Journal of Microbiological Methods 181 (2021) 106130

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Exploring the value of MALDI-TOF MS for the detection of clonal outbreaks of *Burkholderia contaminans*

Selali Fiamanya [a,*], Lucía Cipolla [b], Mónica Prieto [b], John Stelling [c,d]

[a] Oxford University Clinical Academic Graduate School, John Radcliffe Hospital, Oxford OX3 9DU, United Kingdom
[b] Servicio Bacteriología Especial, Instituto Nacional de Enfermedades Infecciosas 'Dr. C. G. Malbrán', Av Velez Sarsfield 563, 1281 Ciudad Autónoma de Buenos Aires, Argentina
[c] Department of Medicine, Division of Infectious Diseases, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, USA
[d] Department of Medicine, Harvard Medical School 25 Shattuck Street, Boston, MA 02115, USA

ARTICLE INFO

Keywords:
MALDI-TOF
Burkholderia
Rapid diagnostics
Outbreak

ABSTRACT

Background: Molecular genetics has risen in both output and affordability to become the gold standard in diagnosis, however it is not yet available for most routine clinical microbiology due to cost and the level of skill it requires. Matrix assisted laser desorption/ionisation – time of flight mass spectrometry (MALDI-TOF MS) approaches may be useful in bridging the gap between low-resolution phenotypic methods and bulky genotypic methods in the goal of epidemiological source-typing of microbes. *Burkholderia* has been shown to be identifiable at the subspecies level using MALDI-TOF MS. There have not yet been studies assessing the ability of MALDI-TOF MS to source-type *Burkholderia contaminans* isolates into epidemiologically relevant outbreak clusters.
Methods: 55 well-characterised *B. contaminans* isolates were used to create a panel for analysis of MALDI-TOF MS biomarker peaks and their relation to outbreak strains, location, source, patient, diagnosis and isolate genetics. Unsupervised clustering was performed and classification models were generated using biostatistical analysis software.
Results: B. contaminans spectra derived from MALDI-TOF MS were of sufficiently high resolution to identify 100% of isolates. Unsupervised clustering methods showed poor evidence of spectra clustering by all characteristics measured. Classification algorithms were discriminatory, with Genetic Algorithm models showing 100% recognition capability for all outbreaks, the pulsed-field gel electrophoresis (PFGE) typeability model, and 96.63% recognition for the location model. A consistent peak at *m/z* of approximately 6943 was identified in all non-typeable strains but in none of the typeable strains.
Conclusions: MALDI-TOF MS successfully discriminates *B. contaminans* isolates into clonal, epidemiological clusters, and can recognise isolates non-typeable by PFGE. Further work should investigate this capability, and include peptide studies and genomic sequencing to identify individual proteins or genes responsible for this non-typeablity, particularly at the peak weight identified.

1. Introduction

Microbial diagnostics aim to identify the causative bacterial, fungal and parasitic agents of infectious diseases. Usually this is a lengthy process relying on the biochemical profiling of metabolic traits (Clark et al., 2013a). Although molecular genetics has risen in both output and affordability to become the gold standard in diagnosis, it is not yet available for most routine clinical microbiology due to cost and the level of skill it requires (Barco et al., 2013). Matrix assisted laser ionisation/desorption – time of flight mass spectrometry (MALDI-TOF MS) approaches may be useful in bridging the gap between low-resolution phenotypic methods and bulky genotypic methods with a goal of epidemiological source-typing of microbes.

MALDI-TOF MS is a high-throughput technology which uses the mass to charge ratio (*m/z*) of molecules as an analytic tool in microbial identification. It can use simple, whole cell samples from colonies grown overnight on culture plates without pre-fractionation, digestion, separation or clean-up (Dieckmann et al., 2008; Wieser et al., 2012). This means that colonies can be sampled a few hours after inoculation of a mixed culture and still provide a successful pure culture identification

* Corresponding author.
E-mail address: s.fiamanya@gmail.com (S. Fiamanya).

https://doi.org/10.1016/j.mimet.2020.106130
Received 14 November 2020; Received in revised form 20 December 2020; Accepted 21 December 2020
Available online 29 December 2020

