Supporting Information

Fusion data from FT-IR and MALDI-TOF MS result in more accurate classification of specific microbiota

Wenjing Gao,[‡a] Ying Han,[‡b] Liangqiang Chen,[b] Xue Tan,[b] Jieyou Liu,[c] Jinghang Xie,[a] Bin Li,[a] Huilin Zhao,[a] Shaoning Yu,[a] Huabin Tu,[b] Bin Feng*[a] and Fan Yang*[b]

[a] Institute of Mass Spectrometry, School of Material Science and Chemical Engineering, Ningbo University, Ningbo, Zhejiang 315211, China

[b] Kweichow Moutai Group, Renhuai, Guizhou 564501, China

[c] Zhuhai DL Biotech Co., Ltd, Zhuhai, Guangdong 519041, China.

‡, Wenjing Gao and Ying Han contributed equally.

*Corresponding address:

Bin Feng, Institute of Mass Spectrometry, School of Material Science and Chemical Engineering, Ningbo University, Ningbo, Zhejiang 315211, China. Electronic address: fengbin@nbu.edu.cn.

Fan Yang, Kweichow Moutai Group, Renhuai, Guizhou 564501, China. Electronic address: yangfanmt@189.cn.

Table S1.   Concentration and MALDI-TOF identification of single bacteria

Short name	Concentration ($\times 10^8$ CFU/mL)	Identification result		
		Genus	Species	Score
		Escherichia	*coli*	2.11
E.c	9.1	*Escherichia*	*coli*	2.13
		Escherichia	*coli*	2.17
		Bacillus	*cereus*	2.15
B.c	4.2	*Bacillus*	*cereus*	2.29
		Bacillus	*cereus*	2.46
		Staphylococcus	*aureus*	2.03
S.a	5.3	*Staphylococcus*	*aureus*	2.11
		Staphylococcus	*aureus*	2.03
		Citrobacter	*youngae*	2.32
C.y	8.1	*Citrobacter*	*youngae*	2.26
		Citrobacter	*youngae*	2.05
		Pseudomonas	*aeruginosa*	2.39
P.a	7.5	*Pseudomonas*	*aeruginosa*	2.33
		Pseudomonas	*aeruginosa*	2.34
		Klebsiella	*aerogenes*	2.39
K.a	7.6	*Klebsiella*	*aerogenes*	2.02
		Klebsiella	*aerogenes*	2.23
		Citrobacter	*freundii*	2.32
C.f	8.3	*Citrobacter*	*freundii*	2.32
		Citrobacter	*freundii*	2.37
		Acinetobacter	*baumannii*	2.22
A.b	8.1	*Acinetobacter*	*baumannii*	2.34
		Acinetobacter	*baumannii*	2.30
		Klebsiella	*pneumoniae*	2.15
K.p	6.8	*Klebsiella*	*pneumoniae*	2.07
		Klebsiella	*pneumoniae*	2.35

