



**HAL**  
open science

## Results on identification of bacteria aging in complex environmental samples using Raman spectroscopy

Véronique Rebuffel, J.-C. Baritoux, Emeric Bergmann, Pierre Marcoux,  
Isabelle Espagnon, Sophie Morales

► **To cite this version:**

Véronique Rebuffel, J.-C. Baritoux, Emeric Bergmann, Pierre Marcoux, Isabelle Espagnon, et al.. Results on identification of bacteria aging in complex environmental samples using Raman spectroscopy. European Conferences on Biomedical Optics, Jun 2019, Munich, Germany. pp.18, 10.1117/12.2527141 . cea-04559050

**HAL Id: cea-04559050**

**<https://cea.hal.science/cea-04559050>**

Submitted on 25 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Results on identification of bacteria aging in complex environmental samples using Raman spectroscopy

V. Rebuffel<sup>\*a</sup>, J-C. Baritoux<sup>a</sup>, E. Bergmann<sup>a</sup>, P. Marcoux<sup>a</sup>, I. Espagnon<sup>b</sup>, S. Morales<sup>a</sup>

<sup>a</sup>Univ. Grenoble Alpes, CEA, LETI, F-38000 Grenoble, France

<sup>b</sup>CEA, LIST, F-91191 Gif-sur-Yvette, France

## ABSTRACT

Spontaneous Raman scattering is a reliable technique for fast identification of single bacterial cells, when spectra are acquired in laboratory conditions where bacteria growth and state are controlled. We have developed a multi-modal system combining Raman spectroscopy and darkfield imaging, aiming at analysing environmental samples, typically in the field context of biological pathogens detection. Such samples are heterogeneous, both in terms of phenotype content and environmental matrix, even after a preliminary purification step. In this paper, we report a study on the identification of *Bacillus Thuriengensis* (BT) mimicing pathogen bacteria, embedded in a real-world matrix: a sample of surface water enriched with environmental bacterial species. The purpose is to evaluate both the detection limit of aging BT over time and the false alarm rate, in the conditions of our experiment.

**Keywords:** Raman spectroscopy, single bacterial cell identification, environmental samples, starvation.

## 1. INTRODUCTION

Vibrational methods, and among them, confocal Raman micro-spectroscopy, are very promising techniques for fast and robust identification of microorganisms. Raman spectroscopy has been demonstrated on laboratory cultures to identify bacteria at the strain level with biomass ranging from micro-colonies down to single cells<sup>1</sup>. In the context of pathogen threat detection the technique has to be deployed in the field, applied on environmental sample without growth. It is well known that the Raman spectra recorded on microorganisms depend on growth and environmental conditions. For bacteria cultivated under various conditions it was previously shown that identification at the species or even strain level is possible providing a comprehensive database is employed<sup>2,3</sup>. Yet, establishing a database for complex environmental samples would be prohibitive. Besides, as the amount of variability included in the database increases, the classes progressively merge, resulting in poor identification results. When dealing with environmental samples, one is faced with the aging of bacteria in hostile conditions. While most microorganisms enter a quiescent state and ultimately die, some will sporulate. This respectively translates in a loss of the Raman signal, or in drastic spectral changes. All these remarks argue in favor of an identification algorithm using a database composed only of the most relevant conditions, assessing the distance to the base.

We developed a classification algorithm to deal with spectra of vegetative bacteria and spores isolated from environmental matrices. It uses a hierarchical decision structure with an evaluation of the decision reliability at each stage. In this work we present the results obtained in two experiments. In the first one, *Bacillus thuringiensis* are aging in a poorly nutritive medium. In the second one, the bacteria were inoculated in surface water sampled from a pond. In this case, a set of species associated with false positives in the context of pathogen detection were also added. The goal of these experiments was to assess the limits of a single cell Raman identification for bacteria aging in the presence of biological noise.

## 2. MATERIAL AND METHOD

### 2.1 Sample preparation and Raman acquisition protocol

We used a strain of *Bacillus thuringiensis* (BT): ATCC-10792. Growth is performed in standard lab condition (a few hours in petri dish at 37°). For a first experiment, BT are introduced in water with poorly nutritive elements and stored at 20°C to observe their evolution over time, but no other bacteria were introduced.