(Wieser et al., 2012). Since the time it takes to travel to the detector is proportional to the square root of *m/z* of an ion, particles are separated as they travel and so frequency distributions of detected particles can be plotted based on TOF up to 100 kDa, though most identified biomarkers are small proteins or peptides under 20 kDa, due to limited sensitivity of TOF analysers above this mass(Gatlin-Bunai et al., 2007; Van Belkum et al., 2012). A separation step such as liquid chromatography is not necessary, allowing the crude protein extractions and even direct whole cell samples to be used for analysis (Van Belkum et al., 2012). The resultant plots, or "spectral fingerprints", are characteristic of individual bacterial species, and can be compared with standard or private reference libraries of other spectra of medically relevant bacteria. A cultured sample can be prepared, loaded onto a target plate, and have results generated within hours (Wieser et al., 2012). The easy user interface of MALDI-TOF MS systems and their multifunctionality with regards to broad genera of microbes increase their ability to be integrated into laboratory workflows. Stainless steel plates can subsequently be reused following a simple treatment with ethanol, trichloroacetic acid and mechanical cleansing in a matter of minutes, adding to the ease of maintenance and reducing the cost of consumables (Wieser et al., 2012).

MALDI-TOF MS obtains a proteomic view of a strain, allowing for detection below the species level provided by conventional biochemical identification (Hou et al., 2019). Several studies have explored the discriminatory power of MALDI-TOF MS in strain typing below the species level, with variable results, and more investigation needed (Florio et al., 2018; Sandrin et al., 2013). MALDI-TOF MS has been shown to be as or more effective at discriminating *Escherichia coli* (Clark et al., 2013b; Novais et al., 2014), *Shigella* (Khot and Fisher, 2013) and *Staphylococcus aureus* (Böhme et al., 2012; Josten et al., 2013; Wolters et al., 2011) isolates into phenotypic pathotypes and genetically derived clonal complexes at the sub-species level than some biochemical or genetic methods. Further successes include strain typing of beta haemolytic streptococci, *Vibrio* spp. (Singhal et al., 2015), *Bacteroides fragilis* (Nagy et al., 2011) and discrimination of various *Salmonella enterica* subsp. *enterica* serovars (Dieckmann and Malorny, 2011; Kuhns et al., 2012).

A major clinical utility of higher typing resolution is in the epidemiological tracing of infections caused by clonal bacteria. For example, MALDI-TOF MS has been used in the real-time tracking of a single plasmid which carried carbapenemase genes in a carbapenem-resistant *Enterobacteriacaea* strain responsible for an outbreak (Lau et al., 2014). In another study, MALDI-TOF MS characterised 33 *Klebsiella pneumoniae* outbreak isolates from a neonatal intensive care unit (NICU) outbreak into two clonal groups in complete concordance with multilocus sequence typing (MLST) results, also associated with the location of patients in the NICU and their residence in the city of the study (Bar-Meir et al., 2020). Development of this capability in MALDI-TOF MS would allow for rapid detection and tracing of bacterial outbreaks in real-time.

Burkholderia contaminans is a relatively newly identified subspecies of the *Burkholderia cepacia* complex (Bcc) subgroup (Vanlaere et al., 2009). It is a cause of opportunistic infection, particularly relevant to cystic fibrosis patients in Argentina where prevalence is approximately 10% (Martina et al., 2013). The proportion of *B. contaminans* among patients with cystic fibrosis (CF) with *Burkholderia* isolates (57%) is also uniquely higher in Argentina compared to other *Burkholderia* species (Martina et al., 2013) (e.g. 37% in Spain (Medina-Pascual et al., 2015)). MALDI-TOF MS has been shown to accurately identify Bcc isolates at the subspecies level (Cipolla et al., 2019; Degand et al., 2008; Fehlberg et al., 2013; Vanlaere et al., 2008) and it can be used to detect biomarker peaks for *B. contaminans* specifically (Miňán et al., 2009). To our knowledge there have not yet been studies assessing the ability of MALDI-TOF MS to source-type isolates from any *Burkholderia* isolates into epidemiologically discrete outbreak clusters.

