Fast antibiotic susceptibility testing via Raman microspectrometry on single bacteria: a MRSA case study

Armelle Novelli Rousseau - bioMérieux, R&D Microbiology, 5 rue des Berges, 38024 Grenoble, France

Nicolas Faure - bioMérieux, R&D Microbiology, 5 rue des Berges, 38024 Grenoble, France

Fabian Rol - Bioaster, 40 avenue Tony Garnier, 69007 Lyon, France

Zohreh Sedaghat - Bioaster, 40 avenue Tony Garnier, 69007 Lyon, France

Joël Le Galudec - bioMérieux, R&D Microbiology, 5 rue des Berges, 38024 Grenoble, France

Fréderic Mallard - bioMérieux, R&D Microbiology, 5 rue des Berges, 38024 Grenoble, France

Quentin Josso - bioMérieux, R&D Microbiology, 376 Chemin de l'Orme, 69280 Marcy-l'Etoile

SUPPORTING INFORMATIONS

SI 1. MATERIALS

Staphylococcus aureus (S. aureus) and *Escherichia coli* (*E. coli*) strains were obtained from the American Type Culture Collection (ATCC, Manassas, VA, USA) and from bioMérieux Culture Collection (API; bioMérieux, Marcy l'Etoile, France). See Table S 1 and Table S 2.

Culture medium Muller Hinton (MHB Ref: AEB 110699), Columbia agar with 5% sheep blood (COS Ref : 43041), Trypcase Soja (TSA Ref. 43011) and TSB (Ref. 42100) were provided by bioMérieux. Phosphate buffer saline (PBS Ref. A9162, 0100) was purchased from Applichem (64291 Darmstadt, Germany). Sodium phosphate (ref : 71640 and ref : 71505), Sodium chloride (ref : 27810.295) were respectively purchased from Sigma Aldrich and Prolabo. Gentamicin (Ref. 00768065) was provided by bioMérieux. Cefoxitin (ref: ca 4786) was purchased from Sigma Aldrich. Aminosilane functionalized Coverslips (Ref. 166 6121) were from Schott Nexterion (Müllheim, Germany). Geneframes (Ref. AB-0577) were purchased from Thermo Scientific (Thermo Fisher Scientific, Inc.). Minimal Inhibiting Concentration (MIC) was controlled either with Vitek® 2 or with e-test®. Bacteria suspension concentrations were measured using a biospectrometer from Eppendorf. Correspondence between measured optical density at 600nm and bacteria concentration were calibrated for the tested strains with plate numeration (data not shown).

Strains	Genotype	Cefoxitin MIC (µg/mL)	Susceptibility To cefoxitin	Number of spectra / replicates
ATCC 25923 (SA44)	Wild	2	S	2002/4
API 0610228 (SA86)	Acquired Penicillinase	4	S	1314/3
ATCC 51153 (SA127)	*mecA +*	16	R	2093/2
ATCC 33592 (SA128)	*mecA +*	>256	R	2015/3
API 1108078 (SA126)	*mecC +*	16	R	2005/3

Table S 1 : Details of the S. aureus biological model used for this study . It is composed of 2 susceptible (S) and 3 resistant (R) S. aureus strains. Genotype and MIC are given for each strains. Number of experiments and spectra acquired are also indicated for each strain.

	Gentamicin MIC (µg/mL)	
ATCC 25922 (EC10)	1	S
ATCC 35421 (EC21)	> 256	R

Table S 2: Details of reduced E. coli model used for control. It is composed of one susceptible (S) and one resistant (R) strains. MIC is indicated for each strain.

SI 2. ANTIBIOTIC INCUBATION AND SAMPLE PREPARATION

Efforts were made to minimize variations in sample handling: The two-hours growth and bacterial concentrations were standardized. Strains were stored at -80°C in broth containing glycerol. Before experiments, a first overnight culture was performed on COS culture medium at 37°C. This first culture was stored at 4°C and constituted a "stock culture" which was used as a single source during the experimental campaign (3 weeks of storage at most). The sample preparation consisted in picking up one colony from this stock culture, streaking it on a new COS culture medium for 15 h at 37°C to revitalize the bacteria. MHB and TSB liquid culture media were respectively selected for *S. aureus* and *E. coli* culture. After an overnight culture, colonies were picked up and diluted at a final concentration of $5\cdot10^6$ CFU/mL in 5 mL of liquid culture medium and followed by a 2-h culture at 37°C to get bacteria in early stage of growth with less variations between them. The culture was then centrifuged at $3000\times g$ for 10 min. The pellet was suspended in a phosphate buffer saline solution (Sodium phosphate 20mM, pH 7.2, NaCl 50mM) to a final concentration of $2\cdot10^7$ CFU/mL. 300 µL of the suspension were loaded on an aminosilane coverslip and incubated for 20 minutes at room temperature. After incubation, bacteria were washed with 400µL of capture buffer (3 times). Last washing was done with MHB 10% for *S. aureus* experiment or PBS/TSB10% for the *E. coli* experiment containing the antibiotic at the selected concentration. The coverslip was then set upside down onto a geneframe previously sticked on a microscope glass slide and filled with MHB 10% with antibiotic (or PBS/TSB10% with antibiotic for *E. coli*). For reference, bacteria were prepared using the same protocol but no antibiotic was added to the washing solution and to the medium contained in the gene frame.

SI 3. CELL VIABILITY TEST.

