**Multi-excitation Raman spectroscopy for label-free, strain-level characterisation of bacterial pathogens in artificial sputum media**

Adam. P. Lister†1,5, Callum. J. Highmore†,2,4, Niall Hanrahan1,5, James Read1,5, Alasdair P. S. Munro3,4, Samuel Tan4, Raymond. N. Allan2,3,6, Saul N. Faust\*3,4,5, Jeremy. S. Webb\*2,5 and Sumeet Mahajan\*1,5

1: School of Chemistry, Faculty of Engineering and Physical Sciences, University of Southampton, SO17 1BJ, Southampton, United Kingdom

2: School of Biological Sciences, Faculty of Environmental and Life Sciences, University of Southampton, SO17 1BJ, Southampton, United Kingdom

3: NIHR Southampton Clinical Research Facility and Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust, Southampton, UK SO16 6YD

4: Faculty of Medicine and Institute for Life Sciences, University of Southampton, Southampton, United Kingdom

5: National Biofilms Innovation Centre (NBIC) and Institute for Life Sciences, University of Southampton, Southampton, United Kingdom

6:School of Pharmacy, Faculty of Health and Life Sciences, De Montfort University, Leicester, LE1 9BH, UK

\*Corresponding author(s): Sumeet Mahajan, [s.mahajan@soton.ac.uk](mailto:s.mahajan@soton.ac.uk); Jeremy S Webb, [j.s.webb@soton.ac.uk](mailto:j.s.webb@soton.ac.uk)

**Abstract**

The current methods for diagnosis of acute and chronic infections are complex and skill-intensive. For complex clinical biofilm infections, it can take days from collecting and processing a patient’s sample to achieving a result. These aspects place a significant burden on healthcare providers, delay treatment and can lead to adverse patient outcomes. We report the development and application of a novel multi-excitation Raman spectroscopy based methodology for the label-free and non-invasive detection of microbial pathogens that can be used with unprocessed clinical samples directly and provide rapid data to inform diagnosis by a medical professional. The method relies on the differential excitation of non-resonant and resonant molecular components in bacterial cells to enhance the molecular finger-printing capability to obtain strain level distinction in bacterial species. Here we use this strategy to detect and characterise the respiratory pathogens *Pseudomonas aeruginosa* and *Staphylococcus aureus* as typical infectious agents associated with cystic fibrosis. Planktonic specimens were analysed both in isolation and in artificial sputum media. The resonance Raman components, excited at different wavelengths, were characterised as carotenoids and porphyrins. By combining the more informative multi-excitation Raman spectra with multivariate analysis (support vector machine) the accuracy was found to be 99.75% for both species (across all strains), including 100% accuracy for drug-sensitive and drug-resistant *S. aureus.* The results demonstrate that our methodology based on multi-excitation Raman spectroscopy can underpin the development of a powerful platform for the rapid and reagentless detection of clinical pathogens to support diagnosis by a medical expert, in this case relevant to cystic fibrosis. Such a platform could provide translatable diagnostic solutions in a variety of disease areas and also be utilised for the rapid detection of anti-microbial resistance.

**Introduction**

Across clinical scenarios the diagnosis of bacterial infection is dependent on traditional culture techniques and, in the case of biofilms, fluorescence *in-situ* hybridisation 1. This is primarily due to diagnostic samples comprising complex biological materials and heterogeneous microbial communities. The methodologies used are labour, cost, and time intensive, thereby delaying effective treatment by days.

The shortcomings in medical diagnostic capabilities are exemplified in cystic fibrosis (CF), where diagnosis of infection relies on the manipulation of complex sputum samples. CF is an autosomal recessive genetic disorder, affecting 1.37 births per 10,000 in the UK 2. Affected individuals are extremely susceptible to bacterial infection of the lower respiratory tract due to impaired innate immune function in the lung and the consequent overproduction of mucus 3. Even with the advent of new disease modifying treatment, recurrent infections cause a dramatic reduction of quality and length of life4.

The pathophysiology of the CF lung impedes the administration of effective treatment. Intensive antibiotic treatment is reported to be beneficial in eradicating *Pseudomonas aeruginosa* at its early stages of colonisation 5-6. When established in the CF lung, there is evidence that *P. aeruginosa* exists as biofilms, microbial communities protected by a self-produced polymeric matrix 7-8. The protection afforded to bacteria by a biofilm reduces the efficacy of antibiotic treatment 9, and is a driver of antimicrobial resistance 10, so effective and rapid treatment is of great importance. In CF, *P. aeruginosa* lineages display variation in virulence and antibiotic sensitivity 11 and undergo recombination events to further increase phenotypic diversity and antibiotic resistance 12. Taken together, these attributes necessitate a rapid and specific alternative to the current slow and resource intensive methods of diagnosing infection in CF patients.

Raman spectroscopy offers many advantages over resource intensive culture-based methods and techniques such as fluorescence, allowing rapid and label-free analysis. It is advantageous over techniques such as ELISA, mass spectrometry, and infrared spectroscopy as well because it is reagentless, avoids complex sample-preparation steps, does not require sophisticated equipment and is water-insensitive. For Raman analysis, bacterial samples require no culturing step, offering results within minutes while culture methods typically require 24 - 48 hours before detection of a specific biomarker can be achieved. Raman spectroscopy relies on the inelastic scattering of light to probe the molecular vibrations present in a sample. This allows ‘fingerprinting’ that can be used to identify molecular, biotic and abiotic components in a sample.

Previously, Raman spectroscopy has been used to examine microbiological samples in clinical scenarios. Kloß *et al* were able to characterise respiratory pathogens at the species level using Raman spectroscopy 13, and Ghebremedhin *et al* used the technique to differentiate between 31 clinical isolates of *Acinetobacter baumannii* at the strain level 14. Raman spectroscopy has been applied to analysis in microbiology, e.g. the detection of human pathogens inoculated into ascitic fluid 15; however, this can often involve long acquisition times that slow down analysis because of the relative scarcity of Raman-scattering events.

Surface-Enhanced Resonant Raman scattering (SERS) offers advantages over spontaneous Raman in terms of speed and an improved limit of detection due to enhancement of signals by electromagnetic and/or chemical mechanisms typically observed with nanoscale metallic materials. SERS has been used to distinguish *Escherichia coli* isolates based on their sensitivity to carbapenem antibiotics 16, and to identify common CF pathogens in pellets with silver nanoparticles 17, and to map *P. aeruginosa* colonies by acquiring SERS spectra via laser-scanning and using the intensity of a CH bond associated with pyocyanin 18-22. Despite the capabilities and advantages of SERS, it requires the introduction of exogenous materials, such as metallic nanoparticles, to the sample to provide signal enhancement. Additionally, this exogenous material must interact closely and reproducibly with the analyte, and the reproducibility and sensitivity can be severely affected by any variations in the number or intensity of plasmonic hotspots, or from changes to the frequency of the plasmon resonance arising from different degrees of aggregation, as might be seen with colloidal gold nanoparticles.

Here, we wanted to develop a method that was rapid and did not require the use of nanoscale materials for enhancing signals. We utilised the principle that the cross-section of Raman-active modes varies with wavelength, and improves significantly, as the excitation nears pre-resonance or resonance with an electronic state of the sample. The use of resonant Raman spectroscopy has been applied to the study of cytochrome *cd1* from bacteria 23, and UV resonance has been applied to whole bacteria and endospore biomarkers 24. The use of the phenomena of resonance Raman with multiple excitation wavelengths in the deep UV spectral region in explosive materials through an interplay of resonant enhancement, self-absorption and laser penetration depth has also been demonstrated25. While the aforementioned use resonance enhancements, it is well known that pre-resonant enhancements can also provide large boosts in the intensity of certain vibrational modes within a Raman spectrum 26.

**Figure 1:** The envisaged workflow for a device using multi-excitation Raman detection technology. A patient arrives with a complaint and a biofluid sample is taken. The sample is prepared by a method analogous to a common blood smear, and placed into the device for Raman analysis and subsequent classification. This classification then informs the clinician’s final diagnosis and treatment options. Inset bottom centre: Representation of the Raman analysis. Raman spectra are taken at multiple excitation wavelengths and concatenated into a spectral matrix. This spectral-matrix is then fed into a trained PCA or SVM (or equivalent classification technique) model of known spectra for classification.

Treatment

Diagnosis

Sample

preparation

Patient arrival

& sample collection

Analysis

E1

Measurement variable [n]