MALDI-TOF MS has the advantage that reference spectra can be created and quality controlled by laboratory personnel to reflect microbes of local interest, and these can be shared between facilities. This is useful for reflecting trends in the community, research interests, and potentially for resistance and outbreak control. Therefore, if MALDI-TOF MS can consistently differentiate below the species level, it could allow for inexpensive, detailed and dynamic epidemiological analysis of clinical and environmental isolates. Real-time clinical diagnosis will be more feasible, leading to faster commencement of appropriate treatment which is vital in cases of drug-resistant bacteria, while focussed epidemiological interventions for outbreak control could also be developed. This could replace simple threshold measures for outbreak detection that are often poorly specific and sensitive, and time consuming and complex genotyping methods such as polymerase chain reaction (PCR), pulsed-field gel electrophoresis (PFGE), multi-locus sequence typing (MSLT), whole-genome sequencing (WGS)and restriction fragment length polymorphism analysis that are hard to apply at scale and have a significantly higher financial cost (Grundmann et al., 2011). This project therefore sought to determine if MALDI-TOF MS could be used to classify *B. contaminans* isolates into source-specific epidemiological clusters.

2. Methods

2.1. Data collection

2.1.1. Strains

Fifty-five strains were obtained with permission from a previous study which characterised circulating *B. contaminans* strains across Argentina between 2011 and 2014, in collaboration with the Instituto Nacional de Enfermedades Infecciosas, the national reference public health laboratory for Argentina, located in Buenos Aires. Eighteen strains belonged to three outbreaks, and there were a further 37 sporadic strains. Strain numbers were as follows: Outbreak 1 (7 isolates); Outbreak 2 (7 isolates); Outbreak 3 (4 isolates). *XbaI*-enzyme restricted PFGE had previously been performed on all samples. There were 6 PFGE patterns across the 55 strains, with strains from Outbreaks 1 and 2 showing homology, and strains from Outbreak 3 being non-typeable (Table 1). In the sporadic strains from Cordoba, 18/23 (78%) showed the BX3 pattern (Table 2), and all of the Outbreak 1 isolates, also from Cordoba, exhibited this pattern, consistent with the samples from Cordoba being clonal in origin. Each sporadic strain and representatives of each of the outbreak strains had existing main spectral projections (MSPs) – highly characterised reference spectra of individual samples. 11 strains collected serially from five patients (patients A-E) with cystic fibrosis (CF), and one further strain from a patient with CF were used for the patient-level analysis.

2.1.2. Sample preparation

MSPs were created for the Outbreak strains which did not previously have one recorded. The protocols used are adopted from those suggested by the manufacturer Bruker Daltonik Inc., to ensure maximal compatibility with established Bruker Biotyper workflows. The author was trained by a Bruker representative before commencing the laboratory work. Briefly, bacteria were taken from frozen stock cultures and plated onto Tryptic Soy Agar (TSA) and grown at 37 °C overnight. Single colonies from these plates were re-plated on TSA as confluent cultures and left to grow overnight at 37 °C. Samples were prepared for MALDI-TOF MS analysis using the Direct, Extended or Extraction Methods a maximum of 24 h after culturing using the Bruker Extraction Protocol (Bruker, 2020).

2.1.3. Bruker Daltonik sample analysis, data acquisition and cleaning

Following this they were placed into the Bruker Microflex instrument (Bruker Daltonik, GmbH, Leipzig, Germany) for analysis. Bruker FlexControl 3.0 software (Bruker Daltonik) was used to collect mass spectra in the range of 2 to 20kD. Data were collected in an automated fashion using random sampling over individual samples to minimize the effects of operator bias. FlexAnalysis software was used after spectra collection

Table 1
Characteristics of included *B. contaminans* strains collected from Argentinian patients.

