Supporting Information

# Rapid Characterization and Identification of Clinically Relevant Microorganisms Using Rapid Evaporative Ionization Mass Spectrometry

Nicole Strittmatter[1], Monica Rebec[2], Emrys A. Jones[1], Ottmar Golf[1], Alireza Abdolrasouli[2], Julia Balog[1], Volker Behrends[1], Kirill A. Veselkov[1], Zoltan Takats[1]*

1: Section of Computational and Systems Medicine, Department of Surgery and Cancer, Imperial College London, UK

2: Department of Microbiology, Imperial College Healthcare NHS Trust, Charing Cross Hospital, London, UK

*To whom correspondence should be addressed: Zoltan Takats, Department of Surgery and Cancer, Sir Alexander Fleming Building, Imperial College London, London SW7 2AZ, UK. Tel: +44 (0)207 5942760; E-mail: z.takats@imperial.ac.uk.

This supporting material includes more detailed information on data-preprocessing and on the sample sets that were analyzed using REIMS (culture conditions, taxonomical information). It also contains further plots of multivariate statistical analysis of the presented data. Mass spectra and schemes shown are a result of different experimental setups (monopolar vs. bipolar) and settings (positive vs. negative ion mode). The spectral features identified using exact mass and tandem MS experiments are further described and experimental data shown.

**Table S-1.** Instrumental parameters of Orbitrap Discovery and Xevo G2-S instruments used in this study.

| Thermo Orbitrap Discovery Exactive | | | Waters Xevo G2-S | |
|---|---|---|---|---|
| **Parameter** | **Setting** | **Setting** | **Parameter** | **Setting** |
| Injection time | 1000 ms | 1000 ms | Scan time | 1000 ms |
| Microscans | 1 | 1 | Scan Mode | Sensitivity |
| Mass analyser | FTMS[a] | FTMS[b] | Mass analyser | TOF |
| Ion mode | negative | negative | Ion mode | negative |
| Mass range | 150-2000 | 150-2000 | Mass range | 150-2000 |
| Tube Lens Voltage | -120 V | -160 V | Sampling Cone | 30 V |
| Capillary Voltage | -40 V | -50 V | Source Offset | 80 V |
| Skimmer Voltage | - | -24 V | Source Temperature | 150 °C |
| Capillary Temperature | 250 °C | 250 °C | | |
| Automatic Gain Control | Off | On | | |

a: Orbitrap Discovery instrument is working at a resolution of 30,000 at *m/z* = 400, b: Mass analyser was used at a resolution of 50,000 (m/z = 200)

## Data analysis

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/). A typical total ion chromatogram as obtained during data acquisition for a database entry is shown in Figure S-1. Individual measurements are represented by the rise and subsequent fall in total ion current. This process takes on average 0.8-1.2 minutes with an individual measurement lasting for 3-6 seconds (see spikes in Figure S-1). All spectra acquired for an individual sample were extracted from the imzML format and subsequently summed up to lead to one sum spectrum per isolate. Spectra were chosen based on an average total ion current threshold of 120% as indicated by the blue line in Figure S-1.

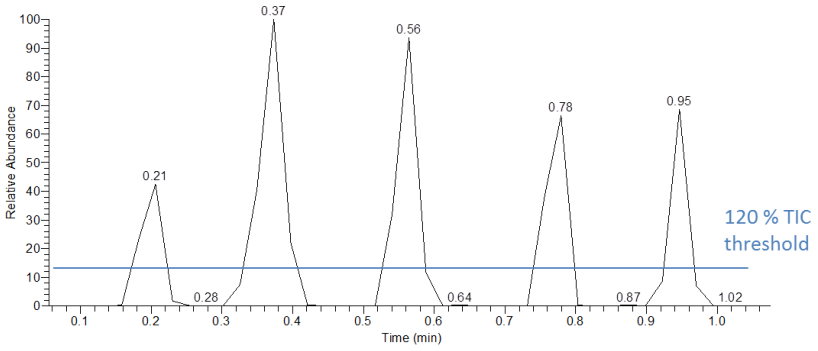


**Figure S-1**. Scheme for selection of spectra using a TIC threshold.

All REIMS spectra were linearly interpolated to a common sampling interval of 0.01 Da. Since a binning strategy was used in this study, recursive segment wise peak alignment was subsequently used to remove small mass shifts in peak positions across spectral profiles.[3] The aligned data were subjected to data normalization (median, mean or TIC normalization as stated in text) to account for sample to sample variation in overall signal intensity unrelated to molecular patterns.[4] Log-based transformation was subsequently applied to stabilize variance as a function of

increased signal intensity (variance stabilization normalization). This ensured that the noise structure was consistent with the downstream application of multivariate statistical techniques.[4]

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. A recursive maximum margin criterion algorithm[5,6] was subsequently applied to derive components with enhanced capacity for discriminating between bacterial types by taking into account the microbiological assignment of specimens. Alternatively, linear discriminant analysis on the components derived from PCA was performed for supervised analyses. The final reduced set of discriminating components was equal to the number of bacterial species (classes) minus one. The discriminating models were validated using cross-validations (CV).

Hierarchical Cluster Analysis (HCA) was performed using Euclidean pairwise distance calculation with a complete linkage metric. 3x3 strains of the original dataset shown in Figure 3A+B were averaged for each bacterial species to form the dataset which was then subjected to HCA. This step was undertaken in order to facilitate visualization while still incorporating a maximum of the biological variance among strains of a certain species.

## Comparison of positive and negative ion mode

In REIMS processes, positively and negatively charged species are generated in equal amounts. However, as Figure S-2 shows comparably bad signal intensity is observed for positive ion mode. REIMS predominantly shows molecular species that were present as ions in the sample material already or such species with very high tendency to form ions. Therefore the detection of lipid species is highly favoured in REIMS mechanism. Bacterial membranes mostly consist of phosphatidylglycerols, phosphatidylethanolamines and cardiolipins which all undergo ionization in negative ion mode. The bad signal intensity in positive ion mode is tentatively attributed to the low abundance of lipid species in bacteria that would readily ionize in positive ion mode (absence of phosphatidylcholins, only low abundance of phosphatidylethanolamines). Whereas negative ion mode gives strong intensity spectra for all species analyzed, positive ion mode proves less reliable. Thus, for the generation of a robust identification method, negative ion mode with constantly high signal content was chosen.

