Automated High-Throughput Identification and Characterisation of Clinically Important Bacteria and Fungi using Rapid Evaporative Ionisation Mass Spectrometry (REIMS)

Frances Bolt (1)†; Simon JS Cameron (1)†; Tamas Karancsi (2); Daniel Simon (2); Richard Schaffer (2); Tony Rickards (3); Kate Hardiman (1); Adam Burke (1); Zsolt Bodai (1); Alvaro Perdones-Montero (1); Monica Rebec (3); Julia Balog (2); Zoltan Takats (1)*.

(1) Section of Computational and Systems Medicine, Department of Surgery and Cancer, Imperial College London, London, SW7 2AZ, United Kingdom; (2) Waters Research Centre, 7 Zahony Street, Budapest, 1031, Hungary; (3) Department of Microbiology, Imperial College Healthcare NHS Trust, Charing Cross Hospital, London, W6 8RF, United Kingdom

† Joint first authors / Contributed equally

* Corresponding Author: Professor Zoltan Takats, Division of Computational and Systems Medicine, Department of Surgery and Cancer, Imperial College London, South Kensington Campus, Sir Alexander Fleming Building, London, SW7 2AZ. Telephone: 020 7594 2760. Email: z.takats@imperial.ac.uk

Abstract

Rapid Evaporative Ionisation Mass Spectrometry (REIMS) has been shown to quickly and accurately speciate microorganisms based upon their species-specific lipid profile. Previous work by members of this group showed that the use of a handheld bipolar probe allowed REIMS to analyse microbial cultures directly from culture plates, without any prior preparation. However, this method of analysis would likely be unsuitable for a high-throughput clinical microbiology laboratory. Here, we report on the creation of a customised platform which enables automated, high-throughput REIMS analysis, which requires minimal user input and operation; and suitable for use in clinical microbiology laboratories. The ability of this high-throughput platform to speciate clinically important microorganisms was tested through the analysis of 375 different clinical isolates, collected from distinct patient samples, from 25 microbial species. After optimisation of our data analysis approach, we achieved substantially similar results between the two REIMS approaches. For handheld bipolar probe REIMS a speciation accuracy of 96.3% was achieved, whilst for high-throughput REIMS, an accuracy of 93.9% was achieved. Thus, high-throughput REIMS offers an alternative mass spectrometry based method for the rapid and accurate identification of clinically important microorganisms in clinical laboratories without any pre-analysis preparative steps.

Introduction

The introduction of mass spectrometry based microbial identification has revolutionised clinical microbiology laboratories. Historically, species delineations were often reliant on phenotypic and biochemical characteristics, which were laborious, costly, and often extended the time required for results by days [1-3]. Whilst mass spectrometry platforms including the MALDI Biotyper (Bruker) and VITEK MS (bioMérieux) have reduced the time to identification, they do require the addition of a matrix to assist in lysis and ionisation and, in some instances, additional extraction steps are required for accurate species classification [2,4].

Rapid evaporative ionisation mass spectrometry (REIMS) provides accurate speciation of bacteria and yeasts directly from colonies without the need for additional preparative steps [5,6]. In contrast to many of the commercially available instruments, which are based upon the fingerprinting of proteins within the mass range 2−20 kDa [7], REIMS allows the lipidomic profile of bacteria and fungi to be determined. The use of lipid analysis for microbial taxonomic classifications is well established and, gas chromatography fatty acid profiling, has been widely used for over fifty years [8-12]. Lipids were also important in differentiating bacteria using Pyrolysis MS; however, in early studies this technique was limited as only small molecules and fragments of larger molecules were observed due to the high energy ionisation process [13-16]. Fast atom bombardment (FAB-MS) based upon the desorption of phospholipid molecules also allowed the differentiation of bacterial species [17]. More recently, desorption electrospray ionization (DESI) provided strain level resolution on intact bacteria based upon analysis within the lipid region [18].

REIMS works by applying a radiofrequency electrical current to the microorganism, which causes rapid heating of microbial biomass. The resulting vapour, containing gas phase ions of metabolites and structural lipids, is channelled to a mass spectrometer for MS analysis. We have previously

demonstrated that lipidomic spectral profiles obtained using REIMS could be used for the accurate speciation of yeast and 28 clinically important bacteria [6].

To increase throughput and reduce costs, clinical microbiology laboratories have undergone rapid automation. This has been the result of a range of factors, including increased financial pressures, centralisation of laboratory diagnostics, and substantially increased number of specimens [19]. Current methods of automation are increasingly being paired with mass spectrometry platforms for microbial identification and have been shown to reduce the time to identification by approximately 24 hours and improve antibiotic prescribing practices [20]. Due to the high-throughput nature of clinical microbiology laboratories [21,22], the REIMS platform has since been modified using a customised TECAN Freedom EVO instrument integrated with a Pickolo system (SciRobotics, Israel) to enable automated analysis. This provides a single platform for automated colony picking, REIMS analysis, and molecular processing thereby minimising user related errors and thus making it highly applicable to a modern clinical microbiology laboratory's workflow. As described here, the method is simple, requires minimal user input, and can easily be adapted to fit into a full laboratory automated workflow [20]. Furthermore, because the technique requires no pre-analysis preparation, it could further reduce analysis times and improve patient outcomes.