Viability assays were performed on an aliquot of the bacteria that were prepared for the experiment with reagent solutions also prepared for Raman experiments. Bacteria were incubated in a liquid broth format (in Eppendorf) in the incubation solution (MHB 10% or PBS/TSB 10% respectively for *S. aureus* and *E. coli*) in presence or absence of the antibiotic (at the concentration that was

tested for the Raman experiment) and then numerated on agar culture medium. This incubation allowed us to confirm that in our conditions the antibiotic was effective on bacterial cells and lead to a loss of viability or no loss of viability depending on susceptibility or resistance phenotype of the bacteria. For numeration on agar culture medium, an aliquot of each sample was serially diluted in water immediately after incubation. 100 μL of dilution were plated onto a COS culture medium without antibiotic and incubated for 24 h at 37°C. Bacterial colonies were numerated to evaluate the concentration of bacteria alive in the different samples. For results see Table S 3.

Strain	Exp #	CFX (μg/mL)	No CFX (cfu/mL)	With CFX (cfu/mL)	% Viability
SA44	1	8	$2,4 \cdot 10^7$	$3,0 \cdot 10^6$	130
	2	8	$2,0 \cdot 10^7$	$2,4 \cdot 10^6$	60
	3	8	$2,1 \cdot 10^7$	$8,0 \cdot 10^4$	11
	4	8	$1,1 \cdot 10^7$	$4,0 \cdot 10^5$	21
SA86	5	16	$1,7 \cdot 10^6$	$5,0 \cdot 10^4$	1
	6	16	$5,3 \cdot 10^7$	$1,5 \cdot 10^6$	12
SA126	7	8	$3,4 \cdot 10^7$	$8,5 \cdot 10^6$	100
SA127	8	8	$1,0 \cdot 10^7$	$1,0 \cdot 10^7$	143
SA128	9	16	$7,0 \cdot 10^7$	$4,0 \cdot 10^7$	200

Table S 3: Viability results. Percentages higher than 100% indicate that bacteria have grown after incubation.

SI 4.   RAMAN SPECROMETER.

Raman spectra were acquired using the prototype described in previous publication [1] that is based on a Raman Xplora microspectrometer (HORIBA JOBIN-YVON) equipped with a 600 lines/mm grating, a VENTUS (LASER QUANTUM) 532 nm CW laser and a ZEISS 100x oil immersion objective with a 1.3 numerical aperture. The confocal pinhole was set to 400 μm resulting in a confocal volume of the order of a single bacterial cell. Typical laser power on the sample was chosen to be 30mW to avoid degrading bacteria and spectra were integrated during 2 seconds for each single cell. The scattered light was analyzed by a monochromator and collected on a 1024×256 Synapse CCD (HORIBA JOBIN YVON) cooled at -70°C. To guaranty an optimal temperature for bacterial growth, the geneframe holding the functionalized coverslip on which bacteria are captured was loaded on a transparent stage thermalized at 37°C. Finally, the sample could be moved manually or automatically by xyz Marzhauser translation stages and an additional piezoelectric focalization device was added before the microscope objective for precise displacement along z.

SI 5.   BACTERIA LOCALIZATION.

To precisely move each individual bacteria in the confocal collection volume (overlapping the laser spot), they had to be precisely localized. This was done by using an in-line holographic imaging module mounted on the Raman micro-spectrometer [1]. Instead of using focused images, holograms of bacteria (corresponding to out-of-focus images in an in-line holographic microscope) were acquired and digitally refocused. The algorithm calculated the back propagation of the E-M field and corrected for the twin image artifact (manuscript in preparation). This procedure allowed us to localize the bacteria in xy plane (as could be done with a focused image) but also in the z direction (information only available with holograms without mechanical change of the focal distance). A first calibration of the system was necessary to calculate the transformation matrix converting a bacteria position given in pixel (xy on the CCD camera and z extracted from hologram) into a position in the XYZ coordinate units of the motorized stage. The calibration procedure is fully described in the patent [WO2019115901 - METHOD FOR CALIBRATING AN ANALYSIS DEVICE, AND ASSOCIATED DEVICE].

SI 6.   SPECTRA ACQUISITION.

The program was able to precisely determine the xyz position of each individual bacteria out of a single holographic image. However, when bacteria started to grow, individual cells in a cluster were more difficult to resolve by this method and could lead to mistakes in positioning. For that reason, an average z position of the bacteria was calculated for each field (and should correspond to the glass /medium interface).

The bacteria where then successively automatically placed in the laser spot by the instrument, and the related Raman spectra were acquired immediately. Each spectrum was acquired for 2 seconds on a different single cell. Five single bacteria spectra were acquired

followed by one spectrum of the background (acquired between bacteria). The background corresponds to the Raman signal coming from the glass and the growth medium; it needs to be precisely known to be separated from the Raman signal of bacteria during data processing. Usually, we measured a maximum of 50 bacteria per field in order to probe a large enough area of the glass slide. A run could take up to 4 hours, and the rate of acquisition was typically 400 spectra per hour.

SI 7. DATA PREPROCESSING.

The raw data is hindered with the presence of various artifacts, namely: (i) random noise; (ii) presence of a baseline due to the contribution of glass and buffer, in variable proportions; (iii) Presence of an offset, that can be considered as constant; (iv) presence of cosmic peaks; (v) variation in relative contribution of the microorganisms, depending on their size and on the relative position of the confocal beam; (vi) presence of outlier spectra, for instance due to the contribution of impurities or instrumental artifacts (shutter errors).