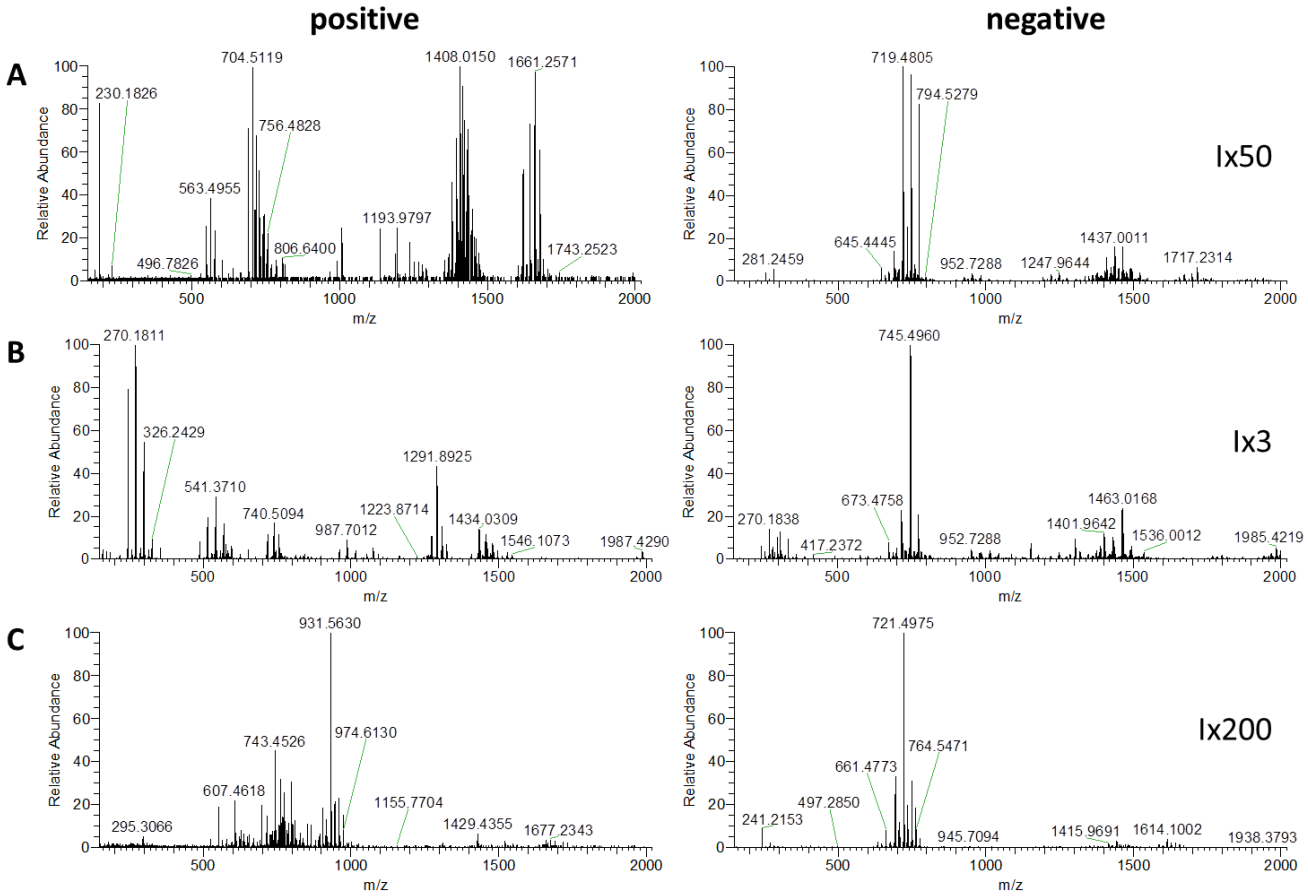


**Figure S-2.** Comparison of REIMS profiles obtained in positive and negative ion mode for A) *Escherichia coli*, B) *Pseudomonas aeruginosa*, C) *Staphylococcus aureus*. I=Intensity.

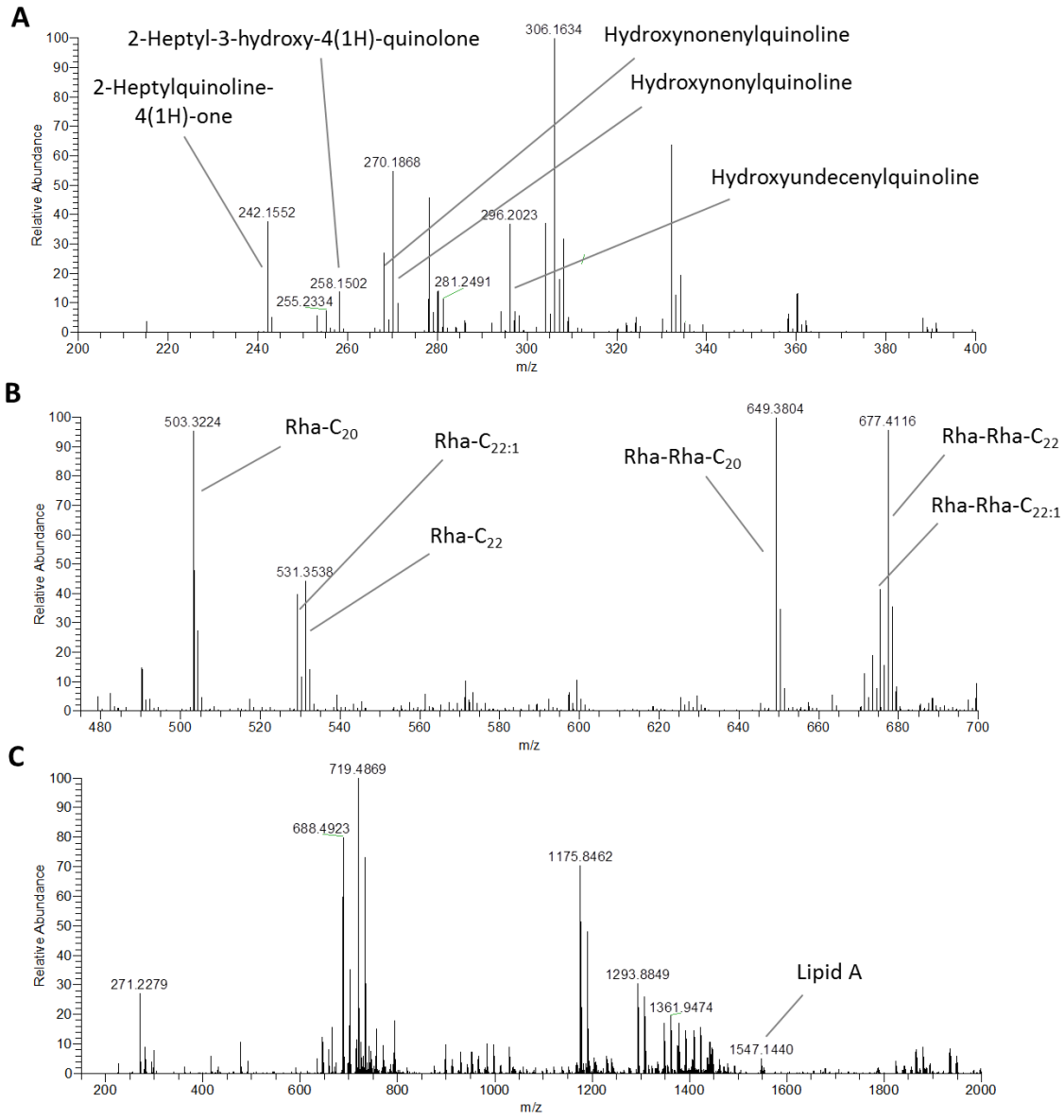


**Figure S-3.** Compounds identified in this study. A) Quorum-sensing molecules in *Pseudomonas aeruginosa*, B) rhamnolipids in *P. aeruginosa*, C) intact Lipid A species in *Helicobacter pylori.*

