Check for updates

MicrobiologyOpen

WILEY

MicrobiologyOpen Open Access

ORIGINAL ARTICLE OPEN ACCESS

Discrimination of *Klebsiella pneumoniae* and *Klebsiella quasipneumoniae* by MALDI-TOF Mass Spectrometry Coupled With Machine Learning

Mari Nishikawa[1] | Wenhao Tang[2] | Markus Kostrzewa[3] | Jonah Rodgus[4] | Frances Davies[4,5,6] | Yi Liu[1] | Elita Jauneikaite[4] | Gerald Larrouy-Maumus[1]

[1]Faculty of Natural Sciences, Department of Life Sciences, Centre for Bacterial Resistance Biology, Imperial College London, UK | [2]Faculty of Natural Sciences, Department of Mathematics, Imperial College London, London, UK | [3]Bruker Daltonics GmbH&Co. KG, Bremen, Germany | [4]NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, Department of Infectious Disease, Imperial College London, London, UK | [5]Imperial College Healthcare NHS Trust, London, UK | [6]Department of Microbiology, North West London Pathology, London, UK

Correspondence: Elita Jauneikaite (e.jauneikaite@imperial.ac.uk) | Gerald Larrouy-Maumus (g.larrouy-maumus@imperial.ac.uk)

Received: 31 March 2025 | **Revised:** 10 June 2025 | **Accepted:** 2 July 2025

Funding: This work was supported by the UK Research and Innovation (105603/Z/14/Z) and Brain Tumour Charity (GN-000595).

Keywords: identification | *Klebsiella* | lipids | MALDI

ABSTRACT

Klebsiella species, including *Klebsiella pneumoniae* and *Klebsiella quasipneumoniae*, present significant challenges in clinical microbiology due to their genetic similarity, which complicates accurate species identification using established methods, including matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) on the protein/peptide level. Although the treatment choice for infections caused by these pathogens is often similar, precise species characterization enhances our epidemiological understanding. While whole-genome sequencing can accurately distinguish *Klebsiella* species accurately, those analyses are time-consuming, requiring specialized expertise, and are not currently used in routine clinical laboratories. Therefore, developing a timely and accurate pathogen characterization method is essential for effective treatment, management, and infection control measures. This study combined MALDI-TOF MS in negative ion mode with machine learning techniques to identify potential lipid biomarkers as a novel method to distinguish between *K. pneumoniae* and *K. quasipneumoniae*. Using this method, we identified discriminative features between the species, with peaks at *m/z* 2157, *m/z* 1931, *m/z* 1964, *m/z* 2042, and *m/z* 1407 highlighted as potential biomarkers for species identification. Our findings suggest that the lipid profiles of the species obtained from MALDI-TOF MS can serve as effective biomarkers for distinguishing *Klebsiella* species. Further research should focus on the structural identification of these biomarkers and expand the data set to include more isolates for each of the species. This approach holds promise for developing more cost-effective and rapid diagnostic tools in clinical microbiology, ultimately improving patient outcomes and infection control.

1 | Introduction

Klebsiella species are common Gram-negative bacteria with complex population structures. *Klebsiella pneumoniae* is a major cause of nosocomial infections, including bloodstream infections, urinary tract infections, and pneumonia. Infections are especially prevalent among immunocompromised or critically ill patients, such as those in intensive care units (Hafiz et al. 2023; Choby et al. 2020).

Mari Nishikawa and Wenhao Tang contributed equally to the work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *MicrobiologyOpen* published by John Wiley & Sons Ltd.

In 2014, whole-genome sequencing has identified two new species, namely *Klebsiella quasipneumoniae* and *Klebsiella variicola*, from isolates that were originally classified as *K. pneumoniae* (Holt et al. 2015; Long et al. 2017). Being highly genetically similar to *K. pneumoniae*, *K. quasipneumoniae* isolates are often misidentified as *K. pneumoniae*; indeed, it is hard to distinguish the two species using routine laboratory techniques (Long et al. 2017; Martinez-Romero et al. 2018). Although whole-genome sequencing provides accurate species identification, its time-consuming nature and requirement for expertise limit its routine use in diagnostic laboratories. Multiplex polymerase chain reaction (PCR) can distinguish these species using chromosomally encoded genes (bla_{SHV}, bla_{LEN} and bla_{OKP}) and their flanking gene (*deoR*) (Fonseca et al. 2017); however, its sensitivity is limited especially when detecting low levels of bacterial DNA in clinical samples. Nevertheless, since *Klebsiella* species isolates can possess different virulence and antibiotic resistance profiles (Holt et al. 2015), accurate species identification is essential for our improved understanding of the epidemiological and clinical trends of these two species.

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) is an analytical technique widely used in biochemistry and molecular biology for rapid and accurate determination of molecular masses of biomolecules, such as proteins, peptides, nucleic acids, and carbohydrates, which can be used for species identification. While ribosomal proteins serve as common biomarkers in conventional MALDI-TOF MS, they lack sufficient resolution for closely related species like *Klebsiella*, leading to species misclassification (Long et al. 2017; Solntceva et al. 2020; Dinkelacker et al. 2018). An alternative approach targets lipid fingerprints, specifically lipid A—the hydrophobic anchor of lipopolysaccharides in the outer membrane of Gram-negative bacteria (Solntceva et al. 2020; Furniss et al. 2020). The diverse modification patterns of lipid A can be regulated by environmental conditions, such as pH (Raetz and Whitfield 2002; Saha et al. 2022; Mozaheb et al. 2023). These heterogeneous lipid fingerprints represent promising biomarkers for precise species identification for *Klebsiella* species, which is crucial for effective infection management and control.

