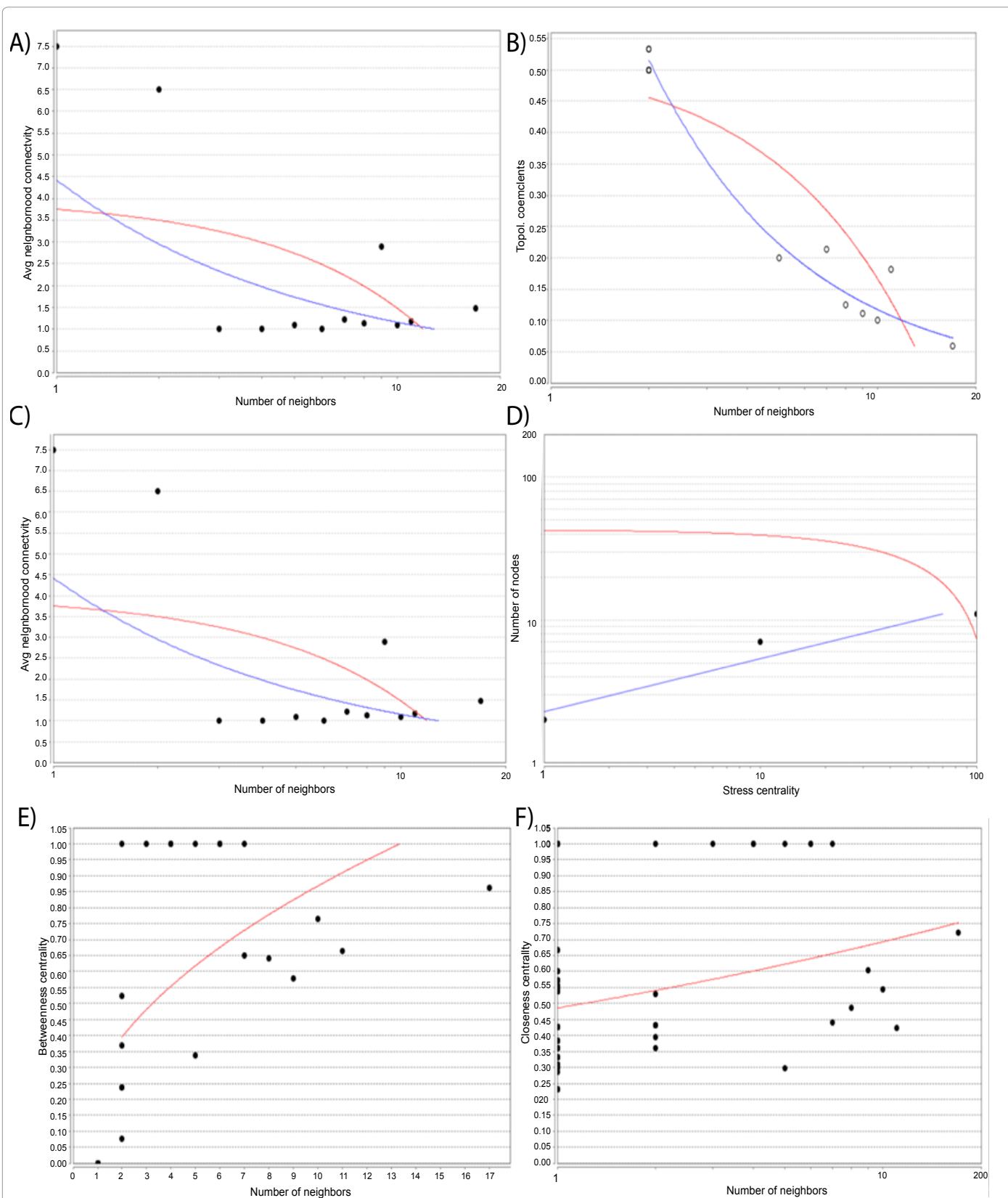
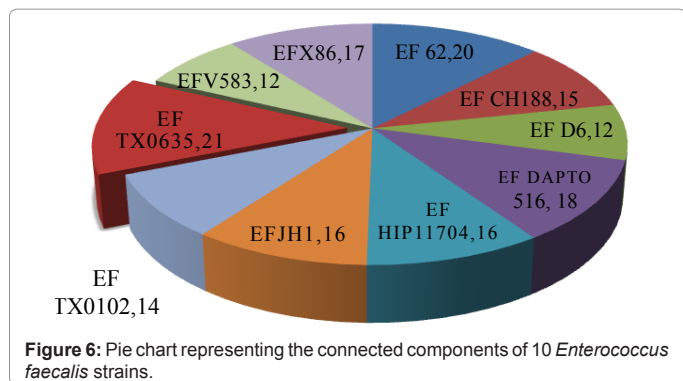