**Table S-2.** Information to identified compounds shown in mass spectra depicted in Figure S-2.

| Compound | Sum formula | Exact mass | Exp. mass | Mass Deviation | MS/MS |
|---|---|---|---|---|---|
| 2-Heptylquinoline-4(1H)-one | $C_{16}H_{21}NO$ | $[M-H]^-$ = 242.1550 | 242.1552 | -0.8 ppm | Yes |
| 2-Heptyl-3-hydroxy-4(1H)-quinolone (PQS) | $C_{16}H_{21}NO_2$ | $[M-H]^-$ = 258.1499 | 258.1502 | -1.2 ppm | Yes |
| Hydroxynonenylquinoline | $C_{18}H_{23}NO$ | $[M-H]^-$ = 268.1707 | 268.1711 | -1.5 ppm | Yes |
| Hydroxynonylquinoline | $C_{18}H_{25}NO$ | $[M-H]^-$ = 270.1863 | 270.1868 | -1.9 ppm | Yes |
| Hydroxyundecenylquinoline | $C_{20}H_{26}NO$ | $[M-H]^-$ = 296.2020 | 296.2023 | -1.0 ppm | Yes |
| Rha-$C_{20}$ | $C_{26}H_{48}O_9$ | $[M-H]^-$ = 503.3225 | 503.3224 | 0.2 ppm | Yes |
| Rha-$C_{22:1}$ | $C_{28}H_{50}O_9$ | $[M-H]^-$ = 529.3382 | 529.3384 | -0.4 ppm | Yes |
| Rha-$C_{22}$ | $C_{28}H_{52}O_9$ | $[M-H]^-$ = 531.3539 | 531.3538 | 0.2 ppm | Yes |
| Rha-Rha-$C_{20}$ | $C_{32}H_{58}O_{13}$ | $[M-H]^-$ = 649.3805 | 649.3804 | 0.2 ppm | Yes |
| Rha-Rha-$C_{22}$ | $C_{34}H_{62}O_{13}$ | $[M-H]^-$ = 677.4118 | 677.4116 | -0.3 ppm | Yes |
| Rha-Rha-$C_{22:1}$ | $C_{34}H_{60}O_{13}$ | $[M-H]^-$ = 675.3961 | 675.3965 | -0.6 ppm | Yes |
| Lipid A in *Helicobacter pylori* | $C_{84}H_{162}N_3O_{19}P$ | $[M-H]^-$ = 1547.1467 | 1547.1440 | 2.5 ppm | No |
| Lipid A in *Escherichia coli* | $C_{94}H_{177}N_2O_{25}P_2$ | $[M-H]^-$ = 1796.2121 | 1796.2110 | 0.6 ppm | No |

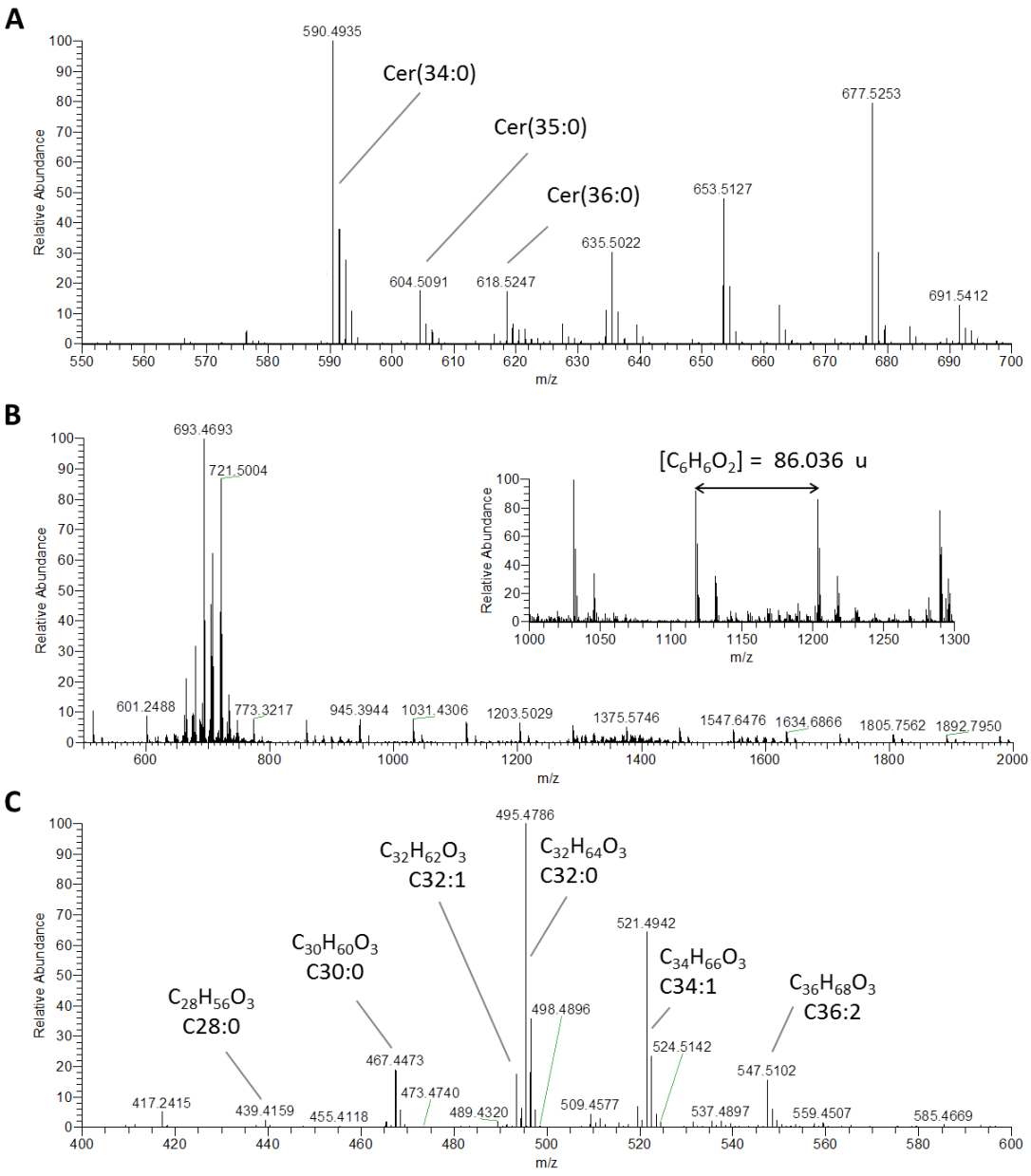


**Figure S-4.** Compounds identified in this study. A) Ceramides as chloride adducts in *Bacteroides fragilis*, B) polyhydroxybutyrate polymer in *Bacillus cereus*, C) short chain mycolic acids in *Corynebacterium striatum*.

