Electronic Supplementary Material (ESI) for Chemical Communications

Electronic Supplementary Information

Analysis of Intact Bacteria Using Rapid Evaporative Ionisation Mass Spectrometry

Nicole Strittmatter,[a] *Emrys A. Jones*,[a] *Kirill A. Veselkov*,[a] *Monica Rebec*,[b] *Jacob G. Bundy*,[a] *and Zoltan Takats**,[a]

[a] Section of Computational and Systems Medicine, Department of Surgery and Cancer, Imperial College London

[b] Department of Microbiology, Imperial College Healthcare NHS Trust, Charing Cross Hospital, London

* Corresponding author E-mail address: z.takats@imperial.ac.uk

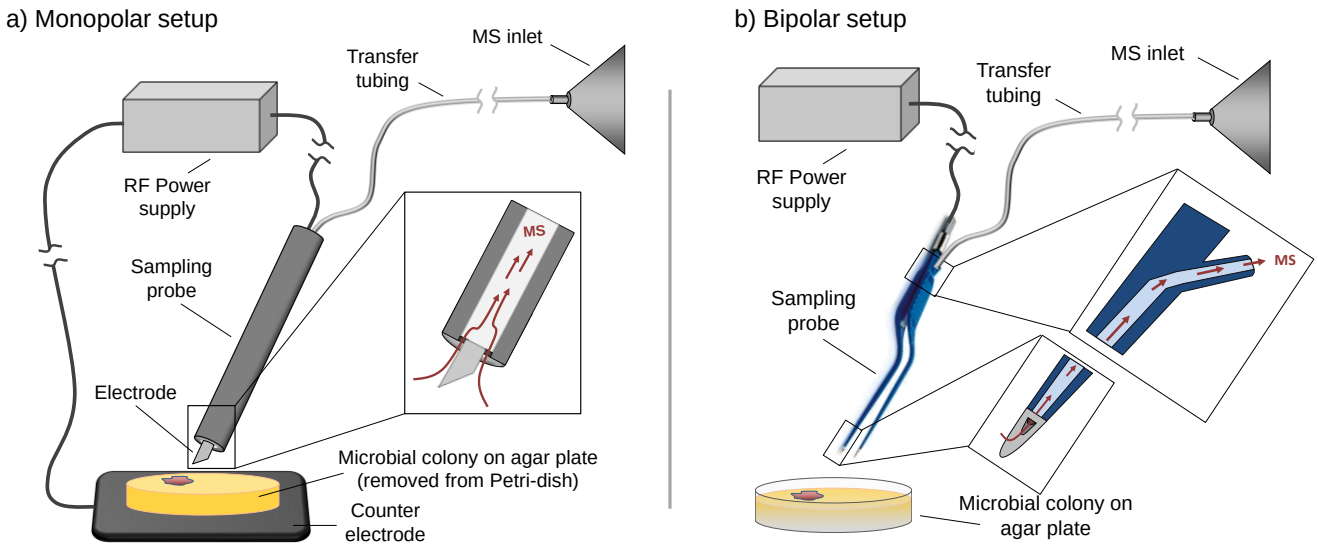

Fig. S1 Monopolar and bipolar setups tested for the analysis of microorganisms using REIMS. Both setups were also tested with external transfer line

Table S1 Details about sample set analysed in this study using REIMS.

Gram-stain	Family	Genus	Species	Growth media[a]
negative	Pseudomonadaceae	Pseudomonas	aeruginosa	LB, BAB, MCC
	Enterobacteriaceae	Citrobacter	koseri	CBA, LB, TS, BHI, BAB, MCC
		Klebsiella	pneumoniae	CBA, LB, TS, BHI, BAB, MCC
		Escherichia	coli	CBA, LB, TS, BAB, MCC
		Proteus	mirabilis	CBA, LB, TS, BHI, BAB, MCC
		Serratia	marcescens	CBA, TS, BHI, BAB, MCC
positive	Staphylococcaceae	Staphylococcus	aureus	LB, TS, BHI, CBA
	Streptococcaceae	Streptococcus	agalactiae	CBA
			pyogenes	CBA, BHI

[a]Abbreviations: LB = Lysogenic broth, BAB = Blood agar base, MCC = McConkey agar, CBA = Columbia blood agar, BHI = Brain-heart infusion agar, TS = Trypticase soy agar.

