

Investigating Time Series Classification Techniques for Rapid Pathogen Identification with Single-Cell MALDI-TOF Mass Spectrum Data

Christina Papagiannopoulou¹, René Parchen², and Willem Waegeman¹

¹ Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium

{christina.papagiannopoulou, willem.waegeman}@ugent.be

² BiosparQ B.V., Leiden, the Netherlands
parchen@biosparq.nl

Abstract. Matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF-MS) is a well-known technology, widely used in species identification. Specifically, MALDI-TOF-MS is applied on samples that usually include bacterial cells, generating representative signals for the various bacterial species. However, for a reliable identification result, a significant amount of biomass is required. For most samples used for diagnostics of infectious diseases, the sample volume is extremely low to obtain the required amount of biomass. Therefore, amplification of the bacterial load is performed by a culturing phase. If the MALDI process could be applied to individual bacteria, it would be possible to circumvent the need for culturing and isolation, accelerating the whole process. In this work, we briefly describe an implementation of a MALDI-TOF MS procedure in a setting of individual cells and we demonstrate the use of the produced data for the application of pathogen identification. The identification of pathogens (bacterial species) is performed by using machine learning algorithms on the generated single-cell signals. The high predictive performance of the machine learning models indicates that the produced bacterial signatures constitute an informative representation, helpful in distinguishing the different bacterial species. In addition, we reformulate the bacterial species identification problem as a time series classification task by considering the intensity sequences of a given spectrum as time series values. Experimental results show that algorithms originally introduced for time series analysis are beneficial in modelling observations of single-cell MALDI-TOF MS.

Keywords: MALDI-TOF MS · single-cell spectrum · single-ionization-event · classification · bacterial species identification · time series

In the diagnostics of infectious diseases, matrix-assisted laser desorption/ ionization-time-of-flight mass spectrometry (MALDI-TOF-MS) is used to identify the causative

Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

organism of an infection as a first step in establishing an antibiotic therapy. Owing to its ease of use, its reliability and the low cost of ownership, the introduction of MALDI-TOF-MS revolutionized the diagnostics of infectious diseases during the last decade [1]. BiosparQ in the Netherlands developed an instrument, called Cirrus D20, together with the appropriate protocols that is able to produce an information-rich signature of bacteria based on MALDI-TOF-MS technology. This abstract is based on [3], in which we evaluate the single-cell MALDI-TOF MS methodology developed by BiosparQ, demonstrating the use of single-particle spectra for the application of pathogen (bacterial species) identification.

The classification of single-cell bacterial fingerprints is not a trivial process even for human annotators. Thus, MALDI-TOF single cell spectrum analysis should be combined with statistical and machine learning methods. In [3], we focus on the analysis of MALDI-TOF single-cell spectra for rapid species identification using machine learning techniques. Instead of only applying general purpose machine learning techniques [2], we also experimented by framing the problem as a time series classification task. In particular, by mapping mass-overcharge (M/Z) ratios to the time axis, we consider the sequences of the different intensities in a spectrum as time series values. This way, standard time series classification methods can be applied. To the best of our knowledge, this is the first time that machine learning approaches and time series classification methods are being applied on single-cell MALDI-TOF data.

The contribution of our work is two-fold. Based on the implementation of the MALDI-TOF-MS methodology in a single-cell setting, we (i) experimentally prove that the single-cell signatures, produced by this MALDI-TOF-MS implementation, are informative in distinguishing different bacterial species by using machine learning data analysis, and (ii) find that algorithms originally introduced for time series analysis are beneficial in modelling observations of single-cell MALDI. As such, we believe that the use of single-cell MALDI-TOF-MS data combined with an accurate modelling approach comprises a solid framework that strives to solve the problem of fast pathogen identification (in terms of minutes or seconds), revolutionizing current state-of-the-art approaches.

References

1. van Belkum, A., Chatellier, S., Girard, V., Pincus, D., Deol, P., Jr, W.M.D.: Progress in proteomics for clinical microbiology: MALDI-TOF MS for microbial species identification and more. *Expert Review of Proteomics* **12**(6), 595–605 (2015). <https://doi.org/10.1586/14789450.2015.1091731>
2. De Bruyne, K., Slabbinck, B., Waegeman, W., Vauterin, P., De Baets, B., Vandamme, P.: Bacterial species identification from maldi-tof mass spectra through data analysis and machine learning. *SYSTEMATIC AND APPLIED MICROBIOLOGY* **34**(1), 20–29 (2011), <http://dx.doi.org/10.1016/j.syapm.2010.11.003>
3. Papagiannopoulou, C., Parchen, R., Waegeman, W.: Investigating time series classification techniques for rapid pathogen identification with single-cell maldi-tof mass spectrum data. *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD'19)* (accepted) (2019)