RESEARCH ARTICLE

Label-free differentiation of clinical *E. coli* and *Klebsiella* isolates with Raman spectroscopy

Amir Nakar[1,2,3] | Aikaterini Pistiki[1,2,3] | Oleg Ryabchykov[1,2] |
Thomas Bocklitz[1,2,3] | Petra Rösch[2,3*] | Jürgen Popp[1,2,3,4,5]

[1]Leibniz Institute of Photonic Technology Jena—Member of the Research Alliance "Leibniz Health Technologies", Jena, Germany

[2]Institute of Physical Chemistry and Abbe Center of Photonics, Friedrich Schiller University Jena, Jena, Germany

[3]Research Campus Infectognostics, Jena, Germany

[4]Jena Biophotonics and Imaging Laboratory, Jena, Germany

[5]Center for Sepsis Control and Care, Jena University Hospital, Jena, Germany
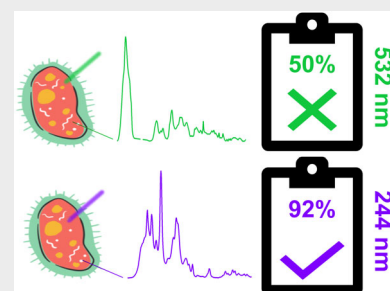
*Correspondence
Petra Rösch, Institute of Physical Chemistry and Abbe Center of Photonics, Friedrich Schiller University Jena, Helmholtzweg 4 07743 Jena, Germany.
Email: petra.roesch@uni-jena.de

Abstract

Raman spectroscopy is a promising spectroscopic technique for microbiological diagnostics. In routine diagnostic, the differentiation of pathogens of the *Enterobacteriaceae* family remain challenging. In this study, Raman spectroscopy was applied for the differentiation of 24 clinical *E. coli*, *Klebsiella pneumoniae* and *Klebsiella oxytoca* isolates. Spectra were collected with two spectroscopic approaches: UV-Resonance Raman spectroscopy (UVRR) and single-cell Raman microspectroscopy with 532 nm excitation. A description of the different biochemical profiles provided by the different excitation wavelengths was performed followed by machine-learning models for the classification at the genus and species levels. UVRR was shown to outperform 532 nm excitation, enabling correct classification at the genus level of 23/24 isolates. Furthermore, for the first time, *Klebsiella* species were correctly classified at the species level with 92% accuracy, classifying all three *K. oxytoca* isolates correctly. These findings should guide future applicative studies, increasing the scope of Raman spectroscopy's suitability for clinical applications.

KEYWORDS

bacteria, diagnostics, machine learning, Raman spectroscopy, SVM

1 | INTRODUCTION

Enterobacteriaceae are a family of Gram-negative bacteria that are part of the mammalian microflora, especially in the gastrointestinal tract.[1] In many *Enterobacteriaceae*

species, pathogenic strains have developed,[1–3] causing a range of clinical conditions ranging from chronic diseases as irritable bowel syndrome (IBS)[4] to acute, life threatening conditions as bacteremia and sepsis.[5] *Klebsiella spp.* and *Escherichia coli* are among the leading causes of hospital acquired infections.[6] Furthermore, these bacteria often carry resistance genes against antimicrobial therapy, making them difficult to treat.[2, 7, 8] *E. coli* is the most common

Amir Nakar and Aikaterini Pistiki should be considered joint first authors.

cause for hospital acquired infections, in particular it is the cause of more than 25% of urinary tract infections.[9] *K. pneumoniae* is the most common pathogen of the *Klebsiella* genus, followed by *K. oxytoca*, an emerging pathogen that contributes between 13% and 24% of all nosocomial bacteremia infections.[10] In clinical laboratory settings, the differentiation of *Enterobacteriaceae* species, for example, *E. coli* and *Klebsiella spp.* remains challenging, as the bacteria are closely related both in their genome and phenotypic appearance.[11] In order to differentiate *E. coli* from *Klebsiella*, an expensive set of 47 biochemical tests is needed.[12] To discriminate *K. pneumonia* from *K. oxytoca* an additional indole reaction is required that is sometimes difficult to detect and can lead to misclassifications.[13]

In clinical settings, the detection of the pathogen species is of high importance for physicians to choose the appropriate antimicrobial treatment as well as for outbreak tracing and epidemiology.[14, 15] The required time and costs of the microbiological procedures are also important factors. Many microbiological methods are used for this purpose with the gold standard being automated platforms which test the bacteria in a complex set of biochemical reactions for different metabolic activities and matches the profile with an established database.[12] However, these methods require approximately 24 h after the isolation of the pathogen to provide results and are not very cost effective. Matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) is a spectroscopic method that has been recently approved for routine clinical laboratory use and can provide results within several hours.[16, 17] Yet, for the *Enterobacteriaceae* group limitations are present as the bacteria are so closely related both in their genome and phenotypic appearance.[18] PCR based methods are also commonly used but require expensive consumables.[19]

