

Fast Pathogen Identification Using Single-Cell Matrix-Assisted Laser Desorption/Ionization-Aerosol Time-of-Flight Mass Spectrometry Data and Deep Learning Methods

Christina Papagiannopoulou, René Parchen, Peter Rubbens, and Willem Waegeman*

Cite This: *Anal. Chem.* 2020, 92, 7523–7531

Read Online

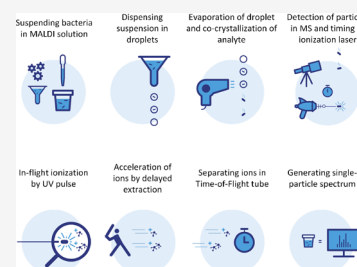
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: In diagnostics of infectious diseases, matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS) can be applied for the identification of pathogenic microorganisms. However, to achieve a trustworthy identification from MALDI-TOF MS data, a significant amount of biomass should be considered. The bacterial load that potentially occurs in a sample is therefore routinely amplified by culturing, which is a time-consuming procedure. In this paper, we show that culturing can be avoided by conducting MALDI-TOF MS on individual bacterial cells. This results in a more rapid identification of species with an acceptable accuracy. We propose a deep learning architecture to analyze the data and compare its performance with traditional supervised machine learning algorithms. We illustrate our workflow on a large data set that contains bacterial species related to urinary tract infections. Overall we obtain accuracies up to 85% in discriminating five different species.



Matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS) is a well-known technology that has become an important tool in clinical microbiology laboratories due to (i) its ease of use, (ii) its reliability, and (iii) its low cost of ownership.^{1,2} Specifically, MALDI-TOF MS can be used to identify the microorganism that causes an infection in order to establish a targeted antibiotic therapy. Its use has been demonstrated for the identification of the causative organism in the case of bacterial infections.^{3–6} The MALDI-TOF MS spectrum of a bacterial species results in a unique signature, allowing for the accurate identification of bacterial organisms at the genus and species levels.^{7,8}

Traditional MALDI-TOF MS has an important limitation. A significant amount of biomass should be considered in order to obtain trustworthy identification results. Some authors have proposed 6×10^3 CFU/spot as the detection limit,⁹ but in practice often the limit of 1×10^5 CFU/spot is considered.^{10,11} Detection limits are formulated in this way because an absolute amount of biomass is required to obtain a mass spectrum with a sufficiently high quality for identification (see, for example, the review memorandum for the 510(k) VITEKMS¹²). However, the combination of CFU load and sample volume is often too low to obtain this amount of biomass, and thus the bacterial load is routinely amplified by culturing, which is a time-consuming process that takes up multiple hours and in some cases even several days. Moreover, species identification in mixed microbial communities is still in an experimental phase.¹³ Isolation of the causative organism is required to identify pathogenic species. To this end, one needs a-priori knowledge of the causative bacteria for determining the

colony(ies) that should be identified, as well as for selecting a suitable culture medium and optimal growing conditions.

The computer algorithms for identifying bacterial species run in terms of seconds or minutes, but the time of the whole MALDI-TOF MS identification process is dominated by the culturing phase. A potential way to alleviate the need for culturing and isolation is the direct application of MALDI-TOF MS to the community as a whole.¹³ This way the time-to-result duration will be drastically decreased. To conduct this type of analysis, BiosparQ has developed Cirrus D20, an instrument able to characterize individual bacterial cells via MALDI-TOF MS, by combining the technologies proposed by Yusof et al.¹⁴ and van Wuijckhuijse et al.¹⁵ Single cells encapsulated in a picoliter sized droplet can be isolated with the technique of Yusof et al.¹⁴ This technique was originally proposed for dispensing individual eukaryotic cells into 96/384 well plates, but it can also be applied to isolate individual bacterial cells from a patient sample. Additionally, one can apply MALDI to individual aerosol particles with the ATOF-MS technology of van Wuijckhuijse et al.,¹⁵ who demonstrated that recognizable spectra could be accumulated from spectra of large numbers of aerosol particles containing pure proteins only. In this way the Cirrus D20 instrument can measure an individual MALDI-TOF MS spectrum for every cell present in

Received: December 24, 2019

Accepted: April 24, 2020

Published: April 24, 2020



a urine sample. To make a clear distinction with traditional MALDI-TOF MS, we will speak of single-cell MALDI-ATOF mass spectrometry when referring to such spectra.

In this paper, we demonstrate for the first time that single-cell spectra obtained with Cirrus D20 allow for a fast and accurate identification of bacterial organisms. Since the technology behind Cirrus D20 has been discussed by Yusof et al.¹⁴ and van Wuijckhuijse et al.,¹⁵ we will focus on the data analysis of single-cell spectra, and the machine algorithms required for this purpose. We implement and compare several machine learning algorithms for identifying species, such as traditional approaches and state-of-the-art deep learning methodologies. Our work serves as a proof-of-concept, showing that it is possible to distinguish a limited number of bacterial species with single-cell MALDI-ATOF MS and machine learning methods. As a concrete use case, we will focus on species that are responsible for urinary tract infections, which are the most common community-acquired and nosocomial bacterial infections.¹⁶ We will prove that pathogens can be accurately identified. This might lead to faster diagnostics when time matters and reduce the need for empirical therapy.

Single-cell MALDI-ATOF MS has two additional challenges compared to traditional MALDI-TOF MS. First, single-cell spectra contain more noise than accumulated spectra, and simple machine learning algorithms might fail to provide accurate identifications. Second, compared to traditional MALDI-TOF MS, training set sizes will be much larger, because a spectrum will be obtained for every cell in a urine sample. Moreover, as samples need to be analyzed in consecutive days to account for instrument variation, one easily obtains training data sets that contain 100 000 spectra. With training data sets of that complexity, we hypothesize that modern deep learning methods might outperform traditional techniques. In this paper, we will compare traditional machine learning algorithms and deep learning approaches, leading to two contributions: (i) we experimentally prove that single-cell MALDI-ATOF MS spectra are informative in distinguishing different bacterial species, and (ii) we find that deep learning methods might be preferred over traditional machine learning algorithms for that task.

