Gene 639 (2018) 1–10

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Comparison of gene co-expression networks in *Pseudomonas aeruginosa* and *Staphylococcus aureus* reveals conservation in some aspects of virulence

Nazanin Hosseinkhan[a,c,*], Zaynab Mousavian[b,c], Ali Masoudi-Nejad[c,**]

[a] Basic and Molecular Epidemiology of Gastrointestinal Disorders Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[b] Department of Computer Science, School of Mathematics, Statistics, and Computer Science, University of Tehran, Tehran, Iran
[c] Laboratory of Systems Biology and Bioinformatics (LBB), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

ARTICLE INFO

Keywords:
P. aeruginosa
S. aureus
Cross species
Gene co-expression network

ABSTRACT

Pseudomonas aeruginosa and *Staphylococcus aureus* are two evolutionary distant bacterial species that are frequently isolated from persistent infections such as chronic infectious wounds and severe lung infections in cystic fibrosis patients. To the best of our knowledge no comprehensive genome scale co-expression study has been already conducted on these two species and in most cases only the expression of very few genes has been the subject of investigation. In this study, in order to investigate the level of expressional conservation between these two species, using heterogeneous gene expression datasets the weighted gene co-expression network analysis (WGCNA) approach was applied to study both single and cross species genome scale co-expression patterns of these two species. Single species co-expression network analysis revealed that in *P. aeruginosa*, genes involved in quorum sensing (QS), iron uptake, nitrate respiration and type III secretion systems and in *S. aureus*, genes associated with the regulation of carbon metabolism, fatty acid-phospholipids metabolism and proteolysis represent considerable co-expression across a variety of experimental conditions. Moreover, the comparison of gene co-expression networks between *P. aeruginosa* and *S. aureus* was led to the identification of four co-expressed gene modules in both species totally consisting of 318 genes. Several genes related to two component signal transduction systems, small colony variants (SCVs) morphotype and protein complexes were found in the detected modules. We believe that targeting the key players among the identified co-expressed orthologous genes will be a potential intervention strategy to control refractory co-infections caused by these two bacterial species.

1. Introduction

With the ever-growing of high-throughput gene expression datasets and sequence similarity data for various species, cross species gene expression studies have recently gained significant attentions. In these type of studies, the conservation and variation of biological processes among different species are subjected to investigation (Stuart et al., 2003). It is believed that during the evolution, the structure and the function of many biological processes have remained unchanged and therefore several biological systems function in a similar manner across different species (Okyere et al., 2014). Consequently, comparative co-expression analysis among several species can help researchers to make conclusions about the possible functions of still uncharacterized genes (Wang et al., 2013). To do this, the expression of orthologous genes, i.e. genes with the high levels of similarity in their sequences that in most cases represent similar functions across different species, can be compared between species (Zarrineh, 2011). Gene sets with the similar expression pattern in all or a subset of experimental conditions build co-expressed gene clusters (modules) in each single organism. Comparing the detected modules across organisms may lead to the identification of conserved co-expressed gene modules.

Conducting cross species gene expression studies on bacterial species which are often present in poly-microbial communities can assist to unravel common conserved biological pathways between them. *Pseudomonas aeruginosa* and *Staphylococcus aureus* are two evolutionary distant opportunistic pathogens simultaneously isolated from the persistent, hard-to-eradicate infections such as those frequently observed in chronic wounds (Korgaonkar et al., 2013) as well as in cystic fibrosis

* Correspondence to: N. Hosseinkhan, Basic and Molecular Epidemiology of Gastrointestinal Disorders Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
** Corresponding author.
E-mail addresses: hosseinkhan@alumni.ut.ac.ir (N. Hosseinkhan), amasoudin@ut.ac.ir (A. Masoudi-Nejad).
URL: http://LBB.ut.ac.ir (A. Masoudi-Nejad).

http://dx.doi.org/10.1016/j.gene.2017.10.005
Received 4 January 2017; Received in revised form 23 September 2017; Accepted 3 October 2017
Available online 04 October 2017

(CF) airways (Baldan et al., 2014; Pernet et al., 2014; Rüger et al., 2014; Filkins et al., 2015). Despite the large evolutionary distance and the large amount of difference between GC-content in *P. aeruginosa* (~66%) and *S. aureus* (~33%), 977 orthologous genes have been already recognized common between these two species.

It has been demonstrated that *P. aeruginosa* and *S. aureus* are both capable of developing a refractory phenotype encountering severe conditions such as long-term exposures to antibiotics. This morphotype which is known as small colony variants (SCVs) is characterized by properties such as small growth and increased resistance against commonly used antibiotics (Besier et al., 2007; Garcia et al., 2013; Wolter et al., 2013; Yagci et al., 2013).

Although *P. aeruginosa* and *S. aureus* have already been the subject of several gene expression studies, to the best of our knowledge, no comprehensive genome-scale study has been conducted to compare the expressional behavior of these two species.