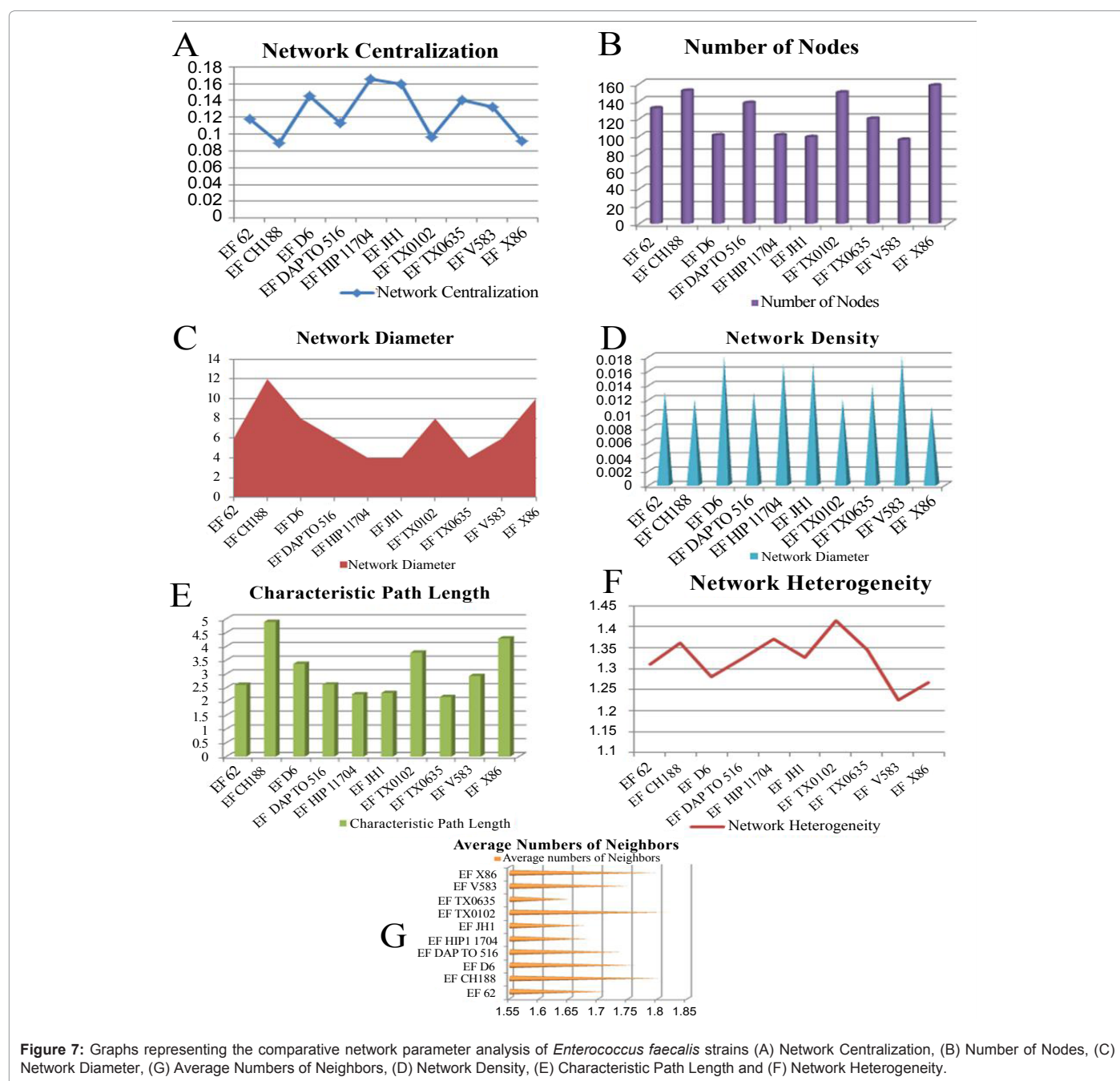
Figure 5: Topological properties of *Enterococcus faecalis* 62 PPI Network analyzed from Cytoscape 2.8.2. (A) Node Degree Distribution (B) Topological coefficients Distribution (C) Neighboring Connectivity Distribution (D) Stress Centrality Distribution (E) Betweenness Centrality Distribution (F) Closeness centrality Distribution.



CH188 (Supplementary Table 5) are high in the comparison of fitted line algorithm ($y = a+bx$). But the values are not significant due to the low value ratio. When the fitted power law ($Y = ax^b$) was applied to the neighborhood connectivity distribution *Enterococcus faecalis* CH188 (Supplementary Table 6) showed a higher Correlation and R-Squared values compared to the other *Enterococcus faecalis* strains. Compared to the fitted power law ($Y = ax^b$) standard (Power law perfect fit Correlation = 1.000 and R-Squared = 1.000) the higher values which were observed during comparative analysis are low.