Table S2. Proportions of 20 model microbiotas

| Microbiota number	Proportion (%)								
	E.c	*B.c*	*S.a*	*K.p*	*C.f*	*A.b*	*P.a*	*C.y*	*K.a*
Q1	13.89	13.89	13.89	13.89	13.89	1.39	1.39	13.89	13.89
Q2	1.39	13.89	13.89	13.89	13.89	13.89	13.89	13.89	1.39
Q3	13.89	13.89	1.39	1.39	13.89	13.89	13.89	13.89	13.89
Q4	13.89	13.89	13.89	13.89	1.39	13.89	13.89	1.39	13.89
Q5	15.87	15.87	1.59	15.87	1.59	15.87	15.87	1.59	15.87
Q6	15.87	1.59	15.87	1.59	15.87	15.87	1.59	15.87	15.87
Q7	15.87	15.87	15.87	15.87	15.87	1.59	1.59	15.87	1.59
Q8	1.59	15.87	1.59	1.59	15.87	15.87	15.87	15.87	15.87
Q9	1.85	18.52	1.85	1.85	18.52	18.52	1.85	18.52	18.52
Q10	18.52	1.85	18.52	18.52	18.52	1.85	1.85	18.52	1.85
Q11	18.52	18.52	1.85	1.85	1.85	18.52	18.52	1.85	18.52
Q12	1.85	1.85	1.85	18.52	18.52	1.85	18.52	18.52	18.52
Q13	2.22	22.22	2.22	2.22	2.22	22.22	22.22	2.22	22.22
Q14	2.22	22.22	22.22	2.22	2.22	22.22	22.22	2.22	2.22
Q15	22.22	22.22	2.22	2.22	2.22	22.22	2.22	2.22	22.22
Q16	22.22	2.22	22.22	22.22	2.22	2.2	2.22	22.22	2.22
Q17	37.04	3.70	3.70	3.70	3.70	3.70	3.70	3.70	37.04
Q18	11.11	11.11	11.11	11.11	11.11	11.11	11.11	11.11	11.11
Q19	3.70	3.70	3.70	3.70	3.70	37.04	37.04	3.70	3.70
Q20	37.04	3.70	37.04	3.70	3.70	3.70	3.70	3.70	3.70
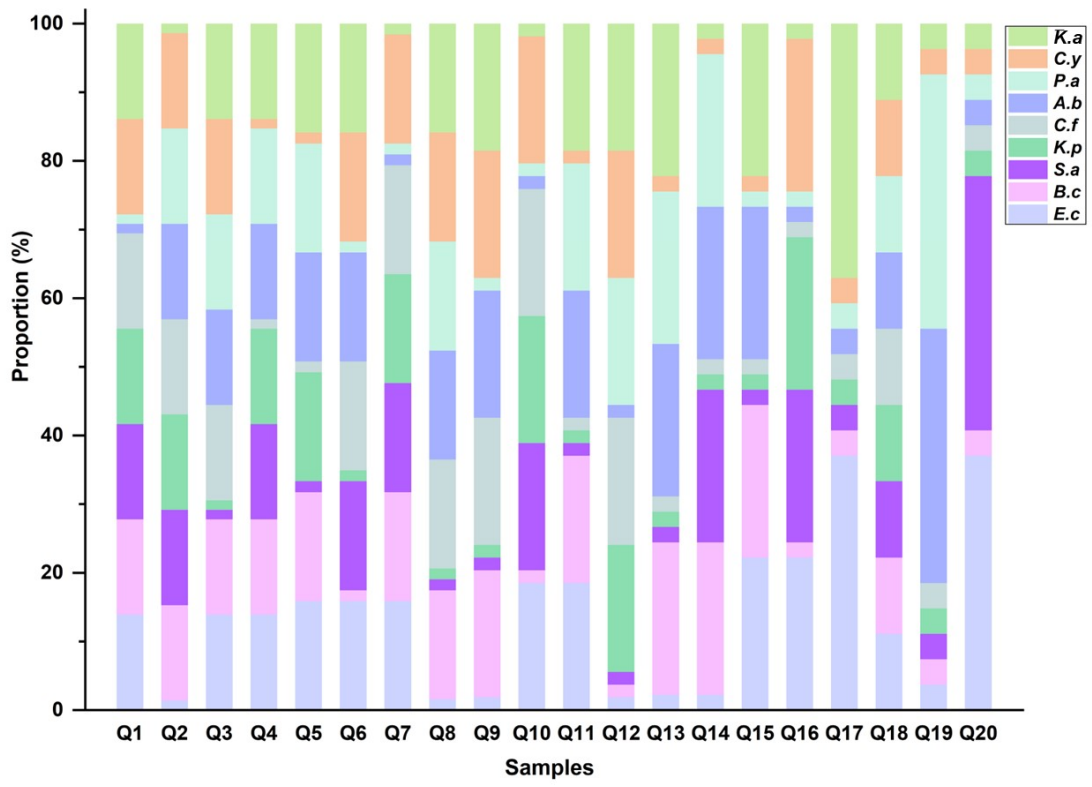
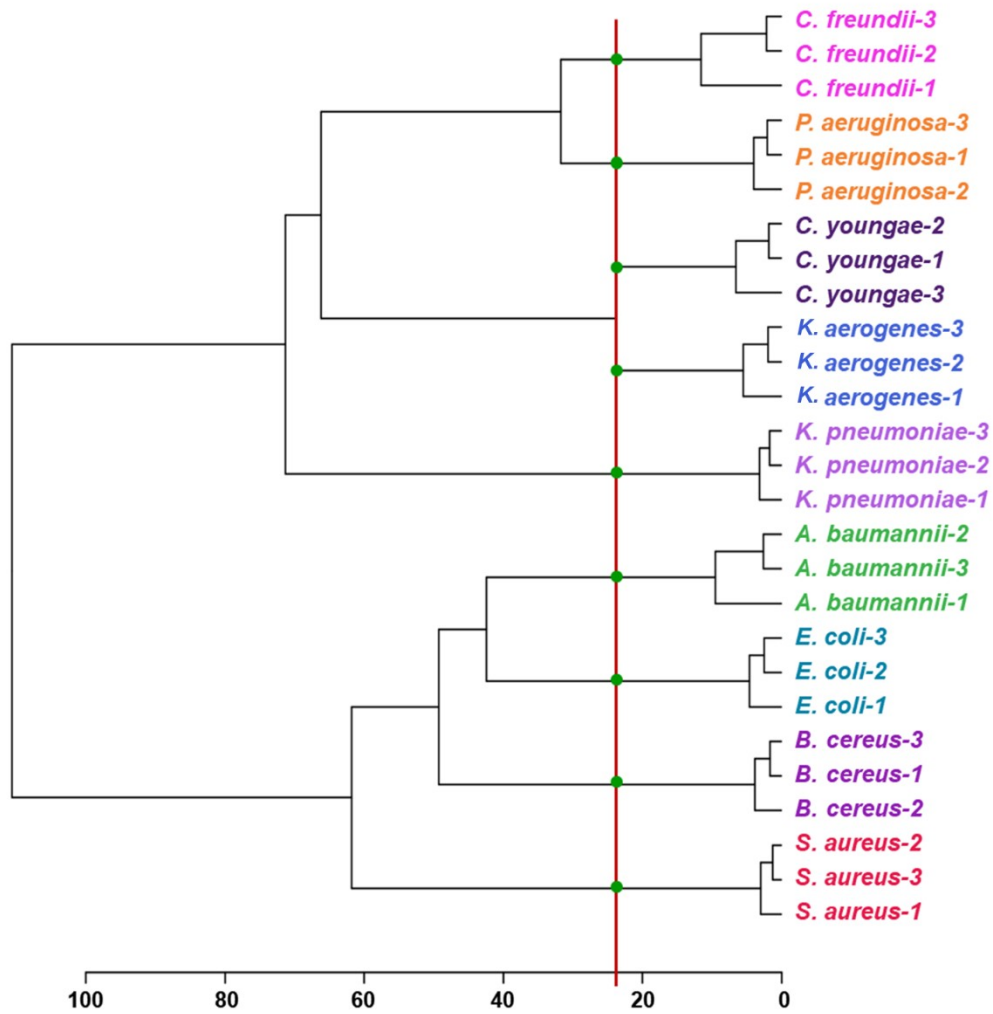**Fig. S1**. Compositions of 20 model microbiotas.