This study aims to accelerate the application of routine MALDI-TOF MS with lipid profiling to address the challenge of rapid discrimination between *K. pneumoniae* and *K. quasipneumoniae*. Analysis using MALDI-TOF MS in the negative-ion mode coupled with a machine learning algorithm demonstrated the feasibility of distinguishing *K. pneumoniae* and K. *quasipneumoniae* based on their lipid profiles.

2 | Material and Methods

2.1 | Bacterial Strains and Species Confirmation From Whole-Genome Sequencing Data

A total of 39 bacterial blood culture isolates, including 20 *K. pneumoniae* and 19 *K. quasipneumoniae* isolates, were used in this study (Table S1). For genomic DNA preparation, all bacterial isolates were grown from frozen stocks onto LB agar (Sigma-Aldrich, Milwaukee, USA) at 37°C overnight. Overnight growth was resuspended in PBS (Sigma-Aldrich, Milwaukee, USA) and processed following the instructions provided by the MicrobesNG (Birmingham, UK). Then the prepared bacterial lysates were sent to MicrobesNG (https://microbesng.com) for whole-genome sequencing. Genomic DNA libraries were prepared using the Nextera XT Library Prep Kit (Illumina, San Diego, USA) following the manufacturer's protocol with the following modifications: input DNA is increased twofold, and PCR elongation time is increased to 45 s. DNA quantification and library preparation were carried out on a Hamilton Microlab STAR automated liquid handling system (Hamilton Bonaduz AG, Switzerland). Libraries were sequenced on an Illumina NovaSeq. 6000 (Illumina, San Diego, USA) using a 250 bp paired end protocol.

FastQC v0.12.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC v0.4 (https://github.com/MultiQC/MultiQC) were used to assess the quality of raw sequence reads. Trimmomatic v0.39 (Bolger et al. 2014) was used to trim raw sequence reads. Reads were trimmed per specific instructions: to remove leading low quality or N bases (below quality three); to remove trailing low quality or N bases (below quality three); to scan the read with a 10-base wide sliding window, cutting when the average quality per base drops below 30, and to drop reads below 50 bases long. Bacterial species were confirmed from trimmed reads using Kraken2 v2.0.7-beta (Wood et al. 2019) and Bracken v2.5.0 (https://github.com/jenniferlu717/Bracken) with the standard bacterial Kraken2 database (last updated September 15, 2023).

2.2 | Sample Preparation for Lipid Profiling

Bacterial strains were cultured in LB medium overnight at 37°C. Bacteria were then pelleted, heat-inactivated (98°C for 1 h), and washed three times with 200 µL ddH_2O. Acid hydrolysis of the bacteria was performed using 1% acetic acid, followed by incubation at 98°C for 2 h. The bacteria were then washed three more times with 200 µL ddH_2O and resuspended in 200 µL ddH_2O. For MALDI-TOF MS analysis, 0.5 µL of the sample and 0.5 µL MBT lipid Xtract matrix (Bruker) were loaded on a MALDI target plate named MSP 96 target polished steel BC (Bruker Part-No. 8280800). The bacterial suspension and matrix were mixed directly on the target by pipetting and then dried gently under a stream of air. This procedure was conducted on three technical replicates of 39 samples comprising a mixture of *K. pneumoniae* and *K. quasipneumoniae*, totaling 117 samples across the two species.

2.3 | MALDI-TOF MS Analysis

MS analyses were performed on a MALDI Biotyper sirius system (Bruker Daltonics, Germany) using the protocol previously described (Pizzato et al. 2022). The mass spectra were scanned in the range of m/z 1000–3000. The mass profiles were acquired using FlexControl 3.4 software (Bruker Daltonics, Germany). The spectra were recorded in the linear negative-ion mode (laser intensity 65%, ion source 1 = 15.00 kV, ion source 2 = 13.70 kV, lens = 4.45 kV, detector voltage = 2698 kV, pulsed

20458827, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mbo3.70035 by Imperial College London, Wiley Online Library on [09/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ion extraction = 200 ns). Each spectrum corresponded to ion accumulation of 2000–5000 laser shots randomly distributed on the spot. The spectra obtained were processed with default parameters using FlexAnalysis v.3.4 software (Bruker Daltonics, Germany).

2.4 | Pre-Processing of Lipid Spectra Data

The bioinformatics analysis pipeline used R version 4.1.2. The method described here used code adapted from a study by Gibb and Strimmer (2015). "MALDIquant" (version 1.21) and "MALDIquantForeign" (version 0.13) packages were used to pre-process the spectra data for all *K. pneumoniae* and *K. quasipneumoniae* samples. First, a square root transformation (sqrt) was performed on the intensities of the spectra. The intensity values were then smoothed using the Savitzky–Golay method. The baseline of the mass spectrometry data was estimated and then removed using the Statistics-sensitive Nonlinear Iterative Peak-clipping algorithm (SNIP). Intensity values were normalized using the Total Ion Current (TIC) method, and then spectra were aligned. A signal-to-noise ratio of 3 (SNR = 3) and a half window size of 20 (HWS = 20) were used to detect peaks above the defined threshold in the mass spectrometry data. Following this, the peak binning function was used to look for similar peaks across different spectra and equalize their mass. Finally, peaks that occurred infrequently within a same species group were removed from the data. After this pre-processing, the result was a two-dimensional feature matrix containing peak intensity information for the spectra of all samples.

2.5 | Machine Learning

The approach used was in agreement as the one previously described (Pizzato et al. 2022). Briefly, after above pre-processing workflow, the feature matrix was converted into both a naïve binary absence–presence matrix (replaced non-negative and missing value with 1 and 0 respectively in feature matrix, true labels were not utilized.) and a dichotomized binary matrix. (For each feature [m/z], a threshold is determined by considering true labels. Intensities above that threshold will be set to 1, otherwise 0, via R packages "binda" [version 1.0.4]; Gibb and Strimmer 2015.) Hierarchical clustering was applied on a naïve binary feature matrix to figure out if different species can be separated in an unsupervised manner.

