Classification of Lactic Acid Bacteria against Cytokine Immune Modulation

Fecha de recepción y aceptación: 22 de noviembre de 2013, 20 de diciembre de 2013

Francisco Torrens Zaragozá

Institut Universitari de Ciència Molecular, Universitat de València,
Correspondencia: Edifici d’Instituts de Paterna, P. O. Box 22085. E-46071. Valencia. España. E-mail: francisco.torrens@uv.es

ABSTRACT

As some functions of lactic acid bacteria (LABs) reside in cytoplasm or cell-wall components, more effects were shown by fresh than by pasteurized yoghurt. Classification algorithms are proposed based on information entropy. They use effects of living and pasteurized LABs on cytokines. Excessive number of results appear compatible with data suffering combinatorial explosion; however, after the equipartition conjecture one gets a criterion: the best classification is that in which entropy production is uniformly distributed. The classification agrees with the principal component analysis.

KEYWORDS: Information entropy, Equipartition conjecture, Principal component analysis, Immune modulation, Lactic acid bacterium, Probiotic.

RESUMEN

Ya que algunas funciones de bacterias ácido-lácticas (BALs) residen en el citoplasma o componentes de la pared celular, más efectos mostró el yogur fresco que pasteurizado. Se propone algoritmos de clasificación basados en entropía informacional. Usan efectos de BALs vivas y pasteurizadas en citoquinas. Aparece excesivo número de resultados compatibles con los datos sufriendo explosión combinatoria; sin embargo, después de la conjetura de equipartición uno obtiene un criterio: la mejor clasificación es aquella en la cual la producción de entropía está uniformemente distribuida. La clasificación concuerda con el análisis en componentes principales.

PALABRAS CLAVE: Entropía informacional, Conjetura de equipartición, Análisis de componentes principales, Inmunomodulación, Bacteria ácido-láctica, Probiótico.

INTRODUCTION

Decays in colony-forming units (CFUs) is due to bacteria clumping/cell death (cf. Fig. 1).
Differences exist in replication cycles of lytic/lysogenic bacteriophages (cf. Fig. 2) [1].

Fig. 1. Two mechanisms by which antibacterial agents may reduce CFU numbers of bacteria in time-kill and minimum bactericidal concentration assays.

Fig. 2. Lytic/lysogenic-bacteriophages replication cycles. (A) Lytic bacteriophages: (1) attachment; (2) injection of bacteriophage deoxyribonucleic acid (DNA) into bacterial host; (3) shutoff of host-components synthesis, bacteriophage-DNA replication and new-capsids production; (4) bacteriophages assembly; (5) mature-bacteriophages release (lysis). (B) Lysogenic bacteriophages: (1) and (2) are similar to lytic bacteriophages; (3) lysogenic bacteriophages can, among other possibilities, initiate a reproductive cycle similar to lytic bacteriophages (a) or integrate their DNA into host bacterium’s chromosome (lysogenization) (b). Lysogenized cells can replicate normally for many generations (1b) or at some point undergo lysogenic induction (2b) spontaneously or due to inducing agents (e.g., radiation, carcinogens) during which time integrated bacteriophage DNA is excised from bacterial chromosome and may pick up bacterial DNA fragments.
Bacteriophages (cf. Fig. 3) kill bacterium to which they infect being antimicrobial agents [2].

![Bacteriophage replicative cycle](image)

**Fig. 3.** Bacteriophage replicative cycle. (1) It finds sensitive bacterium. (2) It joins bacterium surface. (3) It introduces its genetic material inside sensitive bacterium. (4) Genetic material is replicated inside bacterium generating lots of copies. (5) Protein coats are synthesized, inside which bacteriophage genetic material is introduced to form mature bacteriophage particles. (6) New bacteriophages destroy bacterial coat and are released on cell outside to iterate.

Lysis in bacteriophages is performed via two proteins: holin and endolysin (cf. Fig. 4).

![Action-mode scheme of bacteriophage endolysins](image)

**Fig. 4.** Action-mode scheme of bacteriophage endolysins. Surface structure of Gram’ bacterium is indicated formed by cytoplasmic membrane and peptidoglycan layer (cell wall).

