Electronic Supporting Information

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

Naifu Jin[1], Maria Paraskevaidi[2], Kirk T Semple[1], Francis L. Martin[2,*], Dayi Zhang[1,*]

[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}:

$$z_{mi} = w^T y_{mi} \tag{1}$$

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

$$\langle z_m \rangle = w^T \langle y_m \rangle \tag{2}$$

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

$$\langle y_m \rangle = \sum_{j=1}^{n_m} y_{mj} / n_m \tag{3}$$

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

$$B_y = \sum_{m=1}^{G} n_m (\langle y_{mi} \rangle - \langle y \rangle)(\langle y_{mi} \rangle - \langle y \rangle)^T \tag{4}$$

$$E_y = \sum_{m=1}^{G} n_g \sum_{j=1}^{n_m} (\langle y_{mi} \rangle - \langle y_m \rangle)\,(\langle y_{mi} \rangle - \langle y_m \rangle)^T \tag{5}$$

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 λ (the rate of between-groups sum of squares to within-groups sum of squares):

$$\lambda = \frac{w^T B_y w}{w^T E_y w} \tag{6}$$

The solutions of Equation (6) are the eigenvalues $|\lambda|$, which are associated to the eigenvectors $|w|$. In the most cases, the first two ranked λ_1 and λ_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:

$$O_{y,q}|(q = a, b, c) = w^T B'_{y,q} w = w^T \left\{ \sum_{i=1}^{M} \sum_{j=1}^{M} n_m \left(\langle y_{mi} \rangle - \langle y_{qj} \rangle \right) \left(\langle y_{mi} \rangle - \langle y_{qj} \rangle \right)^T \right\} w = \sum_{i=1}^{M} \sum_{j=1}^{M} n_g \left(\langle w^T y_{mi} \rangle - \langle w^T y_{qj} \rangle \right) \left(\langle w^T y_{mi} \rangle - \langle w^T y_{qj} \rangle \right)^T = \sum_{i=1}^{M} \sum_{j=1}^{M} n_g \left(\langle z_{mi} \rangle - \langle z_{qi} \rangle \right) \left(\langle z_{mi} \rangle - \langle z_{qi} \rangle \right)^T \quad (7)$$

$$T_y = w^T E'_y w = \sum_{q=a,b,c} \sum_{i=1}^{M} \sum_{j=1}^{M} n_g (\langle z_{mi} \rangle - \langle z_{qj} \rangle) \left(\langle z_{gi} \rangle - \langle z_{qj} \rangle \right)^T \quad (8)$$

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

$$D_I = \frac{O_{y,a}}{T_y} \quad (9)$$

$$\sum_{q=a,b,c} D_{I,q} = \frac{O_{y,q}}{T_y} = 100\% \quad (10)$$

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 S1. Significant peaks derived from cluster vectors of artificial microbiota.

Microbiota	Significant peaks (cm^{-1})
A. baylyi	1188, 1242, 1508, 1547, 1659, 1744
E. coli	980, 1034, 1501, 1562, 1616, 1740
M. vanbaalenii	1065, 1134, 1192, 1377, 1582, 1744
M1	1223, 1377, 1578, 1612, 1694, 1740
M2	1138, 1188, 1304, 1632, 1678, 1740
M3	1501, 1543, 1612, 1651, 1694, 1728
M4	980, 1188, 1501, 1616, 1694, 1740
M5	1138, 1188, 1447, 1501, 1697, 1740