Binary discriminant analysis was then applied to the dichotomized binary feature matrix to identify and rank the most differentially expressed peaks across the spectra and ascertain whether any of these peaks from the lipid profiles could be used for class prediction, that is, to determine whether the spectra belonged to a sample of *K. pneumoniae* or *K. quasipneumoniae* (by computing *t*-scores between the group means in each species). The top-ranked peaks were used to test their class prediction ability. Data were further split into training and testing data (randomly picked 70% of all samples for training and rest for testing) to study the robustness of the top-ranked features in terms of classification of the two species.

3 | RESULTS

3.1 | *K. pneumoniae* and *K. quasipneumoniae* Showed Distinct Lipid Profile

To evaluate lipid profiling for discriminating *K. pneumoniae* from *K. quasipneumoniae* using a standard MALDI Biotyper Sirius system, we analyzed a panel of 20 *K. pneumoniae* and 19 *K. quasipneumoniae* clinical isolates, with each sample processed in technical triplicate. The samples were prepared for membrane lipid enrichment, and mass spectra were acquired in linear negative ion mode. The m/z 800–2300 range was selected for analysis based on optimal signal-to-noise ratio (S/N > 3) and mass resolution (> 200), ensuring suitability for subsequent data processing.

A total of 117 spectra (60 from *K. pneumoniae* [20 isolates × 3 replicates]; 57 from *K. quasipneumoniae* [19 isolates × 3 replicates]) were pre-processed before further data analysis. As shown in Figure 1, spectra from both species displayed prominent peaks centered at m/z 1824, identified as bis-phosphorylated, hexa-acylated lipid A. Additional common peaks at m/z 1388 (identified as cardiolipin) and m/z 2062 (identified as bis-phosphorylated, hepta-acylated lipid A, resulting from palmitoylation at the C-1 acyl-oxo-acyl position of the m/z 1824 molecule) indicated shared lipid features.

Despite these similarities, peak intensities at m/z 1388 and m/z 2062 differed between species. Furthermore, the peak at m/z 1962 demonstrated higher abundance in *K. pneumoniae*. To determine whether lipid profiles could effectively differentiate the species, we employed machine learning methodologies. This analysis targeted the identification of specific differential markers capable of reliably distinguishing *K. pneumoniae* and *K. quasipneumoniae*.

3.2 | Machine Learning Allows Discrimination of *K. pneumoniae* and *K. quasipneumoniae* Clinical Isolates

Following established data pre-processing protocols (Tang et al. 2019), initial analysis of the lipid profiles revealed unsupervised clustering of the top five ranked features (Figure 2). Subsequently, these top-ranked peaks were identified and their robustness for classification was validated.

Using the dichotomized matrix in a supervised approach, we extracted the top-ranked peaks (Table S2) via multi-class discriminant analysis employing binary predictors (Gibb and Strimmer 2015). Using the "binda" R package, the intensities of the 15 top-ranked peaks efficiently distinguished *K. pneumoniae* and *K. quasipneumoniae* (Figure 3).

3.3 | Validation of the Robustness of Top-Ranked Peaks via Supervised Learning and Using Randomly Selected Features as Control

To validate the classification robustness of the top-ranked peaks for distinguishing the two bacterial species, we partitioned the

20458827, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mbo3.70035 by Imperial College London, Wiley Online Library on [09/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