Raman spectroscopy has been demonstrated as a rapid, label-free and robust tool for the classification and identification of clinically relevant bacteria.[20–22] This method uses the molecular fingerprint of bacterial cells to identify the species. The biochemical information is obtained by exposing the bacterial cells to a laser and measuring the scattered light using a spectrometer. Since microbial species differ in their molecular composition, they provide a distinct spectral fingerprint allowing their identification and classification with the use of chemometrics and machine learning algorithms.[23–25] Single cell Raman microspectroscopy (SC-RMS) was used previously to classify *Legionella*,[26] *Mycobacteria*,[27] *Burkholderia*[28] and other pathogens.[29] The major advantage of using SC-RMS is that it does not require extensive cultivation steps, for example, by measuring cells directly isolated from blood, urine and other body fluids.[21, 29, 30] Alternatively, in UV-Resonance Raman (UVRR) spectroscopy, the bacteria are measured in bulk, where a large biomass of bacteria is exposed to

the light source. This bulk approach is essential for reducing the photothermal damage caused by the destructive UV light and therefore a cultivation step is needed to obtain the required biomass. The main advantage of using UVRR on bacterial samples is that Raman signals originating from nucleic acids and aromatic amino acids are enhanced via a resonance effect, leading to a higher signal-to-noise-ratio in the spectra.[31–33] UVRR spectroscopy was used previously to differentiate clinically relevant yeasts[31] and bacteria.[34] While both methods have been used for microbial diagnostics, few studies have been done to compare their abilities and produce recommendations for choosing the best tool for Raman-based diagnostics.[21]

Previous studies on the application of Raman spectroscopy to *Enterobacteriaceae* have had limited success. In a study on the application of SC-RMS with 532 nm excitation for identification of pathogens in ascetic fluids, all used *Enterobacteriaceae* species were represented as a single group as it was impossible to differentiate them.[35] Another study concluded that the family of *Enterobacteriaceae* are particularly difficult to differentiate using phenotypic methods.[11] Few studies have used examples of *Enterobacteriaceae* but only a small number of strains were used.[11, 26, 30, 36] Therefore, it remains a challenge to differentiate the members of the *Enterobacteriaceae* family using Raman spectroscopy, despite the success shown with other pathogens. Furthermore, to the best of our knowledge, no previous study considered the important emerging pathogen *K. oxytoca* in their dataset.

In this study, 24 clinical isolates of the species *E. coli*, *K. pneumoniae* and *K. oxytoca* were analyzed using Raman spectroscopy. The diagnostic potential of two different Raman approaches was studied: SC-RMS on single cells and UVRR spectroscopy on bulk samples. Both methods were used to describe the biochemical composition of the different species and the classification of the isolates at the genus and species level.

2 | MATERIALS AND METHODS

2.1 | Sample preparation

For this study, 24 clinical isolates were collected in the general University hospital of Larisa, Greece. Identification of the isolates was performed using the Vitek-2 system (BioMérieux, Marcy l'Etoile, France), according to the manufacturer's instructions as described before.[37] The isolates were identified as 15 strains of *Klebsiella pneumoniae*, six strains of *Escherichia coli* and three strains of *Klebsiella oxytoca*.

For Raman measurements, bacteria were cultured from frozen stock on nutrient agar (NA) (Carl Roth,

Karlsruhe, Germany) and incubated overnight at 37°C. A loopful of biomass was then transferred to nutrient broth (NB) (Carl Roth) and incubated at 37°C with shaking of 120 rpm. Each strain was cultivated in three biological replicates on different dates for measuring.

For single cell Raman microspectroscopy (SC-RMS) the cells were grown overnight in 5 mL of NB. The cultures had an optical density of 0.5–1.5 when harvested for measurement. The samples were prepared by adding 100 µL of bacterial culture to 900 µL of distilled water. To remove traces of media, samples were washed three times with deionized water using centrifugation at 5000 g for 5 min. Finally, 10 µL of sample was spotted in 1 µL droplets on a nickel foil disk and allowed to dry at room temperature for 15–60 min. Prior to UV-Resonance Raman (UVRR) measurements, bacteria were grown for 1 h in 20 mL of NB in order to reach the exponential growth state. Three replicates of 1.5 mL each of the inoculum were heat-inactivated at 99°C for 5 min, followed by washing three consecutive times. The final pellet was resuspended in 30 µL of distilled water and allowed to dry on a fused-silica slide at room temperature for 1 h. To ensure heat inactivation was successful, a small amount of biomass was plated on NA agar plates, incubated for 24 h in 37°C and examined for no growth.

2.2 | Raman measurements

SC-RMS spectra were collected from single cells using a Raman microscope (BioParticleExplorer, MicrobioID 0.5, RapID). The microscope was connected to a 532 nm frequency-doubled solid-state Nd:YAG diode pumped laser (LCM-S-111, Laser-Export Company Ltd.). The laser beam was focused with an ×100 magnification objective (MPLFLN ×100, NA: 0.9, Olympus Corporation) onto the sample with approximately 16 mW leading to approximately 3.5 mW on the cells. Backscattered Raman light was focused to a single stage monochromator (HE 532, Horiba Jobin Yvon) equipped with a 920 lines/mm grating. The light was then collected with a thermoelectrically cooled CCD camera (DV401A-BV, Andor Technology). The spectral resolution was approximately 10 cm^{-1}. For each bacterial cell, two consecutive Raman spectra were measured at the same position, which were afterward combined. Integration time was 15 s for each bacterial cell. For each replicate, 50–70 spectra were collected. A total of >4000 spectra were obtained from three biological replications, with an average of 180 spectra per isolate.