Clinical isoltaes were obtained from the central clinical microbiology laboratory located in Charing-Cross Hospital, London and identified during clinical diagnostic routine using a Bruker microflex LT MALDI-TOF mass spectrometer.
Samples were randomly analysed using REIMS and the Orbitrap mass analyser running at instrumental parameters as shown in Table S2.

Table S2 Instrumental parameters used in this study

Parameter	Setting
Injection time	1000 ms
Microscans	1
Mass analyser	FTMS*
Ion mode	negative
Mass range	150-2000
Tube Lens Voltage	-120 V
Capillary Voltage	-40 V
Capillary Temperature	250 °C
Automatic Gain Control	Off

* Orbitrap Discovery instrument is working at a resolution of 30,000 at m/z = 400.

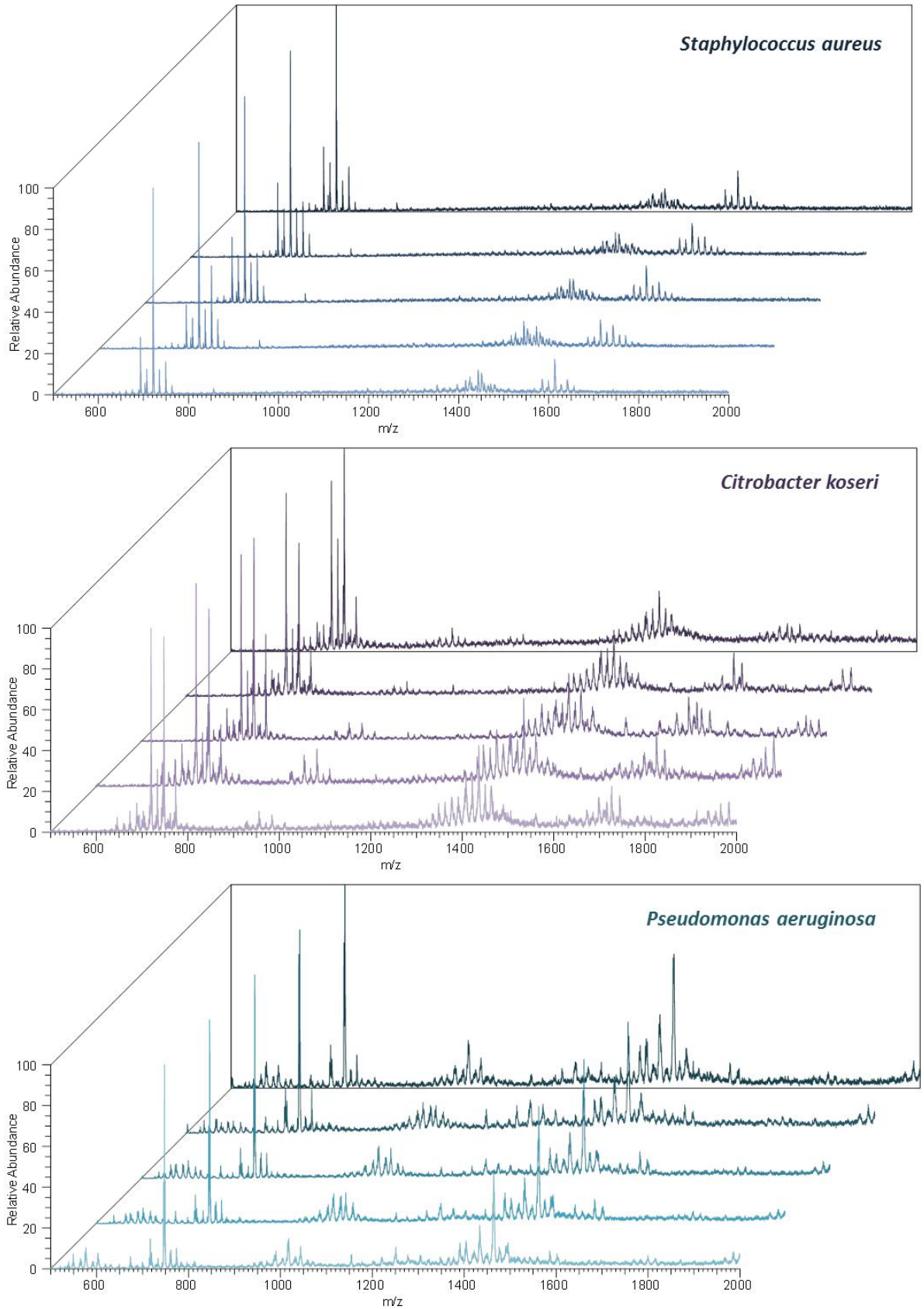

Fig. S2 Mass spectral fingerprints of three different bacterial species recorded using a Paul trap mass analyser in negative ion mode; each 5 individual strains cultured on lysogenic broth, *m/z* 500-2000.

Table S3 Identified phospholipids detected in the mass range *m*/*z* = 600-900 for all analysed species. Only phospholipids with relative abundances >5% and only the most abundant acyl chain combination were included. Solid growth media on which bacteria were grown is given in parentheses.

Nominal mass m/z	*C. koseri* (CBA)	*E. coli* (CBA)	*K. pneumoniae* (LB)	*P. mirabilis* (MCC)	*P. aeruginosa* (LB)	*S. marascens* (MCC)	*S. aureus* (CBA)	*S. agalactiae* (CBA)	*S. pyogenes* (CBA)
645									PA(32:1)*
659			PA(16:0/17:1)	PA(16:0/17:1)		PA(16:0/17:1)			
661							PA(33:0)*		
665									PG(12:0/16:0)
671									PA(34:2)*
673				PA(16:0/18:1)	PA(16:0/18:1)				PA(16:0/18:1)*
675							PG(15:0/15:0-H_2O)		PG(30:0-H_2O)*
688	PE(16:1/16:0)			PE(16:1/16:0)					
691									PG(14:0/16:1)