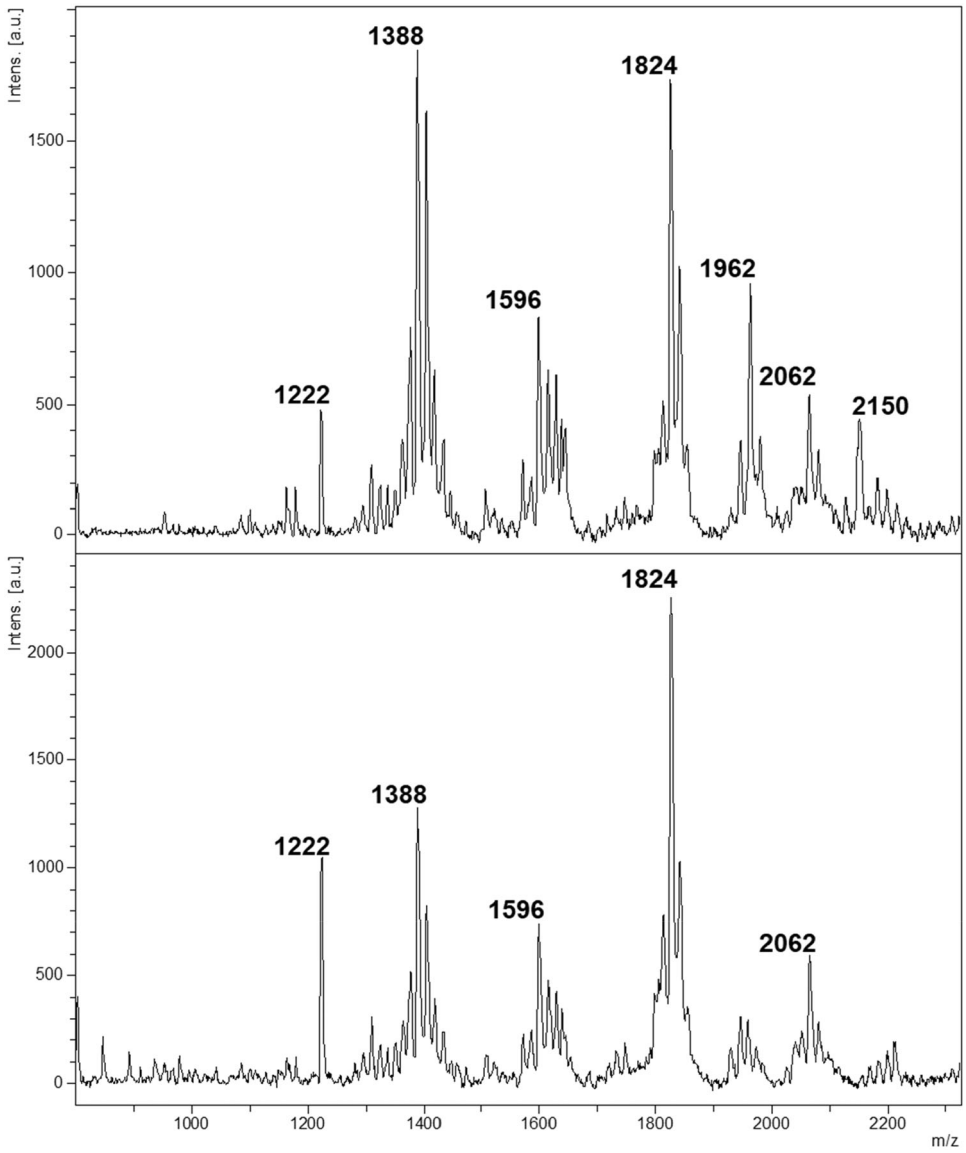


FIGURE 1 | Linear negative ion mode mass spectra of *K. pneumoniae* (top panel) and *K. quasipneumoniae* (bottom panel).

data set into training (70% of 117 samples) and testing subsets. The random separation of training and testing data were repeated 100 times. Various numbers of top-ranked peaks, either all peaks ($n = 44$) or a subset of randomly selected peaks ($n = 15$), were used as control. Consistent with a previous report from Tang et al. (2019), minimal differences in accuracy rates were observed between the groups using different controls. This finding supports previous reports that only a subset of top-ranked peaks is sufficient for classification (Gibb and Strimmer 2015; Conrad et al. 2017). Using top-ranked peaks, we were able to achieve relatively higher performance metrics, in terms of precision, sensitivity, and specificity for identifying *K. pneumoniae* and *K. quasipneumoniae* (Figure 4). In addition, the accuracy rates, based on comparison between *K. pneumoniae* and *K. quasipneumoniae,* enable discrimination of the two species, which is consistent with the data shown in Figure 5.

Overall, the five top-ranked peaks are sufficient for discriminating the species (*m/z* 2157, *m/z* 1931, *m/z* 1964, *m/z* 2042, *m/z* 1407) (Table S2).

4 | DISCUSSION

MALDI-TOF MS is widely applied in clinical microbiology for identification of microorganisms (Singhal et al. 2015; Kostrzewa et al. 2019). Although capsule genotyping and conventional protein-based MALDI-TOF MS are commonly used for *Klebsiella* species identification, their diagnostic accuracy is limited by the high genetic similarities among these closely related strains (Long et al. 2017). This genomic similarity can be quantified using metrics such as average nucleotide identity (ANI). ANI values above 95%–96% generally indicate the same species, while values between 93% and 95% suggest very closely related but distinct species. The primary focus of this study, *K. pneumoniae* and *K. quasipneumoniae*, have an ANI value around 93%–94%, indicating their high similarity and the difficulties in species identification (Brisse et al. 2014). In this project, the negative ion mode of MALDI-TOF MS was used to identify distinctive features in the lipid profiles of *K. pneumoniae* and *K. quasipneumoniae*, which has the potential as a novel species classification method.

20458827, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mbo3.70035 by Imperial College London, Wiley Online Library on [09/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

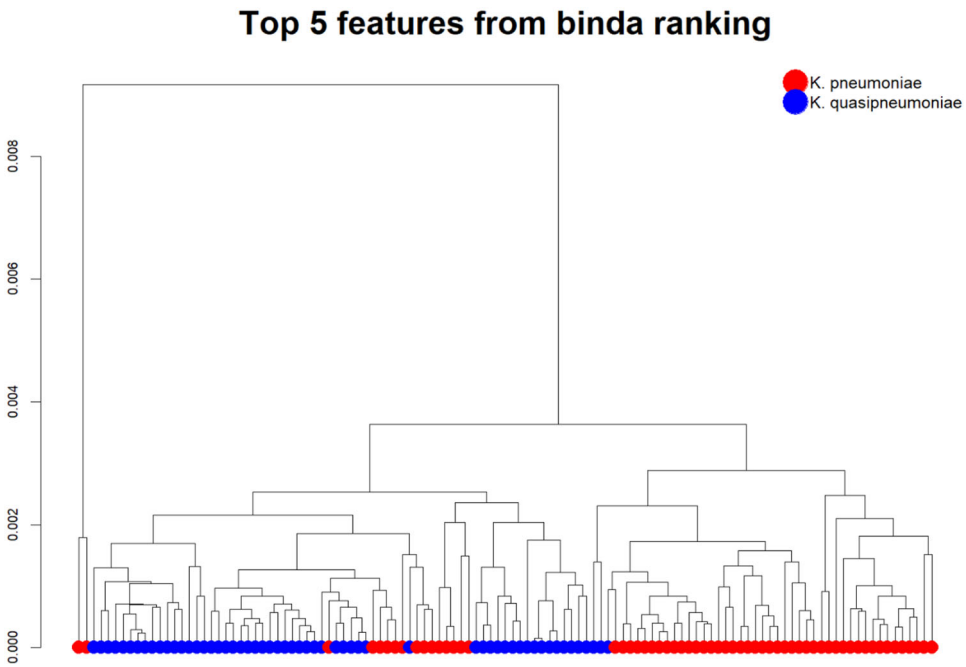


FIGURE 2 | Dendrograms showing hierarchical clustering of *K. pneumoniae* and *K. quasipneumoniae*. Red indicates *K. pneumoniae* isolates and blue indicates *K. quasipneumoniae*. Clustering of species was performed using the naïve binary absence presence matrix.

