Fast and robust identification of single bacteria in environmental matrices by Raman spectroscopy

Jean-Charles Baritaux, Emmanuelle Schultz, Anne-Catherine Simon, Anne-Gaelle Bourdat, Isabelle Espagnon, Patricia Laurent, Jean-Marc Dinten

Fast and robust identification of single bacteria in environmental matrices by Raman spectroscopy

Jean-Charles Baritaux*[a], Emmanuelle Schultz[a], Anne-Catherine Simon[b], Anne-Gaelle Bourdat[a], Isabelle Espagnon[b], Patricia Laurent[a], Jean-Marc Dinten[a]
[a] Université Grenoble-alpes, F-38000 Grenoble, France
CEA, LETI, Minatec-Campus, 17 avenue des Martyrs, 38054 Grenoble Cedex 9, France
[b] CEA, LIST, Gif-sur-Yvette, F-91191, France

ABSTRACT

We report on our recent results on robust identification of single bacterial cells embedded in various environments using Spontaneous Raman Scattering. Five species of bacteria were considered, two of which (*B. Subtilis* and *E. Coli*) were grown under various conditions, or embedded in two real-world matrices. We recorded the Raman spectra of single cells with a confocal instrument developed in our lab, and performed identification at the species level. Our system integrates a Lensfree imaging module that allows fast detection of bacteria over a large Field-Of-View. Identification rates comparable to those obtained on lab cultures were possible using a comprehensive database containing spectra from bacteria in all environments. In addition, *B. Subtilis* was correctly identified in 95.5% of the cases using a database composed exclusively of spectra obtained in standard conditions. This is very promising for pathogen threat detection where the construction of an exhaustive database may be challenging.

Keywords: Raman spectroscopy, Lensfree imaging, Bacteria identification, Single Cell Analysis, Chemometric Analysis

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[1-3]. In the context of pathogen threat detection the technique has to be deployed in the field. Raman spectra aggregate the contributions from all chemical species present in the probing volume. The spectra recorded on microorganisms are therefore dependent on the phenotype under scrutiny, which is largely dictated by growth and environmental conditions. Studies conducted on bacteria cultivated under various conditions (medium, temperature, age) indicate that identification at the species or even strain level is possible, provided a comprehensive database containing spectra from all relevant conditions is employed[4,5]. Similar findings were reported for bacteria measured directly in patients' cerebrospinal fluid[6,7], and tap water[8].

In this work we assess the robustness of identification of single bacteria subjected to various growth conditions or embedded in two real-world environmental matrices. The two chosen matrices are representative of field conditions: river water and ambient air (containing particles) dissolved in water. We use Raman spectra acquired with a short

integration time (10 s), a SVM classifier, and perform identification at the species level. We discuss the choice of a comprehensive reference database containing spectra from all conditions, against a reduced database of standard spectra alone.

2. MATERIALS AND METHODS

2.1 Sample preparation

Six bacterial strains from five species were considered in this study: *B. Cereus* ATCC10702 (BC), *B. Subtilis* ATCC23857 (BS), *E. Coli* ATCC9637 (EC), *E. Coli* ATCC11775 (EC), *S. Epidermidis* ATCC16990 (SE) and *S. Marscesens* ATCC27137 (SM) (see Table 1). Liquid cultures were prepared using a volume of 25 mL, and incubating time of 16 h. For *E. Coli* and *S. Marcescens* 100 uL of the resulting suspension was sampled and further incubated in 10 mL of fresh medium for 4 h. Bacterial growth was monitored by optical density. The cells were then washed in FreeWater using 3500 rpm centrifugation for 2 min and re-suspended in FreeWater at a concentration of about 10^5-10^6 cells/uL. One microliter of suspension was sampled and deposited on a Quartz slide (TedPella Inc. 19x19x0.5 mm). The smear was evaporated for 2 min and immediately taken to our instrument for examination and Raman spectrum collection.