Experimental Section

Culturing of Microbial Isolates

A total of 375 isolates comprising 25 bacterial and fungal species, shown in Supplementary Table S1, were collected from clinical samples routinely processed at the Imperial College Healthcare Trust (London, UK) microbiology diagnostic laboratory. Prior to REIMS analysis, each isolate was cultivated according to the conditions detailed in Table S1. Fifteen isolates were examined for each of the 25 species.

MALDI Biotyper Identification

Microbial species identifications were verified for all isolates using matrix assisted laser desorption ionisation time of flight (MALDI-TOF) mass spectrometry on the Bruker Microflex LT (Bruker Daltonics, Coventry, UK) system, according to the manufacturer's instructions. For bacteria, a single colony of culture was spotted onto a 24-spot steel plate (Bruker Daltonics). After drying, 1 µl of α-cyano-4-hydroxycinnamic acid (HCCA) matrix, (Bruker Daltonics) dissolved in 50% acetonitrile, 47.5% water, and 2.5% trifluoroacetic acid was applied. Once dried, the samples were analysed following manufacturer's instructions. For yeasts, one colony was suspended in 1 mL of 70% ethanol and homogenised. The sample then underwent centrifugation at 15,000 x *g* for 2 minutes and the supernatant was removed. After drying at room temperature for 30 minutes, between 10 µL and 60 µL of 70% formic acid, depending on the size of the culture pellet, was added and mixed through pipetting. Pure acetonitrile was then added in an equal amount to formic acid and incubated at room temperature for 3 minutes. The sample underwent centrifugation at 15,000 x *g* for 3 minutes and 1 µL of supernatant was spotted onto an analysis plate. After drying, the spot was overlaid with 1 µL of HCCA matrix as previously described and allowed to dry before analysis following manufacturer's instructions. Both bacterial and yeast isolates were analysed using the Bruker Biotyper software (version 3.0) and library (version 5.0) for taxonomic classification.

REIMS Analysis using Handheld Bipolar Probe

REIMS analysis with a handheld bipolar probe was performed using a similar protocol to that described earlier[6]. In brief, handheld bipolar probes (Erbe Elektromedizin, Tübingen, Germany) were combined with an ERBE IC 300 electrosurgical generator (Erbe Elektromedizin) used at a 70 W power setting in biopolar mode as a radio frequency alternating current power supply (470 kHz sinusoid). A *ca.* 1.5 m long (inner diameter 1.5 mm, outer diameter 3.2 mm) PTFE tube (Sigma-Aldrich, UK) was used to connect the handheld bipolar probe to the inlet capillary of a Xevo G2-XS Q-ToF mass spectrometer instrument (Waters Corporation, Wilmslow, UK), allowing transfer of the generated aerosol. The vacuum of the mass spectrometer was used for aspiration of the aerosol and it was co-aspirated, at a flow rate of 0.2 mL/s, with 2-propanol (HPLC Chromasolv grade, Sigma Aldrich) containing Leu-enkephaline (VWR) at a concentration of 10 ng/µL. The aerosol and 2-propanol/leu-enkephaline solution was introduced using a T-piece as shown in Supplementary Figure S1. For each analytical repeat, between one and five colonies, was removed from the culture plate using one side of the handheld bipolar probe. The handheld bipolar probe was then closed and the power supply was triggered using a manual foot switch, which caused the biomass to rapidly evaporate and produce an analyte-containing aerosol. Five individual measurements were performed for each isolate.

High-Throughput REIMS Analysis

To allow for high-throughput REIMS analysis, a commercially available colony picker robot (Freedom EVO 75 (TECAN, Switzerland) equipped with a Pickolo platform (SciRobotics, Israel)) was used. As shown in Figure 1, the user places the agar plates in the rack, (a), and loads the probes in position, (b). High-throughput REIMS uses a modified command script whereby the user initially selects the number of culture plates to be analysed. The robotic arm of the instrument then picks up the first plate and places it on the imaging platform, Figure 1c. The operator is then able to see microbial colonies and select ones of interest, or the software is able to automatically detect individual colonies for analysis.

In this study, only the operator select mode was used for colony identification. After the selection of colonies, a second arm picks up a stainless steel electrode probe as demonstrated by Figure 1d. The bespoke stainless steel probe is a modified version of a 200 µL pure tip (TECAN). Once the probe touches the microbial colony, an ERBE IC 300 electrosurgical generator, set at a 17 W power setting in monopolar mode, applies an electrical current. The rapid heating of the microbial biomass results in the formation of aerosol which is channelled from the probe through a *ca.* 1.5 M long (inner diameter 1.5 mm, outer diameter 3.2 mm) PTFE tube (Sigma-Aldrich) to a Xevo G2-XS Q-ToF instrument (Waters Corporation). The analyte aerosol is co-aspirated with a 10 ng/µL solution of leu-enkephaline in isopropanol as described for handheld bipolar probe REIMS analysis. For each isolate, a total of five individual measurements, were performed with one microbial colony used for each.