Unique identifier	Outbreak	PFGE pattern	Origin	Diagnosis	Location	Patient
310-12	1	BX3	Blood	Other	Cordoba	
311-12	1	BX3	Blood	Other	Cordoba	
312-12	1	BX3	Blood	Other	Cordoba	
313-12	1	BX3	Blood	Other	Cordoba	
314-12	1	BX3	Blood	Other	Cordoba	
315-12	1	BX3	Blood	Other	Cordoba	
316-12	1	BX3	Blood	Other	Cordoba	
460-12	2	BX1	Blood	Other	Misiones	
461-12	2	BX1	Blood	Other	Misiones	
462-12	2	BX1	Blood	Other	Misiones	
463-12	2	BX1	Blood	Other	Misiones	
464-12	2	BX1	Blood	Other	Misiones	
465-12	2	BX1	Blood	Other	Misiones	
466-12	2	BX1	Blood	Other	Misiones	
632-13	3	NT	Blood	N/A	CABA	
634-13	3	NT	Blood	N/A	CABA	
635-13	3	NT	Blood	N/A	CABA	
636-13	3	NT	Blood	N/A	CABA	
231-12	Sporadic	BX3	Sputum	CF	Cordoba	A
384-13	Sporadic	BX3	Sputum	CF	Cordoba	A
426-13	Sporadic	BX3	Sputum	CF	Cordoba	A
476-14	Sporadic	BX3	Sputum	CF	Cordoba	A
512-13	Sporadic	BX3	Sputum	CF	Cordoba	A
477-14	Sporadic	BX3	Sputum	CF	Cordoba	B
545-13	Sporadic	BX3	Blood	CF	Cordoba	B
371-14	Sporadic	NT	Sputum	CF	Cordoba	C
452-14	Sporadic	NT	Sputum	CF	Cordoba	C
506-13	Sporadic	BX4	Sputum	CF	CABA	D
575-13	Sporadic	BX4	Sputum	N/A	CABA	D
311-11	Sporadic	NT	Sputum	CF	La Pampa	E
465-13	Sporadic	NT	Sputum	CF	La Pampa	E
026-12	Sporadic	BX4	Sputum	CF	BA	
154-12	Sporadic	BX7	Blood	Other	Rio Negro	
173-13	Sporadic	BX3	Blood	Other	Cordoba	
191-14	Sporadic	BX6	Blood	Other	BA	
193-14	Sporadic	BX6	Blood	Other	BA	
211-13	Sporadic	BX1	Blood	Other	Cordoba	
212-13	Sporadic	BX2	Blood	N/A	Cordoba	
265-11	Sporadic	BX3	Sputum	CF	Cordoba	
341-12	Sporadic	BX3	Blood	Other	Cordoba	
341-14	Sporadic	BX7	Sputum	CF	CABA	
351-12	Sporadic	NT	Sputum	CF	BA	
353-12	Sporadic	BX3	Urine	Other	BA	
369-14	Sporadic	NT	Sputum	CF	Cordoba	
385-13	Sporadic	BX3	Blood	N/A	Cordoba	
386-13	Sporadic	BX3	Blood	N/A	Cordoba	
421-12	Sporadic	BX1	N/A	N/A	Misiones	
423-14	Sporadic	NT	Sputum	CF	BA	
425-12	Sporadic	NT	Sputum	CF	BA	
447-13	Sporadic	BX3	Blood	N/A	Cordoba	
450-12	Sporadic	BX3	Sputum	CF	Cordoba	
472-13	Sporadic	BX3	Blood	N/A	Cordoba	
480-14	Sporadic	BX3	Sputum	CF	Cordoba	
497-13	Sporadic	BX3	Blood	N/A	Cordoba	
546-13	Sporadic	BX3	Blood	N/A	Cordoba	

PFGE: pulsed-field gel electrophoresis, NT: non-typeable, CF: cystic fibrosis, N/A: not available, BA: Buenos Aires province, CABA: Ciudad Autonoma de Buenos Aires.

to perform smoothing, normalisation, baseline subtraction and peak picking. Following this, null, low signal to noise ratio (S/N) and outlier spectra were discarded after visual inspection. Spectra were then combined to make MSPs.

2.2. Data analysis

2.2.1. Confirmation of species-level discrimination

Isolates were initially tested to confirm the ability of Bruker Real-Time Classification (Bruker Daltonik) to correctly classify them at the species level. This is a software feature of the Microflex instrument which identifies isolates by matching test spectra to spectra in Bruker libraries and in user generated *Burkholderia* libraries. Quality control was assured if all isolates were correctly classified as *B. contaminans*.

2.2.2. Assessment of sub-species-level discrimination

Samples were analysed according to outbreak (1, 2 or 3), location, clinical source, patient diagnosis, PFGE subtype, and patient ID. Due to the small number of isolates, sub-species variables were collapsed into binary variables to improve discriminatory power: outbreaks were tested per outbreak vs all other samples, location was tested as Cordoba vs all other samples, and PFGE subtype was tested as non-typeable vs all others.

2.2.2.1. Principle component analysis (PCA). ClinProTools3.1 was used to generate PCA graphs. PCA condenses the many dependent variables (i.e. mass peaks) of spectra into three axes (Wold et al., 1987). Each principle component (PC) is a linear combination of the original variables, with the first PC representing the combination with the largest amount of variability. The second and third linear combinations are then created and plotted orthogonal to the previous one meaning there is no redundant information. PC plots show the possible variance of samples for each PC, and more similar spectra are expected to cluster in the same three-dimensional space. The results from this analysis were then used to create MSP dendrograms using Bruker BioTyper3.0. Top Hat Baseline correction was employed, as was Savitzky-Golay smoothing and recalibration.