To deal with this, we carry out a cleaning and normalization of the spectra, with the following steps:

- Restrict the spectra to 650-3050 cm^{-1} range.
- Class as "outlier" (and therefore disregard for the rest of the analysis), any "bacterium" spectrum with no peak at 3050 cm^{-1} , and any "background" spectrum with a peak at 3050 cm^{-1}. This peak corresponds to CH bound, supposed to be present in any bacterium and absent or negligible in any background spectrum.
- Class as "outlier" any spectrum with an average signal in the 1800-2750 cm^{-1} range (henceforth noted S $(1800 - 2750\ cm^{-1})$ being outside of a manually set interval. This in general corresponds to an instrumental artifact when exposure time has been different from the setpoint, leading to either overexposure (risk of "burned" bacteria, yielding an altered Raman signal) or underexposure (too low signal/noise ratio).
- Class as "outlier" any spectrum with a ratio $S(1280 - 1500\ cm^{-1}) \big/ S(1800 - 2750\ cm^{-1})$ below a threshold set as 1.65. This usually corresponds to a confocal beam badly adjusted with respect to the position of the bacterium.
- Remove cosmic peaks as described in [2]
- Correct a possible spectral shift using 2 reference peaks, as described in [2]
- Smooth spectra using Savitsky-Golay algorithm, using 9 neighboring points and a 4th degree polynomial [3].
- Class as 'outlier" any "background" spectrum with a Euclidean distance in the spectrum space being superior to N times the observed standard deviation in the considered group of spectra. This step is carried out twice, with N = 2 then 1. The excess of "background spectra" allows us to be stringent on this step.
- Subtract residual low-frequency baseline, mainly due to fluorescence, using SNIP-Clayton algorithm [4], with 1 neighbor and 600 iterations.
- Class as "outliers" any "bacterium" spectrum with a too large Euclidean distance, similarly to two steps before for "background" spectra. We take N = 3.5 then N = 3.
- The resulting spectra are finally normalized by their mean intensity in the region of interest [650-1750] cm-1, resulting in so-called "normalized net spectra" that are used for all of the further studies (visualization, classification).

In the following Figure S 1 and Figure S 2 we provide examples of raw spectra and preprocessed spectra:
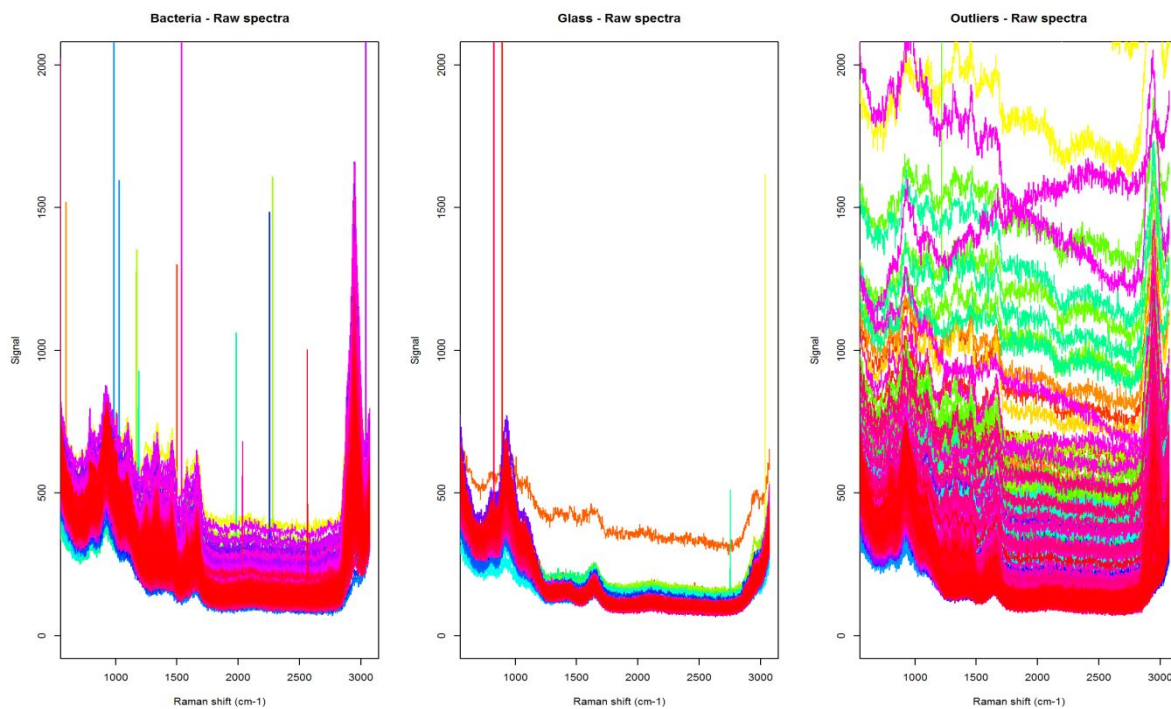
Figure S 1. Examples of Raw Spectra classified as "bacteria", "glass" and "outliers" (from left to right, respectively) after the first preprocessing step.
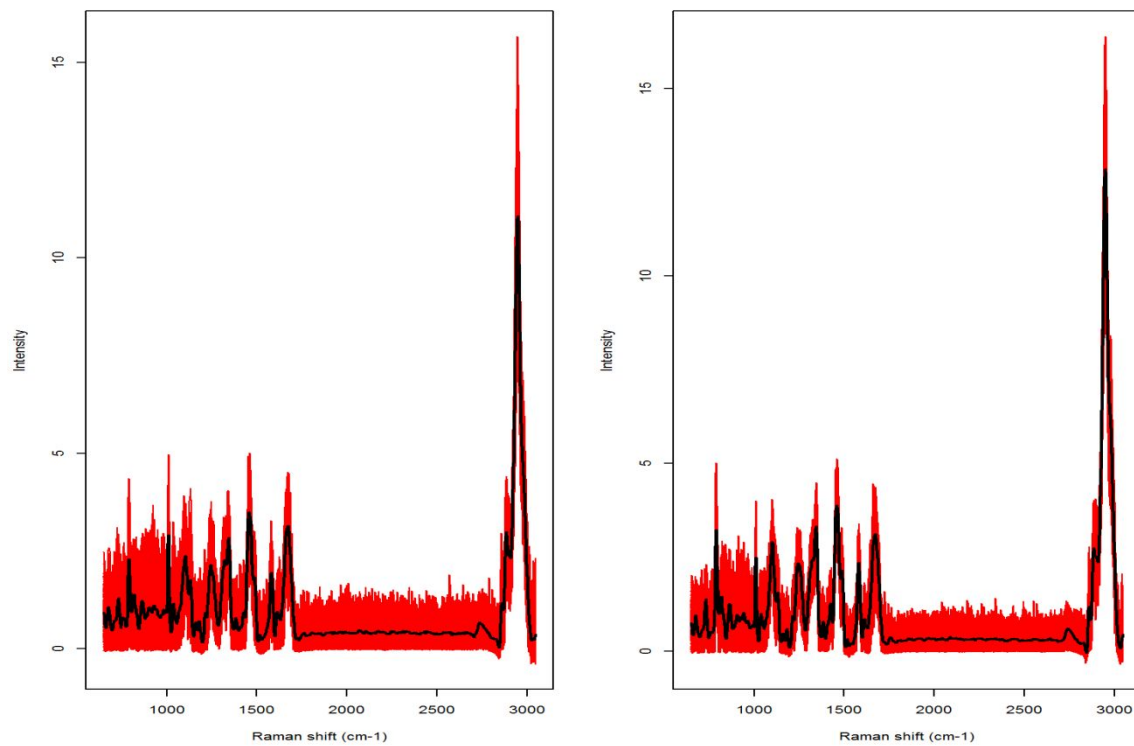