MATERIALS AND METHODS

Bacterial Strains. In this study, we generated single-cell MALDI-ATOF spectra for species that are commonly present in urinary samples, namely, *Staphylococcus epidermidis*, *Klebsiella pneumoniae*, *Enterococcus faecalis*, *Escherichia coli*, and *Staphylococcus aureus*. We used single strains for all five species. These strains were clinical isolates provided by the Leiden Centre for Applied Bioscience. Their identity was established by evaluation of the cultures on a bioMérieux Vitek MS MALDI instrument. An ATCC strain was used to calibrate the instrument (*E. coli* ATCC 25922).

The strains were cultured in a liquid medium (TSB) overnight in a stove at 37 °C shaking at 100 rpm until an OD₆₀₀ ≈ 4 was reached. This bacterial suspension was aliquoted in 300 μL aliquots 1.5 mL Eppendorf tubes. Subsequently, the bacteria were washed twice by consecutively centrifuging the suspension at 21 000g, removing the supernatant, and resuspending it in Milli-Q. The resulting suspension was centrifuged for one more time at 21 000g and the supernatant was removed. The resulting pellet was either stored at −20 °C for future use or used directly. The

pellets produced were resuspended in a MALDI matrix solution consisting of 0.1 mg/mL propyl (E)-2-cyano-3-(4-hydroxyphenyl)acrylate (synthesized in-house), 70% (vol) demi water, 25% (vol) ethanol, and 5% (vol) formic acid.

Single-Cell MALDI-ATOF Mass Spectrometry. MALDI-TOF MS can only be applied on individual bacterial cells by modifying two main aspects of the conventional version of the process. Specifically, while in conventional MALDI-TOF MS a large number of cells are prepared on a target plate, for single-cell MALDI-ATOF MS, cells need to be (1) individually prepared and (2) individually presented to the mass spectrometer. To achieve these two modifications, the following methodologies have been applied: (1) individual cells were prepared using a single-cell dispensing technology introduced by Yusof et al.,¹⁴ and (2) individual cells were presented to the mass spectrometer and ionized using an aerosol TOF (ATOF) technology introduced by van Wuijckhuijse et al.¹⁵

Next, we describe the single-cell MALDI-ATOF MS procedure more in detail (see also the abstract graphic). A volume of 50 μL of the suspension fills in a disposable dispenser cartridge supplied by Cytena GmbH. This dispenser cartridge produces droplets of approximately 50 pL, containing at least one cell, according to the technique described by Yusof et al.¹⁴ Provided that the concentration of cells is sufficiently low, each droplet contains a single cell. The volatile fraction of the droplets is allowed to evaporate, yielding particles consisting of (the remains of) a cell and the crystallized MALDI matrix. These particles are subsequently fed to a MALDI-ATOF mass spectrometer built by BiosparQ BV, according to van Wuijckhuijse et al.¹⁵

In the MALDI-ATOF mass spectrometer, each individual particle is excited in flight with a pulsed UV laser (337 nm). The ions, produced by the MALDI process, are accelerated into the time-of-flight tube using delayed extraction and an extraction potential of 30 kV. An electron multiplier, fixed at the end of the time-of-flight tube, detects the ions. For each particle, the signal produced by the electron multiplier is recorded using a Gage 14 bit digitizer card operated at 200 MHz. For each excited particle, data is recorded as a time series in binary format. The conversion of time-of-flight to mass-to-charge ratio of the ions, m/z , is given by the following formula:

$$m/z = \left(\frac{T_{\text{TOF}} - C_2}{C_1} \right)^2 \quad (1)$$

where C_1 and C_2 are calibration coefficients established by calibrating the mass spectrometer using a sample containing a known organism (*E. coli*, ATCC 25922) and aligning the resulting spectrum with a reference spectrum of *E. coli* ATCC 25922 recorded on a Vitek MALDI-TOF mass spectrometer.

Data Preprocessing. Using eq 1 and the associated calibration coefficients, the single-particle time series are converted into mass spectra. However, to apply machine learning methods, a fixed-length feature representation is needed, where values correspond to predefined m/z ratios. We used two resampling procedures to generate such a feature representation, namely, binning and interpolation.

Binning consists of the following steps: (i) the m/z axis is split into equal intervals (bins), and (ii) in each interval the maximum intensity is preserved. Consecutive bins might have an overlapping subinterval to account for potential left–right

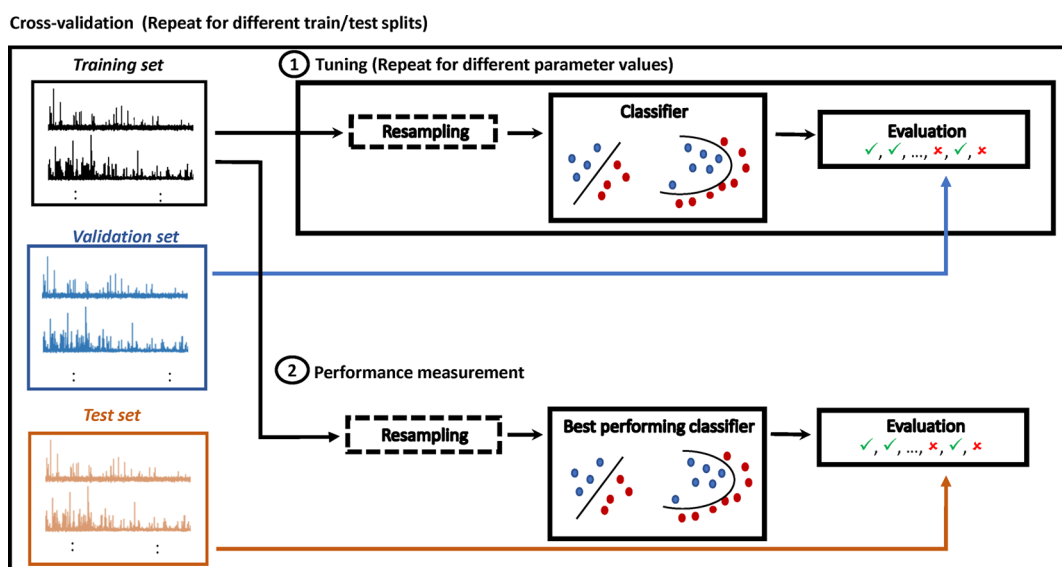


Figure 1. Pipeline of model training and evaluation. The data set is split into train, validation, and test sets. During the tuning phase (first), the training set is used for training a given classifier, and the validation set is used to measure the performance of the classifier. This is repeated for different parameter values. Resampling can be omitted (thus in a dashed line), and the raw data can be used instead, especially in the deep learning approaches. The number of bins in the resampling phase is also a tunable parameter. The parameter values with the best performance on the validation set are selected for the second phase. The model is trained again on the training set by using the best performing parameter values. The model performance is assessed based on the test set. The whole procedure is repeated six times for different train/validation/test splits. The average performance (mean accuracy) as well as the standard deviation over the six splits are reported.

shifts in the spectra, e.g., due to cell-to-cell and instrument variability. The number of m/z values (bins) defined in step (i) is a tunable parameter that is optimized during the learning process (model training).