UVRR spectra were collected using a Raman setup (HR800, Horiba/Jobin-Yvon) with a focal length of 800 mm. A 244 nm frequency-doubled argon-ion laser (Innova 300, FReD) was used to excite the sample. The laser was focused with an ×40 antireflection-coated objective (LMU, NA: 0.5, UVB). Backscattered Raman light was collected through a 400 µm slit into a 2400 lines/mm grating and detected by a nitrogen-cooled CCD camera. The spectral resolution was 2 cm^{-1}. For each spectrum measured, 15 s of illumination time, and a maximum laser output power of 18 mW was used leading to about 0.5 mW on the sample. During measurement, the sample stage was rotated constantly on a spiral path to reduce sample burning. In each measurement, 10 spectra were obtained and averaged to reduce noise. A total of 25 measurements for each replicate were obtained. A total of >1800 measurements were collected from three biological replications, with an average 75 per isolate.

2.3 | Data analysis

Preprocessing and data analysis were done using the RAMANMETRIX software (Version 0.3.4, Leibniz Institute of Photonic Technology). In order to prepare the data for further analysis, several preprocessing steps were taken. First the spectra were de-spiked as described before.[38] Then, the spectra were wavenumber calibrated, and background corrected using a Sensitive Nonlinear Iterative Peak (SNIP) algorithm with 40 iterations. Lastly, spectra were vector normalized and truncated to the relevant range (500–1900 cm^{-1} for UVRR spectra and 400–3050 cm^{-1} for SC-RMS data). Also, for the SC-RMS data, the silent region (1850–2750 cm^{-1}) was removed to reduce noise.

Wavenumber calibration was done with a polynomial fit function. It was based on 4-acetamidophenol and polystyrene spectra for SC-RMS and UVRR data, respectively. The polynomial degree was 3 for the 4-acetamidophenol standard spectra and 2 for the polystyrene standard spectra. A new reference spectrum was used on each measurement date.

The different classification models were calculated using RAMANMETRIX software. A Principal Component Analysis combined with Support Vector Machine (PCA-SVM) approach was used for all models and the number of principal components used was optimized based on the results of a leave-one-strain-out cross validation (LOSOCV) as described by Guo et al.[39] This validation method calculated a model repeatedly based on a dataset, excluding one strain, which is then predicted by the constructed model. For all SVM models, a Radial Basis Function (RBF) kernel was used, the model cost was set to 10 and gamma defined as 1 divided by number of variables (PCs). For the UVRR data, the PCA-SVM model was calculated, based on four PCs. For SC-RMS data, a PCA-SVM model was calculated, based on 20 PCs. In addition, burned SC-RMS spectra were removed

automatically from the SC-RMS dataset using an in-house R script.[40] Moreover, the data was cleansed by a correlation filter to remove any remaining outliers. This filter discards any spectra which have <0.9 correlation with the mean preprocessed spectrum of the entire dataset. More than 99.5% of spectra passed the filter. Once the classification models were calculated, a majority vote was taken to classify each strain individually. This was done to reduce in-sample heterogeneity, which causes some spectra to classify incorrectly. A "vote" is conducted within each isolate, and the class which has the majority of spectra is chosen as the "decided class." Balanced accuracy was calculated as the sum of the model sensitivity and specificity divided by 2. The spectra were visualized using Origin (Pro), version 2018b (OriginLab Corporation, Northampton). For further investigation of the intra-replicate variations, the results for the "vote" within each replicate (three per isolate) were produced (see Tables S3–S6).

For the classification a two-step approach was used in order to produce an algorithm suitable for decision-making. First a classification of the genera was generated in order to differentiate *E. coli* from *Klebsiella spp.* (*Klebsiella pneumoniae* and *Klebsiella oxytoca*). Then, in a second model was trained, with the same parameters in order to differentiate *Klebsiella pneumoniae* form *Klebsiella oxytoca* at the species level. This multi-level approach has been used before[26, 41] and provides models which are complementary, as they can be run one after the other, and fine-tuned to the specific classes.

3 | RESULTS AND DISCUSSION

3.1 | Descriptive analysis of the bacteria using Raman spectroscopy

In this study, we collected Raman spectra of 24 different clinical *Enterobacteriaceae* of the species *E. coli*, *K. pneumoniae* and *K. oxytoca* from the Larissa University hospital, Greece. Two different approaches were used to collect Raman spectra: Single cell Raman microspectroscopy (SC-RMS) and UV-Resonance Raman (UVRR) spectroscopy on bulk samples. The two different approaches were used because they highlight different elements of the bacterial cell. In SC-RMS, we expect a holistic view of the cell components, with signals coming from proteins, lipids, nucleic acids and carbohydrates. On the other hand, we expect UVRR spectroscopy to enhance the signals originating from nucleic acids and aromatic amino acids, providing a less comprehensive signal, but also a less noisy one. Since the methods provide different information, we were interested to find out which one has more diagnostic potential for differentiating the different species. The mean spectra of each
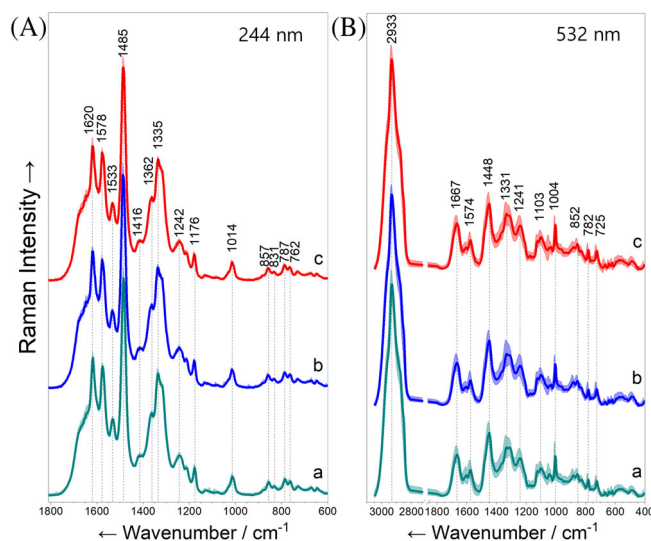