Figure S 2. "normalized net spectra" obtained after preprocessing. The spectra have been split in two graphs according to the class (left: SA44 in the presence of 8 µg/ml Cefoxitin; right: SA44 in the absence of Cefoxitin). Normalized net spectra from individual cells are displayed in red; while their average is displayed in black.
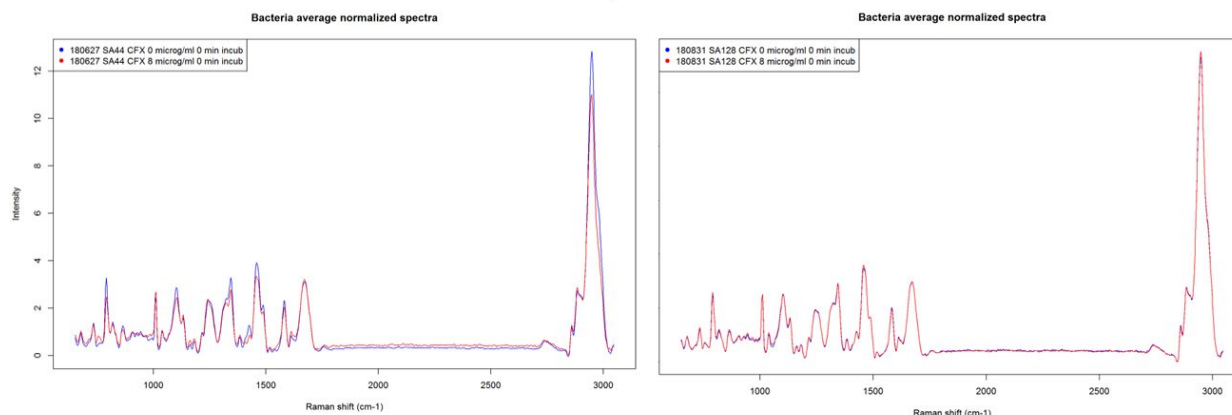
Susceptible strain SA44   **Resistant strain SA128**

Figure S 3 : spectral variations in the presence (red curves) or absence (blue curves) of antibiotics. The average net normalized spectra similar to those of Figure S 2 are shown. No clear spectral variations can be seen in the case of resistant strains (Resistant SA128 / right panel), while spectra in the presence or absence of cefoxitin can be seen in the case of a susceptible SA44 strain (left panel).

SI 8.   PENALIZED LINEAR DISCRIMINANT ANALYSIS FOR BIOLOGICAL RESPONSE VIZUALIZATION.

In previous publications, we used Principal Components Analysis (PCA) to visualize the spectra in reduced dimension and ridge-regularized Support-Vector Machine (SVM) with a Gaussian kernel to evaluate our ability to discriminate between different conditions [3].

In the case of *Staphylococcus aureus*, we had more numerous spectra, but faced a more difficult case with a lower signal/noise ratio, which can be due to a fainter biological effect (more limited spectral response of *Staphylococcus aureus* to Cefoxitin), and/or to a more limited capability of the instrument to capture this effect (noticeably in link to the smaller size of the cell). In such a case, PCA visualization was relatively difficult to interpret. In addition, we wanted to visualize and interpret the biological signature, namely the response or non-response to antibiotics of different strains, which is quite difficult to do on the basis of the results of non-linear kernel SVM.

