Short communication

SpectraBank: An open access tool for rapid microorganism identification by MALDI-TOF MS fingerprinting

Authors: *Karola Böhme[1], Inmaculada C. Fernández-No[1], Jorge Barros-Velázquez[1], Jose M. Gallardo[2], Benito Cañas[3] and Pilar Calo-Mata[1,*]*

Affiliations:[1]Department of Analytical Chemistry, Nutrition and Food Science, School of Veterinary Sciences, University of Santiago de Compostela, Lugo, Spain, [2]Department of Food Technology, Institute for Marine Research (IIM-CSIC), Vigo, Spain, [3]Department of Analytical Chemistry, University Complutense of Madrid, Madrid, Spain

***Correspondence:** Pilar Calo-Mata, Department of Analytical Chemistry, Nutrition and Food Science, School of Veterinary Sciences, University of Santiago de Compostela, Lugo, Spain, Tel: +(34)647344274; Fax: +(34)982252195; E-mail: p.calo.mata@usc.es

Running title: SpectraBank: Microorganism identification by MALDI-TOF MS

Abbreviations: ATCC, American Type Culture Collection; CECT, Spanish Type Culture Collection; MALDI-TOF-MS, Matrix-Assisted Laser Desorption Ionization-Time of Flight-Mass Spectrometry; α-CHCA, α-Cyano-4-Hydroxycinnamic Acid; PCA, Plate Count Agar; BHI, Brain Heart Infusion

Abstract

Matrix-assisted laser desorption/ionisation (MALDI) time-of-flight (TOF) mass spectrometry (MS) has proved to be an accurate, rapid and cost-effective technique for microbial identification in which the spectral fingerprint of an unknown strain can be compared to a database of spectra from reference strains. Most of the existing databases are private and often costly to access, and little spectral information is shared among researchers. The objective of the present communication is to introduce the SpectraBank database (www.spectrabank.org), which provides open access MALDI-TOF mass spectra from a variety of microorganisms. This work aims to familiarise readers with the SpectraBank database, from the sample preparation, data collection, and data analysis to how the spectral reference data can be used for microbial species identification. The database currently includes more than 200 MALDI-TOF MS spectra from more than 70 bacterial species and links to the freely available web-based application SPECLUST (http://bioinfo.thep.lu.se/speclust.html) to allow comparisons of the obtained peak mass lists and evaluate phyloproteomic relationships. The SpectraBank database is intended to be expanded by the addition of new spectra from microbial strains, obtained in our laboratory and by other researchers.

Keywords:

MALDI-TOF MS fingerprinting, Microbial identification, Microbial typing, phyloproteomics, SPECLUST, SpectraBank, Spectral Library of Microorganisms

Matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) mass spectrometry (MS) has been introduced for microorganism identification, demonstrating to be an accurate, rapid and cost-effective method [1-4]. Compared to conventional biochemical and molecular techniques for microorganism identification, MALDI-TOF MS fingerprinting requires minimal sample preparation and achieved more than 92 % of correct species identification [5-7]. Bacterial identification by MALDI-TOF MS fingerprinting is approximately two-thirds less expensive than conventional methods, when taking into account the cost of materials and staff [8, 9]. The highly selective and specific spectral fingerprints obtained by this technique allow the classification and identification of microorganisms at the genus, species, and even strain level [10, 11]. In addition, MALDI-TOF MS fingerprinting is a useful typing tool in developing phyloproteomic relationships. In this sense, it extends phenotypic and genotypic approaches, allowing a much more ample classification of microbial strains. Furthermore, some bacterial species are difficult to distinguish using the commonly applied DNA-based methods due to the high similarity of sequences in species from the same genus. A greater discriminating potential has been described for MALDI-TOF MS fingerprinting, which allows the differentiation and correct identification of much closer bacterial species and even strains of the same species [12-15].

For microbial identification, spectral profiles are compared to a previously constructed database of reference spectra [16]. Several private databases have been created and demonstrated to be suitable for high-throughput routine analysis in clinics and microbiological laboratories [5-7]. The Spectral Archive And Microbial Identification System (Saramis™; AnagnosTec GmbH, Potsdam, Germany) [17] and the MicrobeLynx bacterial identification system (Waters Corporation, Manchester, UK) [14, 16] search against spectral libraries of more than 500 microbial strains each. The MALDI Biotyper 2.0 (Bruker Daltonics) is the largest and most elaborate spectral database created to date and includes more than 1800 bacterial species [18]. One critical drawback of these private databases is their limited availability to other researchers and high costs for access. In addition, although microbial identification is carried

out with a high percentage of correct identification, little spectral information is provided for use in further studies. Mazzeo et al. (2006) constructed a freely available spectral library containing spectral profiles and peak mass lists for 24 food-borne bacterial species (http://bioinformatica.isa.cnr.it/Descr_Bact_Dbase.htm). However, this work was not continued, and the library remains small [19].