Samples of all five species were prepared using the standard culture conditions reported in Table 1. In addition, non-standard samples of *B. Subtilis* and *E. Coli* were obtained by varying culture temperature, medium and re-suspension matrix. Note that a single parameter was varied for each non-standard sample type. The three sets of altered conditions are summarized in Table 2. First, we considered three culture media of increasing nutrient content. Medium 1 was composed of 0.5 g/L NaCl (SIGMA S5886), 0.186 g/L KCl (PROLABO 26764.298), 4.8 g/L MgSO4 (SIGMA M2643), 3.603 g/L alpha-D-glucose (ALDRICH 15,896-8). Medium 2 was the same composition with the addition of 20 g/L Soy Peptone (enzymatic digest FLUKA 87972). Medium 3 was the same as Medium 2, plus 5 g/L yeast extract (FLUKA 70161). In the second set of non-standard conditions, culture temperatures and then media of *B. Subtilis* and *E Coli* were exchanged. And finally, instead being re-suspended in FreeWater, the cells were embedded in two real-world matrices: river water (Seine River), or ambient air dissolved in water.

2.2 Instrument

The database of Raman spectra was built using the setup depicted in Figure 1. A 532 nm laser (Spectra Physics Excelsior 532-50-CDRH) is used as the Raman excitation source. A razor-edge filter (Semrock LPD01-532RS-25) steers the laser beam at 45° into a 100x microscope objective (Olympus LMPLFLN, NA 0.8). A beam expander was added in the excitation path to adapt the beam diameter to the size of the objective entrance pupil, to therefore focus light within a spot with less than 1um in diameter onto the sample. Raman scattered light arising from individual bacterial cell is collected by the same microscope objective, filtered from Rayleigh scattering thanks to both the razor edge plate and a notch filter (Semrock NF03-532E), and finally focused into the spectrometer optical fiber. Here, a customized Hyperflux U1-532 platform (Tornado Spectral systems, Toronto Canada) equipped with a TE-cooled Mity CCD was chosen because of its efficient throughput (comprehensive Raman spectra are acquired in 10 seconds only), good spectral resolution (7 cm^{-1}), and large bandwith (500-3400 cm^{-1}).

In addition to the Raman spectroscopy modality, the setup implements forward lensless imaging and darkfield microscopy to make bacteria detection and targeting rapid and easy. Lensless imaging is embedded in the sample holder.

It consists in an 8-bit 2592x1944 pixels CMOS sensor (MT9P031, Aptina Imaging) placed one millimeter under the substrate that must be transparent to visible light. Figure 2(a) shows a lensless image obtained when the laser beam is defocussed so that the spot size matches the size of the cmos sensor: the whole droplet and bacteria aggregates are imaged, making the selection of a region of interest straightforward. Darkfield microscopy is achieved when the LED ring attached to the objective is switched on, and the laser beam is focused onto the sample. As shown in Figure 2(b), both the bacterial cells and laser spot are clearly seen. Fine alignment of sample and laser beam is realized using translation stages.

Recently, this setup has been integrated in a transportable instrument, the so-called BACRAM system, depicted Figure 3. This novel system is designed to answer the need for rapid and specific identification of bacterial pathogens.

2.3 Data Analysis

Data analysis (spectra pre-processing, calculation of indicators and classification) was performed using the R software environment, with existing functions or routines specifically developed for this use.

Pre-processing of spectra consisted in cosmic spikes removal[9], smoothing, restriction to a region-of-interest (ROI), and finally, normalization by the mean. Smoothing was performed using Savitzky-Golay polynomial filters (degree 4, on 9 points). Given the inter-channel distance of 2 cm^{-1}, a filter width of 9 points corresponds to 18 cm^{-1}. In comparison, typical peak full-width-at-half-maximum ranges from 20 cm^{-1} to 60 cm^{-1}. This smoothing approach enables to reduce the noise in the signal without peak distortion and loss of intensity. We chose a ROI composed of the two regions 650 cm^{-1}-1800 cm^{-1} and 2600 cm^{-1}-3200 cm^{-1}.

