Original Research Paper

Volume : 5 | Issue : 12 | December-2016 • ISSN No 2277 - 8179 | IF : 3.508 | IC Value : 78.46

Differential gene expression analysis of virulence factors in Enterococcus faecalis using clustering algorithm

Biological Science

KEYWORDS: Bacteria, Biological data, Java, Urinary tract, Virulence factors.

Daulatabad Swapna Vidhur	Department of Biotechnology, JB Institute of Engineering & Technology, Yenkapally, Moinabad Mandal, Ranga Reddy District, Hyderabad, Telangana 500075.
Kavya Sri Sai.C	Department of Biotechnology, JB Institute of Engineering & Technology, Yenkapally, Moinabad Mandal, Ranga Reddy District, Hyderabad, Telangana 500075.
Aysha Sherieff	Department of Biotechnology, Joginpally B.R Engineering college, Yenkapally, Moinabad Mandal, Ranga Reddy District, Hyderabad, Telangana-500075.
Zahoorullah. S.MD	Department of Biotechnology, Joginpally B.R Engineering college, Yenkapally, Moinabad Mandal, Ranga Reddy District, Hyderabad, Telangana-500075.

ABSTRACT

Enterococci rank among leading causes of nosocomial bacteremia and urinary tract infection and are also a leading cause of community-acquired subacute endocarditis. Enterococcal factors that contribute to the pathogenesis of disease have been identified. Differences in abundance of virulence genes of mRNA from exponential phase cells grown in serum or urine are compared to baseline expression in 2×YT and the genes related to expression of the enterococcal cytolysin small environment-dependent differences in the abundance of mRNA for cytolysin structural genes. These are cylLL and cylLS and regulatory genes cylR1 and cylR2 of Enterococcus faecalis have been studied by many researchers. Phylogenetic trees are used to represent evolutionary relationships among biological species or organisms. The construction of phylogenetic trees is based on the similarities or differences of their physical or genetic features. Traditional approaches of constructing phylogenetic trees mainly focus on physical features. The recent advancement of high-throughput technologies has led to accumulation of huge amounts of biological data, which in turn changed the way of biological studies in various aspects. In the present study, the differential expression of cytolysin structural (cylLL and cylLS) and regulatory genes of Enterococcus faecalis were analyzed using Cluster and Java Treeview.

Introduction:

Enterococci, which are normal components of human gastrointestinal flora, can cause serious infections such as urinary tract infections, endocarditis, bacteremia and wound infections. In addition, they are now recognized as significant causes of nosocomial infections and are resistant to many antimicrobial agents. Bacterial adherence to host cells is recognized as the initial event in the pathogenesis of many infections. However, little information is available on the factors that promote adhesion of *Enterococcus faecalis* to host tissues. Several studies have shown that E. faecalis strains can adhere to human urinary tract epithelial cells (B. Kreft et al., 1992) and Girardi heart (GH) cells (C.A. Guzman et al.,1991). Potential enterococcal adherence factors involved could be cell surface carbohydrates (C.A. Guzman et al.,1989), the Efa A protein, a homologue of cell surface adhesins found on a number of streptococcal species (A.M. Lowe et al., 1995), the Ace (adhesin of collagen from enterococci) protein, which contains features characteristic of cell surface proteins of Gram-positive bacteria and that can act as a collagen adhesin or the aggregation substance (R.L. Rich et al., 1999). Other factors, such as cytolysin (hemolysin) and gelatinase have been proposed as possible virulence factors of E. faecalis strains. But the role of these factors in enterococcal pathogenicity remains unclear. Cytolysin has hemolytic activity (by lysing a broad spectrum of cells including human,horse, and rabbit erythrocytes) and bacteriolytic activity against Gram-positive bacteria. Cytolysin enhances the virulence of E. faecalis in animal models (J.W. Chow et al.,1993). Enterococci rank among leading causes of nosocomial bacteremia and urinary tract infection and are also a leading cause of community-acquired subacute endocarditis (Jett, B. D.et al, 1994). It was recently shown to be autoinduced by a quorum-sensing mechanism involving a two-component regulatory system (Haas, W. et al., 2002).