Operation of Xevo G2-XS Q-ToF Mass Spectrometer

The Xevo G2-XS Q-ToF instrument was operated under the instrument settings detailed in Supplementary Table S2 for both handheld bipolar probe REIMS and high-throughput REIMS analysis. Prior to its use each day, the mass spectrometer was calibrated using sodium formate in negative electrospray ionisation mode, following the manufacturer's standard instructions. All mass spectrometry analysis was performed in negative ionisation mode across the 50 to 2500 *m/z* range.

Data Analysis

All acquired raw mass spectrometry files were imported into the Offline Model Builder (OMB) software (Waters Research Center, Budapest, Hungary) for preprocessing and multivariate analysis. First the data scans were combined into spectra, resulting in five individual replicate spectra of each strain. The spectra were background subtracted, and lockmass corrected on external compound Leucin-encephalin (negative ionisation = *m/z* 554.2516). After lockmass correction, spectra were TIC normalised and rebinned to 0.1 Da. After preprocessing, the data was subjected to multiple multivariate algorithms. First a principle component analysis (PCA) was calculated, and using the first

50 principle components, a supervised linear discriminant analysis (LDA) was calculated for classification keeping the first 15 discriminant components. Both PCA and LDA attempt to identify linear combinations of variables within a data set. PCA is an unsupervised data analysis approach which does not take into account class seperation of samples [23]. LDA however, is a supervised approach which explicity looks to identify linear relationships between variables taking into account class groupings of samples [24]. Here, we initially used PCA modelling to reduce the dimensionality of our data set, and then used LDA to identify relationships between variables to form the basis of classification models. The resulting mutivariate 'models' were validated with leave-one isolate out cross-validation, i.e. each data file containing spectra acquired from one isolate was left out of the training set, a model was built on all other datasets and the data file left out was classified using the training model. Each spectrum was classified to the closest class in the LDA space based on the Mahalanobis distances between objects [25]. The correct classification rate was calculated based on the number of spectra classified correctly divided by all spectra of the full dataset. For all species analysis, a model built using all isolates from 25 species was created and used in cross-validation. For taxonomic groupings analysis, individual models were constructed for distinct groups at each taxonomic hierarchy, namely phylum, class, order, family, genus, and species. A probability matrix was then constructed to calculate the likely probability that an isolate would be correctly identified to its taxonomic grouping at each level, shown on Supplementary Figure S2. For Random Forest analyses, a total of 400 decision trees were constructed for each model based on a mass spectral matrix exported from the OMB software after data normalisation, background removal, and lock-massing to leu-enkaphaline. For each model, a mass range of 600 to 900 *m/z* was used, with mass binning to 0.1.

Safety considerations

In this work, all microorganisms were treated as Hazard Group 2 organisms and were therefore manipulated within containment level 2 facilities. All cultures and REIMS analyses were performed

within a class 2 biological safety cabinet. All solvents, such as 2-propanol and methanol, were handled according to the material safety data sheet provided by their respective manufacturer.

Results and Discussion

The use of REIMS for the successful classification and characterisation of bacteria and yeasts has previously been reported [5,6]. However, this approach employed a handheld surgical tool (electrosurgical bipolar forceps) for analysis and thus may not be suitable for use in a high-throughput, clinical microbiology laboratory. Here, an automated instrument incorporating a colony picker robot and a mass spectrometer was developed to allow for high-throughput REIMS processing. This study explored its use for the identification of microbes by analysing a taxonomically diverse collection of bacteria and fungi. A total of 15 isolates from 25 species were analysed using both REIMS approaches. This represents a total of 375 distinct clinical isolates of which 8% were fungi, 52% were Gram-positive bacteria, and 40% Gram-negative bacteria. All of the species examined are commonly encountered within a clinical diagnostic laboratory. Some species, including *Lactobacillus jensenii*, are frequently isolated but rarely cause disease whilst others, including *Escherichia coli* and *Staphylococcus aureus*, more frequently result in severe infections.

Comparison Between Handheld Bipolar and High-Throughput Monopolar Probe REIMS

Similar classification accuracies were obtained at Gram-stain, genus, and species level using both manual and high-throughput automated REIMS approaches. However, high-throughput REIMS showed a slightly lower (94% to 96%), genus and species level accuracy compared to handheld bipolar probe REIMS. As Figure 2 demonstrates, this may result from differences in the temperature gradient during the rapid heating of the bipolar (handheld) and monopolar (high-throughput) probe. Because the handheld bipolar probe generates a radiofrequency electrical current between two probe surfaces, the temperature is lower, but more consistent through the microbial biomass, compared to the monopolar probe. The latter may result in reduced ionisation of some microbial metabolites due to the higher localised temperature at the point of contact between the electrical probe and microbial biomass. This may in part, explain the slightly reduced level of genus and species level classification accuracy observed. Optimisation of the high-throughput REIMS's monopolar probe will be pursued to

modify the temperature at the point of contact to improve biomass ionisation and thus species-level accuracy.

Although the high-throughput REIMS approach has a slightly reduced level of genus and species-level accuracy, it is capable of analysing substantially higher numbers of microbial colonies, with reduced variability, and user input. At present, this method is potentially capable of analysing approximately 3,000 to 4,000 microbial colonies over a 24 hour period. The increased analysis throughput, reduced variability, and user input makes the high-throughput REIMS approach, reported here, well suited to implementation in a clinical microbiology laboratory.