**Table S-3.** Information to identified compounds shown in mass spectra depicted in Figure S-3.

| Compound | Sum formula | Exact mass | Exp. mass | Mass Deviation | MS/MS |
|---|---|---|---|---|---|
| Mycolic acid C28:0 | $C_{28}H_{55}O_3$ | $[M-H]^-$ = 439.415669 | 439.4159 | -0.5 ppm | Yes |
| Mycolic acid C30:0 | $C_{30}H_{59}O_3$ | $[M-H]^-$ = 467.446969 | 467.4473 | -0.7 ppm | Yes |
| Mycolic acid C32:1 | $C_{32}H_{61}O_3$ | $[M-H]^-$ = 493.462619 | 493.4634 | -1.6 ppm | Yes |
| Mycolic acid C32:0 | $C_{32}H_{63}O_3$ | $[M-H]^-$ = 495.478269 | 495.4786 | -0.7 ppm | Yes |
| Mycolic acid C34:2 | $C_{34}H_{63}O_3$ | $[M-H]^-$ = 519.478269 | 519.4788 | -1.0 ppm | Yes |
| Mycolic acid C34:1 | $C_{34}H_{65}O_3$ | $[M-H]^-$ = 521.493919 | 521.4942 | -0.5 ppm | Yes |
| Mycolic acid C36:2 | $C_{36}H_{67}O_3$ | $[M-H]^-$ = 547.509569 | 547.5102 | -1.2 ppm | Yes |
| Ceramide Cer(34:0) | $C_{34}H_{69}NO_4$ | $[M+Cl]^-$ = 590.4921 | 590.4935 | -2.4 ppm | Yes |
| Ceramide Cer(35:0) | $C_{35}H_{71}NO_4$ | $[M+Cl]^-$ = 604.5077 | 604.5091 | -2.3 ppm | Yes |
| Ceramide Cer(36:0) | $C_{36}H_{73}NO_4$ | $[M+Cl]^-$ = 618.5234 | 618.5247 | -2.1 ppm | Yes |
| Polyhydroxybutyrate polymer | $[C_6H_6O_2]_n$ | $\Delta m$ = 86.0368 | 86.0366 | 2.3 ppm | Yes |

Compounds 2-Heptylquinoline-4(1H)-one and 2-Heptyl-3-hydroxy-4(1H)-quinolone (PQS) have been confirmed by comparison with tandem mass spectra of standard compounds. Hydroxynonenylquinoline (*m/z* 268), hydroxynonylquinoline (*m/z* 270) and hydroxyundecenylquinoline (*m/z* 296) show similar fragmentation patterns and can thus be ascribed to structurally similar compounds. Tandem mass spectra of these compounds featured in the literature do only include fragmentation of the $[M+H]^+$ quasi-molecular ion. However, the fragments observed in negative ion mode (*m/z* 157 and 170) seem to correlate with the fragments observed in positive ion mode (*m/z* 159 and 172) and are indicative of 4-hydroxy-2-alkylquinolines.[7]

Rhamnolipids at *m/z* 503 and 649 were confirmed by tandem mass spectra published in literature.[8,9] Other rhamnolipids were assigned based on similar fragmentation patterns (loss of rhamnose moieties, loss of one of the acyl chains).[9]

Mycolic acid compounds follow the general fragmentation pattern reported in literature (fragmentation under neutral loss of β-chain as aldehyde, signals observed reflect α-side chain lengths).[10]

Observed ceramides were oxidised ceramides (+16 Da compared to conventional ceramides). Interpretation of the fragmentation pattern is illustrated for species *m/z* 590 below.

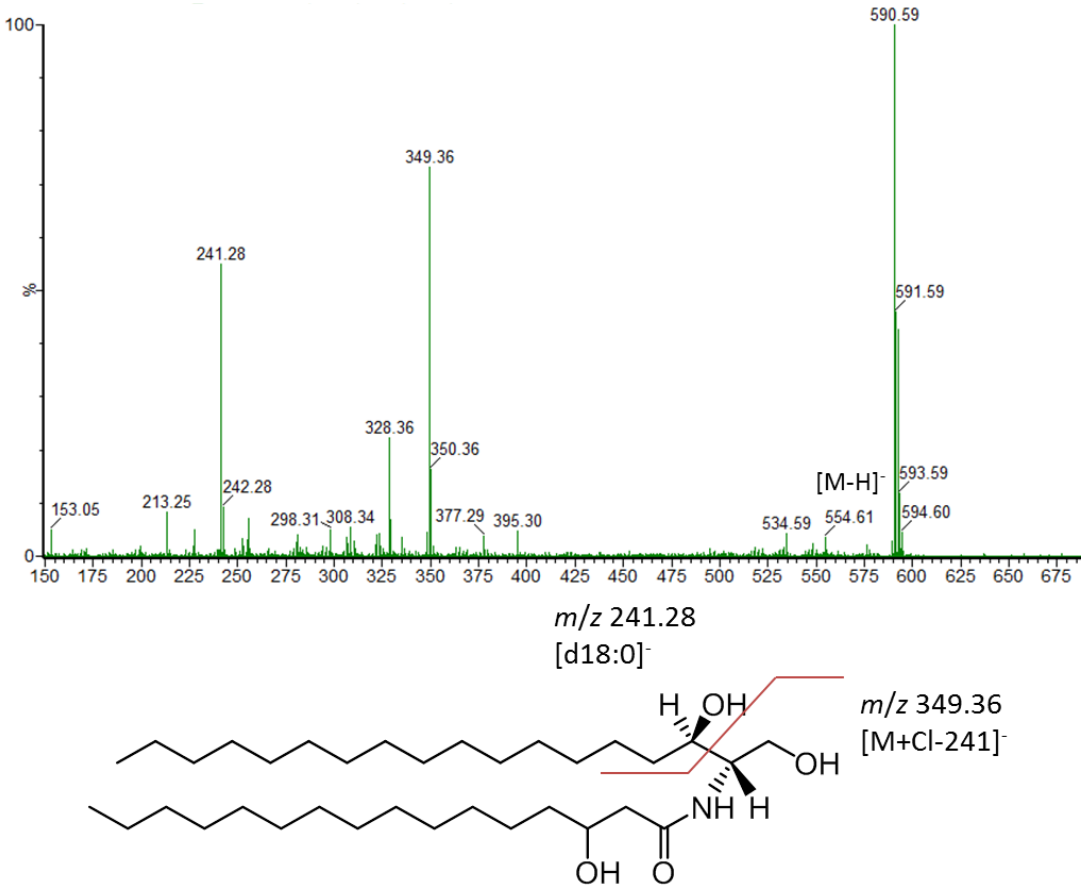


**Figure S-5. Tentative scheme of fragmentation for oxidised ceramide signal at m/z 590.**

Polyhydroxybutyrate peaks were observed to fragment under loss of monomers, dimers etc..