In this sense, the aim of our work is to create an open access spectral library to serve as a reference with which other researchers can carry out spectral comparisons for microbial species identification. The SpectraBank database (www.spectrabank.org, www.spectrabank.eu, www.spectrabank.es), which is linked to http://www.usc.es/gl/investigacion/grupos/lhica/spectrabank, includes spectra and peak mass lists for more than 200 bacterial strains and more than 70 bacterial species, with a focus on species of interest in the food sector. In addition, we demonstrated the utility of this reference library for the correct identification of unknown bacterial strains isolated from commercial seafood products [1].

The difficulty of creating an "in-house" database lies in the requirement for specific algorithms to analyse or compare obtained spectra and perform species identification. In our studies, we used the freely available web-based application SPECLUST (http://bioinfo.thep.lu.se/speclust.html) to compare peak mass lists. The web interface identifies peak masses that are common to different peak mass lists [20]. In this way, genus- and species-specific peaks have been identified for a number of bacterial species [15, 21]. In addition, the application SPECLUST contains a clustering option in which the peak mass lists are grouped based on similarity scores [20]. "Phyloproteomic" clustering has been used successfully for the identification of bacterial strains at the genus and species level [13, 15, 21-26] and has been shown to be a competent typing method for the classification of microbial strains [27, 28].

In this work, we describe in detail the protocol used to create the SpectraBank, including sample preparation, data analysis, and how to use the spectral information available in the SpectraBank for microbial classification and identification.

For comparison of the spectral fingerprints obtained by MALDI-TOF MS for microbial identification, the same sample preparation protocol and the same instrumental parameters should be applied [29, 30]. Figure 1 shows the sample preparation and mass spectrometric analysis procedures used to create the SpectraBank database. The bacterial strains are, in general, grown on PCA and incubated for 24 h at 30ºC. In some cases, special growth conditions (e.g., medium, temperature, or the elimination of oxygen) may be necessary. To obtain bacterial extracts, a 1-µl loop full of each bacterial culture is resuspended in 100 µl of a solution containing 50% acetonitrile (ACN) and 1% aqueous trifluoroacetic acid (TFA). The suspension is then vortexed and centrifuged at 5900 × *g* for 10 min. The supernatant is transferred into a new tube and stored at -20 ºC until analysis [31]. For MALDI-TOF MS analysis, a 1- to 5-µl aliquot of the sample extract is mixed with 10 µl of the matrix solution, which consists of 10 mg α-cyano-4-hydroxycinnamic acid (α-CHCA) in 1 ml of 50% ACN and 2.5% aqueous TFA. A 1-µl aliquot of this sample/matrix solution is then manually deposited onto a MALDI stainless steel plate and allowed to dry at room temperature.

Mass spectra are obtained using a Voyager DE STR MALDI-TOF Mass Spectrometer (Applied Biosystems) operating in linear mode, extracting positive ions with an accelerating voltage of 25000 V and a delay time of 350 ns. The grid voltage is set to 95%. Spectra are generally taken in an m/z range of 1500 – 15000 Da. Every spectrum is the sum of at least 1000 accumulated laser shots obtained in ten different regions and randomly selected in the same sample spot. Spectra are calibrated using an external protein calibration mixture consisting of 2 pmol/µl oxidised insulin B chain and 2 pmol/µl bovine insulin.

Figure 2 shows the data analysis process and the elaboration of spectral data submitted to the SpectraBank database. For spectral analysis, masses of 2000 – 10000 Da are considered due to the good reproducibility of the spectral profiles in that range. First, mass spectra are baseline corrected and noise filtered, and then data lists containing *m/z* values for signals with relative peak areas greater than 2 % are extracted from the mass spectral data with the DataExplorer® software.