The interpolation resampling technique is comprised of four steps: (i) the m/z axis is split into equal intervals, (ii) a cumulative spectrum of each individual spectrum is formed, (iii) linear interpolation is performed on the cumulative spectrum, (iv) the interpolated values are differentiated pairwise (e.g., subtract one value for the next value in the sequence), and the final signal is produced. We also applied the interpolation method directly on the original spectra, resulting in low values for the characteristic peaks of the spectra (see the [Supporting Information](#) for examples).

Machine Learning Methods. In machine learning, species identification can be formulated as a typical multi-class classification task. We evaluate several well-known classifiers that have been broadly applied on traditional MALDI-TOF MS data, namely, random forests (RFs), logistic regression (LR), and k-nearest neighbor (KNN). We also run experiments with deep learning models, testing convolutional neural networks that have been developed for sequential data, e.g., for time series classification tasks.¹⁷

In the first group of experiments, we evaluated the predictive performance of the three aforementioned classifiers (LR, RFs, and KNN). Specifically, each classifier was assessed based on its ability to distinguish single-cell spectra of five species, namely, *S. epidermidis*, *K. pneumoniae*, *E. faecalis*, *E. coli*, and *S. aureus*. We used the same train/validation/test splits to obtain a fair comparison between the various tested algorithms. After the removal of empty particles, we kept a similar number of observations for each species ($\sim 10\,000$) and we performed 6-fold cross validation.¹⁸ The data set consisted of data coming from six different days in total. In particular, the validation/test set consisted of data generated at different days than the

training set, i.e., the validation set contained data produced on a particular day, the test set contained data produced on another day, and the training set included data produced on four different days. Small day-to-day variations in the resulting spectra are expected.

We also test convolutional neural networks for time series data. The main idea of a convolution filter applied on a time series is described as follows: for a given time series T and a 1-dimensional filter f , the filter f slides over the time series T , producing the result of the convolution. For instance, assuming that the weights of f are equal to $[1, -1]$, the result of the convolution will be the difference between two consecutive points of the input time series.¹⁹ The deep learning architectures evaluated in this study consist of convolutional and max pooling modules. We start with the so-called fully convolutional network (FCN) that comprises a convolutional and a max pooling layer followed by a linear layer ([Supporting Information Figure 2a](#)). We also report the performance of a 2-layer FCN ([Supporting Information Figure 2b](#)). Moreover, inspired by the Inception module introduced by Fawaz et al.,²⁰ we experimented with a network consisting of three different convolutional-max pooling blocks that are concatenated to form a single representation (3-concat FCN) ([Supporting Information Figure 2c](#)). The best performing deep learning model has been used in combination with the most promising preprocessing technique, the interpolation-based resampling (see [Results](#) section). Finally, given the variations in the data set and the fact that deep learning models need a lot of data, we trained the best performing model on an augmented data set. We followed the data augmentation technique for spectral data reported in Liu et al.²¹ In brief, we performed left–right shifts of the original observations, inserted variations in the amplitude, and created new observations that were linear combinations of the original observations.

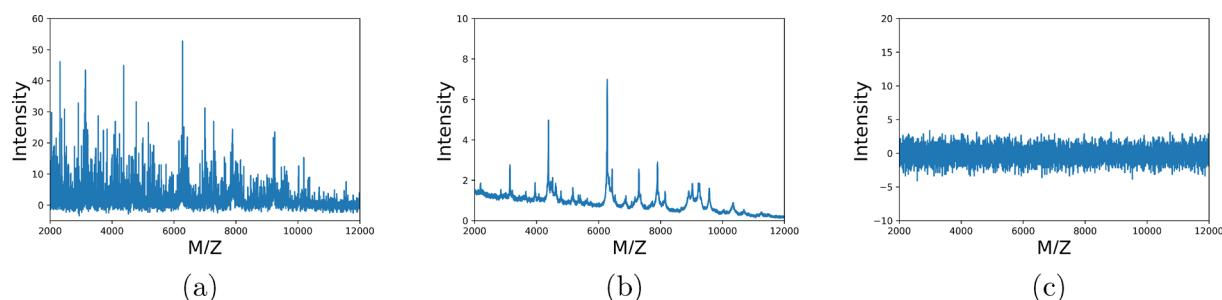


Figure 2. (a) Spectrum of a single-ionization-event of a single *E. coli* particle, (b) accumulated spectrum obtained by averaging the intensities of the individual spectra of *E. coli*, and (c) spectrum of an empty particle.

Table 1. Average Predictive Performance on the Test Sets in Terms of Mean Accuracy for the Traditional Machine Learning Models by Using Different Preprocessing^a

model	raw	interpolation-based	binning	binning with overlap
logistic regression	0.86 ± 0.06	0.87 ± 0.06	0.86 ± 0.06	0.85 ± 0.07
random forests	0.63 ± 0.05	0.73 ± 0.08	0.66 ± 0.05	0.67 ± 0.04
KNN	0.39 ± 0.04	0.68 ± 0.07	0.60 ± 0.06	0.60 ± 0.06

^aStandard deviation is also reported. “Raw” indicates the use of the data without any preprocessing.

Parameter tuning was performed on a separate validation set by means of grid search. The pipeline of the train/test splitting, as well as the tuning of parameters in a validation set, is illustrated in Figure 1. For all the models, the number of bins (when used) was a tunable parameter with a tested range [500, 10000] with steps of 500. For the algorithms LR (regularization parameter range [10^{-5} , 10^2]), RFs (maximum number of features per tree ranges from the squared root of the number of features until the number of features), and KNN (number of neighbors range [1,9]), the sklearn²² python implementation was used, while the deep learning architectures were implemented with the PyTorch²³ library.