FIGURE 1 Mean Raman spectra of different bacterial species: (a) *E. coli*, (b) *K. oxytoca* and (c) *K. pneumoniae*. (A) Spectra from UVRR spectroscopy (B) spectra from SC-RMS. Standard deviations are shown as shades around each spectrum. All spectra are normalized and offset vertically for visualization purposes

species are presented in Figure 1. In Figure 1A, the spectra collected from bulk samples with 244 nm excitation (UVRR), and in Figure 1B the spectra from SC-RMS excited with 532 nm are shown. The standard deviation of each mean spectrum was also calculated and is highlighted.

The mean spectra shown in Figure 1 display known spectral fingerprints of bacteria.[21, 22] The UVRR spectra are dominated by bands enhanced with the resonance effect of excitation with 244 nm light. The bands at 787, 1242, 1335, 1362, 1485, 1533 and 1578 cm^{-1} are resonant bands from nucleic acids (i.e., DNA and RNA).[42–44] The bands at 762, 831, 857, 1014 and 1620 cm^{-1} are resonant bands derived from different aromatic amino acids: tryptophan, tyrosine and phenylalanine, which are essentially protein signals.[42, 43, 45] In addition, the band at 1176 cm^{-1} is a mixed band that can be assigned to both nucleic acids and proteins. A detailed table of band assignment can be found in the supplementary material (Table S1).

Unlike the UVRR spectra, SC-RMS spectra do not show a strong resonance effect. The wide band at 2933 cm^{-1} represents C–H stretching vibrations and the band at 1448 cm^{-1} represents CH_2/CH_3 deformation vibrations. These bands are common in many biomolecules, especially lipids and carbohydrates.[22, 46] In addition, bands at 1667 and 1241 cm^{-1} can be assigned to amide I and amide III vibrations, respectively, and relate to the protein backbone. The bands at 1004 and 852 cm^{-1} (phenylalanine and tyrosine ring breathing vibrations, respectively) also represent the proteins in

the cell. Finally, nucleic acids can be observed at 1574, 1331, 1103, 782 and 725 cm^{-1}.[45, 47–51] As expected, the nonresonant Raman spectrum produces a more comprehensive look into the biochemistry of the bacterial cell, including information from lipids and carbohydrates. More detailed assignments are given in the supplementary material (Table S2).

For both methods, no clear differences can be observed between the species from the mean spectra. This is not surprising, the similarity of all *Enterobacteriaceae* species, and particularly of *E. coli* and *Klebsiella spp.* is well documented.[1, 52] This finding stresses the challenge of classifying these genera both with standard methods and with Raman spectroscopy. The chemical makeup of all three species in overall proteins, lipids, nucleic acids and carbohydrates is almost identical and they share major parts of their core genome.[1, 52]

3.2 | Classification of *E. coli* and *Klebsiella spp.* at the genus level

As a first step, a classification of the genera was performed in order to differentiate between *E. coli* and *Klebsiella* spp. (*K. pneumoniae* and *K. oxytoca* combined).

Principal Component Analysis (PCA) was followed by a Support Vector Machine (SVM), where the number of principal components was optimized based on a leave-one-strain-out cross validation. After the classification model was calculated we used a majority voting approach to provide a classification for each isolate, in order to both remove any heterogeneity within a sample and to produce a clinically relevant decision. The results of classification based on SC-RMS with a 532 nm laser before and after majority voting are presented in Table 1 and those of UVRR spectroscopy in Table 2. For the SC-RMS data, the spectra were not classified accurately, achieving 59% balanced accuracy for spectra and only 67% balanced accuracy per isolate after majority voting. On the other hand, for the UVRR data, the spectra were classified with 78% balanced accuracy, and after using majority voting, isolates were correctly identified with 96% accuracy (23/24 strains correctly identified). Even when corrected for the imbalance of the dataset, the accuracy remains high, with 92% balanced accuracy.

These results demonstrate the importance of choosing the right Raman spectroscopic approach in order to achieve the best possible differentiation. In a previous study aiming to discriminate multi-drug resistant and susceptible *E. coli* strains, the same two Raman

TABLE 1 Summary of the first classification step for differentiation at the genus level using SC-RMS

SC-RMS		Reference		Balanced accuracy/%
Results per spectrum		Escherichia	Klebsiella	
Prediction	*Escherichia*	421	614	59
	Klebsiella	709	2488	
SC-RMS		Reference		Balanced accuracy/%
Results per isolate		*Escherichia*	*Klebsiella*	
Prediction	*Escherichia*	2	0	67
	Klebsiella	4	18	

Note: The top part presents the results per spectrum, that is, before majority voting. The bottom part presents the results per isolate, that is, after majority voting. The true labels are shown by column and the predicted classes by row.