Fig. S2. HCA classification scheme for nine single bacteria using FT-IR. The red line represents the threshold for strain discrimination, and the green spot indicates that the strain is clustered correctly.
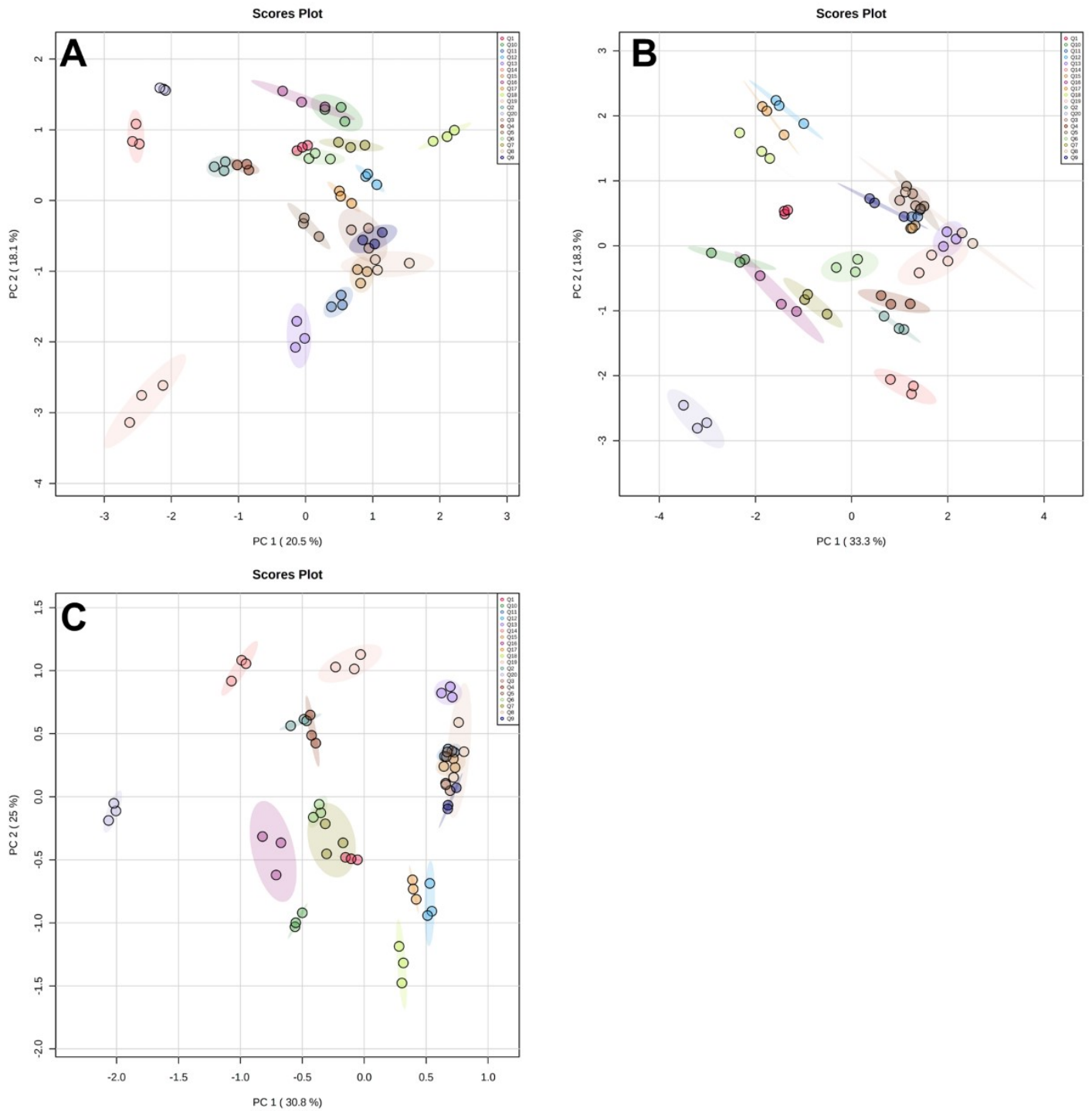
Fig. S3. Principal components analysis. Each set of measurements (spectrum) appears as a single point with the coordinates of the scores on the first two principal components (PCs) chosen as axes for the plot within a 95% confidence interval. A, MALDI-TOF MS; B, FT-IR; C, MS-IR. The distribution points of each strain indicate that the typing accuracy greatly increased by combining the data from MALDI-TOF MS and FT-IR.