For the evaluation procedure, we report the mean accuracy and the standard deviation over the six folds for all the species on the test sets and we present the confusion matrix of the best performing algorithms for further discussion. We also report the so-called precision-recall curves. The precision for a given class is the number of true positives (i.e., the number of correctly classified observations) divided by the total number of observations labeled as belonging to the positive class. Recall is defined as the number of true positives divided by the total number of observations that actually belong to the positive class.

RESULTS

Structure of Single-Cell MALDI-ATOF MS Spectra.

Before delving into quantitative results, we first discuss the structure of single-cell MALDI-ATOF MS spectra. As an example, Figure 2a,b visualize a spectrum of a single-ionization-event of a single *E. coli* particle, and the accumulated spectrum obtained by averaging the intensities of individual spectra of many *E. coli* particles (see the Supporting Information for the other species). Figure 2a is more noisy than 2b, because single-cell MALDI-ATOF MS spectra contain signal caused by (i) charged molecule clusters and (ii) charged molecular debris (due to possible disintegration of molecules). The latter signal is not informative, since it is a result of an imperfect ablation of the analyte/MALDI-matrix mixture, which may be referred to as clutter signal. The amplitude of clutter signal is likely to be of the same order as the informative signal of the analyte, but

the expected value of clutter signal is substantially lower than that of the analyte. The result of imperfect ablation (i.e., see (i) and (ii)) of the mixture is a highly stochastic process, while the presence of the analyte molecules is not. Therefore, by combining a large number of single-ionization-event spectra, an accumulated spectrum is formed. This accumulated spectrum includes high amplitudes at points corresponding to the analyte mass molecules and substantially lower amplitudes at all the other points (clutter) of the spectrum (Figure 2b). For traditional MALDI-TOF MS, these low-amplitude cluttered parts of the signal can be easily removed by the application of routinely used baseline correction algorithms. However, in single-cell MALDI-ATOF MS spectra, the difference in the clutter and analyte stochastics cannot be applied, and thus a different strategy is necessary. It will be the job of the machine learning algorithms to determine, in an automated manner, the parts that correspond to species-specific signal and the parts that are noninformative clutter.

Prior to the identification phase, we remove particles that do not contain any information. To do so, we calculate the variance of each particle. If the calculated variance is low, there are no intensities captured by the ionization procedure. In Figure 2c, an observation of an empty particle is depicted. One can see that it is quite easy to distinguish an empty particle from a particle containing a cell.

Single-Cell MALDI-ATOF MS Spectra Are Informative for Pathogen Identification.

In this section, the performance of the traditional machine learning algorithms is presented. Table 1 shows the predictive performance of the three classification algorithms in terms of the mean accuracy (over the five classes). Overall, the best performing algorithm is logistic regression with a mean accuracy of 0.87. Logistic regression preserves the intensity over the m/z positions and thus, it exploits possible peaks in particular m/z values among the different species. Compared to logistic regression, random forests perform worse. This indicates that Cirrus D20 is able to produce characteristic signals for the various bacterial species, easily distinguishable even by simple methods, such as logistic regression, which just learn weights for the important intensities of the m/z values for each bacterial species. Finally,

KNN is outperformed by the aforementioned methods. This result is not surprising, since nearest neighbor algorithms are not able to generalize well when the number of features is high (due to the curse of dimensionality). This is the reason why the performance of KNN is poor (0.39), when the raw data are used. Therefore, a resampling step is necessary for this model. Note that the optimal number of bins (based on the tuning on the validation set) is low for these models (i.e., always nearly to 500) compared to the corresponding number of bins for other models.

Regarding the resampling techniques applied as a preprocessing step before the model training, we observe that the LR model performs equally well in all cases (i.e., raw data, preprocessed), with a marginal performance increase when the interpolation-based resampling method is used. For the other two models (RFs and KNN), it is clear that the interpolation-based resampling technique is beneficial. By using the interpolation-based resampling, the resulting signal retains the energy of the original signal as described in the section on data preprocessing. Thus, the characteristic peaks for each species become clearer than in the original signal, increasing the performance of the machine learning models. Concerning the type of binning (with or without overlap), there is no clear benefit between the two versions of binning based on the performance of the classifiers.

Finally, in Table 1, the standard deviation of the classifier performance among the different folds is reported. Note that in each of the six different folds, data produced at different dates serve as test sets, while other data (from other different dates) serve as validation sets. It is known that signal variations for a single species between experimental repetitions can be higher than the ones coming from using (e.g.) different strains. This is why we observed that for some test sets the performance of the classifiers is high (~ 0.97), while for others lower (~ 0.78), meaning that for the latter cases the training spectra are not close enough to the test spectra compared to the former cases. Therefore, the standard deviation of ~ 0.06 in most cases can be explained by variations in the data produced at different days (between training and test set or due to the parameter tuning that has been performed on a validation set that can have variations compared to the corresponding test set).

Deep learning approaches are promising for classification of single-cell MALDI-ATOF MS data. In this section the performance of the deep learning models is discussed. Table 2 shows the performance of the various deep

Table 2. Average Predictive Performance on the Test Sets in Terms of Mean Accuracy for the Deep Learning Models on the Raw Data^a

model	logistic regression (SGD)	FCN	2-layer FCN	3-concat FCN
accuracy	0.82 ± 0.07	0.84 ± 0.05	0.85 ± 0.06	0.81 ± 0.09

^aStandard deviation is also reported. For more details see the text.

learning architectures. The performance of a logistic regression algorithm that runs in a stochastic way (model updates are performed by using the stochastic gradient descent (SGD) optimizer and thus, less memory is used) is also reported for reference. The models compared in this section can work on large data sets of thousands or millions of observations, because they run in batch mode (batches of data are loaded

each time to memory) and leverage the GPU power. Therefore, these models can be applied in a real scenario, since the database of spectra is populated at daily basis. That way the training set is growing, allowing modeling approaches to learn potential variations of the data.