Classification was performed using the support vector machine (SVM) implementation “svm” of the R package “e1071”, interfacing the” LIBSVM” library. We used SVM with a linear kernel. Classification performance was assessed by external cross-validation (leave-one-date-out) with training set balancing. To further evaluate classification stability of each spectrum, every cross-validation round was repeated ten times using 90% of the training set sub-sampled randomly. We carried out the classification at the species level. Two options were considered for the training set: a comprehensive reference database containing all conditions and matrices, a standard reference database composed of standard spectra alone.

3. RESULTS AND DISCUSSION

3.1 Database description

Table 2 presents the database acquired in this work. It was assembled in two measurement campaigns, over the course of 4 months. The database contains in total 2056 spectra of the five bacteria BS, BC, EC, SE, SM, cultured in standard (all bacteria), and non-standard conditions (limited to BS and EC). Average spectra in standard conditions are shown Figure 4. As described in Section 2.1, non-standard conditions correspond to media of varying nutrient content, various temperatures, and matrices (Air and Seine). These conditions were chosen to induce heterogeneous phenotypes. In addition, Air and Seine matrices may contain Raman active substances and particles generating background in the Raman spectra. Each sample consists of a smear of bacteria suspension that is evaporated on a slide prior to Raman analysis. Ten spectra were collected in each smear, after which a new smear was examined. In general, strong aggregation was observed on the border of the smears, while the center displayed isolated bacteria. Acquisitions were performed on individuals located in the center. Samples prepared in Air and Seine contained a variety of fluorescent and

Raman active particles unrelated to the investigated bacteria. For the most part, these particles localized in the large aggregates formed at the border of the smears. In the center of the smears, we discriminated between the bacteria of interest and other particles based on their morphology and reflectivity. Unambiguous localization of the ten cells needed for acquisition in each smear was possible because of the high concentration in cultured bacteria of our samples (10^5-10^6 cells/uL). Spatial filtering resulting from the tight laser focus allowed minimum distortion of the bacteria spectra. In particular, we did not observe significant amounts of fluorescence other than bacteria autofluorescence.

3.2 Cluster Analysis

In this Section we intend to assess intra-species variability caused by environmental matrix and growth medium, and compare it to inter-species variability. To that end we computed a supervised dendrogram on all the spectra in the database (Table 2). We started by calculating the average spectrum for each category (one line in Table 2). We then applied the complete linkage clustering algorithm to the averages and obtained the result presented Figure 5. Looking first at the spectra corresponding to the standard conditions, we note that they belong to well-separated clusters, except for EC and SM, which are found very close. Turning now to spectra from non-standard conditions we observe that all BS spectra are assigned to a tight cluster, except for BS Air. Similarly, all EC spectra are grouped together, except for EC Air and EC Seine, which are assigned to the BS cluster with BS Air and BS Seine, respectively. Intra-species variability appears to be lower than inter-species variability in the case of BS. This hints at a possible identification of BS using the standard spectra alone as a reference. The case of EC is more delicate. Already on the standard spectra, EC does not stand out clearly, but is grouped closely with SM. In general, these two species exhibited a lower Raman yield, leading to lower signal-to-noise. The relative contribution of the matrix in their spectrum may be higher for these species, leading to higher dispersion of spectra according to matrices. A comprehensive training set with finer classification method is therefore required for EC.

3.3 SVM Classification

Table 3 to Table 6 present the results of a SVM classifier at the species level. In Table 3 the classifier was run on spectra from standard conditions. In Table 4 toTable 6, culture conditions and environmental matrix of BS and EC are varied. We show side by side the classification results obtained with comprehensive and standard training set. Coming back to Table 3 (standard conditions), the average identification score is 90.1%, with average standard deviation 4.2% per spectrum. We note a better performance for gram positive bacteria (98.6% average on BC, BS, SE), compared to gram negative (76.6% average on EC, SM). Confusion between EC and SM explains the degraded performance. This is in line with the trend observed in the dendrogram of Section 3.2, which attributed a high similarity to EC and SM. We now examine Table 4 toTable 6. The average classification rate with a comprehensive database is 89%. This rate excludes Medium 1 whose nutrient content did not allow bacterial growth, which results in mediocre classification scores. In particular, the combination Medium 1/standard database leads to confusion of BS with the other gram positive. Adding all media to the database is beneficial in this case. Nevertheless, with the exception of severe conditions such as Medium 1 or to a lesser extent TSB 37, BS scores are consistently higher than 95%, regardless of the presence of altered conditions in the reference database (95.5% average rate with a standard database). As a consequence, the use of a comprehensive database does not seem necessary for identifying BS at the species level. In contrast, EC is often confused with SM when the database of standard spectra is employed. In this case identification rate is only 54.9%, including 40% of confusion with SM (excluding Medium 1). This is solved by using a comprehensive database (81.5% average score). Interestingly, Medium 3, Air, and Seine lead to discrimination scores higher than 87%, which improves on standard conditions. Improvement is attributed to an increased distinction of the phenotypes under these conditions.