TABLE 2 Summary of the first classification step for differentiation at the genus level using UVRR

UVRR spectroscopy		Reference		Balanced accuracy/%
Results per spectrum		*Escherichia*	Klebsiella	
Prediction	*Escherichia*	315	134	78
	Klebsiella	168	1229	
UVRR spectroscopy		Reference		Balanced accuracy/%
Results per isolate		*Escherichia*	*Klebsiella*	
Prediction	*Escherichia*	5	0	92
	Klebsiella	1	18	

Note: The top part presents the results per spectrum, that is, before majority voting. The bottom part presents the results per isolate, that is, after majority voting. The true labels are shown by column and the predicted classes by row.

spectroscopy approaches were compared. Results showed similar classification performance of SC-RMS and UVRR spectroscopy.[42, 53] Considering the high similarities present in the different strains of the same species, it is not surprising that UVRR was not sensitive enough to capture the very small differences present in their genome. In this study, the genotypic differences between the different species are significantly larger and could be captured by UVRR, leading to a classification accuracy of 92%.

Similar studies have been performed previously to classify *Enterobacteriaceae* using Raman spectroscopy with 532 nm excitation, however only a limited number of isolates were used.[26, 30] In the present study, a large dataset of isolates was collected leading to more robust conclusions. In another study, using SC-RMS with a 532 nm excitation the *Enterobacteriaceae* could not be differentiated and were therefore classified as a single group.[35] This is in accordance with the findings of the present study where SC-RMS provided poor classification abilities. Similarly, in a large study of pathogenic bacteria using an excitation wavelength of 633 nm, while most species were classified with high accuracy, the *Enterobacteriaceae* were classified with only 69% accuracy and were therefore analyzed as a group rather than different species.[23] This highlights again the importance of the resonant Raman spectra collected with UV excitation, as even at different wavelengths the classification remains difficult. Lastly, several studies using 785 nm excitation on bulk samples were performed successfully in clinical wards to link clonal isolates of *E. coli* and *K. pneumoniae*.[54–57] Yet, these studies have only considered clonality (the direct similarity of one strain to another) and would therefore not be relevant for identifying new strains evolved in a clinical setting. Taken together this shows the limited performance of SC-RMS and the high potential of UVRR in classification of *Enterobacteriaceae*.

3.3 | Classification of *Klebsiella oxytoca* and *Klebsiella pneumoniae* at the species level

As a next step, classification of *K. pneumoniae* and *K. oxytoca* at species level was performed. The spectra obtained for the *Klebsiella* genus group were used and analyzed for the classification of *K. pneumoniae* or *K. oxytoca*. Results are presented in Tables 3 and 4. In Table 3, it can be seen that again SC-RMS performed poorly, providing a balanced accuracy of 50% after majority voting, a result with no discrimination abilities. This is expected, as the two species are almost identical from a biochemical perspective. It is well

known that *K. pneumoniae* and *K. oxytoca* share a large part of their core genes and exchange virulence factor often.[10, 58] Even using biochemical methods, the two species are very difficult to differentiate.[13]

For the UVRR, however, the model produced a balanced accuracy of nearly 70% for spectra and 90% for isolates. This result is of great importance, showing that the resonance effect of nucleic acids and aromatic amino acids present in UVRR spectra improves signal-to-noise ratio and captures the small differences between the two species in their genome and their overall protein composition. It also shows that by reducing the sample's heterogeneity through majority voting we improve the model's accuracy dramatically, as can be seen by the correct classification of all 3 *K. oxytoca* isolates. *K. oxytoca* are known to be exceptionally difficult to differentiate from *K. pneumoniae*, are often misclassified and are a great cause of hospital acquired infections.[10, 13] To the best of our knowledge, this is the first time the differentiation of *K. oxytoca and K. pneumoniae* was performed using Raman spectroscopy. These findings are of great importance and stress the versatility and applicability of Raman spectroscopy in clinical laboratory settings.

Put together with the earlier results at the genus level, we can assert that:

1. UVRR spectroscopy significantly outperforms SC-RMS for the task of classifying *E. coli* and *Klebsiella spp*. isolates at the genus level.
2. UVRR spectroscopy can likely be used not only to differentiate clinical isolates at the genus level, but even to discriminate the cases of *K. oxytoca* from *K. pneumoniae* infections.

While this study shows great potential for Raman-based microbial diagnostics, it is important to note some of its limitations. First, since the dataset was collected in a clinical environment, it is unbalanced and is dominated by *K. pneumoniae* isolates. This is important as only a limited number of *K. oxytoca* isolates were used and the study's conclusions should therefore be considered carefully. Although the unbalanced data sets may introduce bias in the SVM models, it is notable that all 3 *K. oxytoca* isolates were classified correctly, showing that the fine differences present in these two species were successfully captured by UVRR spectroscopy.

As expected, the dataset contains some sample-to-sample heterogeneity. It can be observed that the improvement in the classification accuracy after majority voting per biological replicate for the demonstrated classification models is less prominent (Tables S3–S6) than in the results obtained per isolate (three replicates combined), discussed in the main text. This is partially due to

TABLE 3 Summary of classification step at the species level using SC-RMS

SC-RMS		Reference		Balanced accuracy/%
Results per isolate		K. oxytoca	K. pneumoniae	
Prediction	K. oxytoca	118	387	54.8
	K. pneumoniae	374	2318	
SC-RMS		Reference		Balanced accuracy/%
Results per isolate		K. oxytoca	K. pneumoniae	
Prediction	K. oxytoca	0	0	50
	K. pneumoniae	3	15	

Note: The top part presents the results per spectrum, that is, before majority voting. The bottom part presents the results per isolate, that is, after majority voting. The true labels are shown by column and the predicted classes by row.