Response

Identification

E2

R1

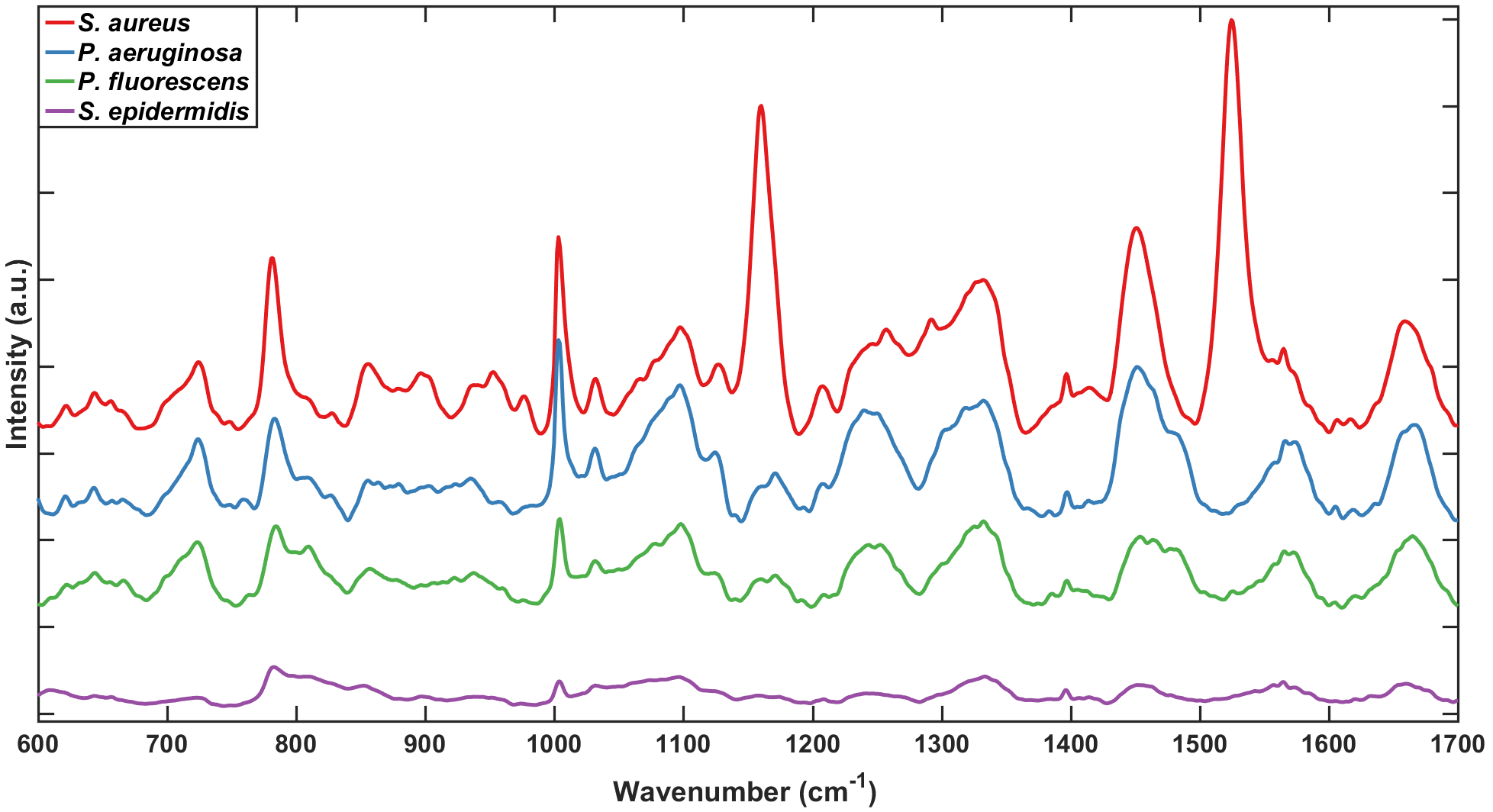
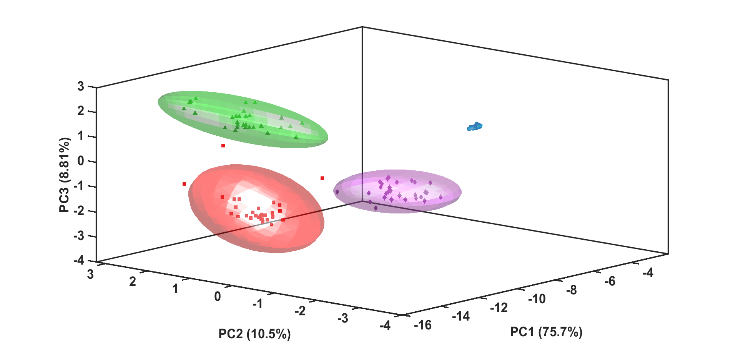
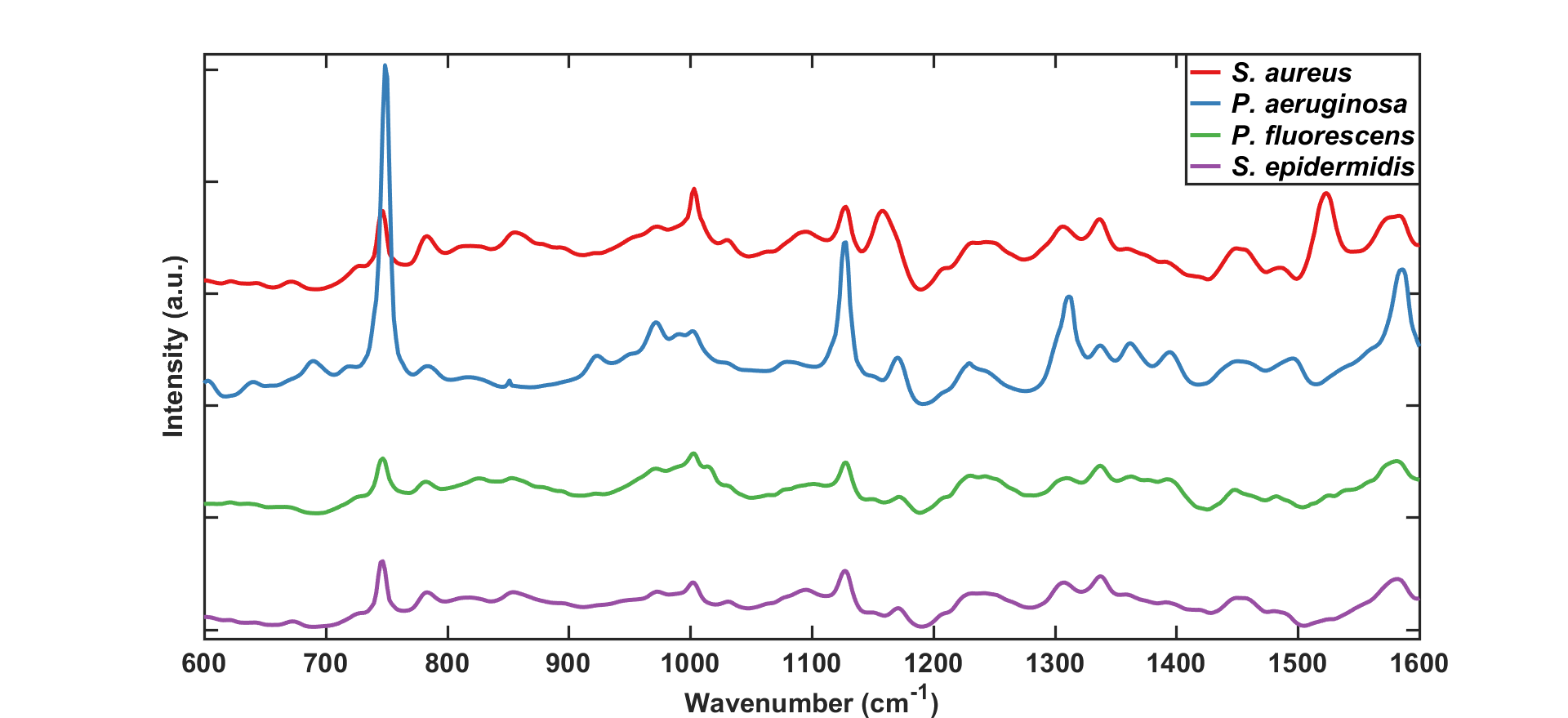
R2

X

Y

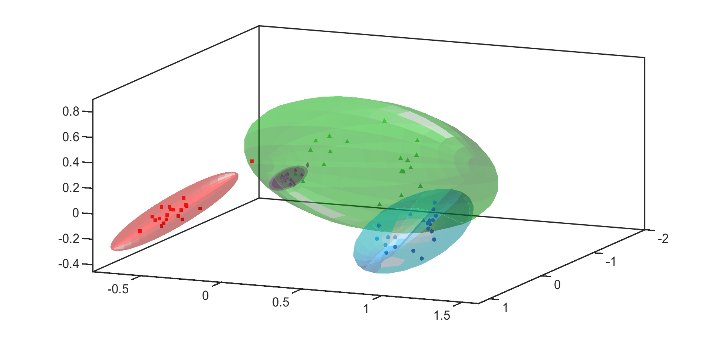
Z

We were additionally motivated by the need to increase the distinction capability of Raman spectral analysis. This would be highly desirable in cases where the differences in Raman spectra are less prominent, e.g. due to different strains of the same bacterial species. While new multivariate methods are increasingly being developed, there is an inherent need to increase the amount of useful information to allow distinction in complex samples with subtle differences. Hence, here we developed a multiple-excitation strategy, acquiring resonant, pre-resonant and non-resonant Raman spectra in the visible and near infrared spectral ranges. This approach allowed us to generate new information (different peak intensities) at the same variables (Raman shifts) due to variation in Raman cross-sections of different vibrations as a function of excitation wavelength. We combined this data with Support Vector Machine (SVM) for increased accuracy in the detection and strain-level identification of common bacterial respiratory pathogens in complex media. Our study demonstrates a step towards a rapid and reagentless diagnostic tool for clinicians, utilising simple and routine sample preparation that could be carried out on a ward. The envisaged workflow of such a device is shown in **Figure 1**. For this work, *Staphylococcus aureus* and *Pseudomonas aeruginosa* were used as Gram-positive and Gram-negative representatives of the primary pathogens present in cystic fibrosis in children and adults, respectively.



**1000 counts**

**A**



**PC1 (57.1%)**

**PC2 (30.5%)**

**PC3 (3.86%)**

**B**

**C**

**D**

**5000 counts**

**Figure 2:** **(A)** Offset, class mean spontaneous Raman spectra of four bacterial species at 785nm excitation: Staphylococcus aureus, Staphylococcus epidermidis, Pseudomonas aeruginosa, and Pseudomonas fluorescens. **(B)** Plot of the first three principal components, showing the clustering of bacterial spectra from **A** along species lines. Coloured envelopes represent the 95% confidence envelope for the corresponding species. Clear grouping based on species is visible. **(C)** Offset, class mean spontaneous Raman spectra of four bacterial species at 532nm excitation: Staphylococcus aureus, Staphylococcus epidermidis, Pseudomonas aeruginosa, and Pseudomonas fluorescens. **(D)** Plot of the first three principal components, showing the clustering of bacterial spectra from **C** along species lines. Coloured envelopes represent the 95% confidence envelope for the corresponding species. Clear grouping based on species is visible.

**Materials and Methods**

Bacterial samples and preparation

Bacteria used were *S. aureus* (SA)strains NCTC 10442, NCTC 11939, NCTC 13143 (EMRSA-16), and ATCC 49230, and *P. aeruginosa* (PA)strains PAO1, PA21, PA30, and PA68 (REC No: 08/H0502/126). *Staphylococcus epidermidis* ATCC 35984 and *Pseudomonas fluorescens* NCTC 13525 were species used for genus differentiation analyses. All bacterial strains were grown in LB broth (Formedium, UK) for 18 hours at 37oC, with shaking at 130 rpm. *Staphylococcus* species were cultured with aeration.