Overall, the 2-layer FCN model wins from all the other models in Table 2, even though the differences are rather small. Specifically, it reaches a similar performance on the raw as well as on the interpolation-based resampled data (0.85) (Figure 3). This means that the consecutive convolutional-max

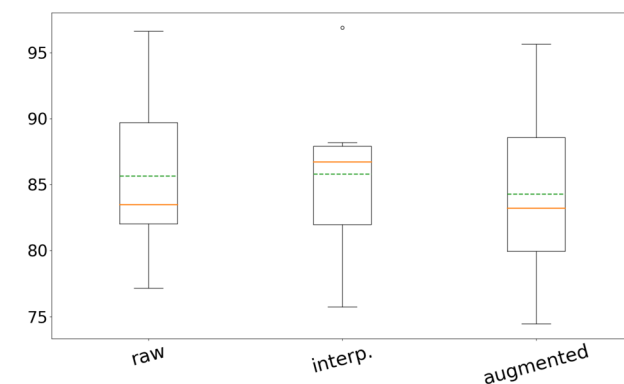


Figure 3. Boxplot for the performance of the 2-layer FCN model using the raw, the preprocessed (using the interpolation-based technique), and the augmented data sets, respectively. The orange lines indicate the median values, while the green lines indicate the mean values over the six folds.

pooling blocks are able to recover the additional information that might occur in the preprocessed data. In addition, there is no performance improvement when variations of the original training examples are added to the training set (Figure 3). This means that these variations (i.e., potential shifts, amplitude fluctuations, etc.) have been already learned from the network (without augmentation). Note that in all the experiments, the optimal max pooling filter size varies from 1 (no max pooling) until 5, confirming the fact that positions of the peaks are informative for the classification of the spectra. The mean accuracy of the FCN architecture (0.84) is similar to that of the 2-FCN architecture, implying that a single convolutional-max pooling block is able to extract informative patterns from the data. This is also confirmed by comparison of these deep learning architectures with the logistic regression (SGD) model. The feature extraction block(s) (convolutional-max pooling) that are used in the deep learning architectures are beneficial for the classification of the bacterial spectra. Recall that the logistic regression (SGD) model uses only the raw data without any feature extraction, similar to the original logistic regression. However, logistic regression (SGD) can be efficiently used in large data sets. Finally, the 3-concat FCN resulted in a decreased performance mainly due to the fact that it contains more tuning parameters than the other models. These parameters are optimized on validation sets that might have differences compared to the test data. Nonetheless, overall, the results confirm that this kind of end-to-end models, which perform feature extraction and classification in a single framework, are useful for the task of bacterial species identification by using single-cell MALDI-TOF mass spectrum data, especially when large data sets are being considered.

Analyzing Predictions Per Bacterial Species. In Figure 4, the confusion matrices of the two best performing methods

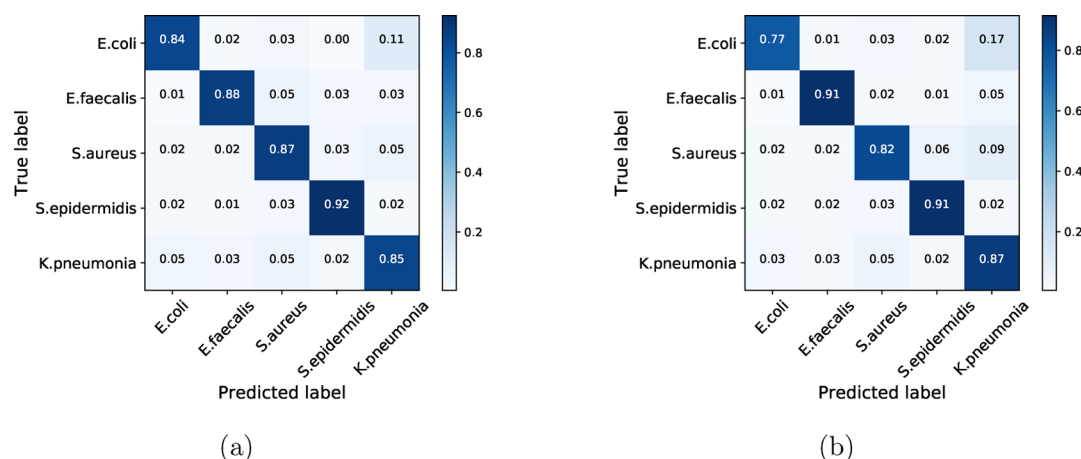


Figure 4. Confusion matrices for (a) logistic regression, and (b) the 2-layer FCN models.

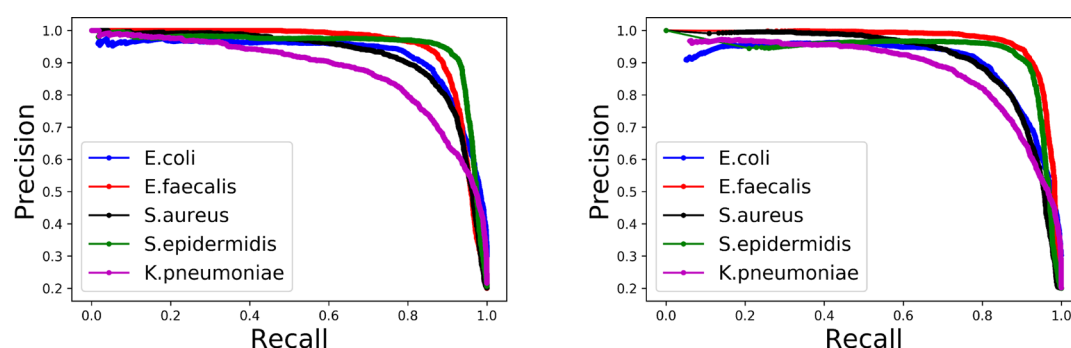


Figure 5. Precision/recall curves for (a) the logistic regression and (b) the 2-layer FCN models.

are presented. Specifically, Figure 4a,b shows the accuracy of logistic regression and the 2-layer FCN model per species, respectively. Logistic regression results in comparable classification accuracies per species, while the 2-layer FCN model reported a lower classification accuracy for *E. coli*. In addition, the spectra of *E. coli* are mostly confused with the species *K. pneumoniae*. The confusion regarding the spectra of the *K. pneumoniae* species can be explained by the fact that the spectra of this species include peaks with low intensities. Thus, these observations with low peaks cannot be easily classified as *E. coli* observations and are spread (and misclassified) to the other classes.