Effect of Mixing Analyte Aerosol with Solvent Matrix

The previously reported handheld bipolar REIMS method did not incorporate a solvent matrix within the analyte aerosol. Here, we used 2-propanol as a solvent matrix based upon preliminary data which revealed that it improves detectability (i.e. by the incremental improvement of MS response factor) of spectral features. Furthermore, it reduces the contamination of the mass spectrometer by removing all solid deposits from the capillary inlet of the instrument and allows an external lock mass compound to be used so that mass drift correction can be applied. This ability would be essential for a platform introduced into a clinical microbiology laboratory, since there is no universal microbial marker compound which could be used as a lock mass reference similarly to human tissue REIMS analysis [26].

Supplementary Figure S3 shows the Gram-stain, genus, and species-level cross-validation scores for both handheld bipolar probe REIMS and high-throughput REIMS with and without 2-propanol infusion. Although Gram-stain accuracy is comparable, there is a marked reduction in genus and species-level accuracy when isolates are analysed using both handheld bipolar probe REIMS and high-throughput REIMS without 2-propanol infusion.

Initial All Species Analysis

Initial data analysis evaluated the entire acquisition range of 50 to 2500 m/z, using a mass bin of 1. An LDA model was created using all 375 isolates and CV scores were determined from this for both REIMS approaches (shown in Figure 3). For handheld bipolar probe REIMS, cross-validation scores to Gram-stain, genus, and species levels were 99.2%, 94.1%, and 88.0% respectively. Cross-validation scores were 98.7%, 91.5%, and 85.6% for each taxonomic level for high-throughput REIMS. Supplementary Table S2 shows the individual species cross-validation scores and suggests that the overall cross-validation score is being reduced substantially by poor scores associated with closely related species, such as members of the *Staphylococcus* genus, and members of the Enterobacteriaceae family.

Comparison of Mass Spectral Ranges

The species-level accuracy of both handheld bipolar probe REIMS and high-throughput monopolar REIMS using an all species analysis model was substantially lower than previously reported accuracies for bipolar REIMS [5,6]. To explore possible optimisations to our data analysis approach, we created eight different all species models using restricted mass ranges of 50 to 2500 *m/z*, 150 to 2500 *m/z*, 250 to 2500 *m/z*, 500 to 2500 *m/z*, 600 to 900 *m/z*, 1000 to 2000 *m/z*, and 1200 to 1800 *m/z*. Figure 4 shows that through a reduction in mass range from 50 to 2500 *m/z* to 600 to 900 *m/z*, species-level accuracy increases from 85% to 90% for handheld bipolar probe REIMS and from 86% to 90% for high-throughput REIMS. This mass range commonly contains ionised lipids [26-28], which prior to the introduction of advanced biochemical, genetic, and mass spectrometric analyses, have historically been used for species-level identification [29-31]. In addition to the improved species-level accuracy, analysis using a reduced mass range could also increase the method's throughput rate. The reduced mass range of 600 to 900 *m/z* equals approximately 12% of the entire acquisition mass range of the Xevo G2-XS Q-ToF. This would correspond to a similar level of reduction in the computational

315 requirements of data analysis and allow for real-time species-level identification, as has been shown
316 in other applications of REIMS including intra-operative tissue assessment [26].
317
318 **Effect of Mass Bin on Species Cross-Validation**
319 We also assessed the effect of bin size on the classification performance of the statistical models. For
320 both REIMS approaches, the mass bin size appeared to have minimal effect on species-level cross-
321 validation accuracy, Figure 5. At mass resolutions below 1, namely 0.01, 0.05, 0.1, and 0.5, no
322 consistent result was visible between the two REIMS approaches. For example, the 0.05 mass bin
323 achieved the highest species accuracy for high-throughput REIMS out of all six bin sizes evaluated,
324 whilst for handheld bipolar probe REIMS it achieved the lowest. In agreement with our earlier studies
325 [5,6], we utilised unit resolution for all subsequent analyses as it achieves a balance between good
326 species level accuracy and computational time. For each tenfold decrease in bin size (for example from
327 1 to 0.1) the number of variables within each spectrum is increased ten-fold, along with the
328 computational time required for analysis. To minimise the time required for bioinformatic analysis
329 and species-level identification, using a mass bin of one will substantially reduce the computational
330 time and resources required; allowing near real-time species-level identification in a clinical
331 laboratory.
332
333 **Species Cross-Validation Using Taxonomy Grouped Models**
334 Members of closely-related taxonomic groups, such as the Enterobacteriaceae family or *Streptococcus*
335 genera, are likely to share substantial similarities in their mass spectral profiles because of the lower
336 level of evolutionary divergence. Therefore, determining species-level cross-validation accuracy using
337 an approach which combines all taxonomic species into one classification model, as described above,
338 may suffer from misclassifications between closely related species. Here, we have explored an
339 alternative approach whereby individual models are created for each taxonomic group at the domain,
340 phylum, class, order, family, genus, and species levels. As Supplementary Figures S4 and S5

demonstrate, this allows for a probability matrix to be constructed to calculate the likelihood that an isolate from each of the 25 species analysed in this study, would be correctly classified within each model of its relevant taxonomic nodes. Using this approach, species-level accuracy was increased from 90.40% to 96.17% for handheld bipolar probe REIMS and from 89.87% to 94.03% for high-throughput REIMS. It has to be noted that hierarchical classification systems have also been employed for species-level identification of bacterial mass spectrometry fingerprints from MALDI-ToF based identification platforms [32].