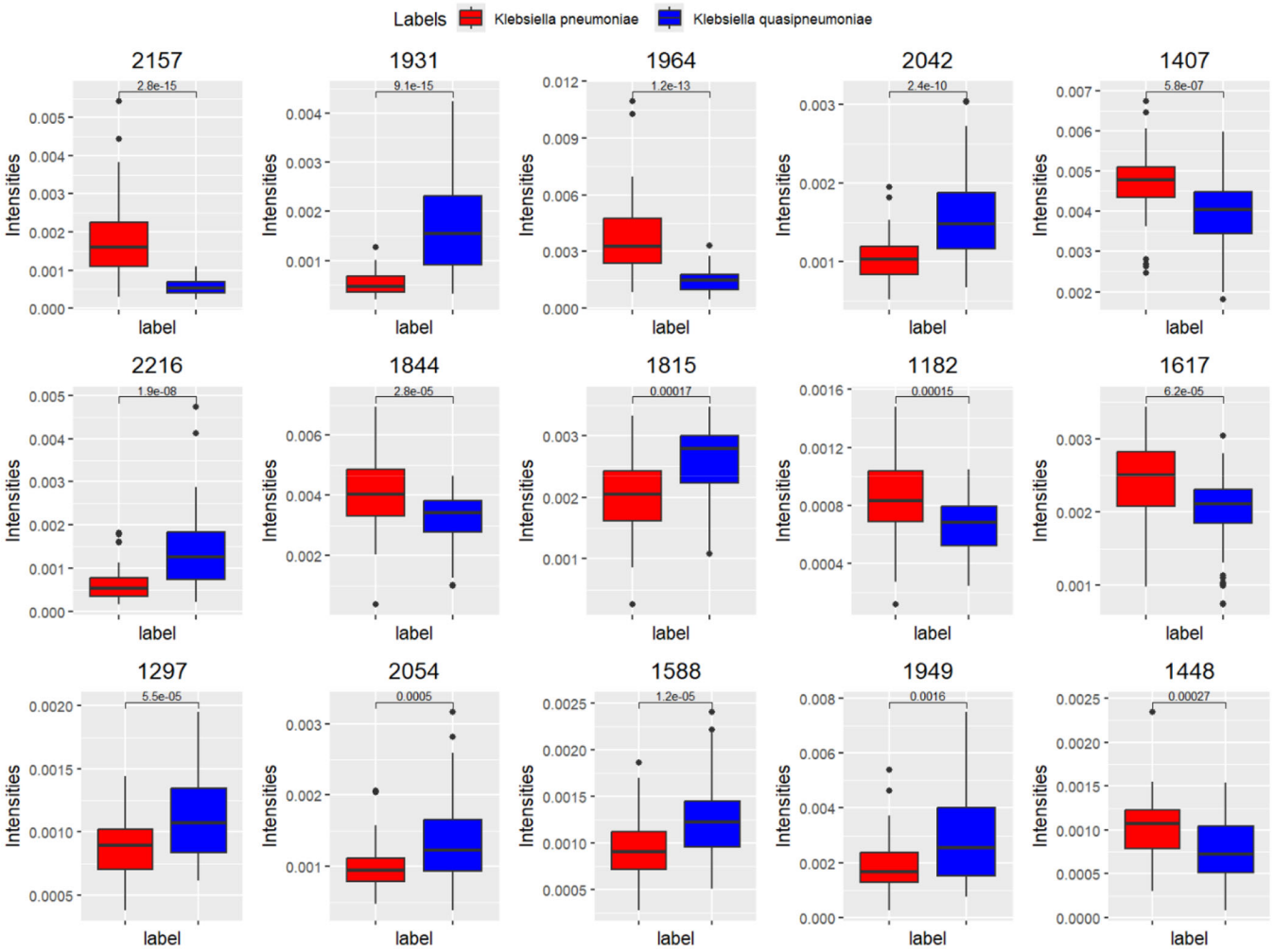


FIGURE 3 | Top 15 ranked features reported from binda (ranked from left to right, from top to bottom). *p* values were calculated from *t*-tests.

20458827, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mbo3.70035 by Imperial College London, Wiley Online Library on [09/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