Bacterial cultures were washed twice in water by centrifugation at 4000 g for 10 minutes in a Heraeus Megafuge centrifuge. The resulting pellet was applied to a fused quartz microscope slide (UQG Optics, UK) and dried by gently heating, for subsequent spectroscopic analysis. Artificial sputum media (ASM) was prepared by adapting the methodology described by Sriramulu 27 (mucin from pig stomach mucosa 5g/l, salmon sperm DNA 4g/l, diethylene triamine pentaacetic acid 5.9mg/l, NaCl 5g/l, KCl 2.2g/l, Tris Base 1.81g/l, egg yolk emulsion 5ml/l, casamino acids 5g/l. Bacterial cultures were washed as above, and the pellet was mixed with ASM at a 1:1 ratio v/v before applying to the fused quartz slide.

Raman microspectroscopy

Raman microspectroscopy experiments were conducted using a Renishaw InVia Raman microscope (Renishaw, UK), with a Leica DM 2500-M bright field microscope and an automated 100nm encoded XYZ stage. Samples were excited using  532nm and 785nm lasers directed through a Nikon 100x air objective (NA = 0.85), with collection after a Rayleigh edge filter appropriate to each excitation wavelength, and a diffraction grating (2400l/mm) that dispersed the Raman scattered light onto a Peltier cooled CCD (1024pixels x 256pixels). Calibration of the Raman shift was carried out using an internal silicon wafer. Spectra were acquired over three accumulations of 20 seconds.

Raman spectral processing and multivariate analysis

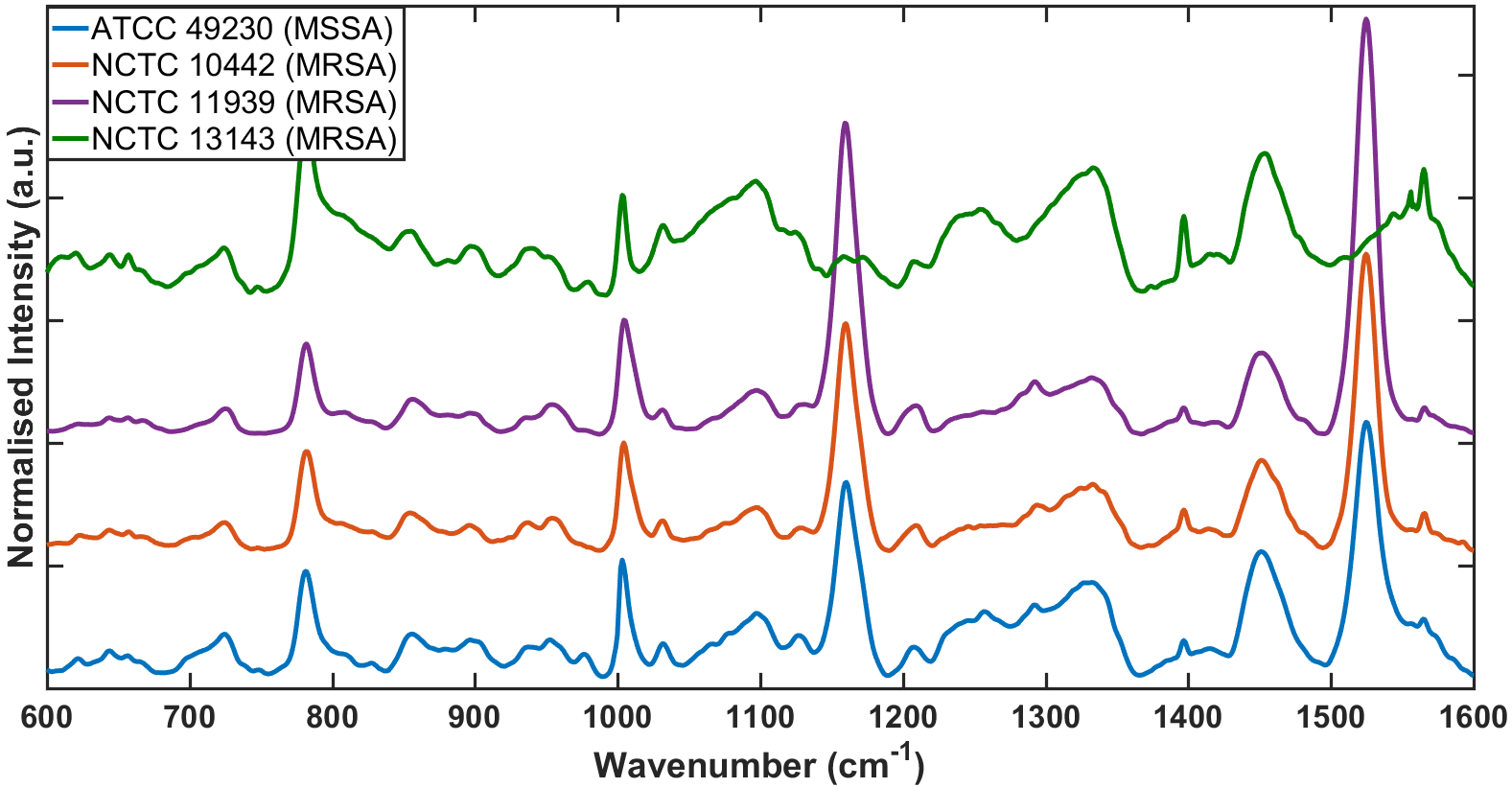
All spectra were cleared of cosmic rays prior to analysis using Renishaw Wire 3.1 software, and then imported into iRootLab version 0.17.8.22-d for Matlab 28 for further processing. Spectra were truncated to the 600-1600cm-1 spectral region and then wavelet denoised to smooth them. To retain fluorescence (background) information, spectra were not background subtracted, but were normalised to their maximum intensity. PCA was applied to all datasets to reduce the dimensionality of the dataset. Either the raw data (for SVM), or the first ten PCs arising from PCA (for PCA-SVM) were fed into a SVM for classification, and ten-fold cross validation was used to validate the classifiers. 240 spectra were used to train the SVM classified (30 for each strain), and the same number of spectra were used for validation. For this study, iRootLab’s in-built PCA, SVM, and k-fold cross-validation functionality were applied to the processed spectra to classify the bacteria by strain. In the case of SVM, the default iRootLab parameters for c and gamma (c = 1, gamma = 1) were used. The full step-by-step methods for the PCA and SVM have been included as **Figure S1**.

**Results & Discussion**

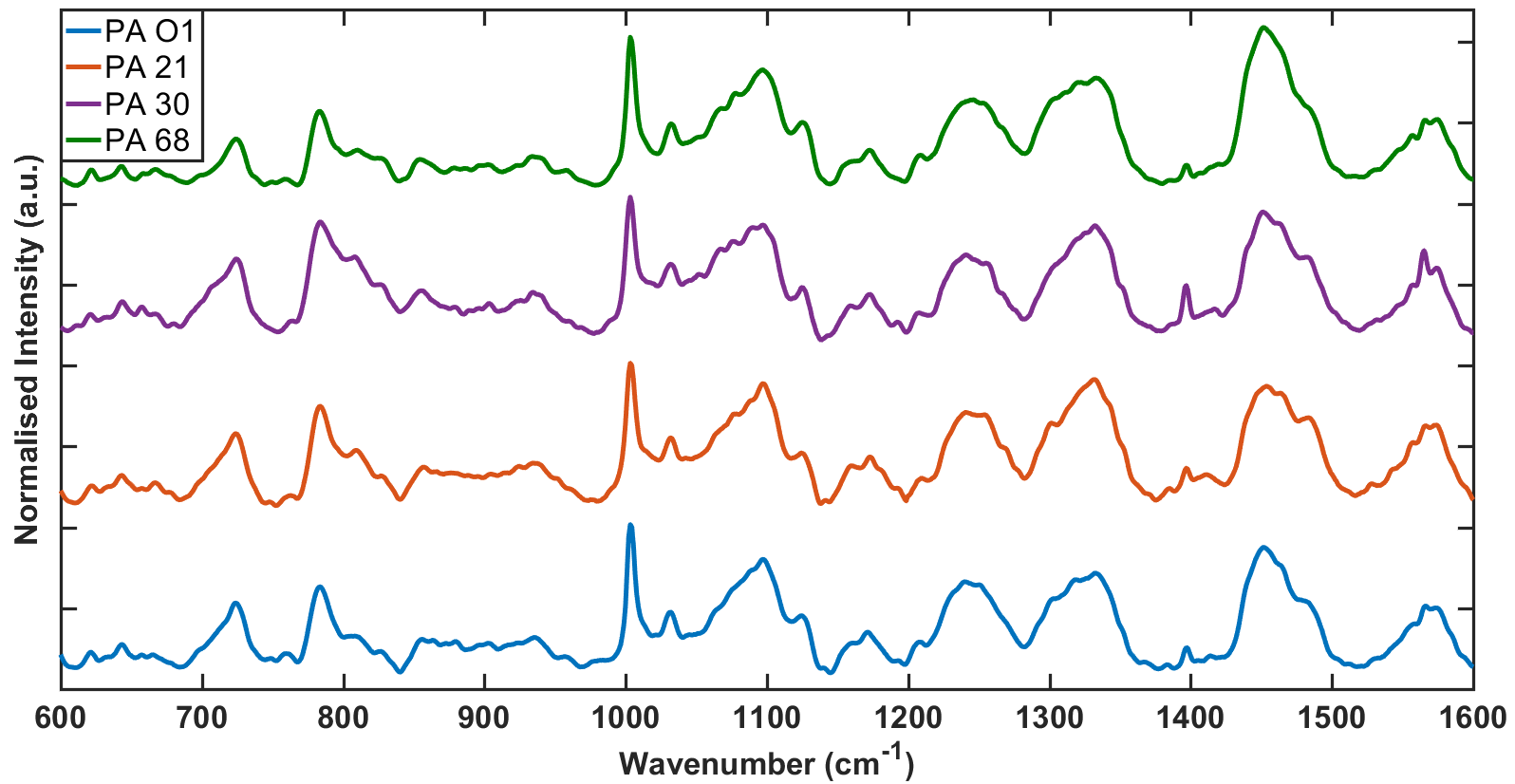
Raman spectroscopy has been shown to be able to readily differentiate between bacteria at the species and genus level 14-15. Using a single excitation wavelength, species often exhibit a large degree of spectral similarity, with many common features arising from shared biomolecules, such as DNA and amino acids. This can be seen for 532nm and 785nm excitation in **Figure 2**. Many bands are common to the spectra obtained with each respective excitation. Despite this similarity, it is possible to easily discriminate between spectra via the use of relatively simple chemometric methods, such as unsupervised PCA. **Figure 2 B** and **D** visually demonstrate this clear clustering of spectra in 3D PCA space based on genus and species. The clustering is better with spectra acquired with 532nm compared to 785nm excitation.