Species Cross-Validation Using Random Forest Models

To this point, REIMS spectral fingerprints have been analysed using a combined PCA/LDA approach. This approach is beneficial as it reduces the dimensionality of data sets and provides good predictive power. However, it is more susceptible than other methods (e.g. Random Forest) to over training and over fitting and has a poor ability to rank variables according to their discriminatory importance [33]. The Random Forest algorithm is a machine learning algorithm based on decision making trees and was originally designed to handle high-dimensional, non-linear ecological data [34], but has found considerable application to mass spectrometry data sets as well [35]. Here. Random Forest constructs decision making trees using mass spectral bins, at 0.1 Da, to construct classification models which are internally validated through the splitting of samples into training and test data sets. Through the construction of hundreds of decision making trees, the importance of individual variables is determined; allowing for a final model using the most discriminatory variables to be constructed. The classification accuracy of this model is then determined using leave one sample out cross-validation.

In the present study we used Random Forest to construct a total of 400 decision making trees to assign species-level taxonomy to all 15 isolates from 25 species. Through this approach, we were able to achieved species-level accuracy of 96.27% and 93.87% to handheld bipolar probe REIMS and monopolar high-throughput REIMS respectively. These results were comparable to those achieved

using taxonomy grouped models, albeit in substantially reduced computational time. The confusion matrices for handheld bipolar probe REIMS, Figure 6a, and for high-throughput REIMS, Figure 6b, show limited misclassification, predominately within the Enterobacteriaceae family and *Staphylococcus* genus. Whilst the classification accuracies for some members of the *Staphylococcus* genus and Enterobacteriaceae group are lower than those described using the Bruker Biotyper system, further platform optimisations, targeting these taxonomic groups, will likely improve classification accuracies substantially [36-38]. Random Forest may have particular utility in a high-throughput diagnostic environment due to its rapid computational ability. Table 1 gives classification accuracy rates for each of the 25 microbial species analysed using both REIMS approaches using three bioinformatic approaches.

Intra-Sample Variation

To assess the intra-sample variability in handheld bipolar probe REIMS and monopolar high-throughput REIMS analysis, we completed species-level cross-validation using each of the five analysis points from each isolate for each REIMS approach. Results were compared against the species-level cross-validation for the combined mass spectra for the five analysis points, as shown in Supplementary Figure S6. Theoretically, the difference between the species-level accuracy of these two approaches would be positively correlated to the level of intra-sample variation. Thus, the greater the difference between the two cross-validation figures, the greater the level of intra-sample variation. As shown in Supplementary Figure S7, minimal differences were observed between the species-level accuracy of the cross-validation. Within the 50 to 2500 *m/z* range a difference of two percentage points was observed for the handheld bipolar probe REIMS approach, and less than one percentage point for the high-throughput REIMS approach. In regards to the 600 to 900 *m/z* range, a difference of two percentage points was observed for handheld bipolar probe REIMS, and one percentage point for high-throughput REIMS. Although not measured directly, the use of leucine enkaphline as an external lock mass compound, allowing for correction of the mass drift associated with time of flight MS, will

minimise the extent of variation between analysis periods. In addition, this work was completed over multiple analysis days, using two different instrument operators. As evident with the reduced intra-sample variation of the high-throughput REIMS platform, the nature of the analytical set up will also likely reduce any potential inter-sample variation introduced as a result of different analysis day or operator.

Conclusions

Here, we have shown that the initial species-level identification and characterisation previously reported by members of this group using bipolar REIMS technology can be comparably achieved using an automated, high-throughput REIMS approach. Both bacterial and fungal isolates are analysed directly from culture, without any prior preparative steps. The method described requires minimal user input and can easily be integrated with automated clinical microbiology workflows. Although not employed here, based on colony analysis time of 15 seconds, this high-throughput workflows allows approximately 3,000 to 4,000 microbial colonies to be analysed over a 24 hour period, including time required for instrument calibration, maintenance, and cleaning, with species-level identifications given in near real-time. Although not directly comparable as the bioinformatic pipeline for clinical microbiology laboratories is in development, this time period is substantially lower than that required for commercial MALDI-ToF platforms, which take approximately six minutes to analyse each isolate [3].

Here, we have discussed commercially available MS-based platforms which are currently used in clinical microbiology laboratories, such as the Bruker BioTyper MALDI-ToF system. These references were used to contextualise our REIMS based platform against familiar MS platforms. Although we used the Bruker BioTyper platform for species level identification of microbial isolates, in line with the standard operation procedures of our collaborating clinical laboratory, we have not drawn direct comparisons between its identification accuracy and that of high-throughput REIMS. The Gram-stain, genus, and species level classifications accuracies reported are used primarily as a tool to compare the handheld and high-throughput REIMS approaches. They could however, be considered a percentage conformity to the Bruker BioTyper platform. To determine the true comparable classification accuracies of the two platforms, and other commercially available ones, a non-MS species identification method, such as 16S/18S/ITS rRNA gene sequencing would be required; which was not conducted.