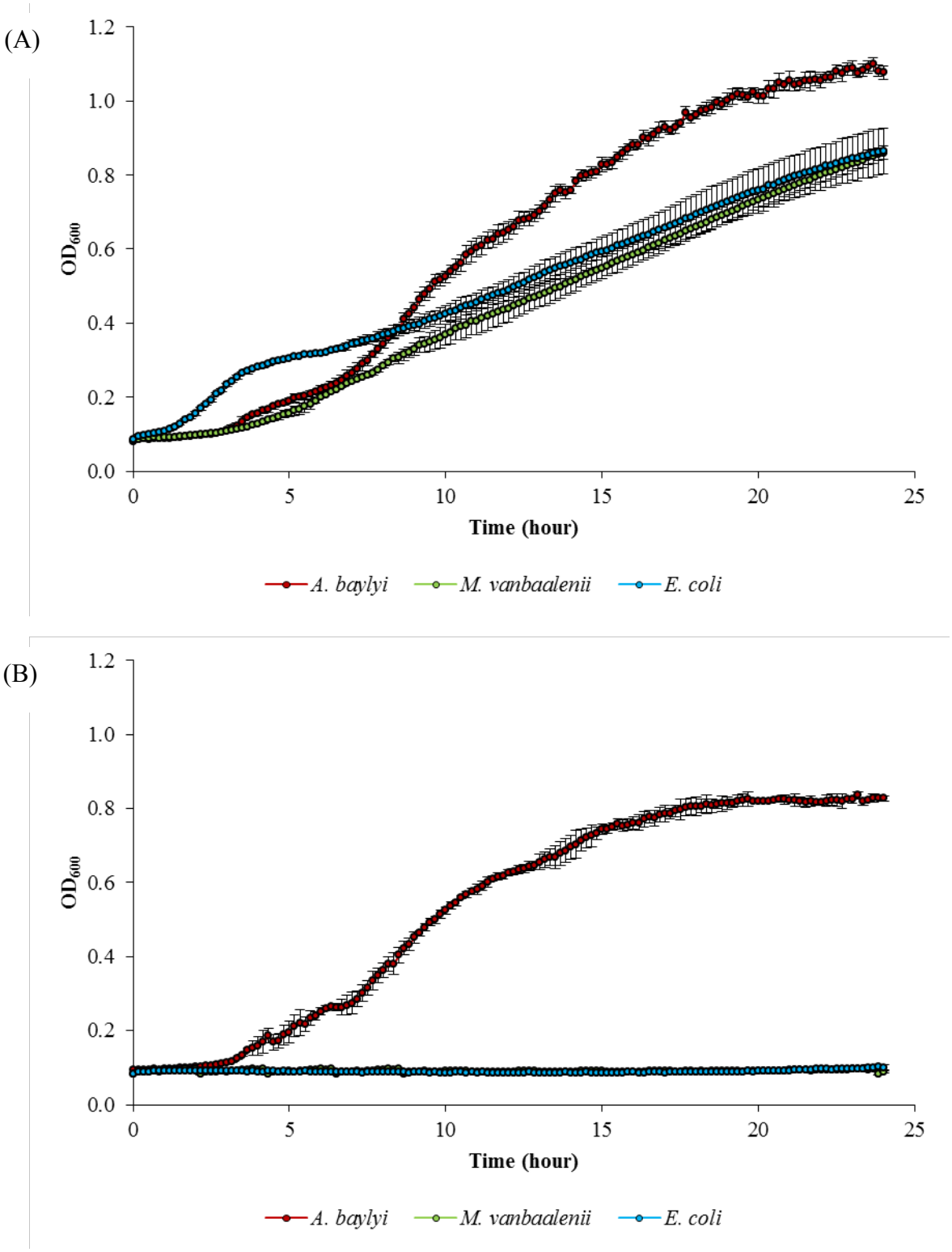

Figure S1. Growth curve of *Mycobacterium vanbaalenii* PYR-1, *Escherichia coli* DH5α and *Acinetobacter baylyi* ADPWH_recA in mineral medium without kanamycin pressure (A) or with 10 mg/L kanamycin (B).