For these reasons, we privileged in our case Linear Discriminant Analysis (LDA), which might be inferior to Gaussian kernel SVM in terms of classification, but has the advantage to allow for visualizing the biological responses of the strains in an orthonormal low-dimension space.

In our specific case, we are in a situation of a relatively high dimensionality problem, where the number of dimensions of the data (channels in the spectra) is comparable to the number of instances (spectra), and a relatively high correlation between neighboring signals. We therefore introduced regularization in the LDA, more specifically Penalized LDA (PenLDA) [5], which uses both lasso and fused lasso regularizations. The rationale of choosing lasso rather than ridge regularization is to favor interpretability, by inducing the coefficients of the obtained axes to be sparse both in absolute value and variation. By default, we use 0.02 for both the lasso and fused-lasso weighting, which according to our tests appears to be a reasonable balance between bias and variance.

SI 9.   CLASSIFICATION.

We carried out a linear ridge-regularized SVM to evaluate our capability to discriminate between resistant and non-resistant microorganisms in the case of *Staphylococcus aureus*.

The classification is carried out on so-called "difference spectra", that is to say the normalized spectra corresponding to the data acquired with antibiotic, minus the mean of normalized spectra corresponding to the data acquired without antibiotic, during the same experiment (same strain, same day). That way, we are discriminating between the different strains according their response to the presence of antibiotic, with as little as possible interference of a potential spectral difference between the strains, independent of the antibiotic.

We also evaluated Gaussian kernel SVM with a restrained dataset, yielding only a minor gain in accuracy (~3%), which can be explained by the fact that in the present case, no heterogeneity of response to antibiotics in the population is observable. Considering this preliminary result, we selected linear SVM as a reference classifier.

For resistance/susceptibility classification evaluation, we successively leave one strain out (typically 1 over 5) as test set. We then optimize the SVM cost hyperparameter on the train set through cross-validation, successively leaving one *experiment* out as the validation set (considering the number of strains in the dataset, we cannot leave another strain out at this stage). In addition, we use

oversampling to correct potential class imbalance effects. The classification results are displayed in % accuracy for each of the stains taken successively as test set, leading to one value per strain.

All of the visualization and classification analyses have been carried out in R environment. Penalized LDA have been carried out with the PenLDA package (https://cran.r-project.org/web/packages/penalizedLDA/penalizedLDA.pdf)[666], and the SVM with the mlr package, which itself calls e1071 package (https://cran.r-project.org/web/packages/mlr/index.html; https://cran.r-project.org/web/packages/e1071/index.html)

SI 10. REPEATABILITY EXPERTIMENT

SUSCEPTIBLE SA44 AND RESISTANT SA128 STRAIN.

We assessed the amplitude and relative variability of Raman information on two strains – one susceptible SA44 strain, and one resistant SA128 strain - exposed to cefoxitin. The resulting spectra were processed as described above. We built up a LDA representation on 4 classes (two strains SA44 and SA128 combined with 0 and 8 µg/ml cefoxitin concentrations). The first 2 dimensions of this 3D LDA are plotted in Figure S 4.
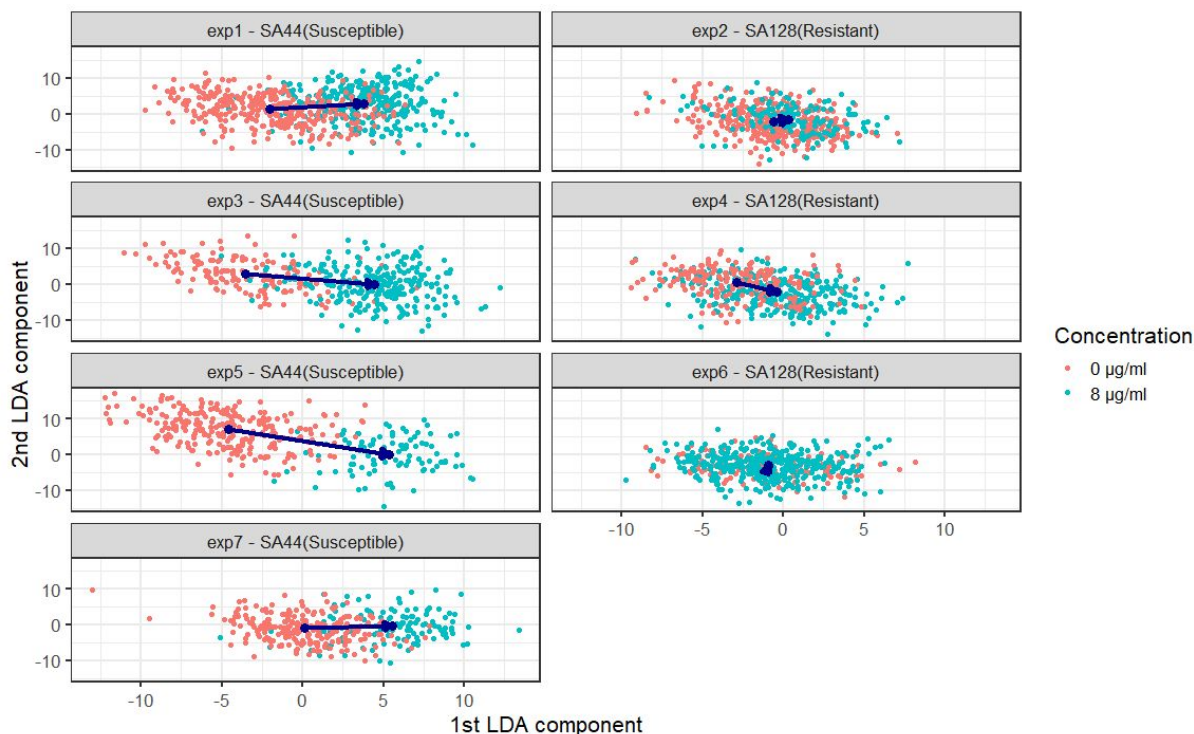