TABLE 4 Summary of classification step at the species level using UVRR

UVRR spectroscopy		Reference		Balanced accuracy/%
Results per spectrum		K. oxytoca	K. pneumoniae	
Prediction	K. oxytoca	194	31	67
	K. pneumoniae	322	850	
UVRR spectroscopy		Reference		Balanced accuracy/%
Results per isolate		K. oxytoca	K. pneumoniae	
Prediction	K. oxytoca	3	3	90
	K. pneumoniae	0	12	

Note: The top part presents the results per spectrum, that is, before majority voting. The bottom part presents the results per isolate, that is, after majority voting. The true labels are shown by column and the predicted classes by row.

TABLE 5 Summary of findings and advantages of the different Raman spectroscopy approaches

Parameter	UVRR	SC-RMS
Wavelength	244 nm	532 nm
Medium	Bulk (bacterial biomass)	Single Cell
Advantages	Enhancement of signals from nucleic acids and proteins	Does not require extensive cultivation steps
Differentiation between *Klebsiella* and *E. coli* (balanced accuracy)	92%	67%
Differentiation of *Klebsiella* species (balanced accuracy)	90% 3/3 correctly classified *K. oxytoca*	50% 0/3 correctly classified *K. oxytoca*

the natural biological variation known to exist among microorganisms[59, 60] and could be addressed by more vigorously standardized sample preparation in future studies.

It has to be noted that the comparison of the reported results for UVRR and SC-RMS has to be interpreted carefully as the sample preparation for single cell and bulk analysis is different. Moreover, the two methods provide spectra with different spectral resolution which could also affect model accuracy. However, the focus of this study was primarily on examining the diagnostic potential of each method, with its own specialized sample preparation method. Future studies should consider both the spectroscopic setup and the measuring medium (single cell or bulk) in their applications.

4 | CONCLUSIONS

This study demonstrates, for the first time, that Raman spectroscopy allows accurate discrimination of clinical *E. coli* and *Klebsiella spp.* as well as

K. pneumoniae and *K. oxytoca* isolates. UVRR spectroscopy yielded better accuracy for differentiation compared with SC-RMS with 532 nm excitation. This is because the resonance effect of nucleic acids in this excitation wavelength could capture the differences present in the genome of the isolates. In addition, for the first time, Raman spectroscopy was applied on the emerging pathogen *K. oxytoca*, showing that UVRR spectroscopy allows correct classification of all used isolates from this species. The advantages of both Raman spectroscopic approaches used are shown in Table 5.

These findings are indicative of Raman spectroscopy's potential as a diagnostic tool. This study has demonstrated the importance of carefully choosing the experimental parameters. The different wavelengths used in this study provided information on the biochemical composition of the studied genus and species. Future studies need to be designed in order to investigate and optimize these parameters on different taxonomic levels. In addition, studies toward a better understanding of the influence of Raman excitation wavelength need to be established in order to generalize the best diagnostic strategy not only for *Enterobacteriaceae* but also for other pathogenic groups. This study is the first step in this direction and should serve as a guide for the future development of Raman spectroscopy as a diagnostic tool.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data presented in this study are available on request from the corresponding author.

ORCID

Amir Nakar https://orcid.org/0000-0002-1342-7471
Oleg Ryabchykov https://orcid.org/0000-0002-4655-8080
Thomas Bocklitz https://orcid.org/0000-0003-2778-6624
Petra Rösch https://orcid.org/0000-0001-6179-3719
Jürgen Popp https://orcid.org/0000-0003-4257-593X

REFERENCES

[1] M. S. Donnenberg, in *220: Enterobacteriaceae* (Eds: J. E. Bennett, R. Dolin, M. J. Blaser), Saunders, Philadelphia **2015**, p. 2503.

[2] J. D. Lutgring, B. M. Limbago, *J. Clin. Microbiol.* **2016**, *54*, 529.

[3] Y. Matsumura, G. Peirano, M. R. Motyl, M. D. Adams, L. Chen, B. Kreiswirth, R. DeVinney, J. D. D. Pitout, *Antimicrob. Agents Chemother.* **2017**, *61*, e02729.

[4] E. J. Giamarellos-Bourboulis, E. Pyleris, C. Barbatzas, A. Pistiki, M. Pimentel, *BMC Gastroenterol.* **2016**, *16*, 67.

[5] C. Gogos, A. Kotsaki, A. Pelekanou, G. Giannikopoulos, I. Vaki, P. Maravitsa, S. Adamis, Z. Alexiou, G. Andrianopoulos, A. Antonopoulou, S. Athanassia, F. Baziaka, A. Charalambous, S. Christodoulou, I. Dimopoulou, I. Floros, E. Giannitsioti, P. Gkanas, A. Ioakeimidou, K. Kanellakopoulou, N. Karabela, V. Karagianni, I. Katsarolis, G. Kontopithari, P. Kopterides, I. Koutelidakis, P. Koutoukas, H. Kranidioti, M. Lignos, K. Louis, K. Lymberopoulou, E. Mainas, A. Marioli, C. Massouras, I. Mavrou, M. Mpalla, M. Michalia, H. Mylona, V. Mytas, I. Papanikolaou, K. Papanikolaou, M. Patrani, I. Perdios, D. Plachouras, A. Pistiki, K. Protopapas, K. Rigaki, V. Sakka, M. Sartzi, V. Skouras, M. Souli, A. Spyridaki, I. Strouvalis, T. Tsaganos, G. Zografos, K. Mandragos, P. Klouva-Molyvdas, N. Maggina, H. Giamarellou, A. Armaganidis, E. J. Giamarellos-Bourboulis, *Crit. Care* **2010**, *14*, R96.