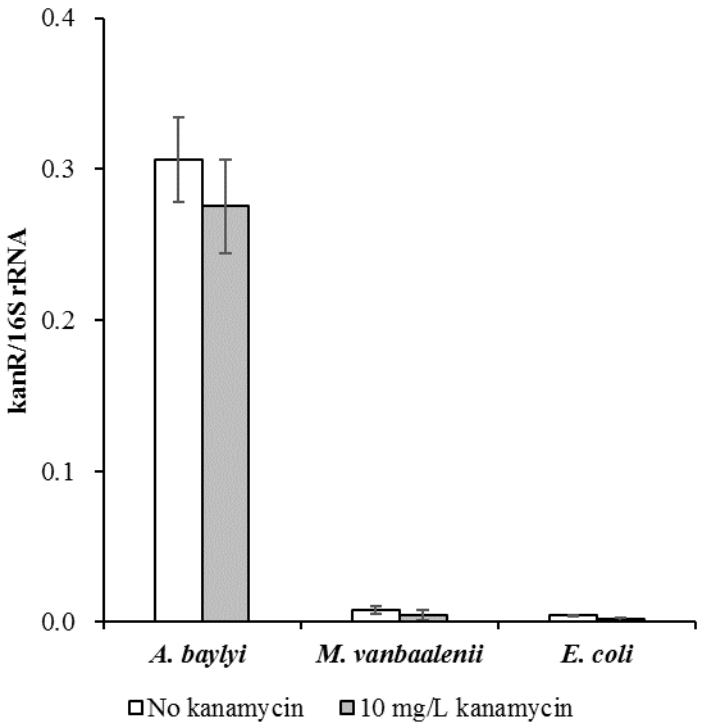

Figure S2. Relative abundance of kanamycin resistance gene (kanR/16S) in *Mycobacterium vanbaalenii* PYR-1, *Escherichia coli* DH5α and *Acinetobacter baylyi* ADPWH_recA after 16-h cultivation without kanamycin pressure or with 10 mg/L kanamycin. Data are presented in mean ± standard error.

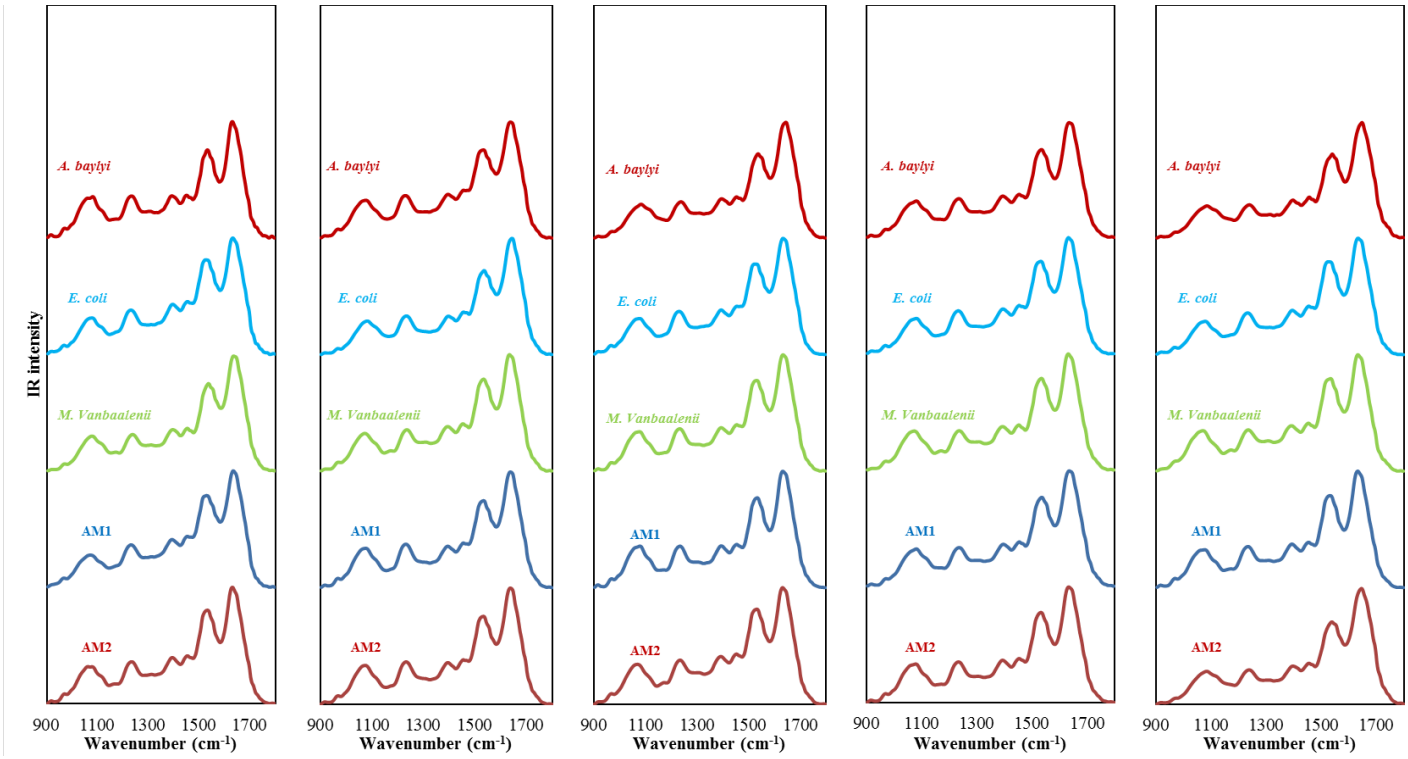

Figure S3. ATR-FTIR spectral dynamics of artificial microbiota.