2.2.2.2. Model generation. ClinProTools3.1 was used to generate isolate classification models, using the spectra generated from all of the study strains. Three algorithms were available for model development: Genetic Algorithm (GA), Supervised Neural Network (SNN) using *3*-nearest neighbours and the QuickClassifier (QC). Recognition capability (RC) is the % of correctly classified spectra for a given model, when all spectra used in the generation of the model are used to test it. Cross-validation (CV) gives an estimate of model reliability and how it will behave in the future and was calculated by randomly selecting 80% of samples to calibrate the model and using 20% to test it, with 10 iterations. If generated models indicated >90% discrimination capacity and further isolates were available, external validation was performed to test model reliability with real-world samples. Due to the large number of hypothesis tests being carried out in model generation, the Benjamini & Hochberg *p*-value adjustment was automatically run by ClinProTools3.1 to overcome the higher risk of type 1 error.

2.2.2.3. Visual inspection. Spectra were analysed visually in ClinProTools3.1 for peak shifts which may relate to changes in peptide weight due to single nucleotide polymorphisms (SNPs) or base substitutions.

3. Results

3.1. Strains

Clinical, geographic, outbreak-association, and PFGE pattern for the studied strains are summarized in Table 1. Spectral analysis by the Bruker RealTime Classification software confirmed that all isolates had been correctly identified *B. contaminans*.

Table 2
PFGE typing of the sporadic *B. contaminans* isolates originating from different territories.

Location	BX1		BX2		BX3		BX4		BX6		BX7		NT		Total	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
BA	0	0%	0	0%	1	14%	1	14%	2	29%	0	0%	3	43%	7	100%
CABA	0	0%	0	0%	0	0%	2	67%	0	0%	1	33%	0	0%	3	100%
Cordoba	1	4%	1	4%	18	78%	0	0%	0	0%	0	0%	3	13%	23	100%
La Pampa	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	2	100%	2	100%
Misiones	1	100%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	1	100%
Rio Negro	0	0%	0	0%	0	0%	0	0%	0	0%	1	100%	0	0%	1	100%
Total	**2**	**5%**	**1**	**3%**	**19**	**51%**	**3**	**8%**	**2**	**5%**	**2**	**5%**	**8**	**22%**	**37**	**100%**

PFGE: pulsed-field gel electrophoresis, NT: non-typeable, N/A: not available, BA: Buenos Aires province, CABA: Ciudad Autonoma de Buenos Aires.

3.2. Unsupervised clustering

3.2.1. Principle component analysis (PCA)

PCA plots show the discriminatory capacity of three orthogonally plotted condensed variables derived from the MS spectra of strains included in the model. Visual groupings of spectra separate from other spectra of a different outbreak signify that they are both similar to others within the group (homologous) as well as different to the other outbreak strains. There was no discrimination of isolates by any sub-species-level variable (Figs. 2-6). Whilst there is evidence of homology of outbreak strains (red) in Fig. 1 and non-CF patients (green) in Fig. 5, this is non-discriminatory, as other strains are present overlap the same area of the graph.

3.2.2. Dendrograms

Statistics from PCA used in dendrogram formation confirm that sub-species-level variables do not cluster together. A composite dendrogram of all variables used in PCA analysis (Fig. 7) displays the heterogenous nature of isolate spectra, with little discriminatory clustering of isolates by any given sub-species-level variable.

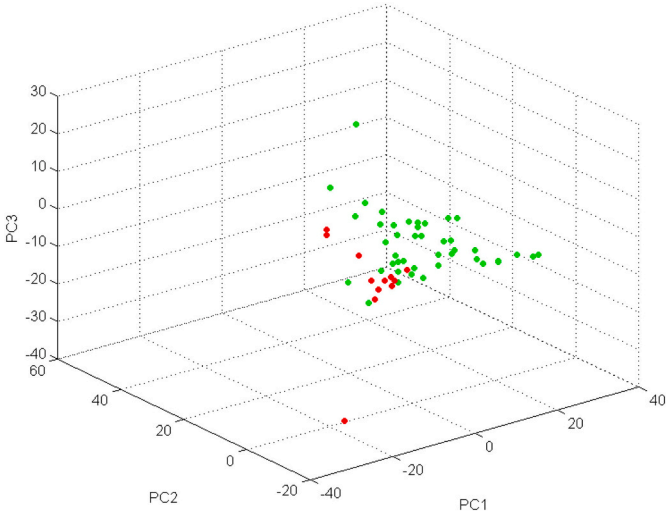


Fig. 2. PCA plot comparing PGFE Typeable strains (green) with PGFE Non-Typeable strains (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