We wanted to increase the spectral information content to improve identification and classification abilities. Hence, we utilised the differential enhancement of different peaks with different laser excitations. During experimentation (as also apparent in Figure 2) we found that Raman spectra with a 785nm excitation laser yielded very strong peaks in SA that were apparently absent in the spectra of PA. On using a 532nm laser, a different set of peaks were of very high intensity in PA but were significantly weaker in the four SA strains. The background subtracted spectral data is shown in **Figure 3** and the differentially excited strong peaks are highlighted. Raw spectra for all three wavelengths used in the study can be found in **Figure S2** and the influence of each pre-processing step can be seen in **Figure S3**. Further, differences in the peak ratios can be seen when the excitation wavelength is changed.

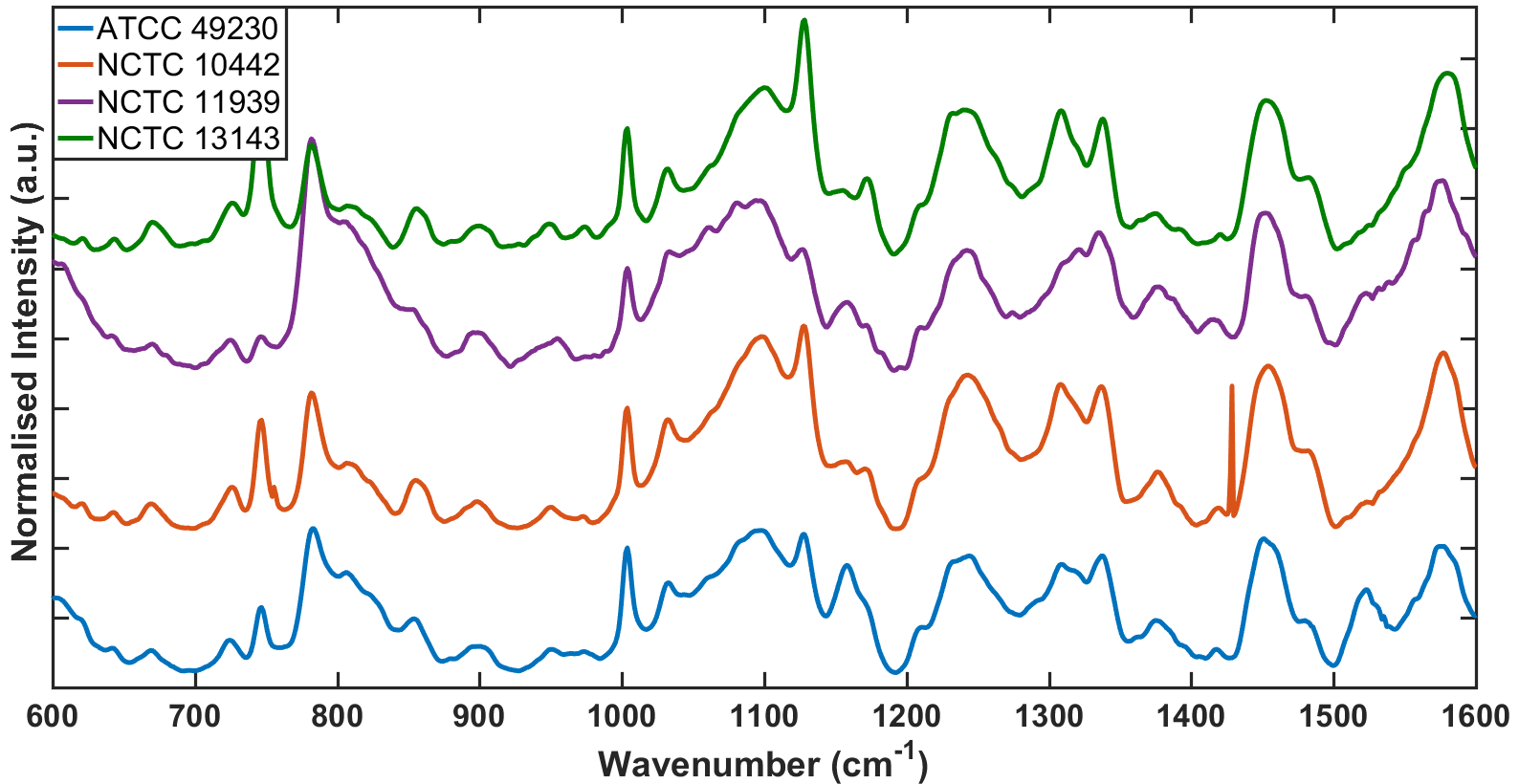
**Figure 3:** Normalised and offset spontaneous Raman spectra of **(A)** S. aureus strains at 785nm excitation **(B)** S. aureus strains at 532nm excitation **(C)** P. aeruginosa strains at 785nm excitation, and **(D)** P. aeruginosa strains at 532nm excitation. Grey highlights show regions of significant peak enhancement for each spectrum, illustrating the change in the magnitude of peaks at 785nm excitation for S. aureus, and at 532nm excitation for P. aeruginosa.