To verify biological and technical reproducibility, at least four replicates are carried out for each sample, including two different extractions. The mass lists of the four replicates of one strain are compared with SPECLUST to identify representative and reproducible peak masses that are present in all spectral profiles of replicates. The web interface calculates the mass differences between two peaks taken from different peak lists and determines if the two peaks are identical after taking into account a certain measurement uncertainty (σ) [20]. In our studies, we set the measurement uncertainty (width of peak match score) to 10 Da. In this way, only those peaks that are present in all spectral replicates obtained for a strain are selected, producing one final peak mass list of *m/z* values and standard deviations that is submitted to the SpectraBank along with the spectral profile of the corresponding strain (Figure 2). Other investigators can download the spectral reference data for their strains of interest and carry out spectral comparison with their own strains, or use the data to identify an unknown strain. For spectral comparison, the final peak mass lists of the strains of interest are selected and analysed by SPECLUST. Common peak masses can be determined, and the peak mass lists can be clustered to visualise the phyloproteomic relationships between the spectra. For clustering, we use a correlation-based metric to calculate the distances between two peak mass lists and use the average linkage method to merge the two clusters with the smallest average of pair-wise distances.

Figure 3 shows how the SpectraBank reference data were used to identify an unknown strain, Sard1, that was isolated from spoiled fish. A final peak mass list was created for the strain Sard1 based on four spectral replicates, as described previously. Subsequently, this peak mass list was added to the archive of reference strain mass lists. The cluster analysis of the peak mass lists correctly identified the strain Sard1 as *Proteus vulgaris* (Figure 3).

In conclusion, SpectraBank is an open access database that permits other investigators to download reference data for spectral comparisons. Comparison of peak mass lists by searching common peak masses and clustering lead to an accurate identification. It should be noted that the described method is applicable to any microbial species. The SpectraBank can be easily enlarged by further strains and a

future objective is the possibility of submitting mass spectral data to the SpectraBank by other investigators.

Acknowledgements: *We thank Pedro Rey for his excellent assistance with the creation of the webpage. This work was funded by project 10PXIB261045PR from Xunta de Galicia and by project AGL2010-19646 from the Spanish Ministry of Science and Technology. The work of K. Böhme and I.C. Fernández-No is supported by a "Maria Barbeito" and "Lucas Labrada" research contract from Xunta de Galicia.*

References

[1] Böhme, K., Fernández-No, I., Gallardo, J., Cañas, B., Calo-Mata, P., *Food Bioprocess Technol.* 2011, *4*, 907-918.

[2] Giebel, R., Worden, C., Rust, S. M., Kleinheinz, G. T., Robbins, M., Sandrin, T. R., Allen, I. L., Sima, S., Geoffrey, M. G., *Adv. Appl. Microbiol.*, Academic Press 2010, pp. 149-184.

[3] Lay Jr., J. O., *Mass Spectrom. Rev.* 2001, *20*, 172-194.

[4] Freiwald, A., Sauer, S., *Nature Protocols* 2009, *4*, 732-742.

[5] van Veen, S. Q., Claas, E. C. J., Kuijper, E. J., *J. Clin. Microbiol.* 2010, *48*, 900-907.

[6] Bizzini, A., Durussel, C., Bille, J., Greub, G., Prod'hom, G., *J. Clin. Microbiol.* 2010, *48*, 1549-1554.

[7] Seng, P., Drancourt, M., Gouriet, F., La Scola, B., Fournier, P.-E., Rolain, Jean M., Raoult, D., *Clin. Infect. Dis.* 2009, *49*, 543-551.

[8] Hsieh, S.-Y., Tseng, C.-L., Lee, Y.-S., Kuo, A.-J., Sun, C.-F., Lin, Y.-H., Chen, J.-K., *Mol. Cell. Proteomics* 2008, *7*, 448-456.

[9] Nassif, X., *Clin. Infect. Dis.* 2009, *49*, 552-553.

[10] Bright, J. J., Claydon, M. A., Soufian, M., Gordon, D. B., *J. Microbiol. Methods* 2002, *48*, 127-138.

[11] Fernández-No, I. C., Böhme, K., Gallardo, J. M., Barros-Velázquez, J., Cañas, B., Calo-Mata, P., *Electrophoresis* 2010, *31*, 1116-1127.

[12] Donohue, M. J., Smallwood, A. W., Pfaller, S., Rodgers, M., Shoemaker, J. A., *J. Microbiol. Methods* 2006, *65*, 380-389.

[13] Vargha, M., Takáts, Z., Konopka, A., Nakatsu, C. H., *J. Microbiol. Methods* 2006, *66*, 399-409.