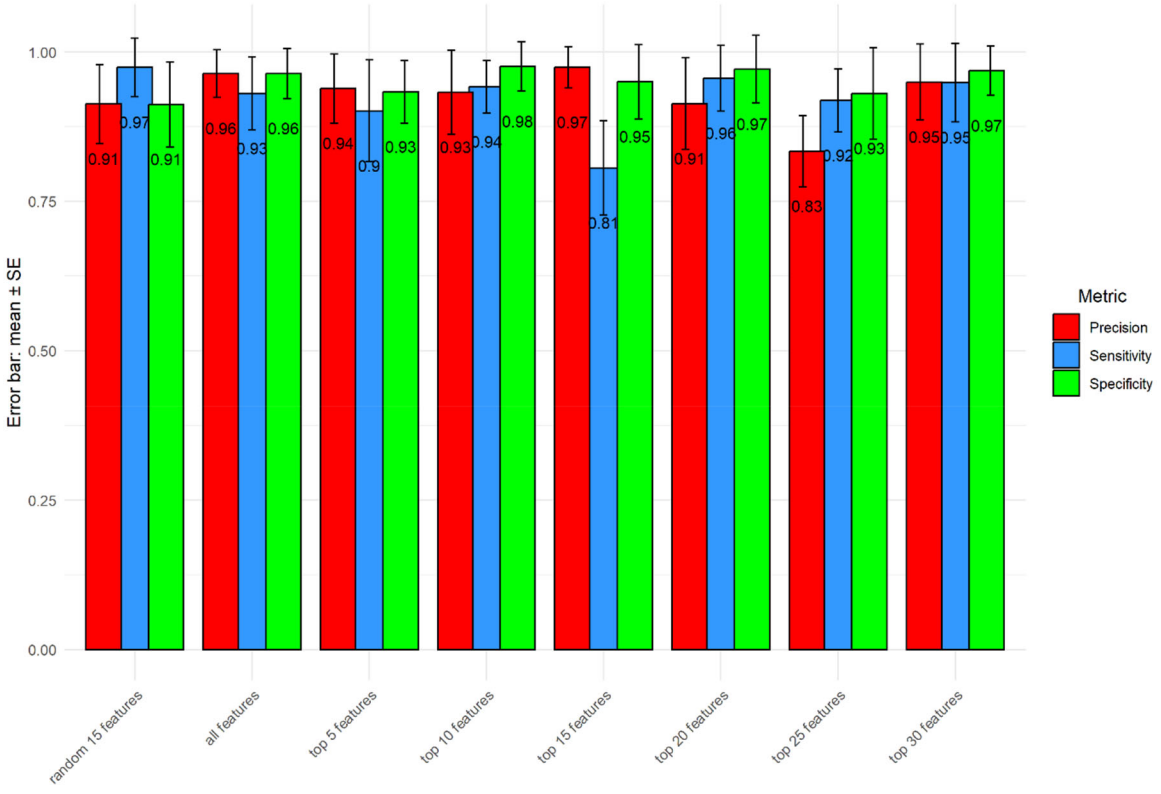


FIGURE 4 | Classification metrics including precision, sensitivity, and specificity.

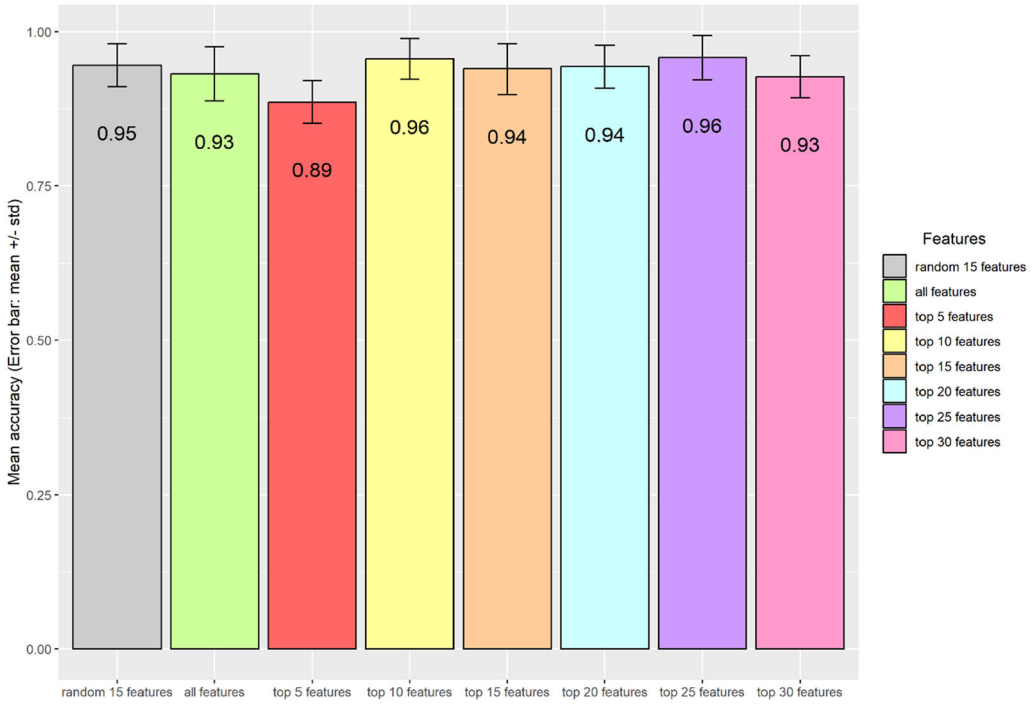


FIGURE 5 | Bar plots showing accuracy values for species prediction (between *K. pneumoniae* and *K. quasipneumoniae*) using top-ranked features; all features and randomly selected features. The accuracy values come from the analysis being repeated 100 times of splitting training and testing data and randomly selection of peaks for control.

To differentiate these closely related *Klebsiella* species, this study focuses on identifying discriminative features between *K. pneumoniae* and *K. quasipneumoniae*. Methodological refinements such as testing alternative mass spectrometry matrices or optimizing preprocessing parameters could enhance result accuracy and reproducibility. These adjustments could further enhance the accuracy and reproducibility of the results. While Monte-Carlo cross-validation was used to test the mean

20458827, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mbo3.70035 by Imperial College London, Wiley Online Library on [09/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

accuracy of the predictions from different rankings, other cross-validation methods, such as *k*-fold cross-validation, could also be used to reduce variability between results.

Based on our data, further investigations should focus on identifying the top-ranked molecules in our supervised studies. For example, techniques such as tandem mass spectrometry (MS/MS) can be used to reveal the biomolecular identities underlying those top-ranked peaks.

