Electronic Supporting Information

Infrared spectroscopy coupled with a dispersion model for quantifying the real-time dynamics of kanamycin resistance in artificial microbiota

Naifu Jin¹, Maria Paraskevaidi², Kirk T Semple¹, Francis L. Martin²*, Dayi Zhang¹*

¹ Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK
² School of Pharmacy and Biomedical Sciences, University of Central Lancashire, Preston PR1 2HE, UK

No. of Pages = 7
No. of Figures = 3
No. of Tables = 1
1. Materials and Methods

1.1 Dispersion indicator model

The initial spectral dataset is an ensemble of multivariate observations partitioned into \( M \) distinct groups (different microbiota composition in this study). For the \( n_m \) observation in each group (\( m \) runs from 1 to \( M \) and refers to the \( m^{th} \) group). The multivariate observation vectors can be written as \( n_{mi} \) where \( i \) is the \( i^{th} \) observation. To search for the linear combination in LDA that optimally separates our multivariate observation into \( M \) groups \(^1\), the linear transformation of \( y_{mi} \) is written as \( z_{mi} \):

\[
z_{mi} = w^T y_{mi}
\]

(1)

Here, \( w^T \) represents the linear transformation matrix, and the mean of the \( m^{th} \) group of the transformed data (\( z_m \)) is:

\[
\langle z_m \rangle = w^T \langle y_m \rangle
\]

(2)

where \( y_m \) is the mean of the observations within a group and defined as:

\[
\langle y_m \rangle = \frac{\sum_{j=1}^{n_m} y_{mj}}{n_m}
\]

(3)

The dispersion among groups (\( B \)) and within groups (\( E \)) are defined in the following equations:

\[
B_y = \sum_{m=1}^{g} n_m (\langle y_{mi} \rangle - \langle y \rangle)(\langle y_{mi} \rangle - \langle y \rangle)^T
\]

(4)

\[
E_y = \sum_{m=1}^{g} n_g \sum_{j=1}^{n_m} (\langle y_{mi} \rangle - \langle y_m \rangle)(\langle y_{mi} \rangle - \langle y_m \rangle)^T
\]

(5)

where \( \langle y \rangle = \frac{1}{M} \sum_{m=1}^{g} \frac{1}{n_m} \sum_{j=1}^{n_m} y_{mj} \) is the total average of the dataset. Using Fisher's linear discriminant, the optimal linear regression in PCA-LDA is to find the vector \( w \) maximizing \( \lambda \) (the rate of between-groups sum of squares to within-groups sum of squares):

\[
\lambda = \frac{w^T B_y w}{w^T E_y w}
\]

(6)

The solutions of Equation (6) are the eigenvalues \( |\lambda| \), which are associated to the eigenvectors \( |w| \). In the most cases, the first two ranked \( \lambda_1 \) and \( \lambda_2 \) account for the most of \( |\lambda| \), and the discriminant functions are obtained as LD1 (\( z_1 = w_1^T Y \)) and LD2 (\( z_2 = w_2^T Y \)) to represent the spectra variables of each community.
To predict the composition of the artificial microbiota, the three control groups (A. baylyi [a], E. coli [b] and M. vanbaalenii [c]) are set as the reference classes. The dispersions of the among groups (B) and within groups (E) (Fig. 2B) are defined in the following equations:

\[ O_{y,q} (q = a, b, c) = w^T B'_{y,q} w = \]

\[ w^T \left\{ \sum_{i=1}^{M} \sum_{j=1}^{M} n_m \left( (y_{mi}) - \langle y_{qj} \rangle \right) \left( (y_{mi}) - \langle y_{qj} \rangle \right)^T \right\} w = \sum_{i=1}^{M} \sum_{j=1}^{M} n_g \left( (w^T y_{mi}) - \langle w^T y_{qj} \rangle \right) \left( (w^T y_{mi}) - \langle w^T y_{qj} \rangle \right)^T \]

(7)

\[ T_y = w^T E'_{y} w = \sum_{q=a,b,c} \sum_{i=1}^{M} \sum_{j=1}^{M} n_g \left( (z_{mi}) - \langle z_{qj} \rangle \right) \left( (z_{mi}) - \langle z_{qj} \rangle \right)^T \]

(8)

Here, we introduced the dispersion indicator \( D_t \) to calculate the composition of antibiotic resistance bacteria (A. baylyi) within the community, defined as:

\[ D_t = \frac{o_{y,a}}{T_y} \]

(9)

\[ \sum_{q=a,b,c} D_{t,q} = \frac{o_{y,q}}{T_y} = 100\% \]

(10)

Reference:

(1) Ami, D.; Mereghetti, P.; Doglia, S. M. Multivariate analysis for Fourier transform infrared spectra of complex biological systems and processes; INTECH Open Access Publisher, 2013.
<table>
<thead>
<tr>
<th>Microbiota</th>
<th>Significant peaks (cm$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A. baylyi</em></td>
<td>1188, 1242, 1508, 1547, 1659, 1744</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>980, 1034, 1501, 1562, 1616, 1740</td>
</tr>
<tr>
<td><em>M. vanbaalenii</em></td>
<td>1065, 1134, 1192, 1377, 1582, 1744</td>
</tr>
<tr>
<td>M1</td>
<td>1223, 1377, 1578, 1612, 1694, 1740</td>
</tr>
<tr>
<td>M2</td>
<td>1138, 1188, 1304, 1632, 1678, 1740</td>
</tr>
<tr>
<td>M3</td>
<td>1501, 1543, 1612, 1651, 1694, 1728</td>
</tr>
<tr>
<td>M4</td>
<td>980, 1188, 1501, 1616, 1694, 1740</td>
</tr>
<tr>
<td>M5</td>
<td>1138, 1188, 1447, 1501, 1697, 1740</td>
</tr>
</tbody>
</table>
Figure S1. Growth curve of Mycobacterium vanbaalenii PYR-1, Escherichia coli DH5α and Acinetobacter baylyi ADPWH_recA in mineral medium without kanamycin pressure (A) or with 10 mg/L kanamycin (B).
Figure S2. Relative abundance of kanamycin resistance gene (kanR/16S) in *Mycobacterium vanbaalenii* PYR-1, *Escherichia coli* DH5α and *Acinetobacter baylyi* ADPWH_recA after 16-h cultivation without kanamycin pressure or with 10 mg/L kanamycin. Data are presented in mean ± standard error.
Figure S3. ATR-FTIR spectral dynamics of artificial microbiota.