Although we have focussed on the lipid region of the mass spectra, additional microbial metabolites identified using the REIMS procedure could have utility in defining phenotypic features including virulence and antibiotic resistance. Furthermore, work is currently underway to extend high-throughput REIMS analysis from pure microbial cultures to analysis of mixed microbial cultures and communities, and complex human bio-fluids including urine and blood using species-specific features for identifications.

Supporting Information

Detailed information regarding bacterial and fungal species analysed in this study, and on additional data analyses completed as referenced within this text is given within the supporting information accompanying this manuscript. This material is available free of charge via the Internet at http://pubs.acs.org.

Author Contributions

Study was planned by ZT, FB, and SJSC. Experiments were conducted by AB, TR, KH, FB, and SJSC. Data was analysed and interpreted by SJSC, FB, ZT, and APM. Technical assistance and method optimisation was provided by ZB, MR, JB, TK, DS, RS, and ZT. The manuscript was written by FB, SJSC, and ZT with input from all authors. All authors have given approval to the final version of the manuscript.

Conflict of Interest Statement

This work was funded by a Biotechnology and Biological Sciences Research Council (BBSRC) grant (reference number BB/L020858/1). Additional support, both technical and financial, was provided by Waters Corporation. TK, DS, RS, and JB are employed by the Waters Corporation. ZT consults for Waters Corporation. The work detailed in this manuscript does not promote any available commercial product from Waters Corporation.

Acknowledgments

The authors would like to thank Imperial College Healthcare Trust for providing isolates and their technical input. We thank the anonymous reviewers for their constructive comments which have made this a better manuscript.

References

(1) Clark, A. E.; Kaleta, E. J.; Arora, A.; Wolk, D. M. *Clin. Microbiol. Rev.* **2013**, *26*, 547-603.

(2) Tan, K. E.; Ellis, B. C.; Lee, R.; Stamper, P. D.; Zhang, S. X.; Carroll, K. C. *J. Clin. Microbiol.* **2012**, *50*, 3301-3308.

(3) Seng, P.; Drancourt, M.; Gouriet, F.; La Scola, B.; Fournier, P. E.; Rolain, J. M.; Raoult, D. *Clin. Infect. Dis.* **2009**, *49*, 543-551.

(4) Cassagne, C.; Cella, A.-L.; Suchon, P.; Normand, A.-C.; Ranque, S.; Piarroux, R. *Med. Mycol.* **2013**, *51*, 371-377.

(5) Strittmatter, N.; Jones, E. A.; Veselkov, K. A.; Rebec, M.; Bundy, J. G.; Takats, Z. *Chem. Commun.* **2013**, *49*, 6188-6190.

(6) Strittmatter, N.; Rebec, M.; Jones, E. A.; Golf, O.; Abdolrasouli, A.; Balog, J.; Behrends, V.; Veselkov, K. A.; Takats, Z. *Anal. Chem.* **2014**, *86*, 6555-6562.

(7) Ryzhov, V.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 746-750.

(8) Purcaro, G.; Tranchida, P. Q.; Dugo, P.; Camera, E. L.; Bisignano, G.; Conte, L.; Mondello, L. *J. Sep. Sci.* **2010**, *33*, 2334-2340.

(9) Buyer, J. S. *J. Microbiol. Methods* **2006**, *67*, 187-190.

(10) Abel, K.; deSchmertzing, H.; Peterson, J. I. *J. Bacteriol.* **1963**, *85*, 1039-1044.

(11) Miller, L. T. *J. Clin. Microbiol.* **1982**, *16*, 584-586.

(12) Welch, D. F. *Clin. Microbiol. Rev.* **1991**, *4*, 422-438.

(13) DeLuca, S.; Sarver, E. W.; Harrington, P. d. B.; Voorhees, K. J. *Anal. Chem.* **1990**, *62*, 1465-1472.

(14) Anhalt, J. P.; Fenselau, C. *Anal. Chem.* **1975**, *47*, 219-225.

(15) Basile, F.; Voorhees, K. J.; Hadfield, T. L. *Appl. Environ. Microbiol.* **1995**, *61*, 1534-1539.

(16) Basile, F.; Beverly, M. B.; Voorhees, K. J.; Hadfield, T. L. *TrAC, Trends Anal. Chem.* **1998**, *17*, 95-109.

(17) Heller, D. N.; Cotter, R. J.; Fenselau, C.; Uy, O. M. *Anal. Chem.* **1987**, *59*, 2806-2809.

(18) Song, Y.; Talaty, N.; Tao, W. A.; Pan, Z.; Cooks, R. G. *Chem. Commun.* **2007**, 61-63.

(19) Bourbeau, P. P.; Ledeboer, N. A. *J. Clin. Microbiol.* **2013**, *51*, 1658-1665.

(20) Mutters, N. T.; Hodiamont, C. J.; de Jong, M. D.; Overmeijer, H. P.; van den Boogaard, M.; Visser, C. E. *Ann. Lab. Med.* **2014**, *34*, 111-117.

(21) Greub, G.; Prod'Hom, G. *Clin. Microbiol. Infect.* **2011**, *17*, 655-660.

(22) Novak, S. M.; Marlowe, E. M. *Clin. Lab. Med.* **2013**, *33*, 567-588.