Stress centrality distribution

In the stress centrality distribution of network analysis, Correlation and R- Squared values of *Enterococcus faecalis* TX0635 (Supplementary



EF Strains	C.C	N.Di	N.R	N.C	S.P	C.P.L	A.N.N	N.N	N.De	N.H
EF 62	20	6	1	0.118	1796(10%)	2.606	1.712	132	0.013	1.311
EF CH188	15	12	1	0.089	4768(20%)	4.892	1.803	152	0.012	1.362
EF D6	12	8	1	0.145	2108(20%)	3.367	1.762	101	0.018	1.280
EF DAPTO 516	18	6	1	0.113	1912(10%)	2.618	1.739	138	0.013	1.325
EF HIP11704	16	4	1	0.166	1226(12%)	2.263	1.683	101	0.017	1.371
EF JH1	16	4	1	0.160	1184(12%)	2.309	1.677	99	0.017	1.326
EF TX0102	14	8	1	0.096	4260(19%)	3.776	1.813	150	0.012	1.414
EFTX0635	21	4	1	0.140	1254(8%)	2.166	1.65	120	0.014	1.345
EF V583	12	6	1	0.132	1618(17%)	2.925	1.75	96	0.018	1.226
EF X86	17	10	1	0.092	4404(17%)	4.288	1.797	158	0.011	1.267

Table 1: Comparison of network parameters in 10 different *Enterococcus faecalis* strains [*F-Enterococcus faecalis*, C.C: Connected Components; N.Di: Network Diameter; N.R: Network Radius; N.C: Network Centralization; S.P: Shortest Paths; C.P.L: Characteristic Path Length; A.N.N: Average Numbers of Neighbors; N.N: Number of Nodes; N.De: Network Density and N.H: Network Heterogeneity].

Table 7) are high in the comparison of fitted line algorithm ($y = a+bx$). But the values are not significant due to the low value ratio. When the fitted power law ($Y = ax^b$) was applied to the stress centrality distribution *Enterococcus faecalis* DAPTO516 (Supplementary Table 8) showed a higher Correlation and R-Squared values compared to the other *Enterococcus faecalis* strains. Compared to the fitted power law ($Y = ax^b$) standard (Power law perfect fit Correlation = 1.000 and R-Squared = 1.000), the higher values which were observed during comparative analysis are high and shows a perfect fit of network. Compared to the topological coefficients fitted power law ($Y = ax^b$) values the stress centrality distribution values are a bit less but considerable for a high network fit.

Betweenness centrality distribution

The correlation and R- Squared value of Betweenness Centrality are very low when compared with the standard (Power law perfect fit Correlation = 1.000 and R-Squared = 1.000). *Enterococcus faecalis* D6 (Supplementary Table 9) showed a high correlation point where as *Enterococcus faecalis* 62 showed a higher R-Squared value when compare with the correlation and R-squared values of other *Enterococcus faecalis* strains.

Closeness centrality distribution

Closeness centrality is a measure of how rapidly information spreads from a given node to other reachable nodes in the network. The closeness centrality distribution of the *Enterococcus faecalis* TX0635 is high (Correlation = 0.431, R-Squared = 0.181) (Supplementary Table 10) this shows that highly connected proteins in the network have a pronounced ability to spread information in the network. The same tendency also exists in other strains, but less when compared with the *Enterococcus faecalis* TX0635 protein interactions network. Nodes with high closeness centrality have potential significance for responding to external perturbations and for maintaining network stabilization. This may be part of the explanation why the highly connected “hub” proteins usually play essential roles in cell processes [28].

Conclusion

The prevailing study presents the entire protein-protein interactions of two component signal transducing proteins where we have comparatively analyzed protein-protein interactions of ten *Enterococcus faecalis* strains. All interactions of proteins which are participating in the two component signal transduction of *Enterococcus*

faecalis has been predicted using STRING database. Host-Pathogen protein interactions have been identified and the Host-Pathogen interactions networks have been analyzed. The two component signal transducing proteins formed a highly linked network. The protein interactions which we have generated are useful for the studies of drug targeting. The protein-protein interactions information was connected to the pathogenic protein secretion pathway and gene regulation. These results will serve as a unique resource for further dissection of signal transduction and infection mechanisms in *Enterococcus faecalis*, as well as for the development of novel drugs.

Acknowledgement

The authors thank UGC/MHRD start up grant for newly joined faculty funds and also acknowledge proteomics facility.

Databases and Software's

NCBI, HPIDB, STRING, NCBI Protein database, Cytoscape 2.8.2.