693	PG(16:0/14:0)		PG(16:0/14:0)				PG(15:0/15:0)	PG(15:0/15:0)	PG(14:0/16:0)
697									PA(36:3)*
699									PA(18:1/18:1)*
701								PG(32:1)-H_2O*	PG(32:1)-H_2O*
702	PE(16:0/17:1)	PE(16:0/17:1)	PE(16:0/17:1)	PE(16:0/17:1)		PE(16:0/17:1)			
707							PG(15:0/16:0)		
716	PE(18:1/16:0)			PE(18:1/16:0)	PE(18:1/16:0)	PE(17:0/17:1)			
717								PG(32:2)*	PG(16:1/16:1)
719	PG(16:1/16:0)	PG(16:1/16:0)	PG(16:0/16:1)	PG(16:0/16:1)	PG(16:0/16:1)	PG(16:0/16:1)		PG(16:0/16:1)	PG(16:0/16:1)
721							PG(15:0/17:0)	PG(15:0/17:0)	PG(16:0/16:0)
725									PA(16:1/18:2)
727									PG(16:1/18:1)-H_2O
729								PG(16:0/18:1)-H_2O*	PG(16:0/18:1)-H_2O
730				PE(16:0/19:1)					
733	PG(16:0/17:1)	PG(16:0/17:1)	PG(16:0/17:1)	PG(16:0/17:1)	PG(16:0/17:1)	PG(16:0/17:1)			
735							PG(15:0/18:0)		
743								PG(16:0/18:3)	PG(16:1/18:2)
745	PG(16:1/18:1)	PG(16:1/18:1)	PG(16:1/18:1)		PG(16:1/18:1)	PG(16:1/18:1)		PG(16:0/18:2)*	PG(16:1/18:1)
747	PG(16:0/18:1)	PG(16:0/18:1)	PG(16:0/18:1)	PG(16:0/18:1)	PG(16:0/18:1)	PG(16:0/18:1)		PG(16:0/18:1)	PG(16:0/18:1)
749							PG(15:0/19:0)	PG(15:0/19:0)	PG(16:0/18:1)*
752									
759		PG(17:1/18:1)	PG(17:1/18:1)		PG(17:1/18:1)	PG(17:1/18:1)			
761		PG(16:0/19:1)	PG(16:0/19:1)	PG(16:0/19:1)	PG(16:0/19:1)	PG(16:0/19:1)			
763							PG(15:0/20:0)		
770									PE(38:2)*
771								PG(36:3)*	PG(18:1/18:1)*
773	PG(18:1/18:1)	PG(18:1/18:1)	PG(17:1/19:1)		PG(17:1/19:1)	PG(18:1/18:1)		PG(36:2)*	PG(18:1/18:1)
775								PG(36:1)*	PG(18:0/18:1)
787			PG(18:1/19:1)						
801			PG(19:1/19:1)						

* Signal intensity not sufficient to obtain meaningful MS/MS data; Abbreviations: PG = phosphatidylglycerol, PE = phosphatidylethanolamine, CBA = Columbia blood agar, LB = lysogenic broth agar, MCC = McConkey agar.