\*veronique.rebuffel @cea.fr

In a second experiment, the medium is sampled from surface water of a small city pond (temperature < 5°). Furthermore we add 6 bacterial species listed in the environmental norm NF X 52-120: 2015-05 (noted Env): *Pseudomonas lurida*, *Pseudomonas rhizosphaerae*, *Pseudomonas fluorescens*, *Pseudomonas jessenii*, *Arthrobacter oxydans*, *Cellulosimicrobium cellulans*, – at concentration 10<sup>7</sup>cfu/mL. BT bacteria are inoculated (10<sup>8</sup>cfu/mL) and the mixture is stored at 20°.

For the acquisition, we used a custom Raman instrument developed in our lab. A detailed description can be found in <sup>4</sup>. Briefly, the system allows targeting of bacterial cells, and the measurement of single-cell Raman spectra using a confocal arrangement. The spectrometer resolution is 7 cm<sup>-1</sup> over the band 500–3400 cm<sup>-1</sup>. The system has been integrated in a transportable instrument called BACRAM, designed to answer the need for rapid identification of bacterial pathogens.

## 2.2 Classification algorithm

Acquired spectra are first preprocessed, then compared to a previously acquired database using supervised multivariate classification algorithms. The database is composed of more than 8000 spectra, and is representative of true pathogens bacteria while being different for obvious facility reasons. It contains four families (Bacillaceae, Burkholderiaceae, Francisellaceae, Enterobacteriaceae) and 14 species (3 or 4 strains per species). Furthermore, 4 *Bacillus* species can sporulate. For database acquisition, all species are grown in standard lab conditions.

Pre-processing of spectra consists successively in cosmic spikes removal, smoothing using Savitzky-Golay polynomial filters, subtraction of baseline by Clayton filtering, channels binning, restriction to region-of-interest, and normalization.

For relatively standardly grown bacteria, an identification using a SVM classifier has shown to provide the best results<sup>3</sup>. When considering more variable conditions, and also bacterial species not included in the database, other classification algorithms were preferred. We developed a hierarchical frame :

- a first step implemented by a PLS (« Partial Least Square ») in addition to the test of a few characteristic peaks, leading to a partition : spore (Sp) / vegetative (Veg) / non exploitable bacteria (NE-Sp or NE-Veg) / outliers (Out).

- a second step performing a LDA decomposition ("Linear Discrimination Analysis") using the Sp database or the Veg one depending on the first step. Classification is performed at family and species levels, followed by a Mahalanobis distance, allowing to provide information on the more probable species and a confidence level.

A threshold allows to tune the algorithms at a more or less conservative level, thus controlling the ratio between false positives / negatives.

## 3. RESULTS AND DISCUSSION

### 3.1 *Bacillus thuringiensis* in water

In the first experiment (illustrated Figure 1), the identification process at T0 gives 97% of Veg-BT. After 5 days, 41% were still identified as Veg-BT, the rest being NE-Veg. After 10 days, only NE were found (Veg and Sp). Thus during the evolution of Veg-BT, they progressively are classified as NE, with a complete evolution in a few days. Also importantly, no other species of the database are found, which would consist in false alarms if the database is considered as pathogens.

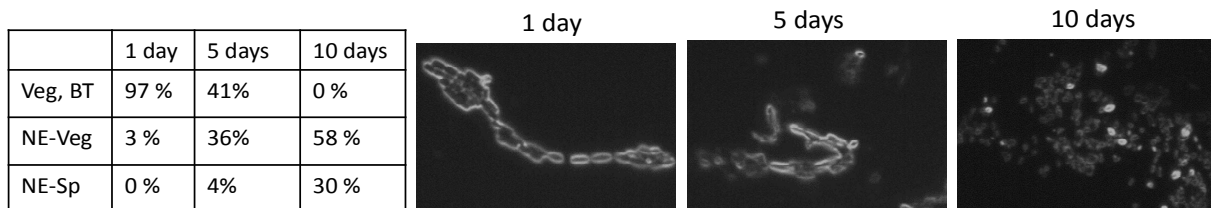


Figure 1. *Bacillus thuringiensis* (BT) after 1, 5 and 10 days in water matrix. Left : Identification results.

### 3.2 *Bacillus thuringiensis* in complex real-world sample

The microscope image shows various microorganisms: unknown ones coming from the lake water, introduced bacteria, both BT and Env ones. Sampling bacteria for spectra acquisition is done manually, statistically representing cells distribution in the image. Ten acquisition time points were performed over 3 days, about 100 spectra each. For the first point, BT is measured in clean water to be used as reference. Figure 2 shows some images, and the results of the first identification step. The red curve (Out+NE\_Veg+NE\_Sp) corresponds to the cases where the system can not consider the spectra for classification with a sufficient confidence level. They could be ignored in final decision step.