Methodology:

Cluster is a program that provide a computational and graphical environment for analyzing data from DNA microarray experiments or other genomic datasets. The program Cluster can organize and analyze the data in a number of different ways. Many of the methods are drawn from standard statistical cluster analysis. Hierarchical clustering methods organizes genes in a tree structure, based on their similarity. Four variants of hierarchical clustering are available in Cluster. In k-means clustering, genes are organized into k clusters, where the number of clusters k needs to be chosen in advance. Self-Organizing Maps create clusters of genes on a two-dimensional rectangular grid, where neighboring clusters are similar. For each of these methods, one of the eight different distance measures can be used. Finally, in Principal Component Analysis, clusters are organized based on the principal component axes of the distance matrix.

The k-means clustering algorithm is a simple, but popular, form of cluster analysis. The basic idea is that you start with a collection of items (e.g. genes) and some chosen number of clusters (k) you want to find. The items are initially randomly assigned to a cluster. The *k-means* clustering proceeds by repeated application of a two-step process where the mean vector for all items in each cluster is computed; Items are reassigned to the cluster whose center is closest to the item.

Since the initial cluster assignment is random, different runs of the *k*-means clustering algorithm may not give the same final clustering solution. To deal with this, the *k*-means clustering algorithms is repeated many times, each time starting from a different initial clustering. The sum of distances within the clusters is used to compare different clustering solutions. The clustering solution with the smallest sum of within-cluster distances is saved.

TreeView is also a program or software that provide a computational and graphical environment for analyzing data from DNA microarray experiments, or other genomic datasets. TreeView allows the organized data taken from Cluster after doing Clustering of given dataset and then to be visualized and browsed.

Results:

Differences in abundance of virulence gene mRNA from log-phase cells grown in serum or urine, compared to baseline expression in 2×YT and the genes related to expression of the enterococcal cytolysin small environment- Dependent differences in the abundance of mRNA for cytolysin structural (*cylL*$_L$ and *cylL*$_S$) and regulatory genes (*cylR*$_1$ and *cylR*$_2$) of Enterococcus faecalis were obtained from Brett D. Shepard etal.,2002 and converted the data in to .txt extension (Table 1).

Gene	Serum	Urine	Serum	Urine
cylR2	2.1	1.8	1.5	24
cylR1	1.6	1.5	1.6	34
cylLl	1.9	1.5	1.9	8
cylLs	1.7	1.5	1.6	12
cylM	3.6	-2.5	-2.5	1.4
cylB	3.8	-2	-1.4	3.4
cyA	7	1.2	1.9	9
cylI	3.7	1.1	-1.1	8
inl	4	6	1.6	162
esp	10	4.9	1.2	24
ace	3.3	1.9	1.1	12
asa	5	6	1.4	19
efa	66	89	5	2195
gls24	15	9	2.1	17
fsrB	28	16	-5	2.1
fsrC	26	11	-3	1.9
gelE	52	7	-12.5	-6.7

Table 1: Different genes showing Serum and Urine results of differential gene expression data (results were obtained from from Brett D. Shepard etal.,2002).

When clustering was performed, there is generation of three files as output, they are .cdt, .gtr, .atr files. Out of these file types only .cdt file was executed in java tree view.

k-Means:

This parameter contains sub parameters where in between genes and arrays adjusted as default. When executed, it shows more clusters than experiments available. In this result there is generation of .kgg file.

Gene	NAME	GWEIGHT	Urine	Serum	Urine
EWEIGHT		0.500000	0.333333	1.000000	0.500000
cylM	cylM	0.071429			0.217290
cylB	cylB	0.071429			0.217290
gelE	gelE	0.071429			0.217290
gls24	gls24	0.111121	0.069966	-0.224090	0.356812
inl like gene	inl like gene	0.154203	-0.784543	-0.911606	
esp	esp	0.286947	-0.328231	0.477556	-0.515425
fsrB	fsrB	0.100000	0.260702		-0.417492
fsrC	fsrC	0.100000	0.260702		-0.417492
cylR2	cylR2	0.078958	0.194087	-0.184934	
cylR1	cylR1	0.078958	0.194087	-0.184934	
cylLl	cylLl	0.078958	0.194087	-0.184934	
cylLs	cylLs	0.078958	0.194087	-0.184934	
cyA	cyA	0.103527	0.135257	-0.220622	0.341727
cylI	cylI	0.111111		0.242462	
ace	ace	0.111111		0.242462	
asa	asa	0.083333	0.260702	-0.184934	
efa	efa	0.083333	0.260702	-0.184934	

SOM (Self Organising Map): The parameters are adjusted, in this mode there is generation of three files, those are .Anf, .Gnf, .data type (Table 2).

Table 2: Enterococcus urine and serum SOM output.