Acyl chain compositions were determined in tandem mass spectrometry experiments using the Paul trap of the Orbitrap Discovery instrument. Mass accuracy for identifications using exact mass was set to <3 ppm.

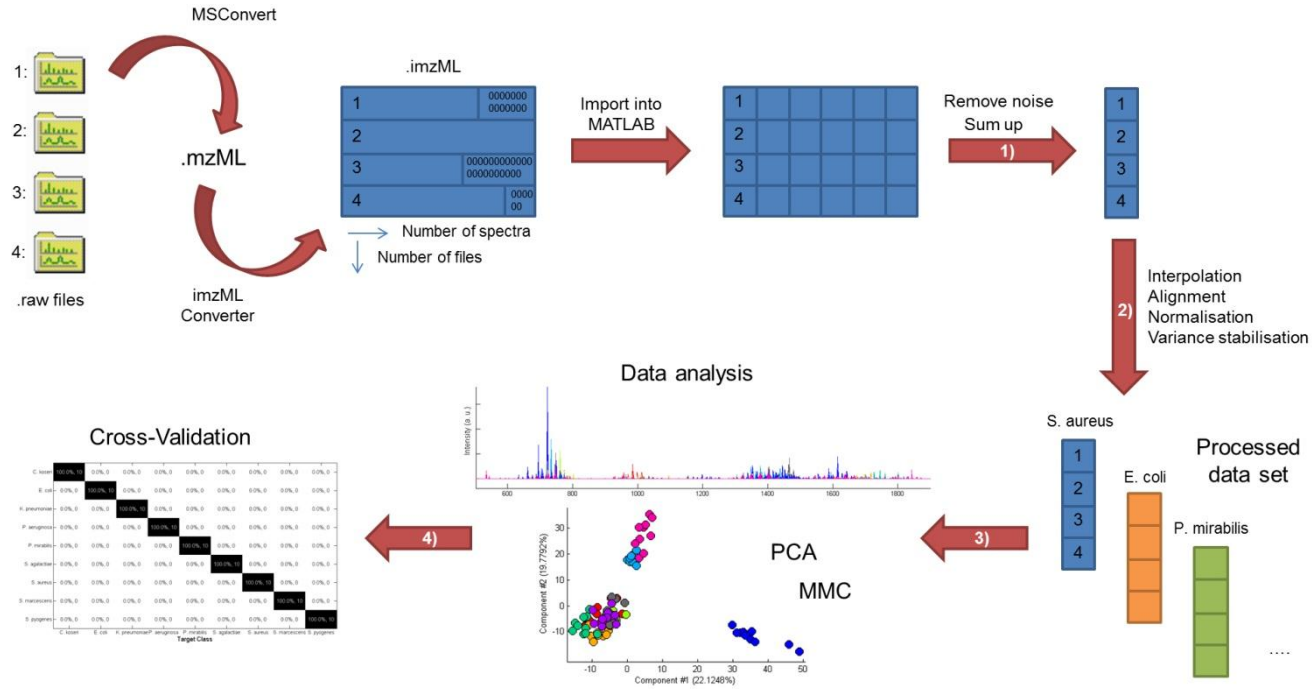

Fig. S3 Scheme depicting data analysis workflow that was followed in this study.

