Analytical Biochemistry 550 (2018) 8–14

Contents lists available at ScienceDirect

Analytical Biochemistry

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

Principal coordinate analysis assisted chromatographic analysis of bacterial cell wall collection: A robust classification approach

Keshav Kumar*, Felipe Cava**

Laboratory for Molecular Infection Medicine Sweden, Department of Molecular Biology, Umeå Centre for Microbial Research, Umeå University, 90187 Umeå, Sweden

ARTICLE INFO

Keywords:
Principal coordinate analysis
Classification
Peptidoglycans
Chromatography
Principal component analysis
Data heterogeneity

ABSTRACT

In the present work, Principal coordinate analysis (PCoA) is introduced to develop a robust model to classify the chromatographic data sets of peptidoglycan sample. PcoA captures the heterogeneity present in the data sets by using the dissimilarity matrix as input. Thus, in principle, it can even capture the subtle differences in the bacterial peptidoglycan composition and can provide a more robust and fast approach for classifying the bacterial collection and identifying the novel cell wall targets for further biological and clinical studies. The utility of the proposed approach is successfully demonstrated by analysing the two different kind of bacterial collections. The first set comprised of peptidoglycan sample belonging to different subclasses of Alphaproteobacteria. Whereas, the second set that is relatively more intricate for the chemometric analysis consist of different wild type *Vibrio Cholerae* and its mutants having subtle differences in their peptidoglycan composition. The present work clearly proposes a useful approach that can classify the chromatographic data sets of chromatographic peptidoglycan samples having subtle differences. Furthermore, present work clearly suggest that PCoA can be a method of choice in any data analysis workflow.

Introduction

Chemometrics is a chemical discipline that essentially uses mathematics, statistics and logics to analyse the large volume of chemical and biochemical data sets generated by single and hyphenated analytical instruments [1–4]. There are several chemometric techniques and each one of them work with different algorithm and involves optimisation of different parameters. Principal coordinate analysis (PcoA) is a chemometric technique that has not been explored so much in the fields of analytical and bio-analytical chemistry [5–9]. PcoA is a type of multidimensional scaling that essentially simplifies the data sets by projecting it in a space spanned by orthogonal axes (i.e. dimensions) that are few in numbers. The use of orthogonal space reduces the amount of information lost while reducing the dimension of the data sets. The projection in orthogonal space allows the visualisation of the multivariate data sets and therefore simplifies the data interpretation process [5–9]. Conceptually, PCoA is similar to principal components analysis (PCA) [2–4,10] however, there is also a substantial difference in the sense that PCoA analyses a matrix of pairwise distances between subjects, whereas PCA analyses a matrix containing the covariance or correlations among the variables. The development of PCoA is mostly attributed to Young and Householder [5], Gower [6], Krzanowski and Marriott [7]. As discussed above, PCoA is not a widely known or used technique; nevertheless, few examples of its application could be seen in the fields related to genetics [11], anthropology [12] and molecular ecology [13].

Peptidoglycan is a polymer consisting of disaccharide unit N-acetyl-glucosamine-β(1 → 4)-N-acetyl-muramic acid that are cross-linked with short peptide chains. It provides structural strength to the cell wall and helps in maintaining cell's morphology, environmental stress and positive turgor pressure. The peptidoglycans are characteristic to the bacteria and its structure differs from one bacterium to other [14–19]. Peptidoglycans compositional study can be helpful in understanding the mechanism of host-pathogen interaction. It can also be helpful in designing the effective antibiotics. It can also be useful towards the identification of certain biomarkers that are specific to a particular bacterium or a group of bacteria [15,16].

The developments in the modern chromatographic instrumentation techniques have attracted the attention of molecular biologist and they have successfully used Ultra-performance liquid chromatography (UPLC)-coupled to UV-visible detector for analysing the peptidoglycan composition of the bacterial cell wall [15,16,20]. However, the manual analyses of large volume of chromatographic are cumbersome and challenging task. The difficulty can be overcome to great extent using a

* Corresponding author.
** Corresponding author.
E-mail addresses: keshavkumar29@gmail.com (K. Kumar), felipe.cava@molbiol.umu.se (F. Cava).

https://doi.org/10.1016/j.ab.2018.04.008
Received 30 January 2018; Received in revised form 28 March 2018; Accepted 8 April 2018
Available online 10 April 2018
0003-2697/ © 2018 Elsevier Inc. All rights reserved.

proper data analysis workflow that is based on suitable chemometric technique.

The present work is aimed towards integrating PCoA on chromatographic data sets to achieve the unsupervised classification of the peptidoglycan samples of certain bacterial collection. To the best of our knowledge, there are no reports in the literature that describes the application of PCoA in biochemistry and more specifically in bacteriology fields. To carry out the present work, two different types of data sets are selected from the literature. The first set consists of peptidoglycan sample belonging to different subclasses of Alphaproteobacteria class [15]. The second set consists of different mutants of wild type *Vibrio Cholerae* where the mutant has subtle differences in their peptidoglycan composition [21,22]. The selected data sets that are quite distinct in nature can really test the discrimination efficiency of PcoA.

Materials and methods

Chromatographic data sets of peptidoglycan samples belonging to Alphaproteobacteria class

Ultra-performance liquid chromatographic (UPLC) data for a set of 18 peptidoglycan samples of Alphaproteobacterium members is taken from the recent work of Espaillat and co-workers [15]. The selected Alphaproteobacterium members are listed in Table 1. The UPLC data set of 12 peptidoglycan samples belonging to *Vibrio Cholerae* mutants are taken form the work reported by Möll and co-workers [21,22]. The selected *Vibrio Cholerae* mutants are listed in Table 1. The UPLC data was collected using the Waters instrument equipped with UV-detector. A detailed discussion on the peptidoglycan isolation and subsequent chromatographic analysis can also be found in the above mentioned work [15,21,22].