4. CONCLUSION

Raman spectra taken with a short acquisition time (10 s) were employed to perform identification at the species level on five bacteria species. The species *B. Subtilis* and *E. Coli* were subjected to varying culture conditions and embedded in real-world environmental matrices. An average identification rate of 89% was achieved using a SVM classifier together with a comprehensive database. This result extends earlier identification reports to the case of two field-relevant environmental matrices: river water and ambient air containing particles dissolved in water. We further observed that species *B. Subtilis* could be robustly identified in 95.5% of the cases using a reduced database composed of standard spectra alone, *i.e.* without incorporating all conditions and matrices in the base. This was not the case of *E. Coli* for which a comprehensive database was needed in order to decrease the confusion with the other gram negative species. *E. Coli* displayed large intra-species variability with environmental conditions. This correlates with lower Raman yield of *E. Coli* spectra in standard conditions. This discussion suggests that in a number of cases a standard laboratory-build database is sufficient to provide satisfying identification of a single bacterium at the species level. This study thus brings a key element in favor of the ability of the technique to be deployed in the field and to answer the needs of first responders.

5. REFERENCES

[1] Rosch, P., Harz, M., Schmitt, M., Peschke, K.-D., Ronneberger, O., Burkhardt, H., Motzkus, H.-W., Lankers, M., Hofer, S., et al., "Chemotaxonomic identification of single bacteria by micro-Raman spectroscopy: application to clean-room-relevant biological contaminations.," Applied and environmental microbiology 71(3), 1626–1637 (2005).

[2] Strola, S.A., Baritaux, J.-C., Schultz, E., Simon, A.C., Allier, C., Espagnon, I., Jary, D., and Dinten, J.-M., "Single bacteria identification by Raman spectroscopy.," Journal of biomedical optics 19(11), (2014).

[3] Huang, W.E., Griffiths, R.I., Thompson, I.P., Bailey, M.J., and Whiteley, A.S., "Raman microscopic analysis of single microbial cells.," Analytical chemistry 76(15), 4452–4458 (2004).

[4] Harz, M., Rosch, P., Peschke, K.-D., Ronneberger, O., Burkhardt, H., and Popp, J., "Micro-Raman spectroscopic identification of bacterial cells of the genus Staphylococcus and dependence on their cultivation conditions.," The Analyst 130(11), 1543–1550 (2005).

[5] Hutsebaut, D., Maquelin, K., De Vos, P., Vandenabeele, P., Moens, L., and Puppels, G.J., "Effect of culture conditions on the achievable taxonomic resolution of Raman spectroscopy disclosed by three Bacillus species.," Analytical chemistry 76(21), 6274–6281 (2004).

[6] Rösch, P., Stöckel, S., Meisel, S., Bossecker, A., Münchberg, U., Kloss, S., Schumacher, W., and Popp, J., "Bacterial identification in real samples by means of micro-Raman spectroscopy," in Proc. SPIE 8087, 808708–808708–6 (2011).

[7] Harz, M., Kiehntopf, M., Stockel, S., Rosch, P., Straube, E., Deufel, T., and Popp, J., "Direct analysis of clinical relevant single bacterial cells from cerebrospinal fluid during bacterial meningitis by means of micro-Raman spectroscopy.," Journal of biophotonics 2(1-2), (2009).