Data pre-processing for improved information recovery: Thermo *raw* mass spectrometric files were transcoded to *mzML* format by the ProteoWizard msconvert tool (version 3.0.4043)[1] and *imzML* format using imzML Converter (version 1.0.5)[2] and imported into MATLAB for data pre-processing, pattern recognition analysis and visualisation (http://www.mathworks.co.uk/). All REIMS spectra were linearly interpolated to a common sampling interval of 0.01Da. Recursive segment wise peak alignment was then used to remove small mass shifts in peak positions across spectral profiles.[3] The aligned data were subjected to the median intensity normalisation to account for sample to sample variation in overall signal intensity unrelated to molecular patterns.[4] This biologically irrelevant variation arises due to a variety of instrumental and sample preparation reasons. The log-based transformation was subsequently applied to stabilise variance as a function of increased signal intensity. This ensured that the noise structure was consistent with the downstream application of multivariate statistical techniques.[4]

Pattern recognition analysis and classification: After the pre-processing step, a set of correlated (redundant) molecular ion variables were transformed by a combination of multivariate statistical techniques into a lower dimensional set of non-redundant components for improved visualisation, explorative analysis and predictive modelling. Principal component analysis (PCA) was initially applied to map high dimensional REIMS data into an uncorrelated set of components capturing the majority of variation in the dataset. Graphical representations of the first few "most informative" components were used to explore the overall similarity/difference in molecular ion composition between bacterial species. To ensure that the derived principal components explain systematic variation not attributable to noise, (3 by 3) bi-cross validation was performed. This procedure leaves out a sub-matrix and then predicts the left out entries by performing PCA on the retained data.[5] Unless the total variation is dominated by the variation between various bacterial species, the derived PCA components are sub-optimal choice for extracting bacterial specific discriminating patterns.[6, 7] Maximum margin criterion[8] was therefore subsequently applied to derive components with enhanced capacity for discriminating between bacterial types by taking into account the microbiological assignment of specimens. The final reduced set of discriminating

components was equal to the number of bacterial types (classes) minus one - eight. The discriminating models were validated using leave-one-out cross-validation. This procedure involves leaving one "test" mass spectrum at a time from the sample set and deriving discriminating components using the retained data. After that, the left out specimen was projected into the derived low dimensional space of discriminating components and classified using 3-nearest neighbour classifier to one of the bacterial species. Confusion matrices were used to show the accuracy of a classification result.