## Comparison monopolar vs. bipolar setup

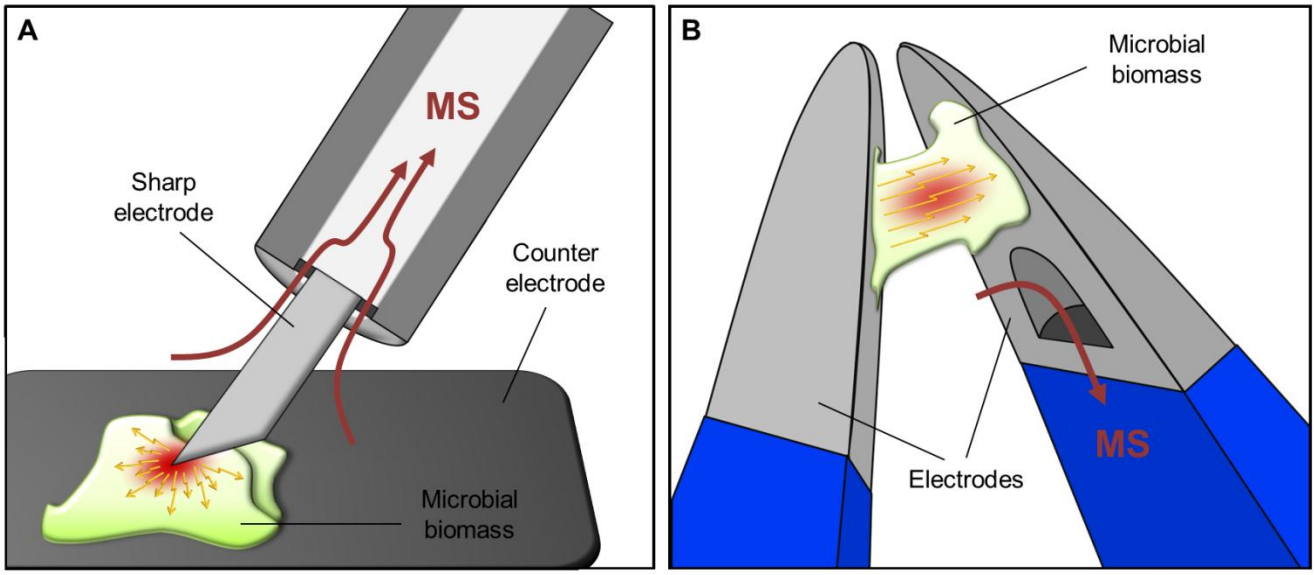


**Figure S-6.** Scheme of experimental setup and electrical current density using A) a monopolar and B) a bipolar tool.

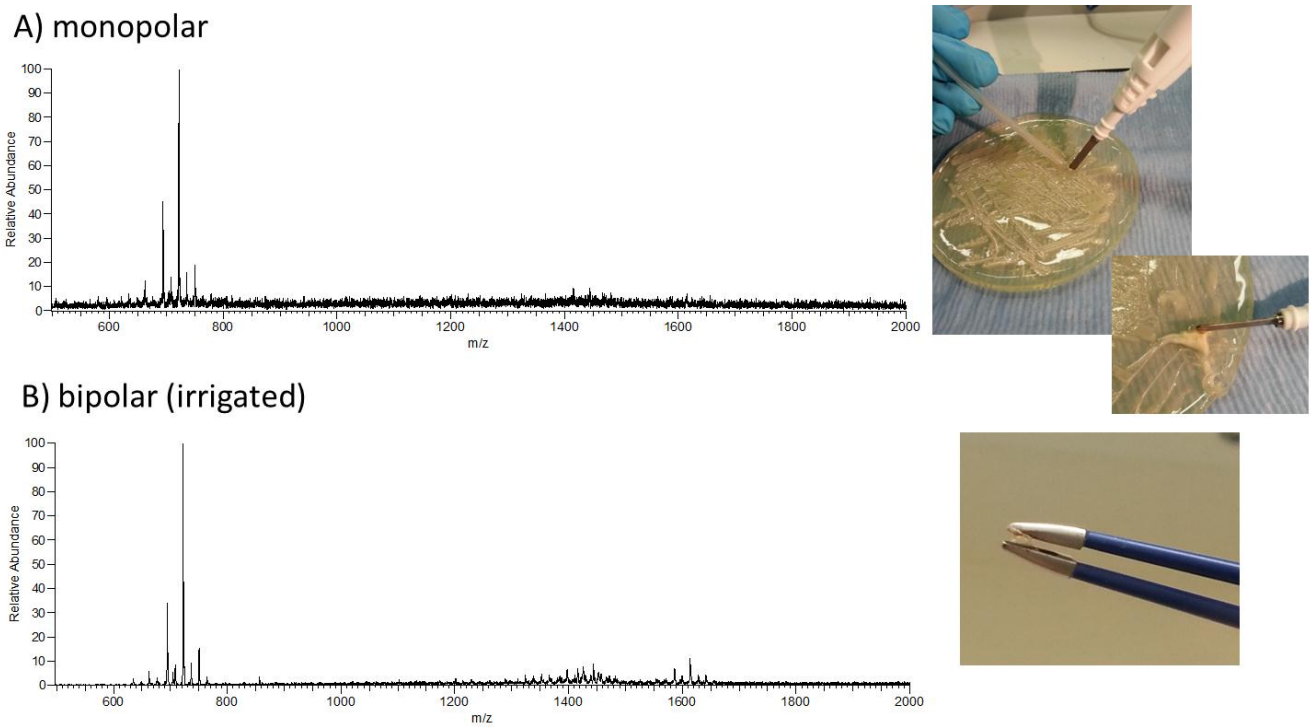


**Figure S-7.** Spectra of *Staphylococcus aureus* obtained using A) monopolar and B) bipolar setup. The use of bipolar setups provides better handling and sensitivity and increased detection of high molecular weight lipid species.

**Table S-4.** Details about sample set analysed in this study using REIMS (Figure 3).

| Gram-stain | Family | Genus | Species | No. of strains | Growth conditions |
|---|---|---|---|---|---|
| negative | *Pseudomonadaceae* | *Pseudomonas* | *aeruginosa* | 15 | CBA, aerobic |
| | *Enterobacteriaceae* | *Citrobacter* | *koseri* | 15 | CBA, aerobic |
| | | *Enterobacter* | *aerogenes* | 14 | CBA, aerobic |
| | | | *cloacae* | 15 | CBA, aerobic |
| | | *Klebsiella* | *oxytoca* | 15 | CBA, aerobic |
| | | | *pneumoniae* | 15 | CBA, aerobic |
| | | *Escherichia* | *coli* | 15 | CBA, aerobic |
| | | *Proteus* | *mirabilis* | 15 | CBA, aerobic |
| | | *Morganella* | *morganii* | 15 | CBA, aerobic |
| | | *Serratia* | *marcescens* | 15 | CBA, aerobic |
| | *Pasteurellaceae* | *Haemophilus* | *influenzae* | 15 | CHOC, aerobic (5 % $CO_2$) |
| | *Burkholderiaceae* | *Burkholderia* | *cepacia complex* | 10 | CBA, aerobic (5 % $CO_2$) |
| | *Xanthomonadaceae* | *Stenotrophomonas* | *maltophilia* | 15 | CBA, aerobic |
| | *Bacteroidaceae* | *Bacteroides* | *fragilis* | 11 | CBA, anaerobic |
| | *Moraxellaceae* | *Moraxella* | *catarrhalis* | 15 | CBA, aerobic |
| | *Neisseriaceae* | *Neisseria* | *gonorrhoeae* | 15 | CBA, aerobic (5 % $CO_2$) |
| positive | *Staphylococcaceae* | *Staphylococcus* | *aureus* | 15 | CBA, aerobic |
| | | | *epidermidis* | 15 | CBA, aerobic |
| | | | *capitis* | 15 | CBA, aerobic |
| | | | *haemolyticus* | 15 | CBA, aerobic |
| | | | *hominis* | 15 | CBA, aerobic |
| | *Enterococcaceae* | *Enterococcus* | *faecalis* | 15 | CBA, aerobic |
| | | | *faecium* | 15 | CBA, aerobic |
| | *Clostridiaceae* | *Clostidium* | *difficile* | 15 | CBA, anaerobic |
| | *Micrococcaceae* | *Micrococcus* | *luteus* | 15 | CBA, aerobic |
| | *Streptococcaceae* | *Streptococcus* | *agalactiae* | 15 | CBA, aerobic |
| | | | *pyogenes* | 15 | CBA, aerobic |
| | | | *pneumoniae* | 15 | CBA, aerobic |