Figure S 4. Response of SA44 (susceptible strain) versus SA128 (resistant strain) to cefoxitin, repeatability experiment: penalized Linear Discriminant Analysis (LDA). The axes are constructed on Strain*Concentration.

As expected, the smaller size and cell shape of *S. aureus* yielded Raman spectra with a higher spectrum-to-spectrum variability and lower signal-to-noise ratio. Despite this variability, Figure S 4 shows that the different classes separated in a repeatable manner. The susceptible and resistant strains behaved differently: the susceptible SA44 strain responded to cefoxitin, mostly along the 1st LDA axis, while the resistant SA128 strain did not exhibit any visible response.

DETECTION OF GENTAMICIN EFFECT ON E.COLI SUSCEPTIBLE (EC10) AND RESISTANT (EC21) STRAINS.

In order to validate the capacity of our system to acquire Raman spectra and detect spectral modifications on single bacterial cells in presence of antibiotic, we performed the experiment on two *E. coli* strains. One susceptible (*E. coli* ATCC25922, here EC10) and one resistant (*E. coli* ATCC 35421, here EC21) for which we evidenced spectral modifications in presence of gentamicin in air drying conditions [3] were captured on an aminosilane functionalized coverslip and incubated in presence of gentamicin 8µg/mL which is 8 times the MIC of EC10 (1µg/mL) and less than the MIC of EC21 (>256µg/mL). Raman spectra were acquired during incubation using our microspectrometer. This has been carried out on a single experiment and on a relatively limited number of spectra: 193 for EC10 (98 and 95 respectively without and with gentamicin 8µg/ml); 73 for EC21 (36 and 37).

On two experiments, we construct a Penalized LDA on Strain*Concentration, that is to say with 4 classes: EC10, no antibiotic; EC21, no antibiotic; EC10, 8 µg/ml Gentamicin; EC21, 8 µg/ml Gentamicin. This allows to construct 3 LDA axes. In Figure S 5 we display (a) the projection on the 2 first LDA axes, and (b) the coordinates of the 1st LDA axis.

Figure S 5 clearly shows that the susceptible strain EC10 has a strong response to antibiotic in terms of Raman spectrum at the cell-level, while the resistant one EC21, does not exhibit any observable change between 0 and 8 µg/ml antibiotic concentration. These results were largely supported by viability tests that exhibited a complete loss of viability for the susceptible strain and not for the resistant one. This is consistent with the assumption that susceptibility to antibiotic induces a measurable shift in Raman spectrum, allowing to discriminate between resistant and susceptible strains.

Figure S 5 also shows that the response of EC10 strain to Gentamicin is essentially oriented along the 1st LDA axis.
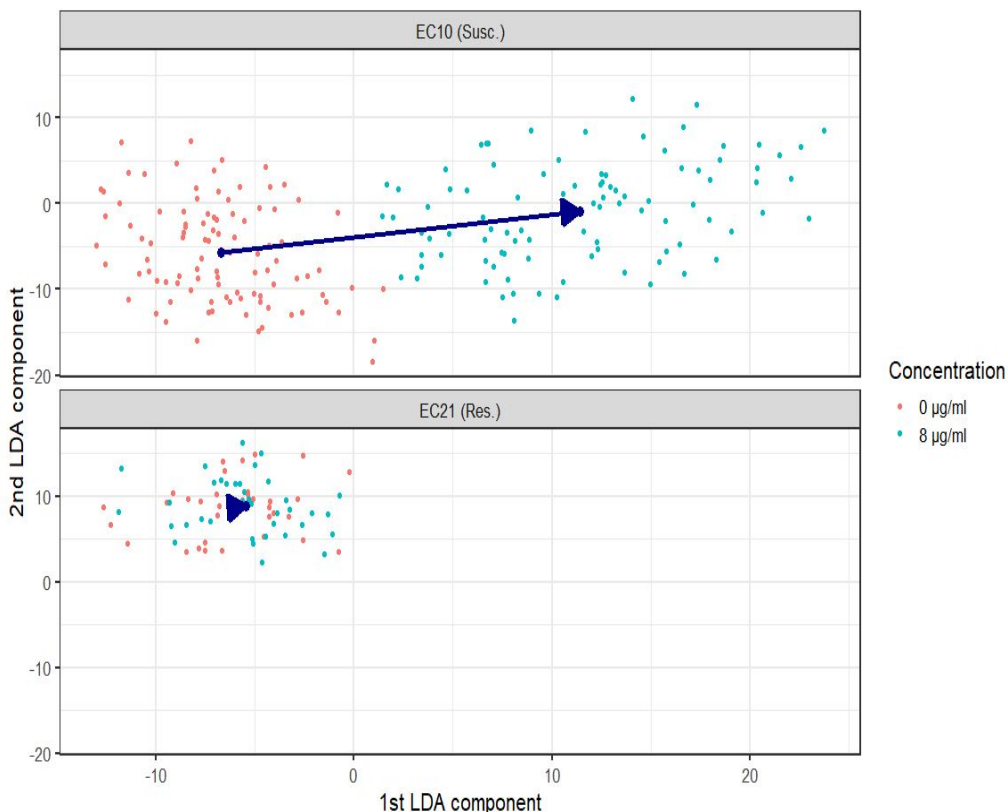