(23) Wold, S.; Esbensen, K.; Geladi, P. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37-52.

(24) Fisher, R. A. *Ann. Hum. Genet.* **1938**, *8*, 376-386.

(25) Mahalanobis, P. C. *Proc. Natl. Inst. Sci. India* **1936**, *2*, 49-55.

(26) Balog, J.; Sasi-Szabó, L.; Kinross, J.; Lewis, M. R.; Muirhead, L. J.; Veselkov, K.; Mirnezami, R.; Dezső, B.; Damjanovich, L.; Darzi, A. *Sci. Transl. Med.* **2013**, *5*, 194ra193-194ra193.

(27) Pulfer, M.; Murphy, R. C. *Mass Spectrom. Rev.* **2003**, *22*, 332-364.

(28) Gibson, B.; Melaugh, W.; Phillips, N. J.; Apicella, M.; Campagnari, A.; Griffiss, J. *J. Bacteriol.* **1993**, *175*, 2702-2712.

(29) Komagata, K.; Suzuki, K.-I. *Methods Microbiol.* **1987**, *19*, 162-207.

(30) Lüderitz, O.; Galanos, C.; Lehmann, V.; Nurminen, M.; Rietschel, E. T.; Rosenfelder, G.; Simon, M.; Westphal, O. *J. Infect. Dis.* **1973**, *128*, S17-S29.

(31) van Baar, B. L. *FEMS Microbiol. Rev.* **2000**, *24*, 193-219.

(32) Simmuteit, S.; Schleif, F.-M.; Villmann, T.; Hammer, B. *Knowl. Inf. Syst.* **2010**, *25*, 327-343.

(33) Gromski, P. S.; Muhamadali, H.; Ellis, D. I.; Xu, Y.; Correa, E.; Turner, M. L.; Goodacre, R. *Anal. Chim. Acta* **2015**, *879*, 10-23.

(34) Cutler, D. R.; Edwards Jr, T. C.; Beard, K. H.; Cutler, A.; Hess, K. T.; Gibson, J.; Lawler, J. J. *Ecology* **2007**, *88*, 2783-2792.

(35) Patterson, A. D.; Bonzo, J. A.; Li, F.; Krausz, K. W.; Eichler, G. S.; Aslam, S.; Tigno, X.; Weinstein, J. N.; Hansen, B. C.; Idle, J. R. *J. Biol. Chem.* **2011**, *286*, 19511-19522.

(36) Dubois, D.; Leyssene, D.; Chacornac, J. P.; Kostrzewa, M.; Schmit, P. O.; Talon, R.; Bonnet, R.; Delmas, J. *J. Clin. Microbiol.* **2010**, *48*, 941-945.

(37) Saffert, R. T.; Cunningham, S. A.; Ihde, S. M.; Jobe, K. E. M.; Mandrekar, J.; Patel, R. *J. Clin. Microbiol.* **2011**, *49*, 887-892.

(38) Zhu, W.; Sieradzki, K.; Albrecht, V.; McAllister, S.; Lin, W.; Stuchlik, O.; Limbago, B.; Pohl, J.; Rasheed, J. K. *J. Microbiol. Methods* **2015**, *117*, 14-17.

582 **Figure Captions**

583

584 **Figure 1. Customised TECAN Freedom EVO automated sample handling machine and sampling probe**

585 A customised TECAN Freedom EVO automated sample handling machine is contained within a class 2
586 microbiological safety cabinet. Before use, agar culture plates to be placed into a plate rack (a),
587 sampling probes to be loaded into a 96 rack (b). Plates are then moved from the plate rack to the
588 Pickolo visualisation platform (c), which also acts as the return electrode, where the user or the
589 automated system can select colonies for analysis. A technical drawing of the stainless steel sampling
590 probe is given in (d).

591

592 **Figure 2. Temperature gradient differences between REIMS approaches**

593 The handheld bipolar probe (a) employed in this study generates a radiofrequency electrical current
594 between two probes, and thus has a lower, but more even, heating temperature throughout the
595 microbial biomass. However, the monopolar probe used in high-throughput REIMS (b) generates a
596 higher, localised heating temperature at the point of contact between probe and microbial biomass.
597 Counter electrode shown by grey bar at base of microbial biomass. Colour gradient of microbial
598 biomass indicates heating temperature, with red showing highest temperatures and green lowest
599 temperatures. The mass spectra beside each representative diagram is of the 600 to 900 *m/z* range of
600 the same *Lactobacillus jensenii* isolate after background removal for each REIMS approach. Figure is
601 not constructed to scale.

602

603 **Figure 3. Gram-stain, genus, and species-level cross-validation accuracy for both REIMS approaches**

604 Cross-validation using a PCA/LDA model over the 50 to 2500 *m/z* range with a mass bin of 1 gives
605 similar Gram-stain level accuracy for both handheld bipolar probe REIMS and high-throughput REIMS.
606 However, at genus and species-level accuracy, handheld bipolar probe REIMS gives slightly higher
607 accuracy levels, by three and two percentage points respectively.