Abbreviations: CBA = Columbia horse blood agar, CHOC = Chocolate agar

Clinical isolates 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. Bacterial specimens were recultured from frozen glycerol broths and incubated for 24hrs using the atmospheric conditions as shown in Tables S-4 and S-5.

Yeast species were isolated during routine clinical mycology work and cultured on Sabouraud agar for 48hrs.

Samples were randomly analyzed using REIMS using the instrumental parameters as shown in Table 1.

**Table S-5.** Cross-validation results for the dataset shown in Figure 3 as a function of the bin size.

| Bin size | Cross-Validation results on Species-level (L1o)[a] | Genus-level (L1o) | Gram-level (Lso)[b] |
|---|---|---|---|
| 0.01 Da | 95.9% | 97.8% | 100.0% |
| 0.02 Da | 95.9% | 97.8% | 100.0% |
| 0.05 Da | 96.1% | 98.0% | 100.0% |
| 0.1 Da | 95.9% | 98.3% | 100.0% |
| 0.2 Da | 94.9% | 98.5% | 100.0% |
| 0.5 Da | 94.6% | 98.3% | 99.8% |
| 1 Da | 94.4% | 97.8% | 99.8% |
| 5 Da | 94.6% | 98.0% | 98.8% |

a: L1o, Leave-one-out, b: Lso, Leave-species-out

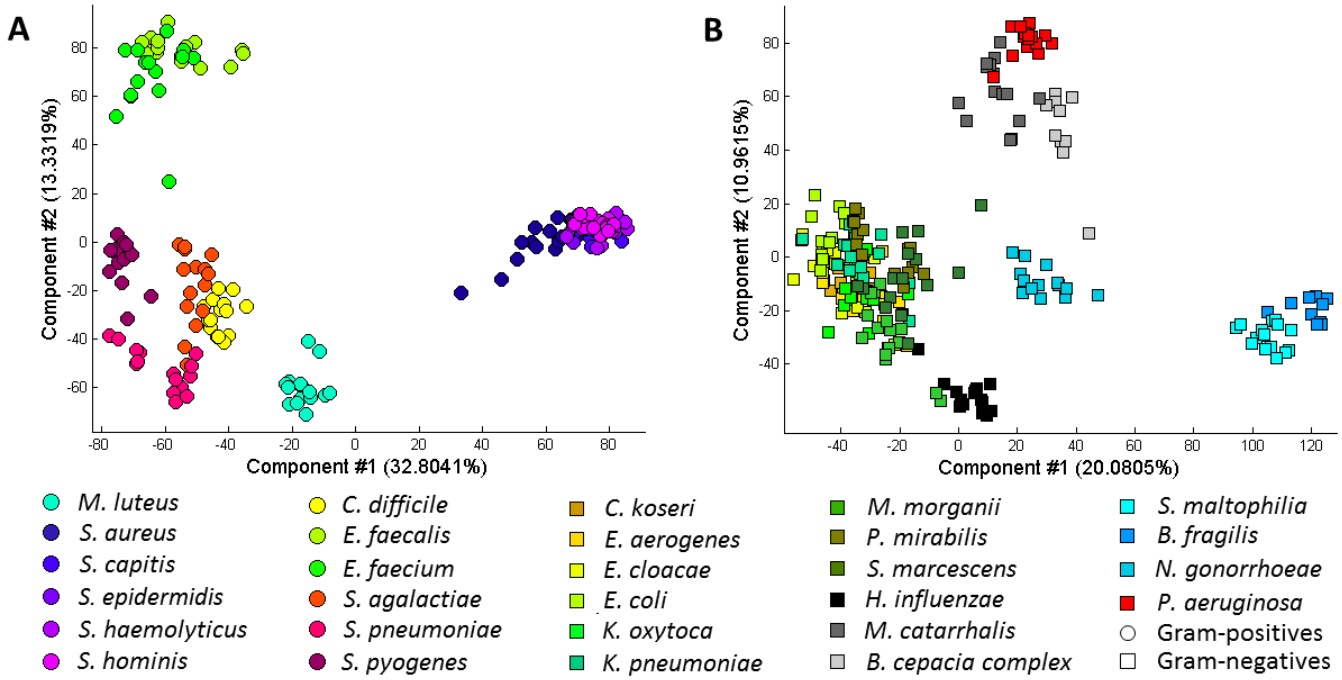


**Figure S-8.** Results of PCA of the dataset shown in Figure 3 for A) Gram-positive species and B) Gram-negative species only.