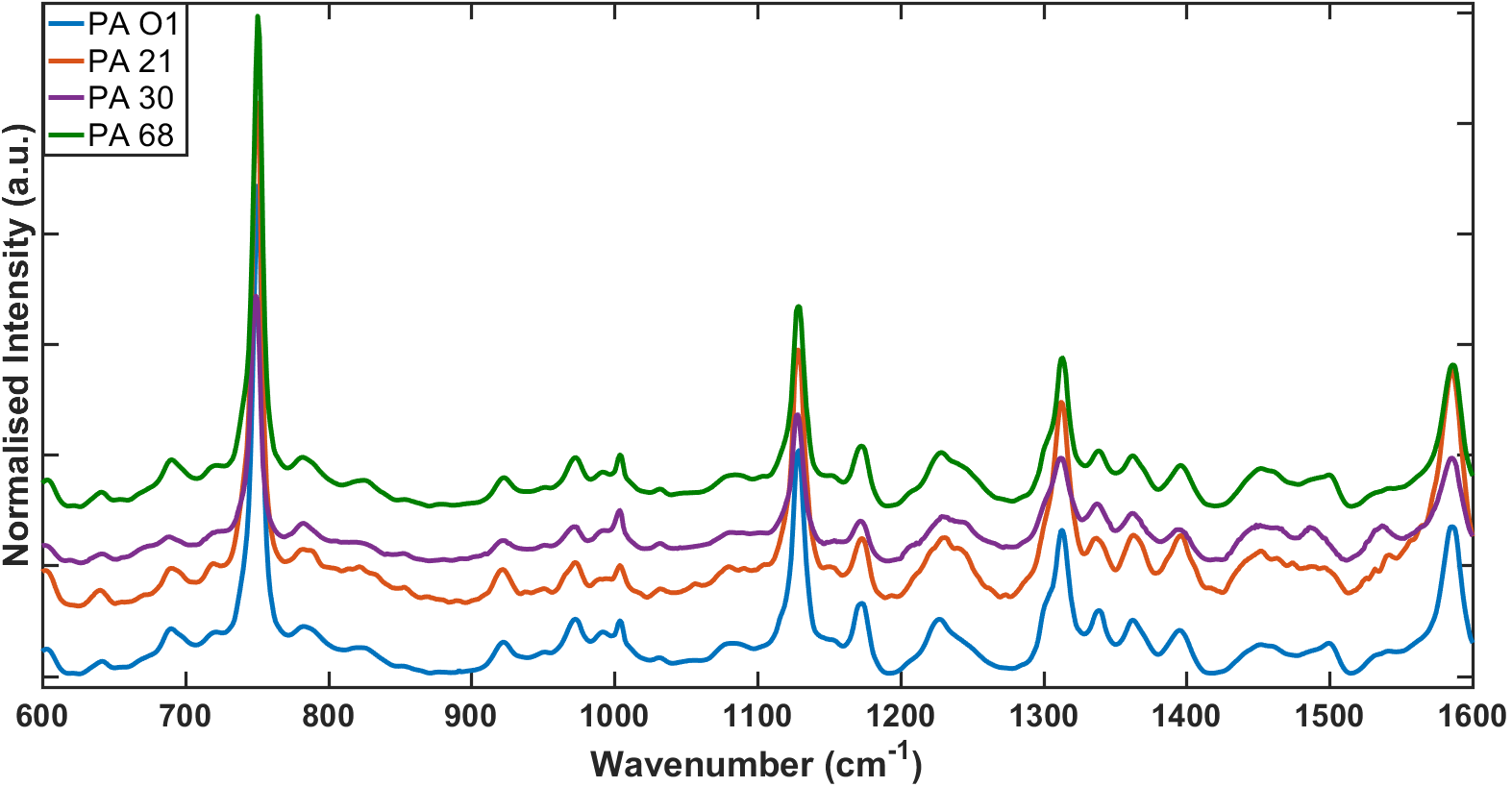
A



C



B

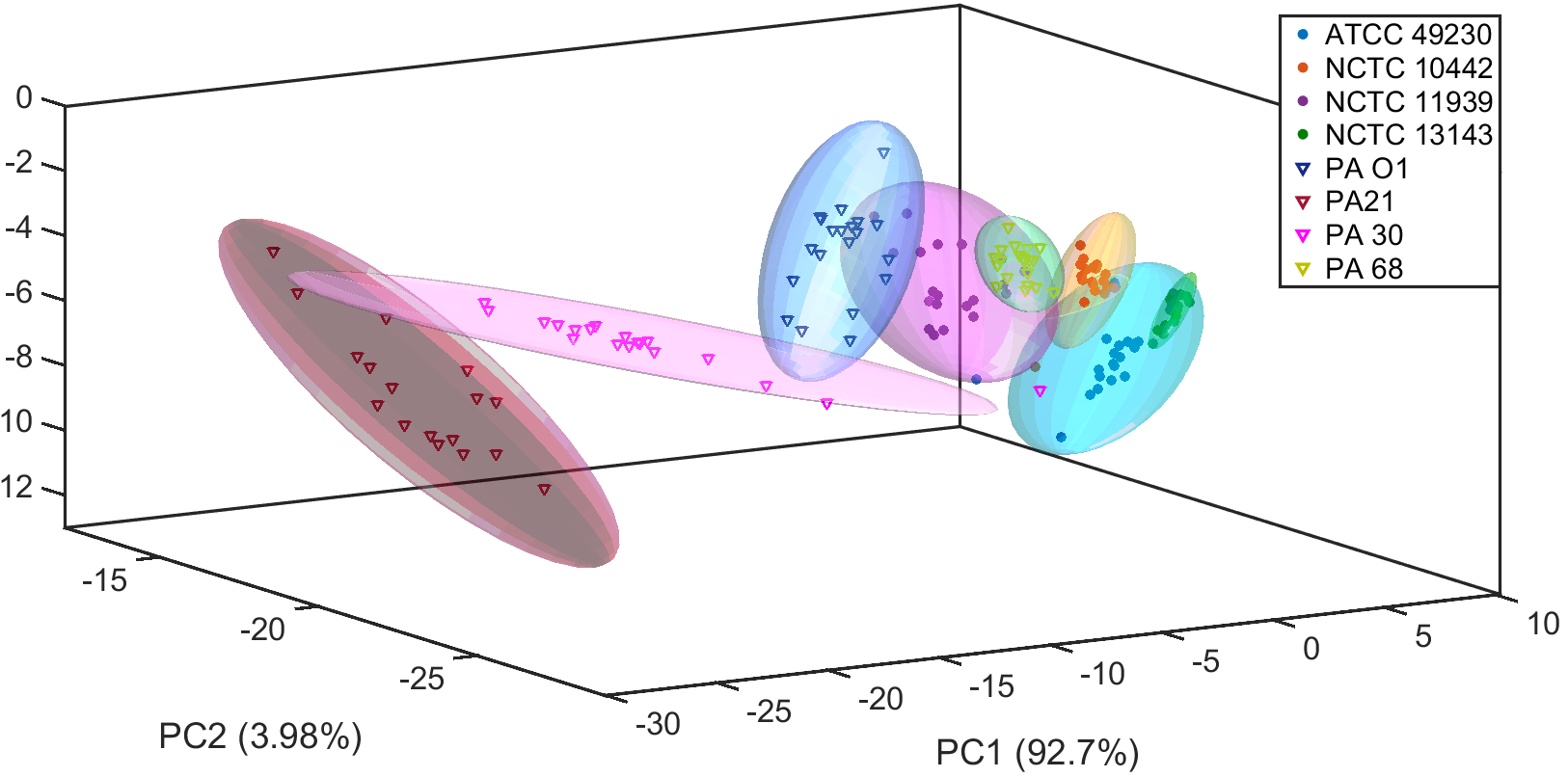
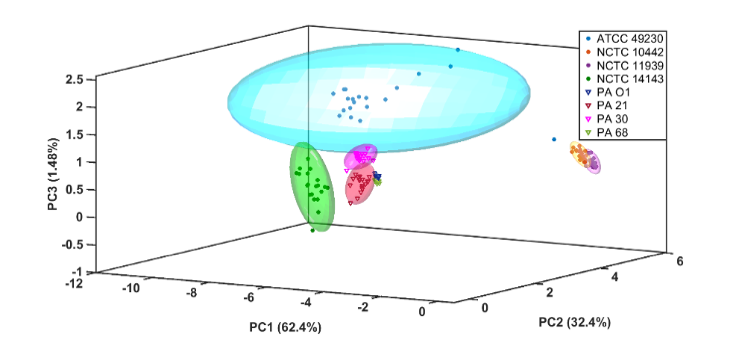
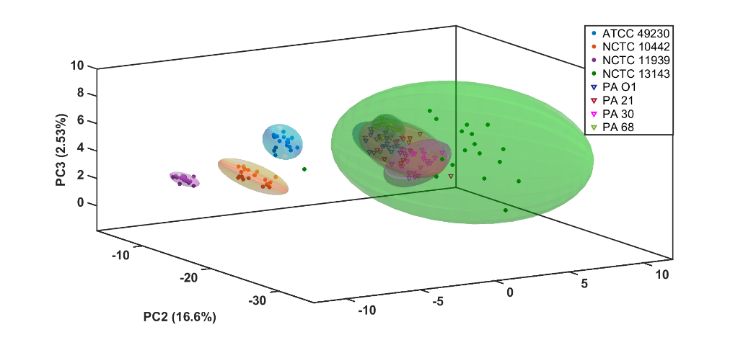
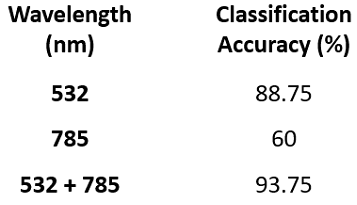


D

This result is explained by the wavelength dependence of the Raman cross-section () of vibrational modes. The enhancement of peaks based on excitation wavelength is due to the resonance or pre-resonance affect, which was first shown by the Albrecht A-term pre-resonance approximation equation for totally symmetric transitions 29.

*K* is a collection of frequency independent factors of Raman molecules, *ω0* is the frequency of the molecular absorption peak, *ωvib* is the vibrational transition energy and *ωPu* is the pump or excitation frequency. Most Raman experiments are performed under non-resonant conditions, when the pump excitation is much lower in energy from the absorption peak (*ω0*). This equation makes it clear that when the pump wavelength is near to, or at, the absorption maximum the Raman cross-section increases, and therefore the signal increases, giving the pre-resonance or the resonant effect, respectively.

The differentially excited peaks, whether due to resonant/pre-resonant or non-resonant excitation, occur in different spectral regions and have different molecular origins. We attribute the strongly-enhanced vibrational modes around 750cm-1 and 1120cm-1 to ring breathing and half ring modes of porphyrins, which is consistent with the findings of Deng 30, and other groups that have probed porphyrin ring vibrations in *Pseudomonas* bacteria 31-32. These are also consistent with our experimental spectra of two different porphyrins (**Figure S4**) which clearly show peaks in these regions confirming that our assignment is consistent with resonance Raman studies on porphyrins 33-34. Similarly, the molecular origins of the enhanced bands at 1150cm-1 and 1520cm-1 in SA are attributed to C-C and C=C vibration modes present in carotenoids. This is also consistent with experimental spectra of typical carotenoids (**Figure S4**), as well as published spectra by others, such as Naumann and Haung, who have observed these bands in biological samples 35-36. In this instance, we ascribe the carotenoid peaks specifically to the presence of the pigment, staphyloxanthin, which is responsible for the yellow colour of SA strains. Further support for this assignment comes from the absence of the peaks in the spectra associated with strain NCTC 13143, which is derived from the drug resistant epidemic strain, EMRSA-16. This strain is white in appearance, and is known not to express staphyloxanthin **(Figure S5)**37. A full assignment of peaks present in the spectra presented here is provided in **Figure S6**.



**A**

**B**

**C**

**D**

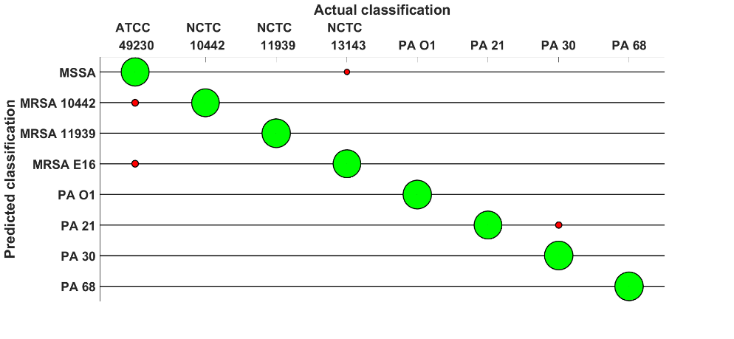
**Figure 4: (A)** Projection of the first three principal components for the spontaneous Raman spectra of 4 strains of S. aureus and 4 strains of P. aeruginosa taken with 785nm excitation. Good separation of the classes belonging to S. aureus, but poor separation of P. aeruginosa (**B)** Projection of the first three principal components of the spontaneous Raman spectra of 4 strains of S. aureus and 4 strains of P. aeruginosa taken with 532nm excitation. **(C)** Combination of the data used for A and B, showing improved separation. **(D)** Table of classification accuracies for projects A-C.

Of interest in these spectra is the observation of carotenoid pre-resonance at 785nm, and its absence at 532nm. It is common in studies of carotenoid species to use excitations in the range of 488nm, as these are known to provide strong enhancement of these molecules; however, the broad absorbance of many carotenoids extends up to around 500nm wavelengths 38-40, and should therefore be pre-resonant with 532nm excitation 38. The use of 785nm excitation for this purpose is uncommon, but large Raman peaks associated with carotenoids in SA and a narrow subset of strains of the related *S. epidermidis* have been detected at 785nm 41.

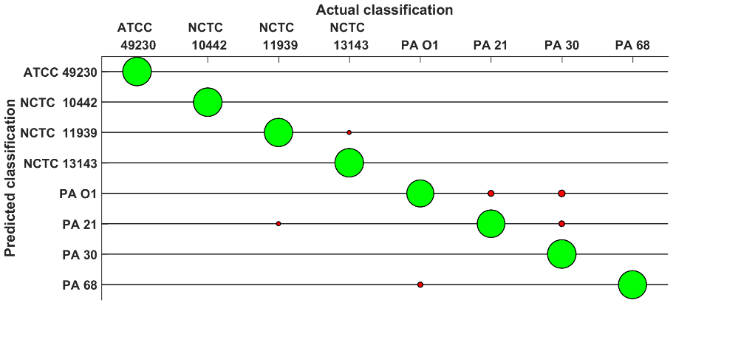
In order to determine whether on using the combination of resonant/pre-resonant and non-resonant excitation there was an improvement in distinction ability, we carried out classification of spectra of different bacterial strains. We recorded as successful classification when a spectrum was within the 95% confidence interval for a strain. Any spectra that appeared outside of the confidence or within multiple confidence intervals of the same or different strains were regarded as unsuccessful. When the spectra were subjected to PCA, it was found that the data with 785nm excitation alone gave clear separation of the four strains of SA but poor separation of PA, leading to a poor classification accuracy of 60% across the 8 strains tested(**Figure 4A**). Spectra taken at 532nm excitation alone gave better separation of PA strains than of SA, but still relatively low classification accuracy, at 88.75% (**Figure 4B**). We then combined the two sets of spectra together in an effort to improve the classification accuracy. Both sets of normalised spectra were concatenated and PCA was performed. It was found that classification accuracies improved to 93.75% (**Figure 4C**). Additionally, in this analyses it was found that accuracy of classification was higher when spectra were classified without polynomial background subtraction (**Figure S7**), indicating that the fluorescent background present in the Raman spectra encodes information that is useful in improving classification accuracy. We attribute the decrease in accuracy after background subtraction to the fact that the luminescent background is unique to each strain (As seen in **Figure S2**), and excitation wavelength. Whilst there is typically broad similarity in the shape of the fluorescence, differences in the background within a species can be indicative of differences in the strain. This is the case with SA NCTC 13143, which shows a different background from other SA strains at the low wavenumber end of the spectrum. Additionally, the level of fluorescence varies between species and strains. Classification of the bacterial spectra based on just the background actually yields a classification of 59.38%, which is comparable to 785nm Raman spectra (**Figure S8**). Hence, the luminescence provides unique information that is lost upon background subtraction.