Acquisition time points:

1	T0
2	2h
3	4h
4	6h
5	1day
6	1day + 4h
7	1day + 7h
8	2days
9	2days + 5h
10	3 days

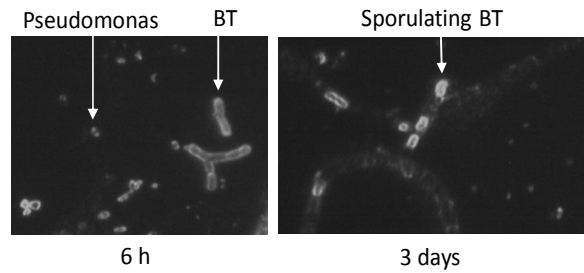
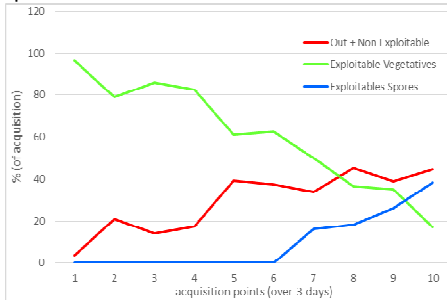


Figure 2. First step of identification process.

For the second step, we have tested two approaches. The environmental species (Env) have been learned previously and can be integrated in the database (case 2 – Figure 3-right) or not (case 1 – Figure 3-left). It is fundamental to remind that the ground truth is unknown. In both cases, identified Veg-BT decrease over time while Sp-BT increase (sporulation). A minimum (time point 6) of 10% bacteria still allow to conclude the presence of BT. In case 1, other bacteria of the database are found (red curve), interpretable in false positive except if these species are present in the sampled water. No reliable ones (orange curve) is an output of the algorithm corresponding to bacteria that are "too far" from any species of the database. In case 2, the black curve records the bacteria that are identified as Env (proposed in the database). Mainly of the spectra of orange and red curves of case 1 are counts as Env of case 2, showing the influence of integrating Env bacteria in the database. Considering visually the images for each acquisition, we observe a strong correlation between the visual aspect and the (BT / Env) classification – this shape criteria is not used in the current decision algorithm.

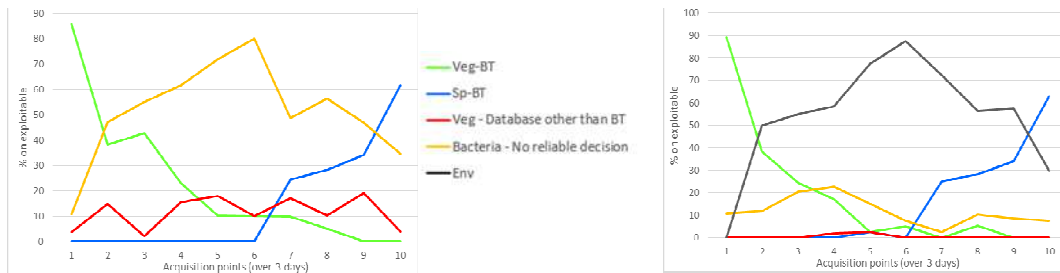


Figure 3. Second step of identification process, environmental species being out of (left) and in (right) the database.

### 3.3 Conclusion

In the context of our experiments, we observe that during the evolution of bacteria, the performances decrease but still allow to conclude after a few days. This result can not be generalized due to all possibilities of real world samples. But the observed trends put the focus on main questions: the database composition, knowing that it is not possible to learn all possible conditions, and the interest to integrate in database a set of representative species. Also the classification algorithm should be designed to deal with over time damaged bacteria –at assessed by Raman spectra. Outputs of the algorithm such as the Non exploitable spectra or Non reliable ones are fundamental to control the false alarms rate and desired confidence level.

### REFERENCES

[1] Huang, W.E., et al., "Raman microscopic analysis of single microbial cells.," *Analyt. chem.* 76(15), 4452–4458 (2004)  
 [2] Harz, M., et al., "Micro-Raman spectroscopic identification of bacterial cells of the genus *Staphylococcus* and dependence on their cultivation conditions.," *The Analyst* 130(11), 1543–1550 (2005).  
 [3] Baritau, J.C., et al., "A study on identification of bacteria in environmental samples using single-cell Raman spectroscopy: feasibility and reference libraries", *Environ Sci Pollut Res* 23,8184–8191 (2016).  
 [4] Strota, S.A., et al., "Single bacteria identification by Raman spectroscopy.," *J. of biomedical optics* 19(11), (2014)