The potential clinical implications of lipid profile-based microbial identification are substantial. The rapid and accurate classification of closely related *Klebsiella* species can distinguish their different clinical pathogenicity, epidemiology, and potential antibiotic resistance, hence improving patient care with timely and appropriate treatments (Watanabe et al. 2022). Although WGS is known for its high accuracy in species identification, its sample preparation process is labor-intensive and time-costing because most WGS is done outside of diagnostic laboratories. Thus, rapid diagnostic methods based on already existing equipment within clinical setting are more practical. The development of diagnostic protocols incorporating these lipid profile-based biomarkers, identified in this study, could reduce the time to diagnosis and improve the management of infections caused by *Klebsiella* species. Clinical implementation of these findings requires addressing practical and regulatory considerations. Standardized protocols that ensure reproducibility must be developed, alongside evaluations of cost-effectiveness and scalability for routine negative-ion mode MALDI-TOF MS diagnostics.

In conclusion, while negative-ion mode MALDI-TOF MS lipid profiling shows initial promise for distinguishing *K. pneumoniae* from *K. quasipneumoniae*, expanded validation using diverse geographical and clinical isolates remains essential. Continued research in this area holds the potential to develop more rapid, accurate, and cost-efficient diagnostic tools for clinical microbiology, ultimately improving the identification and treatment of infections caused by closely related *Klebsiella* species. By expanding the scope of this study and addressing practical considerations, more accurate and rapid microbial diagnostics can be achieved, benefiting patient outcomes and public health.

Author Contributions

Mari Nishikawa: data curation, investigation, methodology. **Wenhao Tang:** data curation, methodology, investigation, software. **Markus Kostrzewa:** resources, writing – review and editing. **Jonah Rodgus:** investigation, data curation, writing – review and editing. **Frances Davies:** resources, writing – review and editing. **Yi Liu:** writing – review and editing. **Elita Jauneikaite:** resources, writing – review and editing, conceptualization, data curation, supervision, investigation, methodology. **Gerald Larrouy-Maumus:** conceptualization, investigation, funding acquisition, writing – original draft, supervision, data curation, methodology.

Acknowledgments

This study was sponsored by the MRC-Confidence in Concept grant number 105603/Z/14/Z (https://mrc.ukri.org/). W.T.'s PhD was funded by a Roth Scholarship from the Department of Mathematics when the tools for this study were developed. He is currently supported by The Oli Hilsdon Foundation through The Brain Tumour Charity, grant number (GN-000595). E.J., J.R., and F.D. are affiliated with the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Healthcare Associated Infections and Antimicrobial Resistance at Imperial College London in partnership with the UK Health Security Agency, in collaboration with Imperial Healthcare Partners, University of Cambridge and University of Warwick. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research, the Department of Health and Social Care or the UK Health Security Agency. Part of the microbiological work and sample preparation for whole-genome sequencing of *Klebsiella* species isolates was undertaken at the Colebrook Laboratory, a facility supported by the NIHR Imperial Biomedical Research Centre (BRC).

Ethics Statement

The isolates were collected in accordance with ethics reference 21/LO/0170 (279677), protocol 21HH6538 Investigation of epidemiological and pathogenic factors associated with infectious diseases.

Conflicts of Interest

M.K. is an employee of Bruker Daltonics GmbH & Co. KG, the manufacturer of the MALDI-TOF MS system used in this study, and MBT Lipid Xtract Kit. The other authors declare no conflicts of interest.

Data Availability Statement

All data are provided in full in the results section of this paper. Whole-genome sequencing raw data (Illumina reads) for isolates used in this study have been deposited in European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/home) under BioProjects PRJEB83319 (*K. pneumoniae*) and PRJEB85157 (*K. quasipneumoniae*), see Table S1.

References

Bolger, A. M., M. Lohse, and B. Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Brisse, S., V. Passet, and P. A. D. Grimont. 2014. "Description of *Klebsiella quasipneumoniae* sp. Nov., Isolated From Human Infections, With Two Subspecies, *Klebsiella quasipneumoniae* Subsp. *quasipneumoniae* Subsp. Nov. and *Klebsiella quasipneumoniae* Subsp. *similipneumoniae* Subsp. Nov., and Demonstration That *Klebsiella singaporensis* Is a Junior Heterotypic Synonym of *Klebsiella variicola*." *International Journal of Systematic and Evolutionary Microbiology* 64: 3146–3152. https://doi.org/10.1099/ijs.0.062737-0.

Choby, J. E., J. Howard-Anderson, and D. S. Weiss. 2020. "Hypervirulent *Klebsiella pneumoniae* - Clinical and Molecular Perspectives." *Journal of Internal Medicine* 287: 283–300. https://doi.org/10.1111/joim.13007.

Conrad, T. O. F., M. Genzel, N. Cvetkovic, et al. 2017. "Sparse Proteomics Analysis–A Compressed Sensing-Based Approach for Feature Selection and Classification of High-Dimensional Proteomics Mass Spectrometry Data." *BMC Bioinformatics* 18: 160.

Dinkelacker, A. G., S. Vogt, P. Oberhettinger, et al. 2018. "Typing and Species Identification of Clinical Klebsiella Isolates by Fourier Transform Infrared Spectroscopy and Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry." *Journal of Clinical Microbiology* 56: e00843-18. https://doi.org/10.1128/JCM.00843-18.

Fonseca, E. L., N. V. Ramos, B. G. N. Andrade, L. L. C. S. Morais, M. F. A. Marin, and A. C. P. Vicente. 2017. "A One-Step Multiplex PCR to Identify *Klebsiella pneumoniae*, *Klebsiella variicola*, and *Klebsiella quasipneumoniae* in the Clinical Routine." *Diagnostic Microbiology and Infectious Disease* 87: 315–317. https://doi.org/10.1016/j.diagmicrobio.2017.01.005.