Figure 4. Species-level classification accuracy using restricted mass ranges

A total of eight PCA/LDA models were constructed using a mass bin of 1 for both handheld bipolar probe REIMS and high-throughput REIMS using restricted mass ranges. For handheld bipolar probe REIMS, the highest species-level accuracy was achieved through using the 500 to 1500 *m/z* range and 600 to 900 *m/z* range, whilst for high-throughput REIMS the highest accuracy was achieved with the 600 to 900 *m/z* range.

Figure 5. Effect of mass bin size on species-level accuracy

Species-level accuracy was measured through cross-validation of PCA/LDA models constructed using the 600 to 900 *m/z* range for mass bins of 0.01, 0.05, 0.1, 0.5, 1.0, and 5.0. For both handheld bipolar probe REIMS and high-throughput REIMS, minimal differences were evident between different mass bin sizes. To achieve a balance between good species-level accuracy and reduced computational time for further analyses, a mass bin size of 1 was subsequently used.

Figure 6. Species-Level confusion matrices from Random Forest analysis

Random Forest analysis was conducted on (a) handheld bipolar probe REIMS and (b) high-throughput REIMS datasets within the 600 to 900 *m/z* range, using a mass bin of 0.1. Random Forest models were constructed using 400 decision making trees. Confusion matrices suggest a marginally higher level of misclassification using the high-throughput REIMS approach than with handheld bipolar probe REIMS, with misclassifications occurring predominately within the Enterobacteriaceae family. Colour scale gradient indicates strength of species-level accuracy with darker gradients showing higher species-level accuracy. Abbreviations used within the True Class and Predicted Class labels are given alongside full species names in Table 1.

Table 1. Individual species-level accuracy for all species, taxonomic groupings, and Random Forest models

The species-level accuracy for each of the 25 microbial species analysed using both REIMS approach is given for each of the three bioinformatic data analysis methods used in this study. For handheld bipolar probe REIMS, the highest overall species-level accuracy was achieved using the Random Forest approach, whilst for high-throughput REIMS the highest was achieved using the taxonomic groupings approach. Although, in both cases, the differences between the taxonomic groupings and Random Forest approaches is less than 0.2 percentage points.

662 **Table 1**

	Isolate Number	Handheld Bipolar Probe REIMS (% Accuracy)			High-Throughput REIMS (% Accuracy)		
		All Species	Taxonomic Groupings	Random Forest	All Species	Taxonomic Groupings	Random Forest
Average	∑ 375	90.4	96.2	96.3	89.9	94.0	93.9
***Candida albicans* (CALB)**	15	100.0	100.0	100.0	100.0	100.0	100.0
***Candida parapsilosis* (CALP)**	15	100.0	100.0	100.0	93.3	100.0	100.0
***Corynebacterium amycolatum* (CAMY)**	15	100.0	100.0	100.0	100.0	97.0	100.0
***Micrococcus luteus* (MLUT)**	15	100.0	100.0	100.0	86.7	97.0	100.0
***Enterococcus faecalis* (EFAC)**	15	60.0	91.2	100.0	86.7	77.6	87.0
***Enterococcus faecium* (EFAM)**	15	66.7	91.2	83.0	86.7	84.4	80.0
***Lactobacillus jensenii* (LACJ)**	15	86.7	98.0	100.0	100.0	93.0	86.0
***Streptococcus agalactiae*(SAGA)**	15	93.3	91.2	100.0	93.3	86.5	87.0
***Streptococcus pneumoniae* (SPNE)**	15	100.0	98.0	97.0	93.3	93.0	97.0
***Streptococcus pyogenes* (SPYO)**	15	100.0	98.0	97.0	93.3	93.0	93.0
***Staphylococcus aureus* (SAUR)**	15	100.0	98.0	93.0	93.3	93.0	85.0
***Staphylococcus epidermidis* (SEPI)**	15	80.0	91.2	100.0	73.3	100.0	97.0
***Staphylococcus haemolyticus* (SHAM)**	15	73.3	78.4	100.0	73.3	80.0	100.0
***Staphylococcus hominis* (SHOM)**	15	86.7	91.2	97.0	73.3	100.0	97.0
***Clostridium difficile* (CDIF)**	15	100.0	99.0	100.0	100.0	100.0	100.0
***Enterobacter cloacae* (ECLO)**	15	100.0	99.0	93.0	86.7	92.1	100.0
***Escherichia coli* (ECOL)**	15	46.7	92.1	97.0	60.0	79.2	93.0
***Klebsiella oxytoca* (KOXY)**	15	100.0	99.0	97.0	86.7	99.0	97.0
***Klebsiella pneumoniae* (KPNE)**	15	86.7	99.0	80.0	86.7	99.0	84.0
***Morganella morganii* (MMORG)**	15	86.7	99.0	97.0	93.3	99.0	90.0
***Proteus mirabilis* (PMIR)**	15	100.0	92.1	87.0	93.3	99.0	83.0
***Serratia marcescens*(SMAR)**	15	93.3	99.0	100.0	93.3	92.1	100.0
***Haemophilus influenza* (HINF)**	15	100.0	100.0	100.0	100.0	99.0	100.0
***Pseudomonas aeruginosa* (PAER)**	15	100.0	100.0	93.0	100.0	99.0	93.0
***Stenotrophomonas maltophilia* (SMAL)**	15	100.0	100.0	97.0	100.0	99.0	97.0

663

664 **Table of Contents Graphic (Actual Size)**

665

666