[8] Tripathi, A., Jabbour, R.E., Treado, P.J., Neiss, J.H., Nelson, M.P., Jensen, J.L., and Snyder, A.P., "Waterborne Pathogen Detection Using Raman Spectroscopy," Applied Spectroscopy 62(1), 1–9 (2008).

[9] Espagnon, I., Ostrovskii, D., Mathey, R., Dupoy, M., Joly, P.L., Novelli-Rousseau, A., Pinston, F., Gal, O., Mallard, F., et al., "Direct identification of clinically relevant bacterial and yeast microcolonies and macrocolonies on solid culture media by Raman spectroscopy.," Journal of biomedical optics 19(2), (2014).

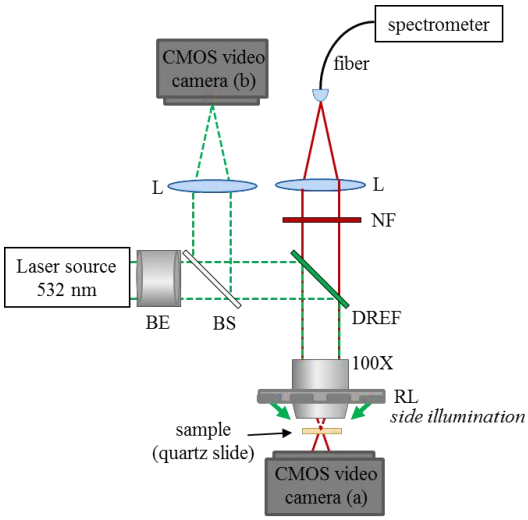

Figure 1 Schematic of the experimental setup. Abbreviations: BE: beam expander (x5), BS: beam splitter, DREF: dichroic razor-edge filter, 100X: microscope objective, NA = 0.8, RL: Ring LED, NF: Notch filter, L: achromatic lens. Video camera (a) enables lensless

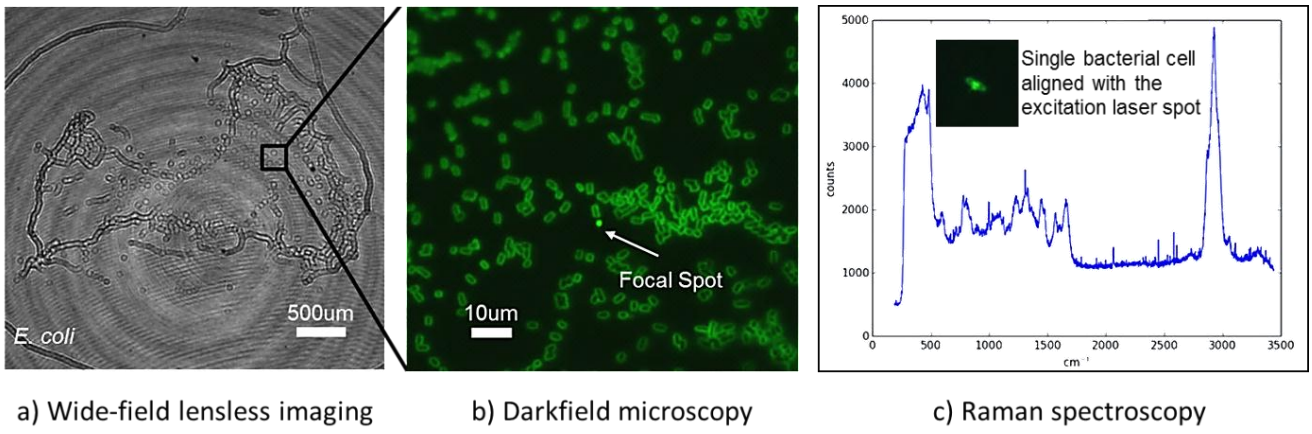

Figure 2 Illustration of the operation steps for a 1uL droplet with E. coli bacteria: a) lensless image giving a wide-field of view of the droplet evaporated onto the quartz substrate, and facilitating the selection of region of interest (square). b) standard microscope image of the bacteria and the laser spot to assist in accurate targeting, and c) typical Raman spectrum obtained on a single bacteria with 10s acquisition time.