PCA (Principal Component Analysis):

Applying PCA to genes and arrays then execution of this mode results in generation of three files. They are Pca_array .coords.txt,Pca_gene.coords.txt,Pca_array.pc.txt,Pca_gene.pc.txt (Table 3,4 and 5).

EIGVALUE	Serum	Urine	Serum	Urine
EWEIGHT	0.500000	0.500000	0.333333	1.000000
11.355694	5.270829	4.683926	-8.829806	-1.124949
3.156275	-0.790575	2.028785	0.873295	-2.111504
2.390608	-1.642087	0.976064	-0.627213	1.293236
0.000000	-0.000000	-0.000000	-0.000000	-0.000000

Table 3: Enterococcus urine and serum Pca_array.coords

Gene	NAME	GWEIGHT	13.962933	3.154031	2.459592
cylR2	cylR2	0.078958	-1.745057	0.381449	-0.415156
cylR1	cylR1	0.078958	-1.745555	0.369697	-0.407576
cylLl	cylLl	0.078958	-1.743228	0.280673	-0.429080
cylLs	cylLs	0.078958	-1.744874	0.350378	-0.414513
cylM	cylM	0.071429	0.828411	-2.198683	0.064392
cylB	cylB	0.071429	-0.168616	-1.643191	-0.332201
cyA	cyA	0.103527	-1.659171	0.007236	-0.609136
cylI	cylI	0.111111	-0.759221	-0.204722	0.394967
inl like gene	inl like gene	0.154203	-1.080102	-0.538184	
esp	esp	0.286947	-1.542920	-0.003120	0.407186
ace	ace	0.111111	-1.668488	0.278542	0.037388
asa	asa	0.083333	-1.284902	0.633046	-0.175868
efa	efa	0.083333	-1.355745	0.641840	-0.203471
gls24	gls24	0.111121	-1.110620	0.369431	-0.288715
fsrB	fsrB	0.100000	3.024191	0.421498	1.670089
fsrC	fsrC	0.100000	1.224088	0.547696	0.821548
gelE	gelE	0.071429	12.531809	0.306415	-0.737952

Table 4: Enterococcus urine and serum Pca_array.pc

Gene	NAME	GWEIGHT	13.962933	3.154031	2.459592
cylR2	cylR2	0.078958	-1.745057	0.381449	-0.415156
cylR1	cylR1	0.078958	-1.745555	0.369697	-0.407576
cylLl	cylLl	0.078958	-1.743228	0.280673	-0.429080
cylLs	cylLs	0.078958	-1.744874	0.350378	-0.414513
cylM	cylM	0.071429	0.828411	-2.198683	0.064392
cylB	cylB	0.071429	-0.168616	-1.643191	-0.332201
cyA	cyA	0.103527	-1.659171	0.007236	-0.609136
cylI	cylI	0.111111	-0.759221	-0.204722	0.394967
inl like gene	inl like gene	0.154203	-1.080102	-0.538184	
esp	esp	0.286947	-1.542920	-0.003120	0.407186
ace	ace	0.111111	-1.668488	0.278542	0.037388
asa	asa	0.083333	-1.284902	0.633046	-0.175868
efa	efa	0.083333	-1.355745	0.641840	-0.203471
gls24	gls24	0.111121	-1.110620	0.369431	-0.288715
fsrB	fsrB	0.100000	3.024191	0.421498	1.670089
fsrC	fsrC	0.100000	1.224088	0.547696	0.821548
gelE	gelE	0.071429	12.531809	0.306415	-0.737952

Table 5: Enterococcus urine and serum Pca_gene cords

EIGVALUE	Serum	Urine	Serum	Urine
MEAN	-0.109700	-0.360033	-1.588277	-0.494270
13.962933	0.019519	-0.018821	-0.897833	-0.439501
3.154031	-0.150412	0.970179	0.064002	-0.178975
2.459592	-0.238143	0.149742	-0.428603	0.858583
0.815566	-0.959314	-0.189671	0.078095	-0.194018

Table6: Enterococcus urine and serum Pca_gene.pc.txt

Java Tree View:

Java Tree view software is used to show differentiation in computational and graphical environment for analysing data taken from output of cluster extensions. The **.cdt** file which is generated in Hierarchical parameter in cluster software is run in this java tree view. Once the file is run then we get results showing the differentiation in between genes and variations in genes

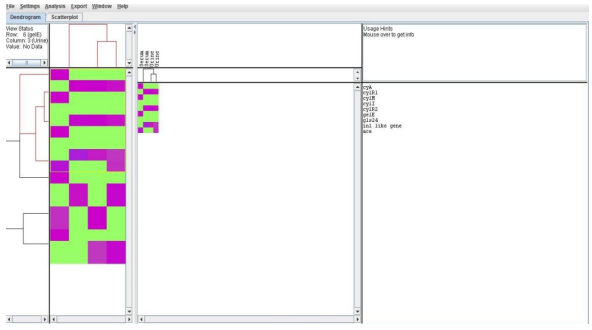
Dendogram:

It displays dendrogram in original TreeView format. Here it shows the node of origin between genes and also the deviation between genes in pictorial representation (Fig1).