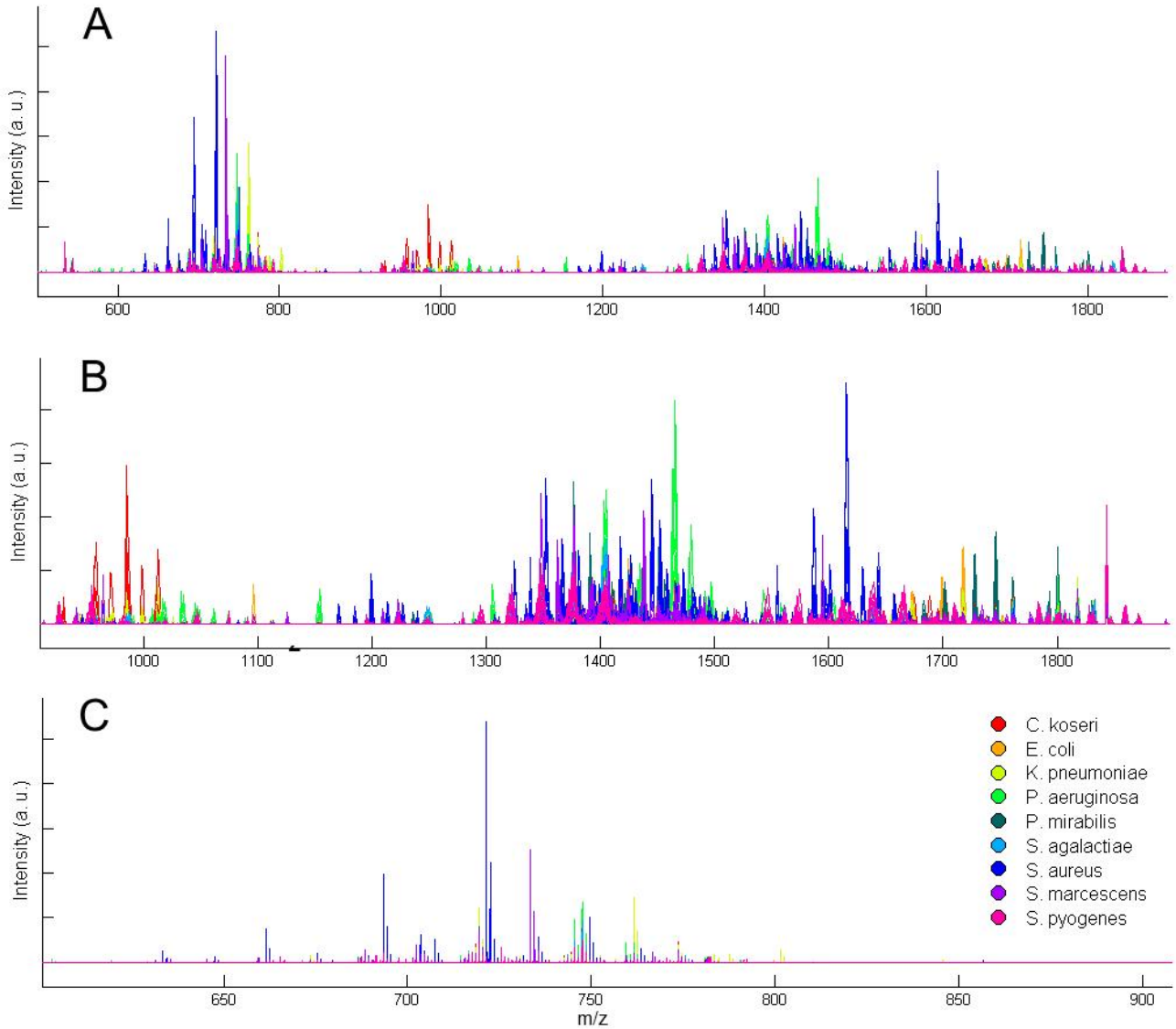

Fig. S4 Processed mass spectral profiles that were used for analysis of different mass ranges. A) *m*/*z* = 500-1900, B) *m*/*z* = 900-1900, C) *m*/*z* = 600-900.