20458827, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mbo3.70035 by Imperial College London, Wiley Online Library on [09/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Furniss, R. C. D., M. Kostrzewa, D. A. I. Mavridou, and G. Larrouy-Maumus. 2020. "The Clue Is in the Lipid A: Rapid Detection of Colistin Resistance." *PLoS Pathogens* 16: e1008331. https://doi.org/10.1371/journal.ppat.1008331.

Gibb, S., and K. Strimmer. 2015. "Differential Protein Expression and Peak Selection in Mass Spectrometry Data by Binary Discriminant Analysis." *Bioinformatics* 31: 3156–3162.

Hafiz, T. A., S. Alanazi, S. S. Alghamdi, et al. 2023. "*Klebsiella pneumoniae* Bacteraemia Epidemiology: Resistance Profiles and Clinical Outcome of King Fahad Medical City Isolates, Riyadh, Saudi Arabia." *BMC Infectious Diseases* 23: 579. https://doi.org/10.1186/s12879-023-08563-8.

Holt, K. E., H. Wertheim, R. N. Zadoks, et al. 2015. "Genomic Analysis of Diversity, Population Structure, Virulence, and Antimicrobial Resistance in *Klebsiella pneumoniae*, An Urgent Threat to Public Health." *Proceedings of the National Academy of Sciences* 112: E3574–E3581. https://doi.org/10.1073/pnas.1501049112.

Kostrzewa, M., E. Nagy, P. Schröttner, and A. B. Pranada. 2019. "How MALDI-TOF Mass Spectrometry Can Aid the Diagnosis of Hard-To-Identify Pathogenic Bacteria - the Rare and the Unknown." *Expert Review of Molecular Diagnostics* 19: 667–682. https://doi.org/10.1080/14737159.2019.1643238.

Long, S. W., S. E. Linson, M. Ojeda Saavedra, et al. 2017. "Whole-Genome Sequencing of Human Clinical *Klebsiella pneumoniae* Isolates Reveals Misidentification and Misunderstandings of *Klebsiella pneumoniae*, *Klebsiella variicola*, and *Klebsiella quasipneumoniae*." *mSphere* 2: e00290-17. https://doi.org/10.1128/mSphereDirect.00290-17.

Martínez-Romero, E., N. Rodríguez-Medina, M. Beltrán-Rojel, et al. 2017. "Genome Misclassification of *Klebsiella variicola* and *Klebsiella quasipneumoniae* Isolated From Plants, Animals and Humans." *Salud pública de México* 60: 56–62. https://doi.org/10.21149/8149.

Mozaheb, N., P. Rasouli, M. Kaur, P. Van Der Smissen, G. Larrouy-Maumus, and M. P. Mingeot-Leclercq. 2023. "A Mildly Acidic Environment Alters *Pseudomonas aeruginosa* Virulence and Causes Remodeling of the Bacterial Surface." *Microbiology Spectrum* 11: e0483222. https://doi.org/10.1128/spectrum.04832-22.

Pizzato, J., W. Tang, S. Bernabeu, et al. 2022. "Discrimination of *Escherichia coli*, *Shigella flexneri*, and *Shigella sonnei* Using Lipid Profiling by MALDI-TOF Mass Spectrometry Paired With Machine Learning." *MicrobiologyOpen* 11: e1313. https://doi.org/10.1002/mbo3.1313.

Raetz, C. R. H., and C. Whitfield. 2002. "Lipopolysaccharide Endotoxins." *Annual Review of Biochemistry* 71: 635–700. https://doi.org/10.1146/annurev.biochem.71.110601.135414.

Saha, S., E. Pupo, A. Zariri, and P. van der Ley. 2022. "Lipid A Heterogeneity and Its Role in the Host Interactions With Pathogenic and Commensal Bacteria." *Microlife* 3: uqac011. https://doi.org/10.1093/femsml/uqac011.

Singhal, N., M. Kumar, P. K. Kanaujia, and J. S. Virdi. 2015. "MALDI-TOF Mass Spectrometry: An Emerging Technology for Microbial Identification and Diagnosis." *Frontiers in Microbiology* 6: 791. https://doi.org/10.3389/fmicb.2015.00791.

Solntceva, V., M. Kostrzewa, and G. Larrouy-Maumus. 2020. "Detection of Species-Specific Lipids by Routine MALDI TOF Mass Spectrometry to Unlock the Challenges of Microbial Identification and Antimicrobial Susceptibility Testing." *Frontiers in Cellular and Infection Microbiology* 10: 621452. https://doi.org/10.3389/fcimb.2020.621452.

Tang, W., N. Ranganathan, V. Shahrezaei, and G. Larrouy-Maumus. 2019a. "MALDI-TOF Mass Spectrometry on Intact Bacteria Combined With a Refined Analysis Framework Allows Accurate Classification of MSSA and MRSA." *PLoS ONE* 14: e0218951. https://doi.org/10.1371/journal.pone.0218951.

Watanabe, N., T. Watari, Y. Otsuka, K. Yamagata, and M. Fujioka. 2022. "Clinical Characteristics and Antimicrobial Susceptibility of *Klebsiella pneumoniae*, *Klebsiella variicola* and *Klebsiella quasipneumoniae* Isolated From Human Urine in Japan." *Journal of Medical Microbiology* 71: e1473-5644. https://doi.org/10.1099/jmm.0.001546.

Wood, D. E., J. Lu, and B. Langmead. 2019. "Improved Metagenomic Analysis With Kraken 2." *Genome Biology* 20: 257. https://doi.org/10.1186/s13059-019-1891-0.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

20458827, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mbo3.70035 by Imperial College London, Wiley Online Library on [09/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License