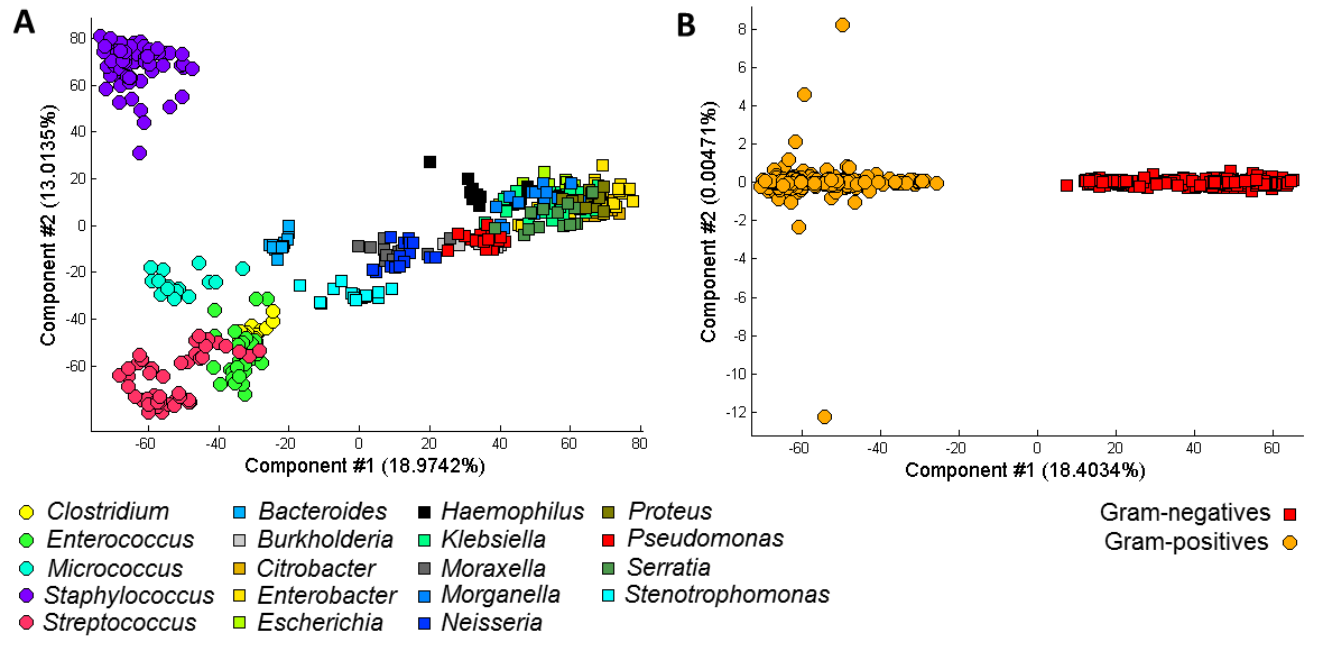


**Figure S-9.** Results of supervised RMMC discriminate analysis of the dataset shown in Figure 3 (main article) on A) Genus-level and B) Gram-level.

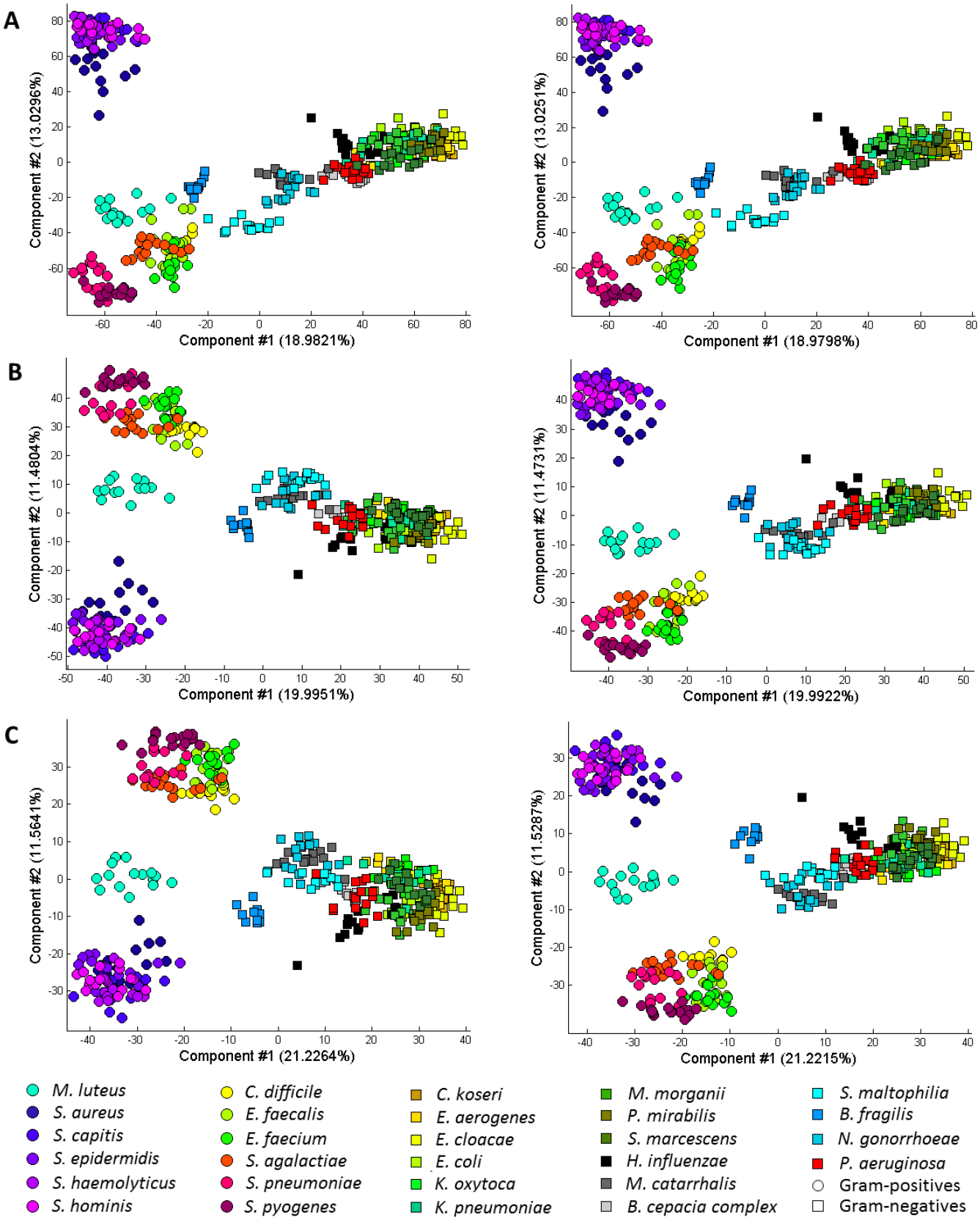


**Figure S-10.** Results of unsupervised (left) and supervised (right) analysis for the dataset shown in Figure 3 using A) 0.01 Da, B) 0.1 Da, and C) 1 Da bin size.

## Inter-Platform Comparison

**Table S-6.** Details of sample set analysed in the cross-platform comparison (Figure 5).

| Gram-stain | Family | Genus | Species | Growth media[a] |
|---|---|---|---|---|
| negative | *Pseudomonadaceae* | *Pseudomonas* | *aeruginosa* | CBA, BAB, CHOC, LB, MCC, ISO, MH, MH+B, TS |
| | *Enterobacteriaceae* | *Escherichia* | *coli* | BHI, BAB, CBA, CHOC, ISO, MH+B, MH, TS |
| | *Moraxellaceae* | *Moraxella* | *catarrhalis* | BHI, CBA, CHOC, MH, MH+B, TS |
| | *Neisseriaceae* | *Neisseria* | *gonorrhoeae* | CBA, CHOC, VCAT, ISON |
| | *Xanthomonadaceae* | *Stenotrophomonas* | *maltophilia* | BHI, CBA, CHOC, MH, MH+B, TS |
| positive | *Staphylococcaceae* | *Staphylococcus* | *aureus* | AZT, BAB, BHI, CBA, ISO, LB, MH, TS |
| | *Streptococcaceae* | *Streptococcus* | *agalactiae* | CBA, ISON |
| | *Clostridiaceae* | *Clostidium* | *difficile* | CBA, Braziers |
| | *Micrococcaceae* | *Micrococcus* | *luteus* | CHOC, BHI, TS, CBA, MH+B, BAB, MH |