In this study, we first investigated co-expressed gene modules in heterogeneous datasets of *P. aeruginosa* and *S. aureus* separately through applying weighted gene co-expression network analysis (WGCNA), which is a computational method for identification of correlation patterns among genes (Zhang and Horvath, 2005). In contrast to homogeneous gene expression datasets which commonly belong to the same experimental conditions and allows for direct comparison between expressional patterns of orthologous genes, the heterogeneous datasets can be used to study conserved mutual expression behavior between pairs of genes across various conditions (Zarrineh et al., 2011). This can also be useful in the detection of core modules representing similar expression pattern regardless of the experienced environmental conditions. The detected modules are representatives of basic cellular biological processes, and can be applied for assigning function to still uncharacterized members of a certain module based on co-expression with functionally known members of that module. Furthermore, analyzing the co-expressed modules in these two species will shed more light on the underlying molecular mechanisms responsible for the better adaptation of *P. aeruginosa* to a variety of environmental conditions compared to *S. aureus*. Moreover, in order to understand the level of co-expression conservations between these two evolutionary distant species, we also conducted cross species co-expression study using WGCNA package. Finding modules of common co-expressed genes in *P. aeruginosa* and *S. aureus* will open new avenues for controlling refractory co-infections caused by these two pathogens.

2. Methods

2.1. Datasets and preprocessing

Very few gene expression studies have been already conducted on *S. aureus* and *P. aeruginosa* strains obtained from lung and chronic wounds infections. This, beside the fact that the aim of the present study was to investigate conserved gene expression behavior of these two organisms, was led us to employ heterogeneous gene expression datasets. We used gene expression compendia of *P. aeruginosa PAO1*, consisting of 255 microarray samples, covering a variety of conditions, deposited in COLOMBOS ("COLections Of Microarrays for Bacterial OrganismS") (Meysman et al., 2014). To construct gene expression compendia for *S. aureus*, 252 Affymetrix microarray samples (most of which consist of methicillin resistant and methicillin sensitive strains) all belong to GPL1339, the most frequently used platform for *S. aureus*, were retrieved from the GEO database (Table S1, Supplementary Materials). The preprocessing steps including normalization with RMA method were carried out using the affylmGUI software package in R (Wettenhall et al., 2006). After preprocessing steps, two expression matrices with genes in rows and samples in columns were generated for each species separately.

2.2. Single network analysis

The microarray compendia of each species were first inspected by the *goodSamplesGenes* function of the WGCNA R package to inspect data for missing values and also for genes with zero-variances. In this study, all genes and samples were identified as good genes and good samples and, consequently no array has been removed from the primary datasets. The microarray samples were also clustered to detect outlier samples which are samples with substantial difference in expression values compared to other samples. To construct a weighted gene co-expression network, correlations between the expression values of all pairs of genes in the gene expression profiles of each species, were calculated. We used the Pearson correlation coefficient measure to compute the correlation between the expression values of each gene pair *cor(i,j)*. Since our goal was to find modules of genes with similar expression pattern in each species, we used the *signed hybrid* network option for network construction in which only the positive correlations among genes (i.e. simultaneous increase or decrease in the level of expressions) are maintained in the network and the links with negative correlations are removed. We used the *adjacency* function of package WGCNA with the input parameters (corFnc = "pearson", type = "-signed hybrid") to calculate the adjacency matrix (a square matrix containing gene-gene correlation values) of co-expression values for each species (Fig. 1, part a).

To construct a gene co-expression network with the scale-free topology property (i.e. the network with few genes having several links versus large number of genes with few connections), which is the main characteristic of most biological networks; we estimated the power β (the lowest integer that its resulting network satisfies approximate scale-free topology) using the *pickSoftThreshold* function of package WGCNA. Then parameter β was used as an input for the *adjacency* function of package WGCNA to which the co-expression values were raised. Using the parameter β, we constructed a weighted gene co-expression network with more emphasis on strong correlations and reducing the effect of weak correlations.

After generating the adjacency matrix, the adjacency values were converted into the similarity values for each pair of genes using a generalized version of Topological Overlap Measure (TOM) which calculates the similarities based on the number of shared neighbors between gene pairs in the resulting co-expression network (Zhao et al., 2010). We used the *TOMsimilarity* function of package WGCNA to perform this conversion.

Following the construction of the similarity matrix, to find groups of genes with similar expression pattern, genes were clustered into modules using the average linkage hierarchical clustering algorithm as implemented in the *flashClust* function. Then we used the *cutreeDynamic* function to cut the branches of the resulted dendrogram which results in the generation of gene modules. In the WGCNA package, each module is represented by a distinct color and the gene expression profiles of each module are summarized by a module eigengene (ME) calculated by the *moduleEigengenes* function. According to the module eigengenes, genes were assigned to modules with different levels of module membership (i.e. MM value), which is calculated by computing the correlation between the module eigengene and the expression profile of each gene. The grey module, represented by ME0, corresponds to genes which could not be assigned to any module and this module was ignored in the downstream analysis. The modules with high level of similarity between their module eigengenes, MEs, were considered as close modules and merged into a single module using the *mergeCloseModules* function.

2.3. Consensus network analysis

The list of common orthologous genes in *P. aeruginosa* and *S. aureus* (i.e. 977 genes) was obtained from the EcoCyc database (Keseler et al., 2005). We searched both compendia for these orthologous genes and