An LAB function in gut microbiota is to improve absorption of nutrients, especially carbohydrates; however, intestine colonization by *Bacteroides thetaiotaomicron* in mice is accompanied by transcriptional changes of a wide number of genes [3]. Gutiérrez-San José et al. reported immune modulation by fresh/pasteurized yoghurt [4]. Surface-layer proteins of LABs are important in
probiotic activity [5,6]. Role of Tyr1 in plantaricin149a to disrupt model membranes was informed [7]. *Lactobacillus crispatus*—bacterial vaginosis interaction was analyzed [8,9]. Cytokines [e.g., tumour necrosis factor (TNF), growth hormone, interleukin (IL)-6] induce insulin resistance. In earlier publications, periodic tables of human immunodeficiency virus (HIV) inhibitors [10]/thiocarbamates protecting vs. HIV [11] were described. The object of the present report is to classify living/pasteurized LABs vs. cytokines; in addition, to compare results with other methods.

**MATERIALS AND METHODS**

The first step in quantifying the similarity concept for LABs is to list their most important cytokine immune modulations. The vector of properties \( \mathbf{i} = \langle i_1, i_2, \ldots, i_k, \ldots \rangle \) is associated with every LAB \( i \), whose components correspond to different characteristics in a hierarchical order according to the expected importance of cytokines production. Components \( i_k \) are 1/0 consistent with whether a similar immune modulation of rank \( k \) is present/absent in LAB \( i \). Analysis includes three cytokine types in LABs: TNF\( \alpha \), IL1\( \beta \), and interferon (IFN)\( \gamma \). Index \( i_1 = 1 \) denotes TNF\( \alpha \)/IL6 (cf. Table 1), \( i_2 = 1 \), IL2 and \( i_3 = 1 \), IL1\( \beta \), IFN\( \gamma \) immune modulations.

<table>
<thead>
<tr>
<th>LAB</th>
<th>Vector property</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Lactobacillus bulgaricus</em></td>
<td>&lt;110&gt;</td>
</tr>
<tr>
<td><em>L. acidophilus</em></td>
<td>&lt;100&gt;</td>
</tr>
<tr>
<td><em>L. casei</em></td>
<td>&lt;101&gt;</td>
</tr>
<tr>
<td><em>L. rhamnosus</em></td>
<td>&lt;110&gt;</td>
</tr>
<tr>
<td><em>L. sakei</em></td>
<td>&lt;101&gt;</td>
</tr>
<tr>
<td><em>Bifidobacterium bifidum</em></td>
<td>&lt;110&gt;</td>
</tr>
<tr>
<td><em>Streptococcus thermophilus</em></td>
<td>&lt;110&gt;</td>
</tr>
<tr>
<td><em>L. johnsonii</em></td>
<td>&lt;000&gt;</td>
</tr>
</tbody>
</table>

The similarity matrix \( \mathbf{R} = [r_{ij}] \), between two LABs \( i = \langle i_1, i_2, \ldots, i_k, \ldots \rangle \) and \( j = \langle j_1, j_2, \ldots, j_k, \ldots \rangle \), is defined by the correlation coefficient between them modified by three weights: 0.5, 0.25 and 0.125. Learning procedures similar to stochastic methods are implemented [12]. Consider a given partition into classes as good from practical observations that corresponds to a reference similarity matrix \( \mathbf{S} = [s_{ij}] \), obtained for an arbitrary number of fictitious properties. Bear in mind the same set of species as in the good classification and actual properties. The similarity degree \( r_{ij} \) is computed from correlation matrix \( \mathbf{R} \). The number of properties for \( \mathbf{R} \) and \( \mathbf{S} \) differs. The learning procedure consists in finding classification results for \( \mathbf{R} \) as close as possible to the good one. The distance between partitions in classes characterized by \( \mathbf{R} \) and \( \mathbf{S} \) results:

\[
D = -\sum_j (1 - r_{ij}) \ln \frac{1 - r_{ij}}{1 - s_{ij}} - \sum_j r_{ij} \ln \frac{r_{ij}}{s_{ij}} \quad \forall 0 \leq r_{ij}, s_{ij} \leq 1
\]