Figure 3 Photograph of the complete integrated Raman analyzer BACRAM

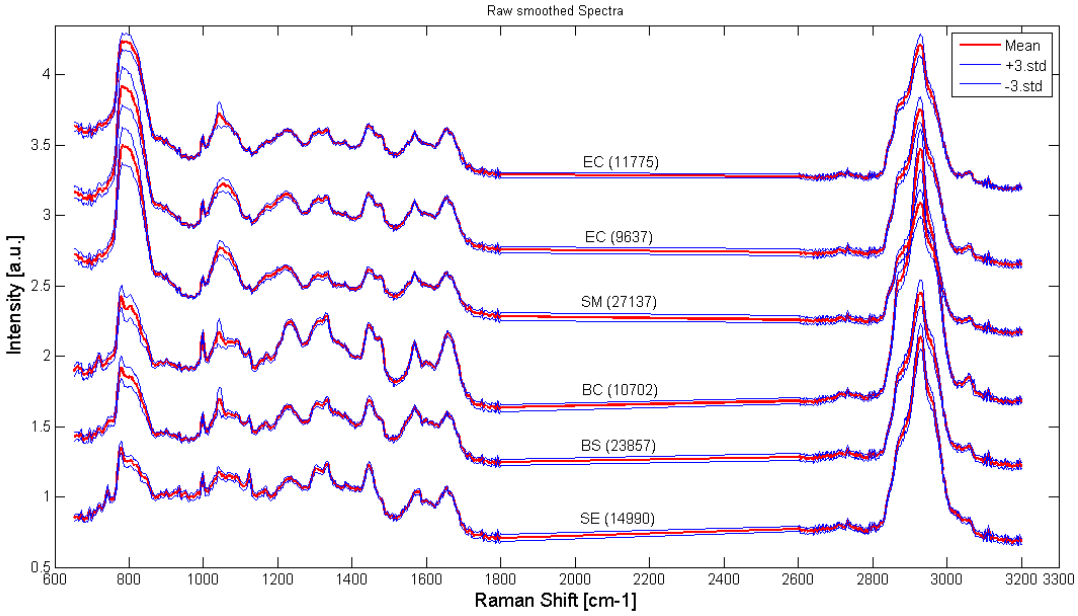

Figure 4 Average and standard deviation of the spectra of *B. Cereus*, *B. Subtilis*, *E. Coli*, *S. Epidermidis* and *S. Marcescen.*measured in standard conditions.

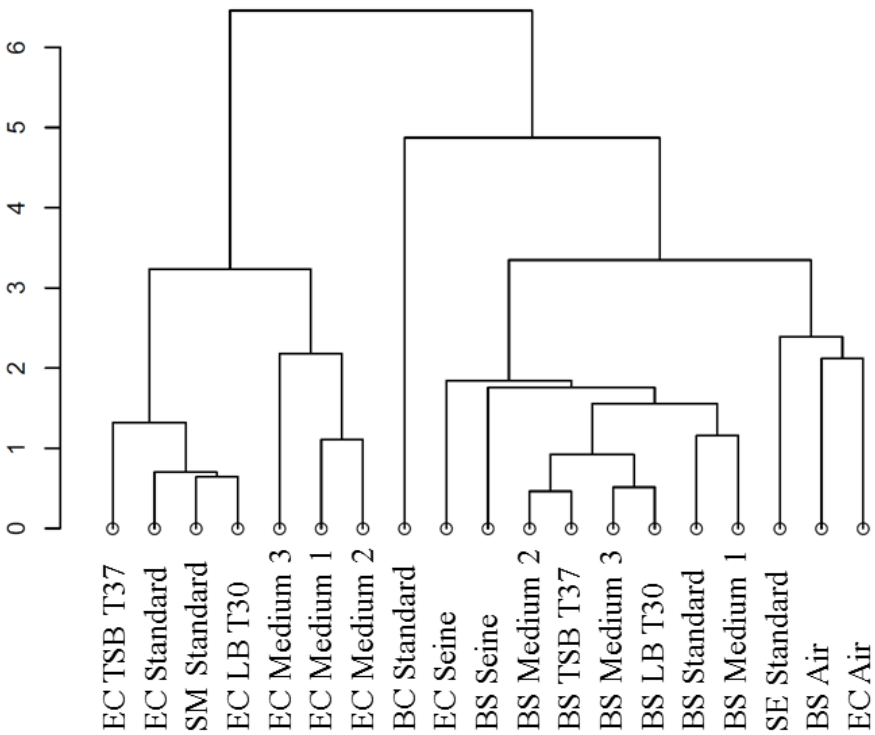

Figure 5 Supervised dendrogram of all spectra acquired in this study.