Figure S 5. Penalized LDA visualization of the response of EC10 (susceptible / upper panel) and EC21 (resistant / lower panel) strains in presence or absence of Gentamicin (8 µg/ml). PenLDA is constructed on Strain*Concentration (4 classes). The dots correspond to the projection of a spectrum on the 2 first LDA axes. The arrows are joining the projections of the mean spectra among each population, with the tip of the arrow corresponding to 8 µg/ml Gentamicin.
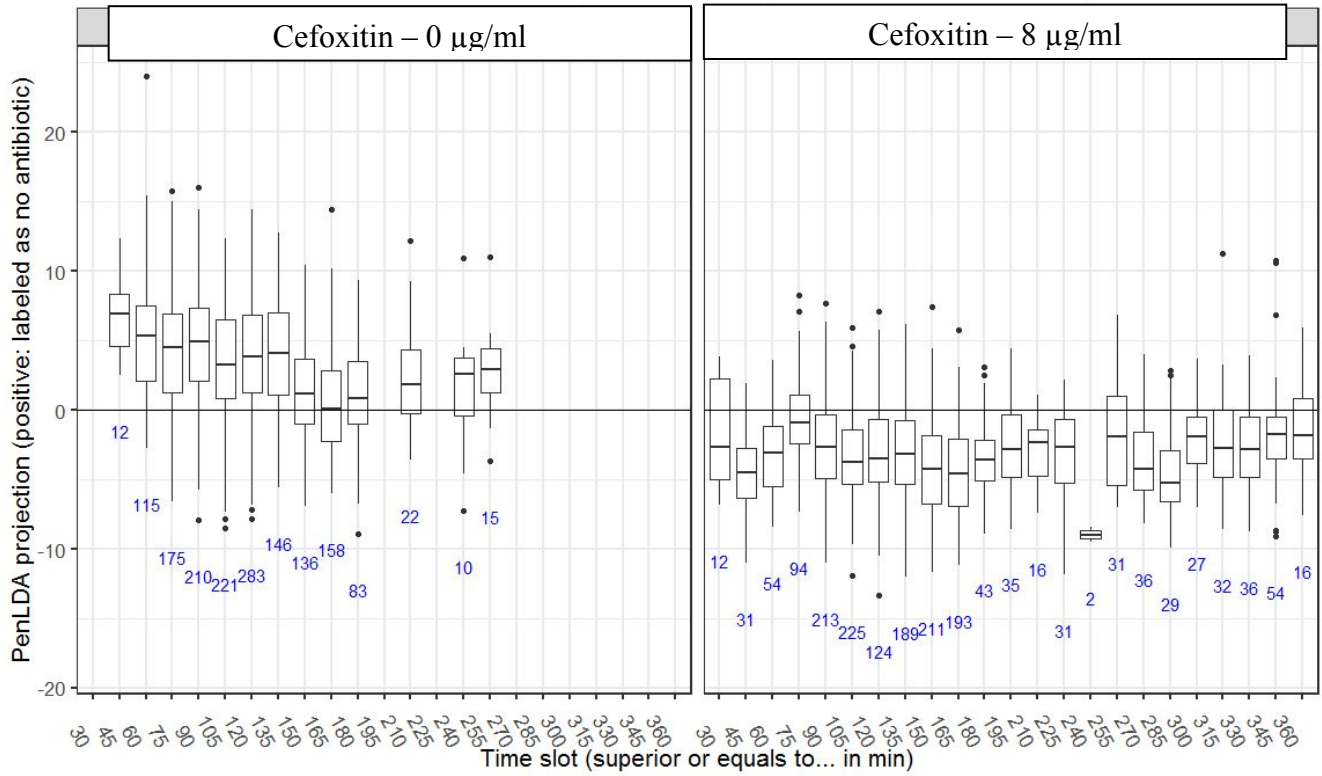
SI 11. VIZUALISATION OF THE TIME DEPENDENCY OF THE ANTIBIOTIC RESPONSE.

In order to evaluate the time dependency of the spectral response to Cefoxitin of *S. aureus*, we built an LDA representation for each strain with 2 classes: presence or absence of cefoxitin. We then evaluated the position of the spectra along the LDA axis as a function of time since the bacteria have been put in contact with cefoxitin. For technical reasons, the first spectra could only be acquired after approximately 30 minutes, so the very first time points are not available.

We observe (see Figure S 6) that for susceptible strains (SA44, SA86), and according to this representation, the spectral response to cefoxitin is relatively constant over time, at least in the 30-270 minutes time slot. This indicates that the response to the antibiotic is already established after 30 minutes.