Software used

All the computational work is carried out on MATLAB platform.

Results and discussion

The pre-processed chromatograms of all the 18 Alphaprotebacterium members and 12 *Vibrio Cholerae* mutants are shown in Fig. 1(a) and (b) respectively. The pre-processing steps are (i) removal of irrelevant segments, (ii) baseline correction, (iii) peak alignment to correct the retention time drift. The irrelevant segments are (i) solvent fronts appearing in the beginning and (ii) column wash appearing towards the end of the chromatogram. As it can be seen that these segments do not contain any bacterial cell wall related information and therefore, must be trimmed out else, it can bias the outcomes of the subsequent chemometric analysis. Baseline correction is achieved using the 'msbackadj' routine of the MATLAB and alignment is achieved using the COW algorithm [16,23]. The pre-processed data set of Alphaprotebacterium is stored in a matrix X of dimension 18 × 1082 (sample (I) × variable (K)). The pre-processed data set of *Vibrio Cholerae* mutants is stored in a matrix Y of dimension 12 × 1082 (I× K).

Table 1
The bacterial samples selected in the Alphaproteobacterium subclasses and Vibrio-Cholerae mutants.

Sample Number	Alphaproteobacteria sub-classes	Vibrio Cholerae mutants
1	G. oxy *(Gluconobacter oxydans)*	ΔamiB
2	A. ace *(Acetobacter aceti)*	ΔnlpD
3	G. fra *(Gluconobacter frateurii)*	ΔenvC
4	A. tet *(Angulomicrobium tetraedale)*	ΔenvC ΔnlpD
5	T. luc *(Thalassospira lucentensis)*	wt
6	H. bal *(Hirschia baltica)*	ΔdacA2
7	A. bip *(Asticcacaulis biprosthecum)*	ΔdacA1
8	C. cre *(Caulobacter crescentus)*	ΔdacB
9	R. den *(Roseobacter denitrificans)*	ΔpbpG
10	A. xyl *(Acetobacter xylinium)*	ΔdacA2 ΔpbpG
11	A. pas *(Acetobacter pasteurianus)*	ΔdacB ΔpbpG
12	A. med *(Acetobacter methanolica)*	ΔdacA2 ΔdacB ΔpbpG
13	E. lit *(Erythrobacter litoralis)*	
14	E. aqu *(Erythrobacter aquamaris)*	
15	A. tro *(Acetobacter tropicalis)*	
16	M. med *(Mesorhizobium mediterraneum)*	
17	M. sep *(Mesorhizobium septentrionale)*	
18	M. hau *(Mesorhizobium haukii)*	

Before proceeding further, it is essential that conceptual difference between PCA and PCoA and the advantage that one can achieve using the former be clearly explained. Some of the striking difference between PCA and PCoA are the followings. PCoA finds the similarity (or dissimilarities) differences between the samples, whereas PCA searches the pattern between the variables. PCoA performs the eigenvalue decomposition on the dissimilarity matrix whereas PCA involves eigenvalue decomposition on the covariance matrix. As discussed above PCA works on the co-variability of the variables and finds the factors that explains the major source of variation in the data set. Thus, one can easily argue and anticipate that PCA is best suited for analysing the samples belonging to a specific group. Whereas, PCoA works on the dissimilarity matrix therefore one can expect that it can be a method of choice for finding the relationship (similarity or dissimilarity) among the samples. PCoA can provide the relationship among the samples whether it belongs to a particular group or not. PCoA finds the set a number of axes such that samples with similar characteristics or properties are close to each other.

Principal coordinate analysis (PCoA) algorithm

The input for the PCoA is obtained using the approach suggested by Legendre and Legendre [9]. The suggested approach involves the calculation of dissimilarity matrix (D) of dimension I×I from the original data matrix W of dimension I×K (Sample ×Variable). The dissimilarity matrix can be obtained using the Euclidean distance metric formula given in equation (1)

$$d_{ij} = \sqrt{\sum_{k=1}^{I} (W_{ik} - W_{jk})^2} \tag{1}$$

where d_{ij} is the element of matrix D describing the dissimilarity between the *ith* and *jth* sample.

Each element of matrix D is transformed using the equation (2)

$$e_{ij} = -\frac{1}{2} d_{ij}^2 \tag{2}$$

Using the elements e_{ij}, a new matrix E of dimension I×I is obtained which is further processed using equation (3)

$$f_{ij} = e_{ij} - \overline{e_i} - \overline{e_j} - \overline{E} \tag{3}$$

$\bar{e}_i$ are the row wise mean of the matrix E, $\bar{e}_j$ is the column wise mean of the matrix E and $\bar{E}$ is the overall mean of E. The elements f_{ij} are stored in matrix F of dimension I×I that is subsequently used in further steps of PCoA. It has been found that transformation of matrix D to F ensures that distance relationships are preserved among the samples [24]. The matrix F cab further be subjected to the eigenvalue eigenvector decomposition, shown in equation (4)

$$F\Psi = \lambda\Psi \tag{4}$$

In the above equation λ is I×I matrix that contains the Eigen values along the diagonal. The matrix Ψ is also of dimension I×I and it contains the Eigenvectors. The amount of information or variance captured by *ith* Eigenvector (*PCoAi*) can be calculated using the equation (5).

Download English Version:

https://daneshyari.com/en/article/7556814

Download Persian Version:

https://daneshyari.com/article/7556814

Daneshyari.com