For diagnostic applications, an accuracy of 100% is highly desirable, whereas with PCA of the combined data we achieved 93.75%. Hence, in order to improve classification accuracy, we used a more sophisticated classifier. We used support vector machine (SVM), and attempted PCA-SVM as a point of comparison. For PCA-SVM, ten principle components were established from the eight strains, at individual excitation wavelengths and the combined multi-excitation data, and these were fed into the SVM for classification. For SVM, 10-fold cross validation was used.

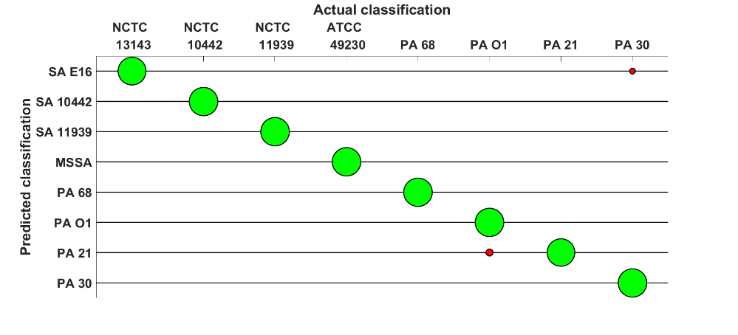
**Figure 5:** Classification accuracies for SVM of pure pellets of bacterial strains at **(A)** 785nm excitation, **(B)** 532nm excitation, and **(C)** for concatenated spectra (785 & 532nm), showing improved classification accuracy for concatenated spectra, particularly in the case of P. aeruginosa strains. The size of green balls correspond to the number of correct identifications for that species. Red balls indicate incorrect classifications.



A



B

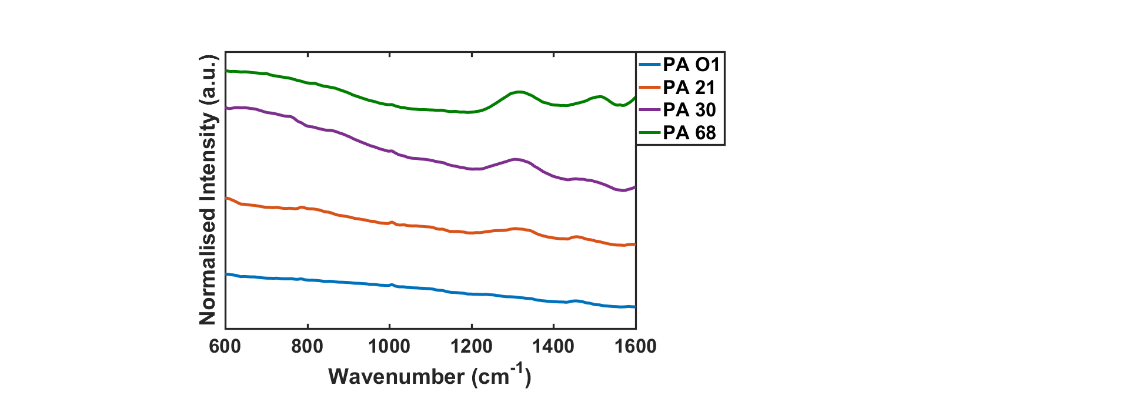


C

The results for the SVM classification can be seen in **Figure 5**. In each figure, the areas of the circles are proportional to the number of samples classified as that strain, with red circles representing misclassifications, and green circles showing correct classifications. Results for PCA-SVM offered lower accuracy than for SVM, and are presented in the SI as **Figure S9**. This loss of accuracy relative to pure SVM is attributed to benefits of extra data points being available to SVM to aid in classification. In PCA, the data set only comprised ten principle components, whereas the number for the raw SVM is closer to 2500 observations at different wavenumbers, offering the trainer far more to learn from. Using this improved method, classifications improved in all cases, with 785 nm achieving drastic improvements (97.52%), and smaller improvements for 532 nm (97.38%). Combining the two datasets also offered some improvement over single excitation wavelength, achieving an accuracy of 98.64%. While the overall accuracy improvement is marginal, it is notable that there are fewer misclassifications with the combined excitation analysis. There were minimal inter-species classification errors, allowing us to preserve the clinically relevant distinction between the Gram-positive (SA) and Gram-negative (PA) species chosen. Intra-species accuracies were also high, which is particularly important for pathogens such as SA. Only one of the four SA strains (ATCC 49230) shows methicillin sensitivity, and is reliably classified as distinct from its methicillin-resistant counterparts. This promising result indicates that the method can potentially identify drug resistance phenotypes in bacterial samples. While further work is required to confirm this, related Raman techniques have previously been used to differentiate between carbapenem resistant and sensitive *E. coli* isolates 16. There is greater genetic diversity across methicillin-sensitive *S. aureus* isolates than there is across MRSA lineages 42-43, which we anticipate will support identification of antibiotic resistance by the described methodology.

**A**

**B**



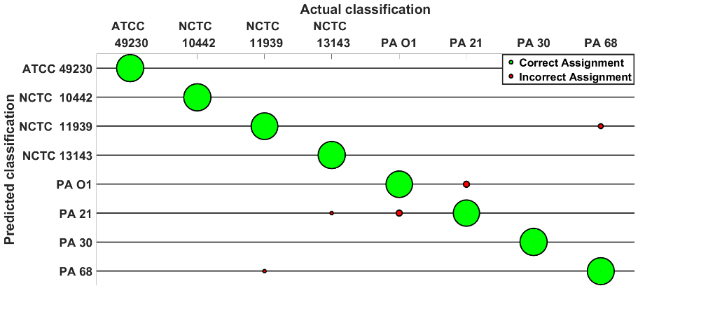
**Figure 6:** Class mean Raman spectra at 785nm excitation, after subtraction of the ASM spectrum for **(A)** S. aureus strains in ASM, and **(B)** P. aeruginosa strains in ASM. The spectra were normalised at 1300cm-1 before subtraction and are offset for clarity.

As sputum is not a sterile sample of pure bacteria, it was necessary to validate the SVM method in a more complex medium. Spectra were acquired with the bacterial strains mixed into artificial sputum medium (ASM) as described in the methods, which was originally formulated to replicate the biochemical composition of CF sputum 44. Our choice of this medium was further informed by the use of this formulation in a variety of microbiological studies pertaining to PA strains 45-46. These mixtures were prepared for analysis using a blood smear method, which is a common sample preparation method used in clinical laboratories (**Figure S10**). Spectra for both ASM and the bacteria-sputum mixtures are included as **Figure S11**. A reduction in the level of enhancement for the pre-resonant vibrational modes was observed in these bacteria-sputum samples, suggesting peaks may have been masked, or that there is a chemical process occurring with the resonant molecules; however, it is unclear what may be causing this effect, as the media does not contain any oxidising or reductive agents that would react with the molecules. Diethylenetriamine pentaacetic acid is present in ASM, and is widely used as a chelating agent for metals. The cytochrome c peroxidase in cells contains two iron atoms, but they are already coordinated to the haem groups of the enzyme, making any interaction with the diethylenetriamine pentaacetic acid unlikely. The broad spectral feature around 1300 cm-1 corresponds to vibrational modes observed in the spectrum of pure ASM. Despite the apparent degradation in spectral quality, the subtle differences remained sufficient for classification. We also subtracted the mean blank ASM spectrum from the mixed-sample spectra to see if the characteristic SA features corresponding to carotenoid peaks remained present and were masked by the ASM spectra. The results are presented in Figure 6, and show that the peaks are present. It is also possible that Raman spectra of the bacteria could be recovered from the mixture using a spectral unmixing technique, such as Multivariate Curve Resolution-Alternating Least Squares, which have been applied to complex biological samples with overlapping peaks47-48.

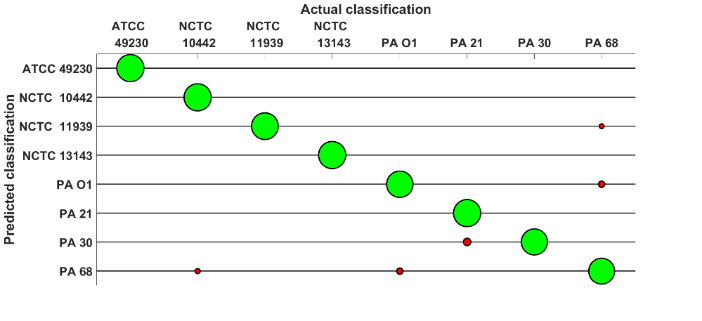
As before, SVM was performed on both the single- and combined-excitation data to determine the performance of the technique on the more realistic samples, and the findings are presented in **Figure 7**. The method provided very high classification accuracies, even in the presence of the interfering media. A total classification accuracy of 94.38% was achieved for 532 nm, and 98.33% at 785nm. Then the two datasets were concatenated as before, a final accuracy of 99.75% was attained. It is important to note that in ASM, the combined dataset showed no inter-species classification errors, and 100% accuracy for SA, showing that the presence of the interfering media was no impediment to the discrimination of the drug-resistant and drug-sensitive strains of this pathogen.