From the confusion matrices of Figure 4, it seems that the models perform relatively well for all the studied species. However, there are false positive examples that are not taken into account in the calculation of the accuracy metric. The ratio of false positives is incorporated in the estimation of the precision-recall curves, see Figure 5. Figure 5 shows the precision-recall trade-off for each of the five species for logistic regression (Figure 5a) and the 2-layer FCN (Figure 5b) models. Figure 5a,b show that for all the species, the precision is above 0.8 for a recall value of 0.8. This means that 80% of the observations (for each species) are identified correctly with an accuracy higher than 80%. Note that the model precision in this application is crucial, because the targeted antibiotic therapy is based on the classification results, i.e., the correct identification of the bacterial species. However, it is clear in both plots that for the *K. pneumoniae* species, the precision increases (0.8) slower compared to the rest of the species. This means that the models will wrongly classify several examples of

other species (i.e., *E. coli*) in order to predict a high ratio of the *K. pneumoniae* observations.

DISCUSSION

Classical machine learning algorithms, such as linear models, support vector machines²⁴ and random forests²⁵ are often used for analyzing traditional MALDI-TOF MS spectra,^{9,26–29} and they are also often deployed on several single-cell technologies, such as flow cytometry^{30,31} and Raman spectroscopy.^{32–34} More recently, deep learning models have been proposed for the identification of bacterial species based on Raman data.^{21,35} These models have the ability to construct higher-level features and perform the classification task in an end-to-end fashion, in addition to the fact that they scale to very large data sets. Especially this last property makes deep learning models very useful in single-cell MALDI-ATOF MS, as we expect that training data sets will become huge in the near future, once the technology will be routinely used.

This paper is a follow-up work of Papagiannopoulou et al.³⁶ Compared to our previous work, we (i) describe more in detail the fundamental principles of Cirrus D20, (ii) collected an extended data set by using data produced in different days, and (iii) exploit the use of the deep learning models. As in the work of Papagiannopoulou et al.,³⁶ by mapping mass-to-charge (m/z) ratios to the time axis, we consider the sequences of the different intensities in a spectrum as time series values and thus, we apply deep learning models for time series classification.¹⁷ It is worthwhile mentioning that we do not compare the performance of our models to the time series classification methods proposed in our previous work,³⁶ since

this kind of methods is computationally more intensive and cannot scale into large collections.

Species identification based on single-cell MALDI-ATOF MS data is much faster than traditional MALDI-TOF MS. When a new sample needs to be analyzed, the time-to-result is dominated by the time spent for generating the single-cell spectra per sample. Owing to the continuous stream of particles into the MS, depending on the level of preconcentration of the sample, typically up to 500 cells per minute can be evaluated. In a second step, the classification phase of the newly generated samples is performed in order of seconds. The training of the machine learning models, as well as the preprocessing of the training data, takes place only once and does not affect the time-to-result in an operational laboratory.

Species identification based on single-cell MALDI-ATOF MS data can be extremely fast, but are the results comparable to traditional MALDI-TOF MS? When the level of detail in the accumulated spectra is similar to that of traditional MALDI-TOF MS spectra, then the taxonomic resolution of single-cell MALDI-ATOF MS will be comparable to that of traditional MALDI-TOF MS. In traditional MALDI-TOF MS, this resolution is generally sufficient to discriminate many species, with the exception of a few cases such as *E. coli* and *Shigella*.^{37,38} The detail in the accumulated single-cell spectra presented in this paper is a little lower than that in traditional MALDI-TOF MS spectra. Hence, the discriminative power is expected to be less than that of traditional MALDI-TOF MS.³⁹

Similarly as with traditional MALDI-TOF MS, different runs of the same organism might yield small variations in spectra, even when exactly the same protocol is used. Those variations are present in the intensity of the peaks of the accumulated spectra, but not in the position. This is due to the fact that bacterial species are subject to multiple intrinsic sources of heterogeneity,⁴⁰ which is why phenotypic differences can manifest themselves within genetically homogeneous communities.^{40,41} One reason for heterogeneity is technical variation caused by conditions that vary from 1 day to another. However, the main reason for technical variation is that on a microscopic scale it is practically impossible to prepare two samples completely identically. This results in differences in e.g. MALDI-crystal morphology and the subsequent cocrystallization of the analyte with the MALDI-matrix, even within a single spot on a traditional target plate or between two particles in the MALDI-ATOF MS instrument.⁴² In our experiments we have shown that this technical variation can be mitigated by collecting data from several days.

Besides technical variability, also biological variability can have a considerable impact on the quality of the spectra, and hence, the classification performance. The biological variability can be further subdivided into cell-to-cell and strain-to-strain variability. Cell-to-cell variability originates from the fact many cells of the same strain are individually analyzed. While traditional MALDI-TOF MS accumulates a large number of "single ionization event/single laser pulse" spectra to reduce variability between these events, due to the "single-cell" ambition of our approach, accumulation of individual spectra is clearly not possible, implying that variation introduced by biological differences between the cells (e.g., life phase) must be handled by the identification scheme. Strain-to-strain variability refers to the differences obtained by analyzing several strains of the same species. In this paper, the cell-to-cell variability is addressed, while the strain-to-strain variability will be further investigated in future work.

In particular, we believe that the effect of strain-to-strain variability to the classification performance needs to be analyzed before our method can be considered as a reliable identification technique. However, based on the vast worldwide experience with traditional MALDI-TOF MS identification of bacteria, we believe that the effect of strain-to-strain variation is limited. Since the level of detail in the MALDI-TOF MS spectra themselves is quite limited (i.e., the number of peaks present in the spectrum is in the order of 50 to 100 compared to the total number of proteins, peptides, lipids, etc. present in a micro-organism), the taxonomic resolution that may be attained is limited as well. Generally, due to the aforementioned lack of detail, it is accepted that MALDI-TOF MS is very capable of distinguishing between species but less capable of distinguishing between strains of the same species. We believe that the same conclusions can be drawn for single-cell MALDI-ATOF spectrum analysis.