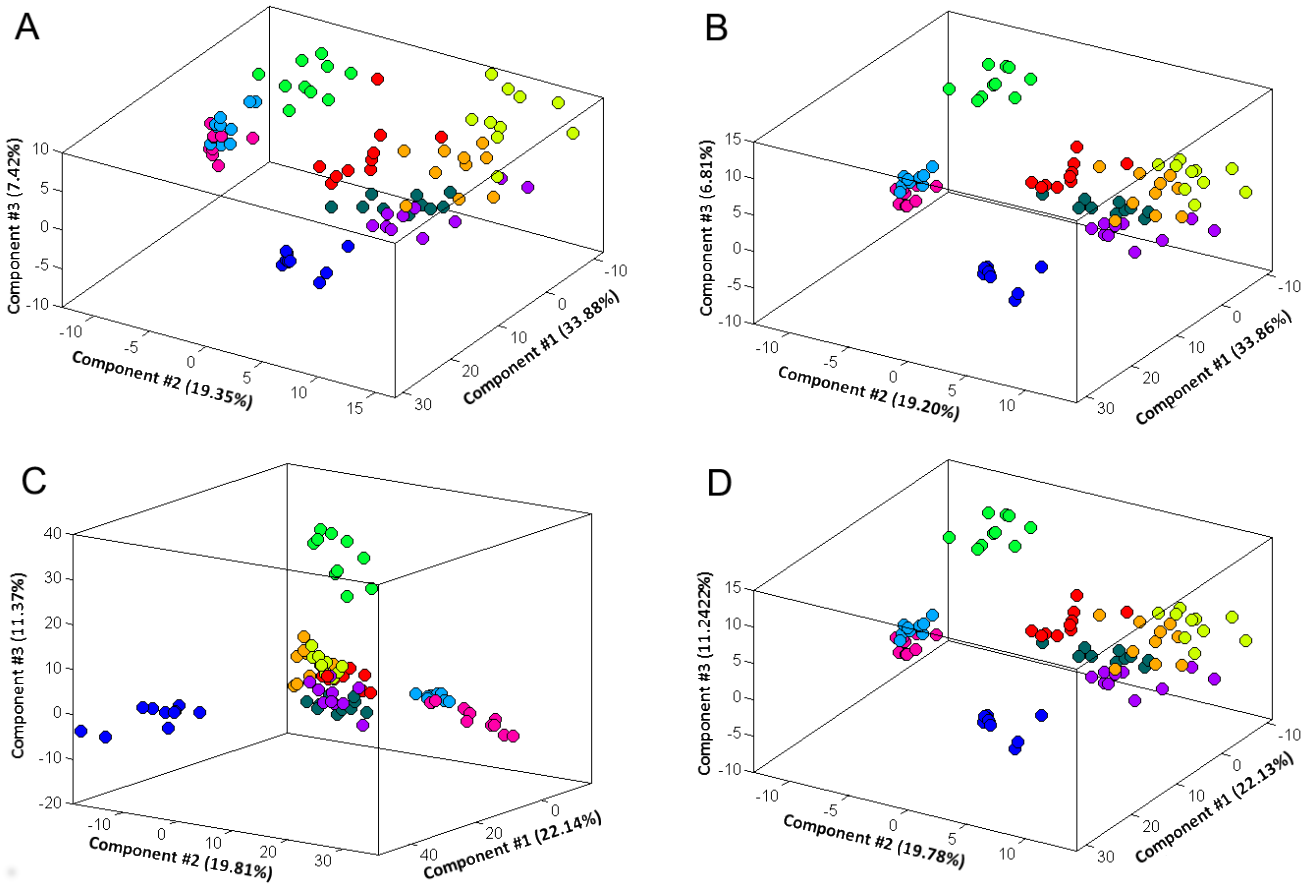

Fig. S5 3D-PCA and MMC plots of the mass ranges A+B) 600-900 and C+D) 900-1900. Model depicted in B gave 90.0% correct classification, model shown in D gave 96.6% correct classification rate when performing leave-one-out cross-validation.

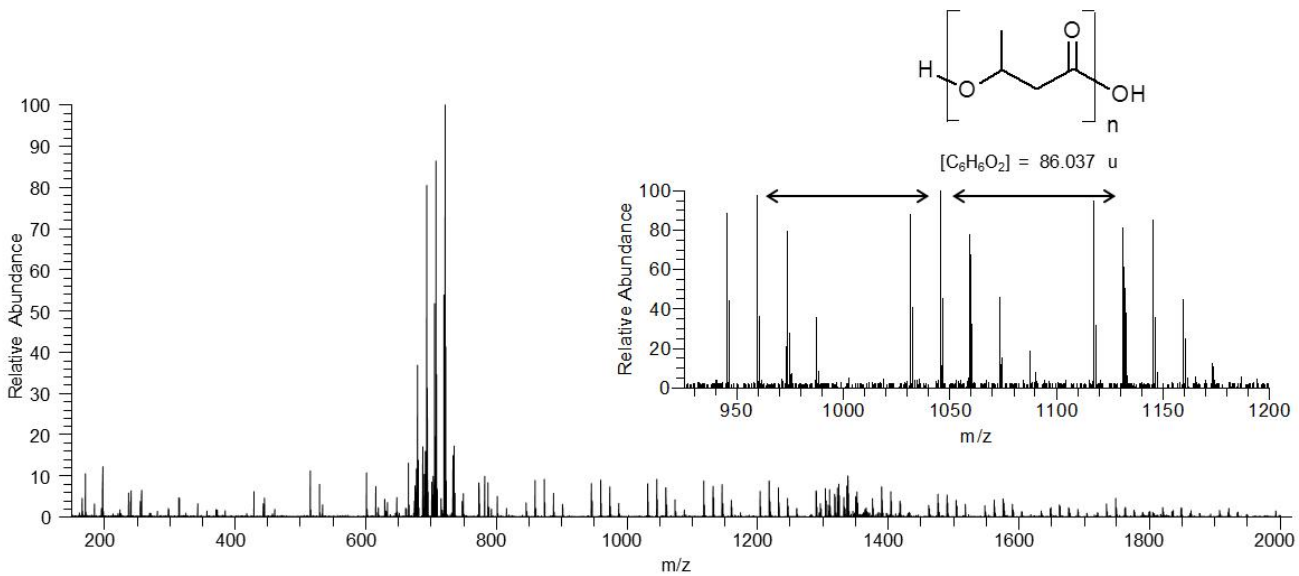

Fig. S6 REIMS profile obtained for Bacillus cereus showing repeating peak distances of 86.037 u corresponding to polyhydroxybutyrate. Additional peakset between *m/z* = 1300-1400 was identified as cardiolipins.