**Discussion and conclusions**

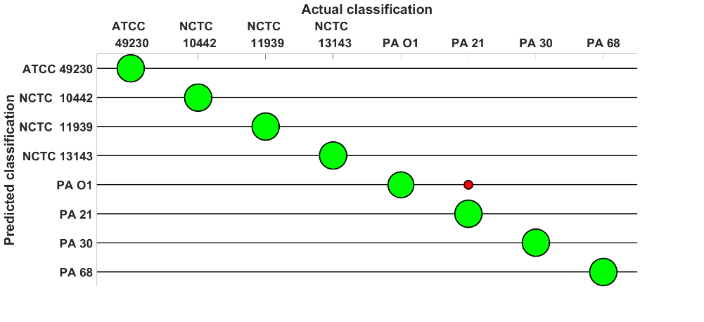
**Figure 7:** Classification accuracies for SVM of bacterial strains in ASM for **(A)** 785nm excitation, **(B)** 532nm excitation, and **(C)** for concatenated spectra (785 & 532nm), showing improved classification accuracy for concatenated spectra, particularly for strains of S. aureus. The size of green balls correspond to the number of correct identifications for that species. Red balls indicate incorrect classifications



A



B



C

Prior work on the use of Raman spectroscopy has used single-excitation to detect pathogens in CF patient samples, but the exposures times involved in the analysis required lengthy photobleaching steps to reduce the fluorescence in the samples 49. Our findings yielded high quality spectra within exposure times of ~1 minute, whilst the aforementioned study required 15 minutes of pre-bleaching before spectral features began to appear, and 30 minutes before they became prominent and enabled the authors to achieve a comparable level of accuracy (<95%). Our results compare well to recent work by Ho *et al*, where deep learning was used for the identification of pathogenic bacteria based on their Raman spectra50. The authors in this study reported an accuracy of 99.7%, which is comparable to our own accuracy; however, our method allows us to relate spectral peaks to the biochemicals that are contributing to the spectrum, so is a chemistry-informed method, rather than a black box.

Other approaches, such as mass spectrometry, have also been applied to the strain level. MS has been widely studied for this application, and excellent reviews exist discussing its ability as a powerful tool for the strain level characterisation of bacteria, such as the one by Sandrin *et al*, which deals with the application of MALDI-TOF Mass spectroscopy for this application 51. By contrast, our Raman spectroscopy method relies only on the vibration of molecular bonds within the sample, which are consistent within the strains. Additionally, our method has considerable advantages in simplicity over MS-based methods, which require a dedicated expert and lab-based working environment. Given the commercial availability of a range of miniaturised and handheld Raman spectrometers, our methodology is feasible for deployment on wards by existing clinical personnel.

At present, the priority of clinicians is most often the identification of the species of the pathogen rather than the strain. With the rising prevalence of drug resistant bacteria this strain-level differentiation ability is becoming increasingly clinically relevant. The high levels of accuracy in our experiments differentiating sensitive and resistant SA strains suggests that strain and serotype-linked anti-microbial resistance profiles of other pathogens will be technically feasible. This technique could help improve clinical outcomes, and enable the efficient targeting of treatments, by allowing for the monitoring of developing drug resistance in ongoing treatment, as well as aiding current efforts in anti-microbial stewardship. Our samples were analysed with a 1:1 ratio of concentrated bacteria to ASM, which, while high, does not result in an unprecedented bacterial quantity. *P*. *aeruginosa* is commonly enumerated in CF sputum at concentrations of 108-1010 CFU/ml, with Stressmann *et al* finding a maximum of 1.8\*1011 CFU/ml, in addition to a total bacterial density mean of approximately 1010 CFU/ml 52-53. While future studies will focus on analyses of samples that better model the polymicrobial nature and bacterial density of CF sputum, our current work represents a strong proof-of-principle for the method. Whilst we have demonstrated this method within the context of a CF model, the principles underlying this method are also applicable to a variety of other clinical scenarios and applications across sectors such as in food safety, where effective testing depends on rapidity and specificity.

**Supporting Information**

Detailed classification methods.

Raw Raman spectra at 532nm, 633nm, and 785nm.

Pre-processed Raman spectra.

UV-visible absorbance spectra and Raman spectra (532nm and 785nm excitation) of β-carotene, xanthophyll, hemin, and protoporphyrin IX.

Images of dried bacterial pellets of all *P. aeruginosa* and *S. aureus* strains used in the study.

Peak assignments for major vibrational modes.

Projections and classification accuracies for PCA of the 8 strains used in this study after polynomial background subtraction.

Comparison of the classification of bacteria using the luminescent background versus Raman spectra.

Classification accuracy for PCA-SVM of pure bacterial pellets.

Sample preparation for bacterial suspensions in ASM.

Raman spectra of blank ASM, and PA and SA strains in ASM, taken at 532nm and 785nm excitations.

Classification accuracy for SVM of bacteria suspended in ASM.

Classification accuracy for PCA-SVM of bacteria suspended in ASM.

**Acknowledgements**

The authors would like to acknowledge industrial CASE studentship funding to Adam Lister by EPSRC (EP/P510646/1) and the United Kingdom Ministry of Defence under contract DSTLX-10000110975. We also acknowledge funding by BRC-NAMRIP University of Southampton and the NIHR Southampton Antimicrobial Resistance Laboratory. SM acknowledges EPSRC grant EP/T020997/1 and ERC NanoChemBioVision (638258). JSW and CH acknowledge funding from BBSRC and Innovate UK, IKC National Biofilms Innovation Centre BB/R012415/1. SNF is an NIHR Senior Investigator.

**References**

1. Høiby, N.; Bjarnsholt, T.; Moser, C.; Jensen, P. Ø.; Kolpen, M.; Qvist, T.; Aanæs, K.; Pressler, T.; Skov, M.; Ciofu, O., Diagnosis of biofilm infections in cystic fibrosis patients. *APMIS* **2017,** *125* (4), 339-343.

2. Farrell, P. M., The prevalence of cystic fibrosis in the European Union. *Journal of Cystic Fibrosis* **2008,** *7* (5), 450-453.

3. Munro, A. P. S.; Highmore, C. J.; Webb, J. S.; Faust, S. N., Diagnosis and treatment of biofilm infections in children. *Current opinion in infectious diseases* **2019,** *32* (5), 505-509.

4. Registry, C. F. F. P., Annual data report. Cystic Fibrosis Foundation Bethesda, MD: 2011.

5. Döring, G.; Flume, P.; Heijerman, H.; Elborn, J. S., Treatment of lung infection in patients with cystic fibrosis: Current and future strategies. *Journal of Cystic Fibrosis* **2012,** *11* (6), 461-479.

6. Gibson, R. L.; Emerson, J.; McNamara, S.; Burns, J. L.; Rosenfeld, M.; Yunker, A.; Hamblett, N.; Accurso, F.; Dovey, M.; Hiatt, P., Significant microbiological effect of inhaled tobramycin in young children with cystic fibrosis. *American journal of respiratory and critical care medicine* **2003,** *167* (6), 841-849.

7. Murray, T. S.; Egan, M.; Kazmierczak, B. I., Pseudomonas aeruginosa chronic colonization in cystic fibrosis patients. *Current opinion in pediatrics* **2007,** *19* (1), 83-8.

8. Singh, P. K.; Schaefer, A. L.; Parsek, M. R.; Moninger, T. O.; Welsh, M. J.; Greenberg, E. P., Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms. *Nature* **2000,** *407* (6805), 762-4.

9. Olsen, I., Biofilm-specific antibiotic tolerance and resistance. *European Journal of Clinical Microbiology & Infectious Diseases* **2015,** *34* (5), 877-886.

10. Høiby, N.; Bjarnsholt, T.; Givskov, M.; Molin, S.; Ciofu, O., Antibiotic resistance of bacterial biofilms. *Int J Antimicrob Agents* **2010,** *35* (4), 322-32.

11. Jorth, P.; Staudinger, B. J.; Wu, X.; Hisert, K. B.; Hayden, H.; Garudathri, J.; Harding, C. L.; Radey, M. C.; Rezayat, A.; Bautista, G.; Berrington, W. R.; Goddard, A. F.; Zheng, C.; Angermeyer, A.; Brittnacher, M. J.; Kitzman, J.; Shendure, J.; Fligner, C. L.; Mittler, J.; Aitken, M. L.; Manoil, C.; Bruce, J. E.; Yahr, T. L.; Singh, P. K., Regional Isolation Drives Bacterial Diversification within Cystic Fibrosis Lungs. *Cell Host Microbe* **2015,** *18* (3), 307-319.

12. Darch, S. E.; McNally, A.; Harrison, F.; Corander, J.; Barr, H. L.; Paszkiewicz, K.; Holden, S.; Fogarty, A.; Crusz, S. A.; Diggle, S. P., Recombination is a key driver of genomic and phenotypic diversity in a Pseudomonas aeruginosa population during cystic fibrosis infection. *Scientific reports* **2015,** *5*, 7649.

13. Kloß, S.; Lorenz, B.; Dees, S.; Labugger, I.; Rösch, P.; Popp, J., Destruction-free procedure for the isolation of bacteria from sputum samples for Raman spectroscopic analysis. *Analytical and Bioanalytical Chemistry* **2015,** *407* (27), 8333-8341.

14. Ghebremedhin, M.; Heitkamp, R.; Yesupriya, S.; Clay, B.; Crane, N. J., Accurate and Rapid Differentiation of <span class="named-content genus-species" id="named-content-1">Acinetobacter baumannii</span> Strains by Raman Spectroscopy: a Comparative Study. *Journal of Clinical Microbiology* **2017,** *55* (8), 2480-2490.

15. Kloß, S.; Rösch, P.; Pfister, W.; Kiehntopf, M.; Popp, J., Toward Culture-Free Raman Spectroscopic Identification of Pathogens in Ascitic Fluid. *Analytical Chemistry* **2015,** *87* (2), 937-943.

16. Li, J.; Wang, C.; Kang, H.; Shao, L.; Hu, L.; Xiao, R.; Wang, S.; Gu, B., Label-free identification carbapenem-resistant Escherichia coli based on surface-enhanced resonance Raman scattering. *RSC Advances* **2018,** *8* (9), 4761-4765.