[14] Keys, C. J., Dare, D. J., Sutton, H., Wells, G., Lunt, M., McKenna, T., McDowall, M., Shah, H. N., *Infect., Genet. Evol.* 2004, *4*, 221-242.

[15] Böhme, K., Fernández-No, I. C., Barros-Velázquez, J., Gallardo, J. M., Cañas, B., Calo-Mata, P., *Electrophoresis* 2011, *32*, 2951-2965.

[16] Dare, D., in: Tang, Y.-W., Stratton, C. W. (Eds.), *Advanced Techniques in Diagnostic Microbiology*, Springer Science+Business Media, LLC, New York 2006, pp. 117-133.

[17] Erhard, M., Hipler, U.-C., Burmester, A., Brakhage, A. A., Wöstemeyer, J., *Exp. Dermatol.* 2008, *17*, 356-361.

[18] Seibold, E., Maier, T., Kostrzewa, M., Zeman, E., Splettstoesser, W., *J. Clin. Microbiol.* 2010, *48*, 1061-1069.

[19] Mazzeo, M. F., Sorrentino, A., Gaita, M., Cacace, G., Di Stasio, M., Facchiano, A., Comi, G., Malorni, A., Siciliano, R. A., *Appl. Environ. Microbiol.* 2006, *72*, 1180-1189.

[20] Alm, R., Johansson, P., Hjerno, K., Emanuelsson, C., Ringnér, M., Häkkinen, J., *J. Proteome Res.* 2006, *5*, 785-792.

[21] Böhme, K., Fernández-No, I. C., Barros-Velázquez, J., Gallardo, J. M., Calo-Mata, P., Cañas, B., *J. Proteome Res.* 2010, *9*, 3169-3183.

[22] Conway, G. C., Smole, S. C., Sarracino, D. A., Arbeit, R. D., Leopold, P. E., *J. Mol. Microbiol. Biotechnol.* 2001, *3*, 103-112.

[23] Dieckmann, R., Graeber, I., Kaesler, I., Szewzyk, U., von Döhren, H., *Appl. Microbiol. Biotechnol.* 2005, *67*, 539-548.

[24] Stackebrandt, E., Päuker, O., Erhard, M., *Curr. Microbiol.* 2005, *50*, 71-77.

[25] Carbonnelle, E., Beretti, J.-L., Cottyn, S., Quesne, G., Berche, P., Nassif, X., Ferroni, A., *J. Clin. Microbiol.* 2007, *45*, 2156-2161.

[26] Dubois, D., Leyssene, D., Chacornac, J. P., Kostrzewa, M., Schmit, P. O., Talon, R., Bonnet, R., Delmas, J., *J. Clin. Microbiol.* 2010, *48*, 941-945.

[27] Siegrist, T. J., Anderson, P. D., Huen, W. H., Kleinheinz, G. T., McDermott, C. M., Sandrin, T. R., *J. Microbiol. Methods* 2007, *68*, 554-562.

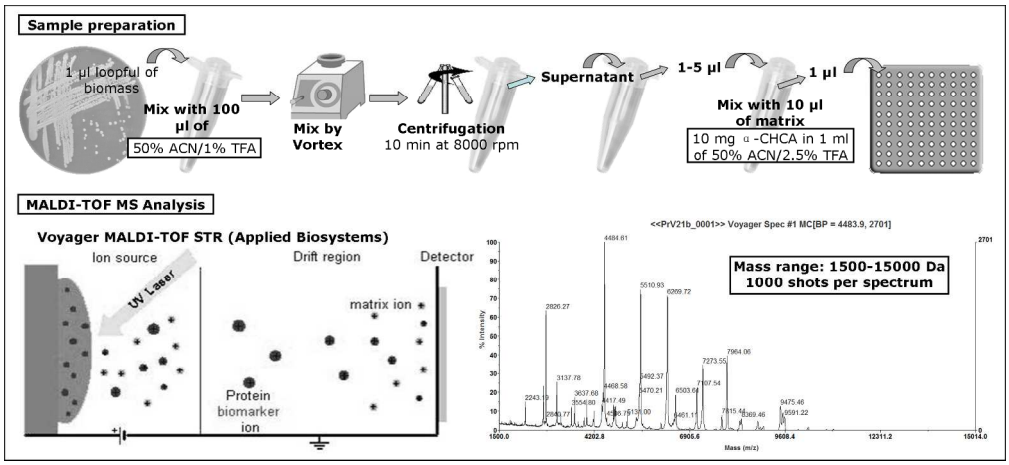
[28] Teramoto, K., Sato, H., Sun, L., Torimura, M., Tao, H., Yoshikawa, H., Hotta, Y., Hosoda, A., Tamura, H., *Anal. Chem.* 2007, *79*, 8712-8719.
[29] Wunschel, S. C., Jarman, K. H., Petersen, C. E., Valentine, N. B., Wahl, K. L., Schauki, D., Jackman, J., Nelson, C. P., White V, E., *J. Am. Soc. Mass. Spectrom.* 2005, *16*, 456-462.
[30] Barbuddhe, S. B., Maier, T., Schwarz, G., Kostrzewa, M., Hof, H., Domann, E., Chakraborty, T., Hain, T., *Applied and Environment Microbiology* 2008, *74*, 5402-5407.
[31] Böhme, K., Fernández-No, I. C., Barros-Velázquez, J., Gallardo, J. M., Cañas, B., Calo-Mata, P., *Anal. Methods* 2010, *2*, 1941-1947.