Supporting Information Available

Only PPI data of *E. faecalis* 62 was shown in this paper, the complete PPI network high resolution images of remaining strains of *Enterococcus faecalis*; Host pathogen interactions graph data of remaining strains of *Enterococcus faecalis*; PPI enrichment analysis, Tables representing the topological analysis and measures of each PPI network; power-law distribution of node degree and topological coefficient; the distribution of protein category and PPI; conserved subnetworks; protein-protein interactions pairs of *Enterococcus faecalis*; are available as a supporting information.

References

- Terradot L, Noiro-Gros MF (2011) Bacterial protein interaction networks: puzzle stones from solved complex structures add to a clearer picture. *Integr Biol (Camb)* 3: 645-652.
- West AH, Stock AM (2001) Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem Sci* 26: 369-376.
- Hoch JA, Silhavy TJ (1995) Two-Component Signal Transduction. ASM Press, Washington DC, USA.
- Stock AM, Robinson VL, Goudreau PN (2000) Two-component signal transduction. *Annu Rev Biochem* 69: 183-215.
- Wolanin PM, Thomason PA, Stock JB (2002) Histidine protein kinases: key signal transducers outside the animal kingdom. *Genome Biol* 3: S3013.
- Pao GM, Saier MH Jr (1995) Response regulators of bacterial signal transduction systems: selective domain shuffling during evolution. *J Mol Evol* 40: 136-154.
- Hancock L, Perego M (2002) Two-component signal transduction in *Enterococcus faecalis*. *J Bacteriol* 184: 5819-5825.

8. Noirot P, Noirot-Gros MF (2004) Protein interaction networks in bacteria. *Curr Opin Microbiol* 7: 505-512.
9. Lebreton F, Riboulet-Bisson E, Serror P, Sanguinetti M, Posteraro B, et al. (2009) ace, Which encodes an adhesin in *Enterococcus faecalis*, is regulated by Ers and is involved in virulence. *Infect Immun* 77: 2832-2839.
10. Coburn PS, Pillar CM, Jett BD, Haas W, Gilmore MS (2004) *Enterococcus faecalis* senses target cells and in response expresses cytolysin. *Science* 306: 2270-2272.
11. Engelbert M, Mylonakis E, Ausubel FM, Calderwood SB, Gilmore MS (2004) Contribution of gelatinase, serine protease, and fsr to the pathogenesis of *Enterococcus faecalis* endophthalmitis. *Infect Immun* 72: 3628-3633.
12. Shankar N, Baghdayan AS, Gilmore MS (2002) Modulation of virulence within a pathogenicity island in vancomycin-resistant *Enterococcus faecalis*. *Nature* 417: 746-750.
13. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, et al. (2009) STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* 37: D412-D416.
14. Münter S, Way M, Frischknecht F (2006) Signaling during pathogen infection. *Sci STKE* 2006: re5.
15. Lengeling A, Pfeffer K, Balling R (2001) The battle of two genomes: genetics of bacterial host/pathogen interactions in mice. *Mamm Genome* 12: 261-271.
16. Kahn RA, Fu H, Roy CR (2002) Cellular hijacking: a common strategy for microbial infection. *Trends Biochem Sci* 27: 308-314.
17. Stebbins CE (2005) Structural microbiology at the pathogen-host interface. *Cell Microbiol* 7: 1227-1236.
18. Forst CV (2006) Host-pathogen systems biology. *Drug Discov Today* 11: 220-227.
19. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389-3402.
20. Kumar R, Nanduri B (2010) HPIDB—a unified resource for host-pathogen interactions. *BMC Bioinformatics* 11: S16.
21. Assenov Y, Ramírez F, Schelhorn SE, Lengauer T, Albrecht M (2008) Computing topological parameters of biological networks. *Bioinformatics* 24: 282-284.
22. Weisstein EW (2012) Least Squares Fitting -- Power Law. *MathWorld - A Wolfram Web Resource*.
23. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL (2002) Hierarchical organization of modularity in metabolic networks. *Science* 297: 1551-1555.
24. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, et al. (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 122: 957-968.
25. Goldstein ML, Morris SA, Yena GG (2004) Problems with fitting to the power-law distribution. *Eur Phys J B* 41: 255-258.
26. Tanaka R, Yi TM, Doyle J (2005) Some protein interaction data do not exhibit power law statistics. *FEBS Lett* 579: 5140-5144.
27. Brede DA, Snipen LG, Ussery DW, Nederbragt AJ, Nes IF (2011) Complete genome sequence of the commensal *Enterococcus faecalis* 62, isolated from a healthy Norwegian infant. *J Bacteriol* 193: 2377-2378.
28. Jeong H, Mason SP, Barabási AL, Oltvai ZN (2001) Lethality and centrality in protein networks. *Nature* 411: 41-42.