17. Allen, D.; Kelly, J.; Gilpin, D.; Bell, S.; Tunney, M., Detection and characterisation of bacteria causing lung infection in people with Cystic Fibrosis (CF) by surface-enhanced Raman spectroscopy (SERS). *Access Microbiology* **2019,** *1* (1A).

18. Polisetti, S.; Baig, N. F.; Morales-Soto, N.; Shrout, J. D.; Bohn, P. W., Spatial Mapping of Pyocyanin in Pseudomonas Aeruginosa Bacterial Communities Using Surface Enhanced Raman Scattering. *Applied Spectroscopy* **2017,** *71* (2), 215-223.

19. Jarvis, R. M.; Goodacre, R., Discrimination of Bacteria Using Surface-Enhanced Raman Spectroscopy. *Analytical Chemistry* **2004,** *76* (1), 40-47.

20. Jarvis, R. M.; Goodacre, R., Characterisation and identification of bacteria using SERS. *Chemical Society Reviews* **2008,** *37* (5), 931-936.

21. Premasiri, W. R.; Lee, J. C.; Sauer-Budge, A.; Théberge, R.; Costello, C. E.; Ziegler, L. D., The biochemical origins of the surface-enhanced Raman spectra of bacteria: a metabolomics profiling by SERS. *Analytical and Bioanalytical Chemistry* **2016,** *408* (17), 4631-4647.

22. Tadesse, L. F.; Safir, F.; Ho, C.-S.; Hasbach, X.; Khuri-Yakub, B.; Jeffrey, S. S.; Saleh, A. A. E.; Dionne, J., Toward rapid infectious disease diagnosis with advances in surface-enhanced Raman spectroscopy. *The Journal of Chemical Physics* **2020,** *152* (24), 240902.

23. Cotton, T. M.; Timkovich, R.; Cork, M. S., Resonance Raman and surface-enhanced resonance Raman studies of cytochrome cd 1. *FEBS Letters* **1981,** *133* (1), 39-44.

24. Manoharan, R.; Ghiamati, E.; Dalterio, R. A.; Britton, K. A.; Nelson, W. H.; Sperry, J. F., UV resonance Raman spectra of bacteria, bacterial spores, protoplasts and calcium dipicolinate. *Journal of Microbiological Methods* **1990,** *11* (1), 1-15.

25. Yellampelle, B.; Sluch, M.; Asher, S.; Lemoff, B., *Multiple-excitation-wavelength resonance-Raman explosives detection*. SPIE: 2011; Vol. 8018.

26. Asher, S. A., UV resonance Raman spectroscopy for analytical, physical, and biophysical chemistry. Part 2. *Analytical chemistry* **1993,** *65* (4), 201A-210A.

27. Sriramulu, D., Artificial Sputum Medium. *Protocol Exchange* **2010**.

28. Trevisan, J.; Angelov, P. P.; Scott, A. D.; Carmichael, P. L.; Martin, F. L., IRootLab: a free and open-source MATLAB toolbox for vibrational biospectroscopy data analysis. *Bioinformatics* **2013,** *29* (8), 1095-1097.

29. Asher, S. A., UV Resonance Raman Studies of Molecular Structure and Dynamics: Applications in Physical and Biophysical Chemistry. *Annual Review of Physical Chemistry* **1988,** *39* (1), 537-588.

30. Deng, J. L.; Wei, Q.; Zhang, M. H.; Wang, Y. Z.; Li, Y. Q., Study of the effect of alcohol on single human red blood cells using near-infrared laser tweezers Raman spectroscopy. *Journal of Raman Spectroscopy* **2005,** *36* (3), 257-261.

31. Berezhna, S.; Wohlrab, H.; Champion, P. M., Resonance Raman Investigations of Cytochrome c Conformational Change upon Interaction with the Membranes of Intact and Ca2+-Exposed Mitochondria. *Biochemistry* **2003,** *42* (20), 6149-6158.

32. Rönnberg, M.; Österlund, K.; Ellfolk, N., Resonance Raman spectra of Pseudomonas cytochrome c peroxidase. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **1980,** *626* (1), 23-30.

33. Gorski, A.; Starukhin, A.; Stavrov, S.; Gawinkowski, S.; Waluk, J., Resonance Raman spectroscopy study of protonated porphyrin. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2017,** *173*, 350-355.

34. Spiro, T. G.; Strekas, T. C., Resonance Raman Spectra of Hemoglobin and Cytochrome c; Inverse Polarization and Vibronic Scattering. *Proceedings of the National Academy of Sciences* **1972,** *69* (9), 2622.

35. Naumann, D., *Infrared and NIR Raman spectroscopy in medical microbiology*. SPIE: 1998; Vol. 3257.

36. Huang, Z.; McWilliams, A.; Lui, H.; McLean, D. I.; Lam, S.; Zeng, H., Near-infrared Raman spectroscopy for optical diagnosis of lung cancer. *International Journal of Cancer* **2003,** *107* (6), 1047-1052.

37. McAdam, P. R.; Templeton, K. E.; Edwards, G. F.; Holden, M. T. G.; Feil, E. J.; Aanensen, D. M.; Bargawi, H. J. A.; Spratt, B. G.; Bentley, S. D.; Parkhill, J.; Enright, M. C.; Holmes, A.; Girvan, E. K.; Godfrey, P. A.; Feldgarden, M.; Kearns, A. M.; Rambaut, A.; Robinson, D. A.; Fitzgerald, J. R., Molecular tracing of the emergence, adaptation, and transmission of hospital-associated methicillin-resistant &lt;em&gt;Staphylococcus aureus&lt;/em&gt. *Proceedings of the National Academy of Sciences* **2012,** *109* (23), 9107.

38. Igor, V. E.; Mohsen, S.; Maia, R. E.; Werner, G., Resonance Raman detection of carotenoid antioxidants in living human tissue. *BIOMEDO* **2005,** *10* (6), 1-18.

39. Bernstein, P. S.; Zhao, D.-Y.; Wintch, S. W.; Ermakov, I. V.; McClane, R. W.; Gellermann, W., Resonance Raman measurement of macular carotenoids in normal subjects and in age-related macular degeneration patients. *Ophthalmology* **2002,** *109* (10), 1780-1787.

40. Merlin, J. C., Resonance Raman spectroscopy of carotenoids and carotenoid-containing systems. *Pure and Applied Chemistry* **1985,** *57* (5), 785-792.

41. Rebrošová, K.; Šiler, M.; Samek, O.; Růžička, F.; Bernatová, S.; Ježek, J.; Zemánek, P.; Holá, V., Differentiation between Staphylococcus aureus and Staphylococcus epidermidis strains using Raman spectroscopy. *Future Microbiology* **2017,** *12* (10), 881-890.

42. Lindsay, J. A., Genomic variation and evolution of Staphylococcus aureus. *International Journal of Medical Microbiology* **2010,** *300* (2), 98-103.

43. Patel, D.; Ellington, M. J.; Hope, R.; Reynolds, R.; Arnold, C.; Desai, M., Identification of genetic variation exclusive to specific lineages associated with Staphylococcus aureus bacteraemia. *Journal of Hospital Infection* **2015,** *91* (2), 136-145.

44. Sriramulu, D. D.; Lünsdorf, H.; Lam, J. S.; Römling, U., Microcolony formation: a novel biofilm model of Pseudomonas aeruginosa for the cystic fibrosis lung. *Journal of Medical Microbiology* **2005,** *54* (7), 667-676.

45. Behrends, V.; Geier, B.; Williams, H. D.; Bundy, J. G., Direct Assessment of Metabolite Utilization by Pseudomonas aeruginosa during Growth on Artificial Sputum Medium. *Applied and Environmental Microbiology* **2013,** *79* (7), 2467.

46. Kirchner, S.; Fothergill, J. L.; Wright, E. A.; James, C. E.; Mowat, E.; Winstanley, C., Use of Artificial Sputum Medium to Test Antibiotic Efficacy Against Pseudomonas aeruginosa in Conditions More Relevant to the Cystic Fibrosis Lung. *JoVE* **2012,** (64), e3857.

47. Horii, S.; Ando, M.; Samuel, A. Z.; Take, A.; Nakashima, T.; Matsumoto, A.; Takahashi, Y. k.; Takeyama, H., Detection of Penicillin G Produced by Penicillium chrysogenum with Raman Microspectroscopy and Multivariate Curve Resolution-Alternating Least-Squares Methods. *Journal of Natural Products* **2020,** *83* (11), 3223-3229.

48. Barzan, G.; Sacco, A.; Mandrile, L.; Giovannozzi, A. M.; Portesi, C.; Rossi, A. M., Hyperspectral Chemical Imaging of Single Bacterial Cell Structure by Raman Spectroscopy and Machine Learning. *Applied Sciences* **2021,** *11* (8).

49. Rusciano, G. C., P.; Pesce, G.; Abete, P.; Carnovale, V.; Sasso, A., Raman spectroscopy as a new tool for early detection of bacteria in patients with cystic fibrosis. *Laser Physics Letters* **2013,** *10* (7).

50. Ho, C.-S.; Jean, N.; Hogan, C. A.; Blackmon, L.; Jeffrey, S. S.; Holodniy, M.; Banaei, N.; Saleh, A. A. E.; Ermon, S.; Dionne, J., Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. *Nature Communications* **2019,** *10* (1), 4927.

51. Sandrin, T. R.; Goldstein, J. E.; Schumaker, S., MALDI TOF MS profiling of bacteria at the strain level: A review. *Mass Spectrometry Reviews* **2013,** *32* (3), 188-217.

52. Palmer, K. L.; Mashburn, L. M.; Singh, P. K.; Whiteley, M., Cystic fibrosis sputum supports growth and cues key aspects of Pseudomonas aeruginosa physiology. *J Bacteriol* **2005,** *187* (15), 5267-5277.

53. Stressmann, F. A.; Rogers, G. B.; Marsh, P.; Lilley, A. K.; Daniels, T. W. V.; Carroll, M. P.; Hoffman, L. R.; Jones, G.; Allen, C. E.; Patel, N.; Forbes, B.; Tuck, A.; Bruce, K. D., Does bacterial density in cystic fibrosis sputum increase prior to pulmonary exacerbation? *Journal of Cystic Fibrosis* **2011,** *10* (5), 357-365.

For Table of Contents only:

E1

Measurement variable [n]

Response

E2

R1

R2