Legends to Figures

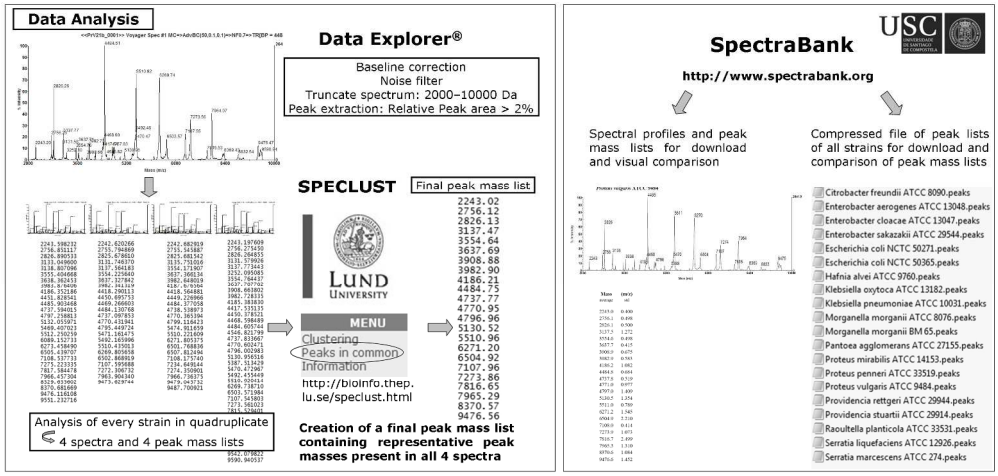
Figure 1. Protocol for sample preparation and mass spectral analysis.

Figure 2. Scheme of data analysis and submission of spectral data to SpectraBank.

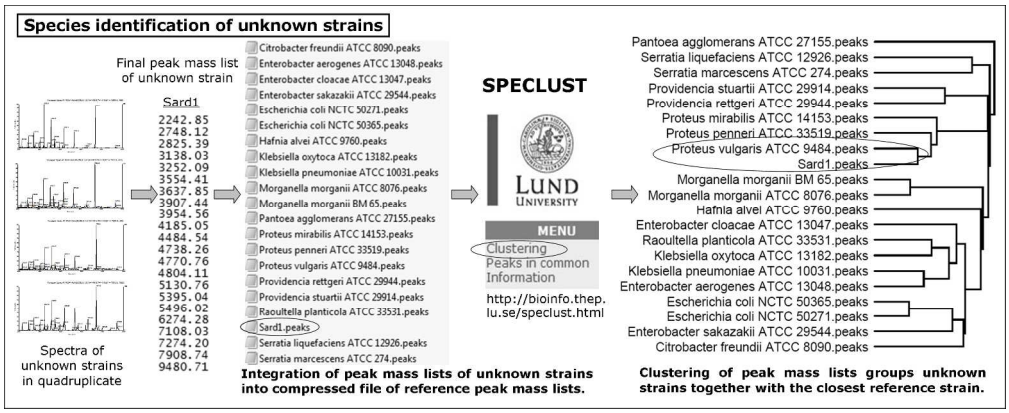
Figure 3. Application of SpectraBank for species identification of the unknown strain Sard1.



422x192mm (150 x 150 DPI)



488x238mm (150 x 150 DPI)



458x184mm (150 x 150 DPI)