It should be clarified that the number of strains used in this study is too low to claim a full urinary tract infection (UTI) diagnostic capability. Typically, approximately 20 species are responsible for the vast majority of UTIs (95–99%).⁴³ Furthermore, since urine samples are generally contaminated with skin flora, the most abundant UTI pathogens and the most abundant contaminant species need to be considered to generate a representative learning set to identify UTIs. The work presented here is meant as a first proof-of-concept for single-cell based diagnosis of infectious diseases such as UTIs. The considered organisms were not chosen based on their abundance (though the most abundant pathogen *E. coli* is represented) but to provide a reasonable representation of the organisms that may be encountered. Thus, the organisms were chosen to represent both typical Gram-negative species, Gram-positive species, and typical representatives for skin flora.

The promise of single-cell MALDI-ATOF MS is that, when identification on a single-cell basis is possible, it would allow for a semiquantitative analysis of patient samples (it is possible to count the number of classified cells). It would also allow for the analysis of samples containing different species of microorganisms. Since, with few exceptions, patient samples are never sterile, diagnostics of infectious diseases is often based on the number of organisms present. When this number exceeds a threshold (e.g., 10^5 CFU/mL is historically used for UTIs), an infection is diagnosed.⁴⁴ When it is significantly lower, no infection is diagnosed. Furthermore, the presence of bacterial contamination in patient samples is highly probable. Depending on the general condition of the patient, urine samples may contain skin flora of the patient. This results in samples containing a mix of organisms.

Traditional MALDI-TOF MS is not quantitative. To evaluate whether a threshold is exceeded, a secondary analysis is required (e.g., quantitative culture, microscopy, FACS,...). Moreover, traditional MALDI-TOF MS has limited capabilities to analyze mixed samples. Both aspects would be covered by single-cell MALDI-ATOF MS, provided that the semiquantitative capability and the mixture analysis capability are verified. The authors consider the work presented in this paper as a first step in developing the required tools to implement both capabilities.

CONCLUSION

In this paper, we demonstrated that single-cell MALDI-ATOF MS data can be used for the identification of pathogenic bacteria in urine samples. The fast response time of single-cell

MALDI-ATOF MS, which is in terms of minutes or seconds, is a clear improvement over traditional MALDI-TOF MS, which takes much more time due to culturing. In contrast, single-cell MALDI-ATOF MS provides additional challenges with respect to the variation of the mass spectra. To account for this variation, we combined single-cell MALDI-ATOF MS with machine learning algorithms, and we experimentally proved that the generated spectra are informative in distinguishing different bacterial species. Traditional machine learning and deep learning methods resulted in a similar performance, but it is expected that the performance of the latter will increase on larger data sets containing thousands or even millions of observations. In the future this will allow the use of databases that are expanded at daily basis (which is a realistic scenario).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.9b05806>.

Additional information on resampling spectra via interpolation, visualizations of the deep learning architectures, and examples of single-cell and accumulated spectra for all analyzed species (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Willem Waegeman – Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent 9000, Belgium; orcid.org/0000-0002-5950-3003;
Email: willem.waegeman@ugent.be

Authors

Christina Papagiannopoulou – Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent 9000, Belgium

René Parchen – BiosparQ B.V., Leiden 2333 BD, The Netherlands

Peter Rubbens – Flanders Marine Institute (VLIZ), Ostend 8400, Belgium

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.analchem.9b05806>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work is funded by BiosparQ B.V., The Netherlands. For this research, W.W. also received funding from the Flemish Government under the “Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen” Programme. The authors sincerely thank Dr. Willem van Leeuwen of the Leiden Centre for Applied Bioscience for providing the strains used in this study. A few strains were provided by Dr. Bas Zaat of the Amsterdam University Medical Center.

■ REFERENCES

- (1) van Belkum, A.; Chatellier, S.; Girard, V.; Pincus, D.; Deol, P.; Dunne, W. M. *Expert Rev. Proteomics* **2015**, *12*, 595–605.
- (2) Tan, K.; Ellis, B.; Lee, R.; Stamper, P.; Zhang, S.; Carroll, K. *Journal of Clinical Microbiology* **2012**, *50*, 3301–3308.
- (3) Ferreira, L.; Sánchez-Juanes, F.; González-Ávila, M.; Cembrero-Fuciños, D.; Herrero-Hernández, A.; González-Buitrago, J. M.;

Muñoz-Bellido, J. L. *Journal of Clinical Microbiology* **2010**, *48*, 2110–2115.

(4) Li, W.; Sun, E.; Wang, Y.; Pan, H.; Zhang, Y.; Li, Y.; Zhang, X.; Li, C.; Du, L.; Wang, C. *Front. Microbiol.* **2019**, *10*, 1182.

(5) Segawa, S.; Sawai, S.; Murata, S.; Nishimura, M.; Beppu, M.; Sogawa, K.; Watanabe, M.; Satoh, M.; Matsutani, T.; Kobayashi, M.; Iwadata, Y.; Kuwabara, S.; Saeki, N.; Nomura, F. *Clin. Chim. Acta* **2014**, *435*, 59–61.

(6) Ceyssens, P.-J.; Soetaert, K.; Timke, M.; Van den Bossche, A.; Sparbier, K.; De Cremer, K.; Kostrzewa, M.; Hendrickx, M.; Mathys, V. *J. Clin. Microbiol.* **2017**, *55*, 624–634.

(7) Seng, P.; Drancourt, M.; Gouriet, F.; La Scola, B.; Fournier, P.; Rolain, J.; Raoult, D. *Clin. Infect. Dis.* **2009**, *49*, 543–551.

(8) Bizzini, A.; Jaton, K.; Romo, D.; Bille, J.; Prod'homme, G.; Greub, G. *Journal of Clinical Microbiology* **2011**, *49*, 693–696.

(9) 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.

(10) Food and Drug Administration. *De novo request for evaluation of automatic class III designation for the VITEK MS, 510(k) Decision Summary*, 2013. https://www.accessdata.fda.gov/cdrh_docs/reviews/K124067.pdf.

(11) Croxatto, A.; Prod'homme, G.; Greub, G. *FEMS Microbiology Reviews* **2012**, *36*, 380–407.

(12) bioMérieux, I. *De novo request for evaluation of automatic class III designation for the VITEKMS*. https://www.accessdata.fda.gov/cdrh_docs/reviews/K124067.pdf.

(13) Yang, Y.; Lin, Y.; Qiao, L. *Anal. Chem.* **2018**, *90*, 10400–10408.