which definition was suggested by Kullback to measure the distance between two probability distributions [13] and applied in the synthesis of complex dendrograms *via* information entropy [14,15]. Code MolClas was written for molecular classification based on equipartition conjecture of entropy production. It punches similarity and difference matrices, and the latter in format NEXUS (.NEX) for codes PAUP, MacClade and SplitsTree. It performs single and complete-linkage hierarchical cluster analyses (CAs) of compounds *via* IMSL subroutine CLINK [16]. Code GraphCor was written for partial correlation diagrams (PCDs). Codes MolClas and GraphCor are available from the author on the Internet (francisco.torrens@uv.es) and are free for academics.

**CALCULATION RESULTS AND DISCUSSION**

Immune modulation data from Gutiérrez-San José *et al.* were used for classification. They analyzed the effects (induction, inhibition and no effect) of eight living/pasteurized LABs on the production of nine cytokines: TNF\( \alpha \), IL-1\( \beta \), IL2/4–6, 10/12 and IFN\( \gamma \). Intercorrelations are shown in PCD that contains high \((r \geq 0.75)\), medium \((0.50 \leq r < 0.75)\), low \((0.25 \leq r < 0.50)\) and zero \((r < 0.25)\) partial autocorrelations. Pairs of LABs with higher partial correlations show similar vector property. The method avoids
the problem of others of continuum variables because Entry 8 (L. johnsonii) with constant vector <000> shows null standard deviation, causing greatest Pearson partial correlations $r = 1$ with any component resulting an artifact. Correlations are illustrated in PCD that contains nine high (cf. Fig. 5, red lines), 12 medium (orange), three low (yellow) and four zero (black) partial correlations. All seven high partial correlations of Entry 8 are corrected: its correlations with Entries 2, 3 and 5 are low, and its correlations with Entries 1, 4, 6 and 7 are zero partial correlations. However, the partial correlations of Entry 8 are, at most, low and it can be an outlier.

![Fig. 5. PCD of LABs: high (red), medium (orange) and low (yellow) partial correlations.](image)

The grouping rule in the case with equal weights $a_k = 0.5$ for $0.76 \leq b \leq 0.87$ allows classes:

$$C_b = (1, 4, 6, 7)(2)(3, 5)(8)$$

Four groupings are obtained with associated entropy $b-R_b = 8.54$ matching to $<i_1, i_2, i_3>$ and $C_b$ (cf. Fig. 6) [17,18]; binary taxonomy (dendrogram) of Table 1 separates classes 4, 1, 2 and 3 with 1, 4, 1 and 2 LABs, respectively [19]. It shows LABs different behaviour depending on genus; however, $C_b$ results should be taken with care because classes (2) and (8) with only one LAB could be outliers. The results are in qualitative agreement with PCD (Fig. 5).
The illustration of the classification above in a radial tree (cf. Fig. 7) shows LABs different behaviour depending on genus. The same classes above are clearly recognized in qualitative agreement with PCD and dendrogram (Figs. 5 and 6).
Program SplitsTree allows examining CA data [20]. Based on split decomposition, it takes as input a distance matrix and produces as output a graph that represents relations between taxa. For ideal data the graph is a tree, whereas less ideal data give rise to a tree-like net that can be interpreted as possible evidence for conflicting data. Furthermore, as split decomposition does not force data onto a tree, it provides good indication of how tree-like are given data. Splits graph for the eight LABs (cf. Fig. 8) shows that *L. bulgaricus*, *L. rhamnosus*, *Bifidobacterium bifidum* and *Streptococcus thermophilus* collapse, and *L. casei* with *L. sakei*. It reveals no conflicting relation between LABs. It illustrates LAB different behaviour depending on genus. The same classes above are clearly distinguished in qualitative agreement with PCD and binary/radial trees (Figs. 5–7).