Scatterplot:

It displays scatterplot of data values or per-gene statistics. Depending upon the number of genes, the scatterplot visibility can be seen. It displays all the points where genes are connected. In the graphical notation between x and y co-ordinates these genes are located (Fig 2).

Fig 1: Enterococcus Dendrogram



Volume : 5 | Issue : 12 | December-2016 • ISSN No 2277 - 8179 | IF : 3.508 | IC Value : 78.46

Fig 2: Scatterplot of Enterococcus genes

Conclusion:

In conclusion the data set of Enterococcus faecalis with differences in abundance of virulence gene mRNA from log-phase cells grown in serum or urine were analysed using the Cluster 3.0 which shows the values for Filter Data, Adjusting Data, Hierarchial, k-MEANs, SOMs and PCA. The output files generated in some of the parameters are ,in Hierarchial three output files with extensions .cdt, .gtr, .atr and in k-MEANS two files ".kgg" and ".xls" extensions , in SOM three files namely .Anf, .Gnf and .data type. In PCA three files "gene.coords", "gene.pc", "array.coords", "array.pc". whereas the Array.coords gives EIvalue& GWEIGHT, Gene.coords gives different genes and its GWEIGHT, Gene.pc gives EIvalue& GWEIGHT, Array.pc gives different genes and its MEAN respectively.

The Java Tree view is used to generate the dendrograms (or phylogenetic trees) for each of the dataset. The output shows the genes aligned as per their Hierarchial values and the scatterplot have been drawn using the same software which displays the data values or per-gene statistics.

References:

1. Brett D. Shepard and Michael S. Gilmore, "Differential Expression of Virulence-Related Genes in Enterococcus faecalis in Response to Biological Cues in Serum and Urine", Infect Immun. 2002 Aug; 70(8): 4344–4352.
2. C.A. Guzman, C. Pruzzo, M. Platè, M.C. Guardati, L. Calegari, "Serum dependent expression of Enterococcus faecalis adhesins involved in the colonization of heart cells", Microb. Pathogen. 11 (1991) 399–409.
3. Chow, J. W., L. A. Thal, M. B. Perri, J. A. Vazquez, S. M. Donabedian, D. B. Clewell, and M. J. Zervos. 1993. "Plasmid-associated hemolysin and aggregation substance production contribute to virulence in experimental enterococcal endocarditis. Antimicrob. Agents Chemother. 37:2474-2477.
4. Guzmàn, C. A., C. Pruzzo, G. LiPira, and L. Calegari. 1989. "Role of adherence in pathogenesis of Enterococcus faecalis urinary tract infection and endocarditis". Infect. Immun. 57:1834-1838.
5. Haas, W., B. D. Shepard, and M. S. Gilmore. 2002. "Two-component regulator of Enterococcus faecaliscytolysin responds to quorum-sensing autoinduction". Nature 415:84-87.
6. Jett, B. D., M. M. Huycke, and M. S. Gilmore. 1994. "Virulence of enterococci." Clin.Microbiol.Rev. 7:462-478.
7. Kreft, B., R. Marre, U. Schramm, and R. Wirth. 1992. "Aggregation substance ofEnterococcusfaecalis mediates adhesion to cultured renal tubular cells." Infect. Immun. 60:25-30.
8. Lowe, A. M., P. A. Lambert, and A. W. Smith. 1995. "Cloning of an Enterococcus faecalis endocarditis antigen: homology with adhesins from some oral streptococci." Infect. Immun. 63:703-706.
9. Rich, R. L., B. Kreikemeyer, R. T. Owens, S. LaBrenz, S. V. Narayana, G. M. Weinstock, B. E. Murray, and M. Hook. 1999. "Ace is a collagen-binding MSCRAMM from Enterococcus faecalis." J. Biol. Chem. 274:26939-26945.