[6] R. Podschun, U. Ullmann, *Clin. Microbiol. Rev.* **1998**, *11*, 589.

[7] L. Poirel, J. D. Pitout, P. Nordmann, *Future Microbiol.* **2007**, *2*, 501.

[8] P. Nordmann, L. Dortet, L. Poirel, *Trends Mol. Med.* **2012**, *18*, 263.

[9] W. Bereket, K. Hemalatha, B. Getenet, T. Wondwossen, A. Solomon, A. Zeynudin, S. Kannan, *Eur. Rev. Med. Pharmacol. Sci.* **2012**, *16*, 1039.

[10] N. Neog, U. Phukan, M. Puzari, M. Sharma, P. Chetia, *Curr. Microbiol.* **2021**, *78*, 1115.

[11] I. Espagnon, D. Ostrovskii, R. Mathey, M. Dupoy, P. Joly, A. Novelli-Rousseau, F. Pinston, O. Gal, F. Mallard, D. Leroux, *J. Biomed. Opt.* **2014**, *19*, 027004.

[12] D. H. Pincus, *Encyclopedia of Rapid Microbiological Methods*, Parenteral Drug Association, Bethesda, MD **2006**.

[13] A. Rave, A. Kuss, G. Peil, S. Ladeira, J. Villarreal, P. Nascente, *Braz. J. Biol.* **2018**, *79*, 555.

[14] N. AlMasoud, H. Muhamadali, M. Chisanga, H. AlRabiah, C. A. Lima, R. Goodacre, *Analyst* **2021**, *146*, 770.

[15] D. Jonas, B. Spitzmüller, F. D. Daschner, J. Verhoef, S. Brisse, *Res. Microbiol.* **2004**, *155*, 17.

[16] A. Croxatto, G. Prod'hom, G. Greub, *FEMS Microbiol. Rev.* **2012**, *36*, 380.

[17] P. M. Luethy, J. K. Johnson, *J Appl Lab Med* **2019**, *3*, 675.

[18] W. Moussaoui, B. Jaulhac, A. M. Hoffmann, B. Ludes, M. Kostrzewa, P. Riegel, G. Prevost, *Clin. Microbiol. Infect.* **2010**, *16*, 1631.

[19] World Health Organization (WHO) Global Antimicrobial Resistance Surveillance System (GLASS): molecular methods for antimicrobial resistance (AMR) diagnostics to enhance the Global Antimicrobial Resistance Surveillance System, (Editor), World Health Organization, **2019**.

[20] S. Pahlow, S. Meisel, D. Cialla-May, K. Weber, P. Rösch, J. Popp, *Adv Drug Deliv Rev.* **2015**, *89*, 105.

[21] B. Lorenz, C. Wichmann, S. Stöckel, P. Rösch, J. Popp, *Trends Microbiol* **2017**, *25*, 413.

[22] S. Stöckel, J. Kirchhoff, U. Neugebauer, P. Rösch, J. Popp, *J Raman Spectrosc* **2015**, *47*, 89.

[23] C. S. Ho, N. Jean, C. A. Hogan, L. Blackmon, S. S. Jeffrey, M. Holodniy, N. Banaei, A. A. E. Saleh, S. Ermon, J. Dionne, *Nat Commun* **2019**, *10*, 4927.

[24] U. Neugebauer, P. Rösch, J. Popp, *Int. J. Antimicrob. Agents* **2015**, *46*, S35.

[25] N. Ali, S. Girnus, P. Rösch, J. Popp, T. Bocklitz, *Anal. Chem.* **2018**, *90*, 12485.

[26] D. Kusić, B. Kampe, P. Rösch, J. Popp, *Water Res* **2014**, *48*, 179.

[27] S. Stöckel, S. Meisel, B. Lorenz, S. Kloß, S. Henk, S. Dees, E. Richter, S. Andres, M. Merker, I. Labugger, P. Rösch, J. Popp, *J Biophoton* **2017**, *10*, 727.

[28] A. A. Moawad, A. Silge, T. Bocklitz, K. Fischer, P. Rosch, U. Roesler, M. C. Elschner, J. Popp, H. Neubauer, *Molecules* **2019**, *24*, 4516.

[29] B. Lorenz, P. Rösch, J. Popp, *Anal. Bioanal. Chem.* **2019**, *411*, 5445.

[30] S. Kloß, B. Kampe, S. Sachse, P. Rösch, E. Straube, W. Pfister, M. Kiehntopf, J. Popp, *Anal Chem* **2013**, *85*, 9610.

[31] A. Silge, R. Heinke, T. Bocklitz, C. Wiegand, U.-C. Hipler, P. Rösch, J. Popp, *Anal Bioanal Chem* **2018**, *410*, 5839.

[32] S. A. Asher, *Anal. Chem.* **1993**, *65*, 59A.

[33] K. Gaus, P. Rösch, R. Petry, K. D. Peschke, O. Ronneberger, H. Burkhardt, K. Baumann, J. Popp, *Biopolymers* **2006**, *82*, 286.

[34] N. Kunapareddy, J. Grun, R. Lunsford, S. Nikitin, Z. Wang, D. Gillis, *Appl. Spectrosc.* **2015**, *69*, 966.