![Radial tree for the effects of lactic acid bacteria on the production of nine cytokines.](image)
In quantitative structure–property relationships (QSPRs), data file contains <100 objects and >1000 variables. So many variables exist that nobody discovers by inspection patterns, trends, clusters, etc. in objects. Principal components (PCs) analysis (PCA) is a technique useful to summarize information contained in Xmatrix and present it in an understandable form [21–26]. The PCA works decomposing Xmatrix as product of matrices P and T. Loading matrix (P) with information about variables contains few vectors, PCs that are obtained as linear combinations of original variables. In score matrix (T) with information about objects, every object is described in terms of projections onto PCs instead of original variables: $X = TP' + E$, where $'$ denotes transpose matrix. Information not contained in matrices remains as unexplained variance in residual matrix (E). Every PC, is a new coordinate expressed as linear combination of the old $x_j$: $PC_i = \sum b_{ij} x_j$. New coordinates $PC_i$ are called scores or factors while coefficients $b_{ij}$ are called loadings. Scores are sorted according to information content with regard to variance among objects. The PCs present two properties: (1) they are extracted by decaying importance; (2) every PC is orthogonal to one another. A PCA was performed for cytokines immune modulation of LABs. First factor $F_1$ explains 54% variance (46% error), $F_{1/2}$ 93% variance (7% error) and $F_{1/3}$ 100% variance (0% error). Instead of eight LABs in space $\mathbb{R}^9$ of nine cytokines consider nine cytokines in space $\mathbb{R}^8$ of eight LABs. The PCD (cf. Fig. 9) contains one low and two zero partial autocorrelations. The pair 1–2 of cytokines (TNFα–IL2) with low partial correlation undergoes similar immune modulation. However, cytokine 3 (ILβ) presents zero partial correlations and can be an outlier.

Fig. 8. Splits graph for the effects of lactic acid bacteria on the production of nine cytokines.
A PCA/CA for cytokines immune modulation corresponding to LABs (cf. Fig. 10) separates IL1β from TNFα and IL-2. Factor $F_1$ explains 53% variance (47% error), $F_{1/2}$ 86% variance (14% error), $F_{1/3}$ 97% variance (3% error), etc. Results are in agreement with PCD (Fig. 9).

Fig. 9. PCD of cytokine immune modulation for LABs: low (yellow) partial correlation.

Fig. 10. Cluster analysis for the cytokines immune modulation corresponding to lactic acid bacteria.
CONCLUSIONS

From the present results and discussion the following conclusions can be drawn.

1. Several criteria to reduce analysis to a manageable quantity of lactic acid bacteria referred to cytokine immune modulation that are ranked: tumour necrosis factorα > interleukin2 > interleukin1β. Many classification algorithms are based on information entropy. For sets of moderate size, excessive number of results appear compatible with the data and suffer combinatorial explosion; however, after the equipartition conjecture one gets a selection criterion, according to which the best configuration is that in which entropy production is most uniformly distributed. The method avoids the problem of others of continuum variables because for L. johnsonii <000>, null standard deviation causes a Pearson correlation coefficient of one. Program MolClas is a simple, reliable, efficient and fast procedure for categorization of lactic acid bacteria.

2. The obtained classification was in agreement with splits graph and principal component analysis. Furthermore, splits graph showed no conflicting relationship between lactic acid bacteria.

3. The classification of the lactic acid bacteria vs. cytokine immune modulation depends on genus; e.g., a trend exists in the lactic acid bacteria from genus Lactobacillus species to be close in the dendrogram. However, genera Lactobacillus, Bifidobacterium and Streptococcus are not separated in different clusters, which is in agreement with the fact that in bacteria the concepts of genus and species present more intravariation than in eukaryotes. The categorization of cytokine immune modulation vs. lactic acid bacteria depends on cytokines production; notwithstanding, tumour necrosis factorα, interleukins and interferonγ are not split into different groupings.

4. Despite the good results obtained by Corell group and many others, their reports are published in the journal Alimentación, Nutrición y Salud of the Danone Institute. Therefore, independent publications are desired.

ACKNOWLEDGEMENT

The author wants to dedicate this manuscript to Dr. Enrique Pérez-Payá, who was greatly interested in this research and would have loved to see its conclusion.

LITERATURE CITED

[16] IMSL. 1989. Integrated Mathematical Statistical Library (IMSL); IMSL: Houston, TX.
Classification of Lactic Acid Bacteria against Cytokine Immune Modulation