Table 1 Bacterial strains and standard conditions used in this study.

Strain	Abbreviation used in text	Standard Culture Conditions*
B. Subtilis ATCC23857	BS	TSB, T30
B. Cereus ATCC10702	BC	TSB, T30
E. Coli ATCC9637	EC	LB, T37
E. Coli ATCC11775	EC	LB, T37
S. Epidermidis ATCC16990	SE	LB, T30
S. Marscesens ATCC27137	SM	LB, T37

*TSB : Trypticase Soy Broth, LB : Luria-Bertani Broth, T30 : Temperature 30°C, T37 : Temperature 37°C.

Table 2 Bacteria database.

Species	Conditions	Total number of acquired spectra	Species	Conditions	Total number of acquired spectra
BC	Standard	158	BS	LB 30	40
BS	Standard	175	EC	LB 30	61
EC	Standard	198	BS	TSB 37	62
SE	Standard	110	EC	TSB 37	60
SM	Standard	169	BS	Air	120
BS	Medium 1	102	EC	Air	100
EC	Medium 1	102	BS	Seine	120
BS	Medium 2	100	EC	Seine	97
EC	Medium 2	92			
BS	Medium 3	100			
EC	Medium 3	90			

Table 3 Confusion Matrix (on “clean” bacteria optimized acquisition conditions)

	BC	BS	SE	EC	SM
BC	**97.9**	2.1			
BS	2	**98.0**			
SE			**100**		
EC		0.1		**79.2**	20.7
SM	1.8			24.2	**74.0**

Average classification rate : 90.1 % ± 4.2%*

*average standard deviation over 10 classifications

Table 4 Confusion Matrix: media of increasing nutrient content

	All spectra in training set					Only standard spectra in training set				
	BC	BS	SE	EC	SM	BC	BS	SE	EC	SM
BS Medium 1	3.9	**89.9**	0.1	6.1		23.9	**37.7**	37.7	0.7	
EC Medium 1		18.5	15.1	**56.6**	9.8	0.9	7.2	23.1	**10.8**	58
BS Medium 2		**99.7**	0.3				**97.2**	2.8		
EC Medium 2	0.2	3.2		**81.4**	15.2	0.9	3.6	0.2	**41.7**	53.6
BS Medium 3		**99.5**	0.4	0.1			**98.1**	1.9		
EC Medium 3	0.3			**88.8**	10.9	1.4			**51**	47.6

Table 5 Confusion Matrix: varying conditions and media

	All spectra in training set					Only standard spectra in training set				
	BC	BS	SE	EC	SM	BC	BS	SE	EC	SM
BS LB 30	0.5	**99.5**				0.7	**99.3**			
EC LB 30	0.7		1.6	**75.9**	21.8	4	1.3	1.6	**68.7**	24.4
BS TSB 37	0.1	**83.7**	3.2	2.3	10.7	0.2	**81.9**	4.7	2.3	10.9
EC TSB 37	3.2	2.8		**67.2**	26.8	5	4		**60.5**	30.5

Table 6 Confusion Matrix: varying environmental matrix

	All spectra in training set					Only standard spectra in training set				
	BC	BS	SE	EC	SM	BC	BS	SE	EC	SM
BS AIR	0.2	**98.8**		0.8	0.2	0.6	**96.7**	0.2	1.8	0.7
EC AIR		1.4	1	**88.3**	9.3	0.3	0.6	2	**55.2**	41.9
BS SEINE	0.4	**98.1**		0.6	0.9	0.3	**99.5**		0.2	
EC SEINE		2.6		**87.3**	10.1		16		**42**	42