[35] S. Kloß, P. Rösch, W. Pfister, M. Kiehntopf, J. Popp, *Anal Chem* **2015**, *87*, 937.

[36] C. Wichmann, M. Chhallani, T. Bocklitz, P. Rösch, J. Popp, *Anal Chem* **2019**, *91*, 13688.

[37] Z. Athanasakopoulou, M. Reinicke, C. Diezel, M. Sofia, D. C. Chatzopoulos, S. D. Braun, A. Reissig, V. Spyrou, S. Monecke, R. Ehricht, K. Tsilipounidaki, A. Giannakopoulos, E. Petinaki, C. Billinis, *Antibiotics* **2021**, *10*, 389.

[38] O. Ryabchykov, T. Bocklitz, A. Ramoji, U. Neugebauer, M. Foerster, C. Kroegel, M. Bauer, M. Kiehntopf, J. Popp, *Chemom. Intel. Lab. Syst.* **2016**, *155*, 1.

[39] S. Guo, T. Bocklitz, U. Neugebauer, J. Popp, *Anal Methods* **2017**, *9*, 4410.

[40] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, **2020**.

[41] S. Meisel, S. Stöckel, P. Rösch, J. Popp, *Food Microbiol* **2014**, *38*, 36.

[42] A. Walter, M. Reinicke, T. Bocklitz, W. Schumacher, P. Rösch, E. Kothe, J. Popp, *Anal Bioanal Chem* **2011**, *400*, 2763.

[43] M. Harz, R. A. Claus, C. L. Bockmeyer, M. Baum, P. Rösch, K. Kentouche, H. P. Deigner, J. Popp, *Biopolymers* **2006**, *82*, 317.

[44] A. T. Tu, *Raman Spectroscopy in Biology: Principles and Applications*, Wiley, New York **1982**.

[45] J. de Gelder, K. de Gussem, P. Vandenabeele, L. Moens, *J Raman Spectrosc* **2007**, *38*, 1133.

[46] K. Czamara, K. Majzner, M. Z. Pacia, K. Kochan, A. Kaczor, M. Baranska, *J Raman Spectrosc* **2015**, *46*, 4.

[47] W. E. Huang, M. Li, R. M. Jarvis, R. Goodacre, S. A. Banwart, *Chapter 5: Shining Light on the Microbial World—The Application of Raman Microspectroscopy*, Vol. 70, Academic Press, Cambridge (MA), USA, **2010**, p. 153.

[48] K. Maquelin, C. Kirschner, L. P. Choo-Smith, N. van den Braak, H. P. Endtz, D. Naumann, G. J. Puppels, *J. Microbiol. Methods* **2002**, *51*, 255.

[49] K. C. Schuster, E. Urlaub, J. R. Gapes, *J Microbiol Methods* **2000**, *42*, 29.

[50] N. Uzunbajakava, A. Lenferink, Y. Kraan, B. Willekens, G. Vrensen, J. Greve, C. Otto, *Biopolymers* **2003**, *72*, 1.

[51] N. Töpfer, M. M. Müller, M. Dahms, A. Ramoji, J. Popp, H. Slevogt, U. Neugebauer, *Integr Biol (Camb).* **2019**, *11*, 87.

[52] H. Hata, T. Natori, T. Mizuno, I. Kanazawa, I. Eldesouky, M. Hayashi, M. Miyata, H. Fukunaga, S. Ohji, A. Hosoyama, E. Aono, A. Yamazoe, K. Tsuchikane, N. Fujita, T. Ezaki, *Microbiol Immunol* **2016**, *60*, 303.

[53] A. Nakar, A. Pistiki, O. Ryabchykov, T. Bocklitz, P. Rösch, J. Popp, *Anal. Bioanal. Chem.* **2022**, *411*, 1481.

[54] D. Willemse-Erix, T. Bakker-Schut, F. Slagboom-Bax, J.-W. Jachtenberg, N. Lemmens-den Toom, C. C. Papagiannitsis, K. Kuntaman, G. Puppels, A. van Belkum, J. A. Severin, W. Goessens, K. Maquelin, *J. Clin. Microbiol.* **2012**, *50*, 1370.

[55] D. F. Willemse-Erix, M. J. Scholtes-Timmerman, J. W. Jachtenberg, W. B. van Leeuwen, D. Horst-Kreft, T. C. Bakker Schut, R. H. Deurenberg, G. J. Puppels, A. van Belkum, M. C. Vos, K. Maquelin, *J. Clin. Microbiol.* **2009**, *47*, 652.

[56] I. T. M. A. Overdevest, M. Heck, K. van der Zwaluw, I. Willemsen, J. van de Ven, C. Verhulst, J. A. J. W. Kluytmans, *J Clin Microbiol* **2012**, *50*, 3999.

[57] K. Guyot, V. Biran, C. Doit, D. Moissenet, T. Guillard, L. Brasme, C. Courroux, K. Maquelin, W. van Leeuwen, H. VuThien, Y. Aujard, C. de Champs, E. Bingen, *Eur. J. Clin. Microbiol. Infect. Dis.* **2012**, *31*, 2827.

[58] D. Moradigaravand, V. Martin, S. J. Peacock, J. Parkhill, *Genome Biol Evol* **2017**, *9*, 574.

[59] M. Ackermann, *Nat. Rev. Microbiol.* **2015**, *13*, 497.

[60] J. Bär, M. Boumasmoud, R. D. Kouyos, A. S. Zinkernagel, C. Vulin, *Sci Rep* **2020**, *10*, 16084.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.