(14) Yusuf, A.; Keegan, H.; Spillane, C. D.; Sheils, O. M.; Martin, C. M.; O'Leary, J. J.; Zengerle, R.; Koltay, P. *Lab Chip* **2011**, *11*, 2447–2454.

(15) van Wuijckhuijse, A.; Stowers, M.; Kleefman, W.; van Baar, B.; Kientz, C.; Marijnissen, J. J. *Aerosol Sci.* **2005**, *36*, 677–687.

(16) Wilson, M. L.; Gaido, L. *Clin. Infect. Dis.* **2004**, *38*, 1150–1158.

(17) Ismail Fawaz, H.; Forestier, G.; Weber, J.; Idoumghar, L.; Muller, P.-A. *Data Mining and Knowledge Discovery* **2019**, *33*, 917–963.

(18) Roberts, D. R.; Bahn, V.; Ciuti, S.; Boyce, M. S.; Elith, J.; Guillera-Aroita, G.; Hauenstein, S.; Lahoz-Monfort, J. J.; Schröder, B.; Thuiller, W.; Warton, D. I.; Wintle, B. A.; Hartig, F.; Dormann, C. F. *Ecography* **2017**, *40*, 913–929.

(19) Cui, Z.; Chen, W.; Chen, Y. *arXiv* **2016**, 1603.06995.

(20) Fawaz, H. I.; Lucas, B.; Forestier, G.; Pelletier, C.; Schmidt, D. F.; Weber, J.; Webb, G. I.; Idoumghar, L.; Muller, P.-A.; Petitjean, F. *arXiv* **2019**, 1909.04939.

(21) Liu, J.; Osadchy, M.; Ashton, L.; Foster, M.; Solomon, C. J.; Gibson, S. J. *Analyst* **2017**, *142*, 4067–4074.

(22) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

(23) Paszke, A. et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems 32 (NIPS 2019)*, Vancouver, Canada, December 8–14, 2019; pp 8026–8037.

(24) Cortes, C.; Vapnik, V. *Machine Learning* **1995**, *20*, 273–297.

(25) Breiman, L. *Machine Learning* **2001**, *45*, 5–32.

(26) Schleif, F.-M.; Lindemann, M.; Diaz, M.; Maaß, P.; Decker, J.; Ellsner, T.; Kuhn, M.; Thiele, H. *Comput. Visualization Sci.* **2009**, *12*, 189–199.

(27) De Bruyne, K.; Slabbinck, B.; Waegeman, W.; Vauterin, P.; De Baets, B.; Vandamme, P. *Syst. Appl. Microbiol.* **2011**, *34*, 20–29.

(28) van Belkum, A.; Lacroix, B.; Veyrieras, J.-B.; Surre, J.; Ramjeet, M.; Arsac, M.; Perrot, N.; Mahé, P.; Mailler, S.; Chatellier, S.; Monnin, V.; Girard, V. *Bioinformatics* **2014**, *30*, 1280–1286.

(29) Dumolin, C.; Aerts, M.; Verheyde, B.; Schellaert, S.; Vandamme, T.; Van der Jeugt, F.; De Canck, E.; Cnockaert, M.; Wieme, A. D.; Cleenwerck, I.; Peiren, J.; Dawyndt, P.; Vandamme, P.; Carlier, A. *mSystems* **2019**, *4*, e00437-19.

- (30) Rajwa, B.; Venkatapathi, M.; Ragheb, K.; Banada, P. P.; Hirleman, E. D.; Lary, T.; Robinson, J. P. *Cytometry, Part A* **2008**, *73A*, 369–379.
- (31) Rubbens, P.; Props, R.; Boon, N.; Waegeman, W. *PLoS One* **2017**, *12*, e0169754.
- (32) Goodacre, R.; Timmins, E. M.; Burton, R.; Kaderbhai, N.; Woodward, A. M.; Kell, D. B.; Rooney, P. J. *Microbiology* **1998**, *144*, 1157–1170.
- (33) Schmid, U.; Rösch, P.; Krause, M.; Harz, M.; Popp, J.; Baumann, K. *Chemom. Intell. Lab. Syst.* **2009**, *96*, 159–171.
- (34) Strola, S. A.; Baritoux, J.-C.; Schultz, E.; Simon, A. C.; Allier, C.; Espagnon, I.; Jary, D.; Dinten, J.-M. *J. Biomed. Opt.* **2014**, *19*, 111610.
- (35) Ho, C.-S.; Jean, N.; Hogan, C. A.; Blackmon, L.; Jeffrey, S. S.; Holodniy, M.; Banaei, N.; Saleh, A. A. E.; Ermon, S.; Dionne, J. *Nat. Commun.* **2019**, *10*, 4927.
- (36) Papagiannopoulou, C.; Parchen, R.; Waegeman, W. Investigating time series classification techniques for rapid pathogen identification with single-cell MALDI-TOF mass spectrum data. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD '19)*, Würzburg, Germany, September 16–20, 2019.
- (37) Croxatto, A.; Prod'homme, G.; Greub, G. *FEMS Microbiology Reviews* **2012**, *36*, 380–407.
- (38) Paauw, A.; Jonker, D.; Roeselers, G.; Heng, J. M.; Mars-Groenendijk, R. H.; Trip, H.; Molhoek, E. M.; Jansen, H.-J.; van der Plas, J.; de Jong, A. L.; Majchrzykiewicz-Koehorst, J. A.; Speksnijder, A. *Int. J. Med. Microbiol.* **2015**, *305*, 446–452.
- (39) Mather, C. A.; Rivera, S. F.; Butler-Wu, S. M. *J. Clin. Microbiol.* **2014**, *52*, 130–138.
- (40) Avery, S. V. *Nat. Rev. Microbiol.* **2006**, *4*, 577–587.
- (41) Govers, S. K.; Adam, A.; Blockeel, H.; Aertsen, A. *Sci. Rep.* **2017**, *7*, 8473.
- (42) Gabriel, S. J.; Schwarzing, C.; Schwarzing, B.; Panne, U.; Weidner, S. M. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1356–1363.
- (43) Flores-Mireles, A.; Walker, J.; Caparon, M.; Hultgren, S. *Nat. Rev. Microbiol.* **2015**, *13*, 269–284.
- (44) Kass, E. *AMA Arch Intern Med.* **1957**, *100*, 709–714.