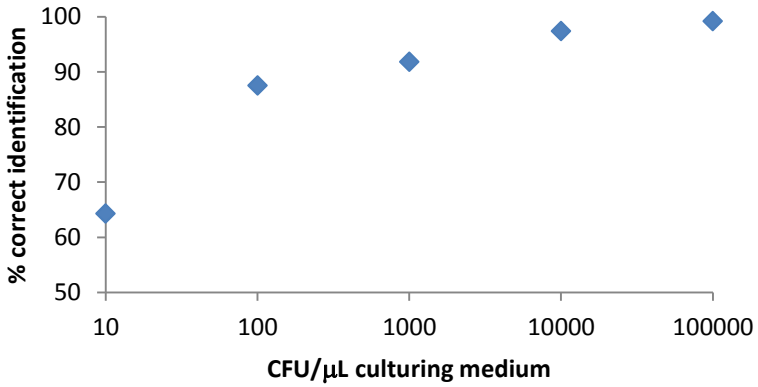

Figure S7 Correct classification rate as a function of bacterial cell concentrations in CFU/µL. One bacterial strain for each of the nine species was grown in liquid culturing medium and resulting sample was sequentially diluted. CFU numbers were determined by direct plating. Dilution series were analysed using the described experimental setup and resulting spectra were identified using the statistical model. Correct classification rates were averaged for each dilution step.

The number of cells required for analysis is dependent on the experimental setup and aerosol transfer efficiency, however, for the setup used following correct classification rates were obtained for different bacterial cell concentrations in liquid culturing medium. These results show that for concentrations exceeding 10^7 CFU/ml onwards the correct classification is better than 97.4 %. Colony counts of on average $2x10^5$ CFU/µL in pus and infected peritoneal fluids, [9] *Gardnerella vaginalis* concentrations above $1x10^5$ CFU/µL detected in bacterial vaginosis[10] and a mean colony count of 1.7×10^4 CFU/µL in children with urinary tract infections[11] clearly show that direct detection of bacteria in infected samples is feasible with the current sensitivity. However, sensitivity is expected to further improve with ongoing development of electrode setup and instrumentation.

1. M. C. Chambers, B. Maclean, R. Burke, D. Amodei, D. L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, K. Hoff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T. A. Baker, M.-Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S. L. Seymour, L. M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E. W. Deutsch, R. L. Moritz, J. E. Katz, D. B. Agus, M. MacCoss, D. L. Tabb and P. Mallick, *Nat. Biotech.*, 2012, **30**, 918-920.
2. A. M. Race, I. B. Styles and J. Bunch, *J. Proteom.*, 2012, **75**, 5111-5112.
3. K. A. Veselkov, J. C. Lindon, T. M. D. Ebbels, D. Crockford, V. V. Volynkin, E. Holmes, D. B. Davies and J. K. Nicholson, *Anal. Chem.*, 2009, **81**, 56-66.
4. K. A. Veselkov, L. K. Vingara, P. Masson, S. L. Robinette, E. Want, J. V. Li, R. H. Barton, C. Boursier-Neyret, B. Walther, T. M. Ebbels, I. Pelczer, E. Holmes, J. C. Lindon and J. K. Nicholson, *Anal. Chem.*, 2011, **83**, 5864-5872.
5. A. B. Owen and P. O. Perry, *Ann. Appl. Stat.*, 2009, **3**, 564-594.
6. P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, *IEEE Trans. Patt. Anal.Mach. Intell.*, 1997, **19**, 711-720.
7. J. Yang and J. Yang, *Patt. Recog.*, 2003, **36**, 563-566.
8. L. Haifeng, J. Tao and Z. Keshu, *IEEE Trans. Neural Netw.*, 2006, **17**, 157-165.
9. C. Koenig, H.-P. Simmen and J. Blaser, *J. Antimicrob. Chemother.*, 1998, **42**, 227-232.
10. A. Swidsinski, W. Mendling, V. Loening-Baucke, A. Ladhoff, S. Swidsinski, L. P. Hale and H. Lochs, *Obstet. Gynecol.*, 2005, **106**, 1013-1023.
11. M. G. Coulthard, M. Kalra, H. J. Lambert, A. Nelson, T. Smith and J. D. Perry, *Pediatrics*, 2010, **125**, 335-341.