[a]Abbreviations: LB = Lysogenic broth agar, BAB = Blood agar base, MCC = McConkey agar, CBA = Columbia blood agar, BHI = Brain-heart infusion agar, TS = Trypticase soy agar, CHOC = Chocolate agar, ISO = Iso-sensitest agar (for antimicrobial susceptibility testing), MH = Mueller-Hinton agar, MH+B = Mueller-Hinton agar containing horse blood, ISON = Iso-sensitest agar with blood and NAD, Braziers = Braziers *Clostridium difficile* selective agar, AZT = Aztreonam blood agar (selective agar for Gram-positives), VCAT = Chocolate agar including Vancomycin, Colistin, Amphotericin and Trimethoprim (selective for *N. gonorrhoeae*)

**Table S-7.** Confusion matrices for blind identification tests using different experimental platforms. Both Orbitrap and Xevo data was classified based on the model recorded on the Exactive instrument.

| Target Class, Exactive | *C. difficile* | *E. coli* | *M. catarrhalis* | *M. luteus* | *N. gonorrhoeae* | *P. aeruginosa* | *S. agalactiae* | *S. aureus* | *S. maltophilia* |
|---|---|---|---|---|---|---|---|---|---|
| *C. difficile* | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *E. coli* | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *M. catarrhalis* | 0 | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 |
| *M. luteus* | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 0 | 0 |
| *N. gonorrhoeae* | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 0 |
| *P. aeruginosa* | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 0 |
| *S. agalactiae* | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 0 |
| *S. aureus* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 |
| *S. maltophilia* | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 8 |

Predicted Class, Xevo

| Target Class, Exactive | *C. difficile* | *E. coli* | *M. catarrhalis* | *M. luteus* | *N. gonorrhoeae* | *P. aeruginosa* | *S. agalactiae* | *S. aureus* | *S. maltophilia* |
|---|---|---|---|---|---|---|---|---|---|
| *C. difficile* | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *E. coli* | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *M. catarrhalis* | 0 | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 |
| *M. luteus* | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 0 | 0 |
| *N. gonorrhoeae* | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 0 |
| *P. aeruginosa* | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 0 |
| *S. agalactiae* | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 0 |
| *S. aureus* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 |
| *S. maltophilia* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 |

Predicted Class, Orbitrap

Ten different strains belonging to a subset of nice species of bacteria named in table S2 were randomly analysed on both the Orbitrap Discovery and the Xevo G2S instrument using instrument parameters as shown in Table S-1. The bacterial strains were non-identical with those used for model building using data acquired on the Exactive instrument. Several different culturing media were used per species (see Table S-6). Only enriched (blood containing) media were investigated in case of *N. gonorrhoeae* and *S. agalactiae* due to limitations in growth on other media. However, chocolated/non-chocolated and selective agars were included in this study. In addition, for the Xevo G2S instrument, a modified atmospheric pressure interface was used to enable ionization by collision induced dissociation of liquid covered droplets.

*Inter-platform data processing*

Thermo .raw files and Waters .raw folders were converted to .mzXML file format using MSConvert (ProteoWizard 3.0.4601 64-bit). Intensity values of Exactive, Orbitrap and Xevo mass spectra were uniformed (interpolated) to a common mass range of $m/z$ = 200 to $m/z$ = 2000 with 0.01 mass resolution. Mass spectra originating from the same replicates were averaged and subjected to peak picking. Peak lists of averaged mass spectra were normalised to their unit vector and uniformed (binned) to a common mass range of $m/z$ = 200 to $m/z$ = 2000 with 1Da mass resolution. Xevo data was denoised by threshold detection according to Donoho and Johnstone.[11] Binned peak lists were then subjected to multivariate analysis.

**Bacteria vs. Yeast**

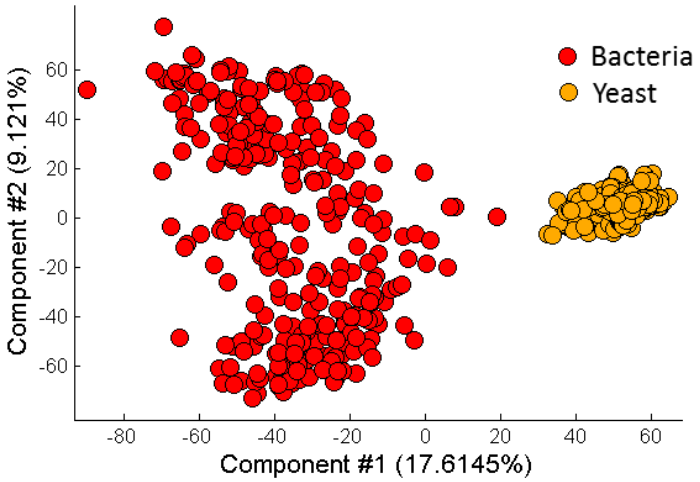


**Figure S-11.** PCA plot generated from a dataset comprising a variety of bacteria and yeast.

Figure S-11 shows results of PCA of 298 strains of bacteria (87 species, both Gram-negative and Gram-positive) and 209 strains of yeasts (16 species).

## References


(1) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M. Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. *Nat Biotechnol* **2012**, *30*, 918-920.

(2) Race, A. M.; Styles, I. B.; Bunch, J. *Journal of Proteomics* **2012**, *75*, 5111-5112.

(3) Veselkov, K. A.; Lindon, J. C.; Ebbels, T. M. D.; Crockford, D.; Volynkin, V. V.; Holmes, E.; Davies, D. B.; Nicholson, J. K. *Anal Chem* **2008**, *81*, 56-66.

(4) Veselkov, K. A.; Vingara, L. K.; Masson, P.; Robinette, S. L.; Want, E.; Li, J. V.; Barton, R. H.; Boursier-Neyret, C.; Walther, B.; Ebbels, T. M.; Pelczer, I.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Anal Chem* **2011**, *83*, 5864-5872.

(5) Li, H., Jiang, T., Zhang, K. *Proc Adv Neural Inf Process Syst* **2003**, 97-104.

(6) Veselkov, K. A.; Mirnezami, R.; Strittmatter, N.; Goldin, R. D.; Kinross, J.; Speller, A. V. M.; Abramov, T.; Jones, E. A.; Darzi, A.; Holmes, E.; Nicholson, J. K.; Takats, Z. *Proceedings of the National Academy of Sciences* **2014**.

(7) Lépine, F.; Milot, S.; Déziel, E.; He, J.; Rahme, L. G. *Journal of the American Society for Mass Spectrometry* **2004**, *15*, 862-869.

(8) G.V.A.O. Costa, S.; Nitschke, M.; Haddad, R.; Eberlin, M. N.; Contiero, J. *Process Biochemistry* **2006**, *41*, 483-488.

(9) Zgoła-Grześkowiak, A.; Kaczorek, E. *Talanta* **2011**, *83*, 744-750.

(10) Song, S. H.; Park, K. U.; Lee, J. H.; Kim, E. C.; Kim, J. Q.; Song, J. *Journal of Microbiological Methods* **2009**, *77*, 165-177.
(11) Donoho, D. L.; Johnstone, I. M. *Comptes rendus de l'Académie des sciences. Série I, Mathématique* **1994**, *319*, 1317-1322.