Time Response of SA44 to Cefoxitin

Projection on PenLDA axis built on the CFX response of this single strain. In blue: number of spectra
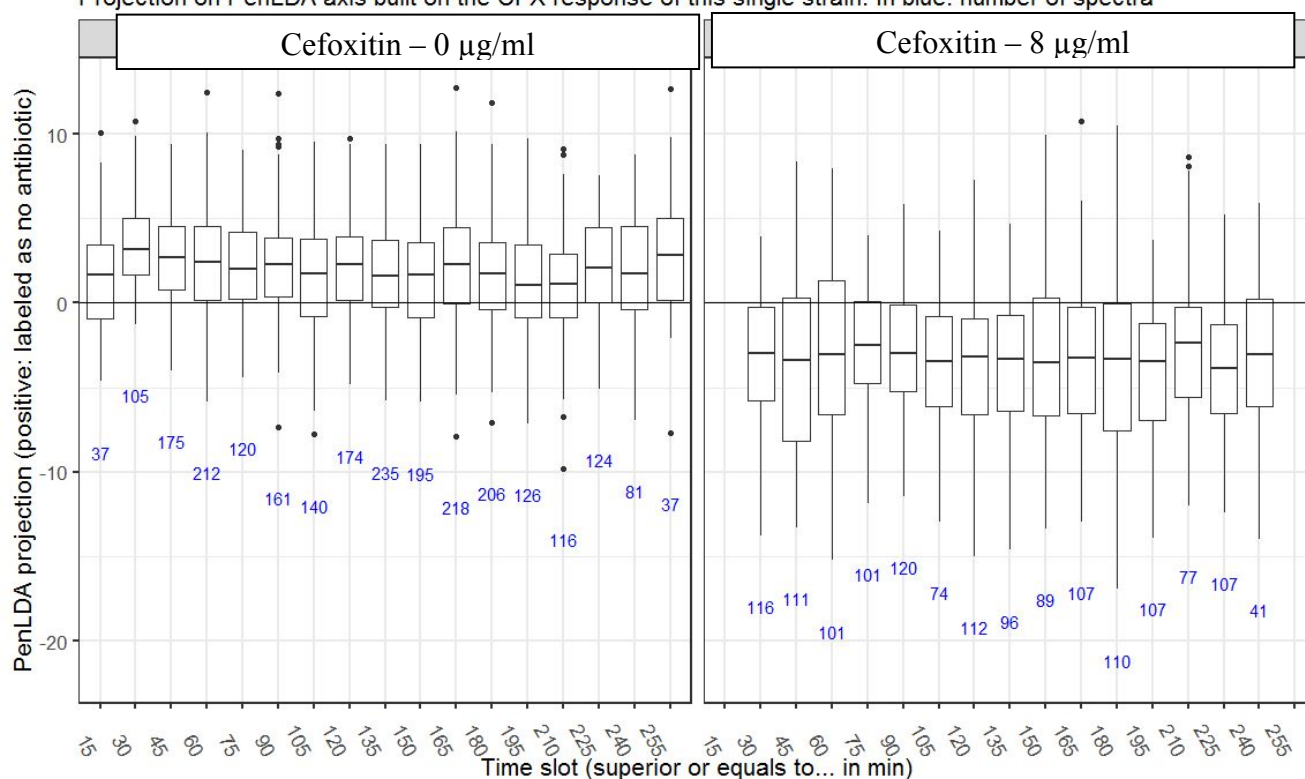
Figure S 6. representation of the time response of 2 susceptible strains (SA44 and SA86) to cefoxitin. The spectra are projected on the LDA axis built on presence/absence of cefoxitin for the considered strain. The projected spectra corresponding to each 15 minutes time of incubation slot, for each strain and each cefoxitin concentration are displayed as box-plots. The total number of spectra for each time slot is displayed in blue below the box plot.

SI 12. BIBLIOGRAPHIC REFERENCES

1.   Douet, A.; Josso, Q.; Marchant, A.; Dutertre, B.; Filiputti, D.; Novelli-Rousseau, A.; Espagnon, I.; Kloster-Landsberg, M.; Mallard, F.; Perraut, F., Fast Raman single bacteria identification: toward a routine in-vitro diagnostic. SPIE: 2016; Vol. 9887.

2.   Espagnon, I.; Ostrovskii, D.; Mathey, R.; Dupoy, M. G.; Joly, P.; Novelli-Rousseau, A.; Pinston, F.; Gal, O.; Mallard, F.; Leroux, D. F., Direct identification of clinically relevant bacterial and yeast microcolonies and macrocolonies on solid culture media by Raman spectroscopy. *Journal of biomedical optics* **2014,** *19* (2), 1-14, 14.

3.   Novelli-Rousseau, A.; Espagnon, I.; Filiputti, D.; Gal, O.; Douet, A.; Mallard, F.; Josso, Q., Culture-free Antibiotic-susceptibility Determination From Single-bacterium Raman Spectra. *Sci Rep* **2018,** *8* (1), 3957.

4.   Ryan, C. G.; Clayton, E.; Griffin, W. L.; Sie, S. H.; Cousens, D. R., SNIP, a statistics-sensitive background treatment for the quantitative analysis of PIXE spectra in geoscience applications. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **1988,** *34* (3), 396-402.

5.   Witten, D. M.; Tibshirani, R., Penalized classification using Fisher's linear discriminant. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **2011,** *73* (5), 753-772.

6.   Raro, O. H. F.; da Silva, R. M. C.; Filho, E. M. R.; Sukiennik, T. C. T.; Stadnik, C.; Dias, C. A. G.; Oteo Iglesias, J.; Pérez-Vázquez, M., Carbapenemase-Producing Klebsiella pneumoniae From Transplanted Patients in Brazil: Phylogeny, Resistome, Virulome and Mobile Genetic Elements Harboring bla (KPC-) (2) or bla (NDM-) (1). *Frontiers in microbiology* **2020,** *11*, 1563.