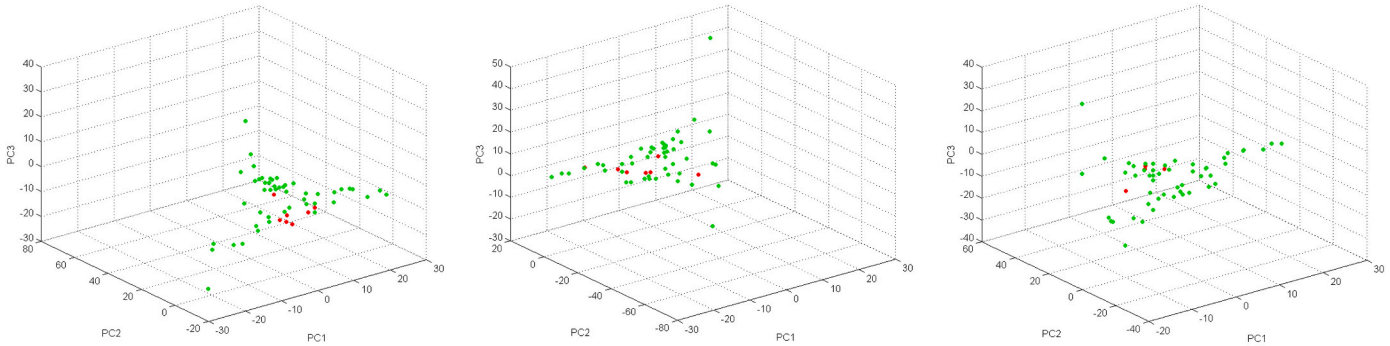
3.3. Models

3.3.1. Classification models

Table 3 demonstrates the ability of the different classification algorithms to recognise individual strain samples used to develop the algorithm at the sub-species level. Cross validation further verified this by randomly selecting 80% of the strains to calibrate the algorithm and 20% to test it. High scores for cross-validation indicate high internal validity. The GA algorithm consistently recognised samples at the sub-species level, with 100% recognition for typeability, location and outbreak class, and 88% to 100% recognition for cross validation of these classes. SNN performed the worst of the three, generally recognising around half of the strains correctly, while the QC algorithm performed reasonable successfully, consistently recognising over 80% correctly for all models.

3.3.2. External validation

A further 25 typeable and non-typeable strains collected from various hospitals in Argentina were used to test the PFGE typeability models. In this case the QC model performed the best, with 100% (95% CI 74% to 100%) sensitivity and 100% (95%CI 75 to 100) specificity on the samples tested (Table 4).



A: Outbreak 1 (red) vs Others (green) B: Outbreak 2 (red) vs Others (green) C: Outbreak 3 (red) vs Others (green)

Fig. 1. PCA plots comparing strains from one outbreak with strains from all other outbreaks.

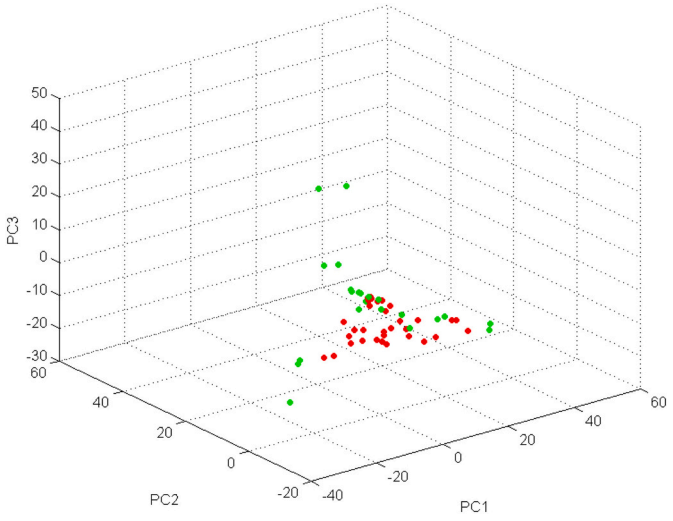


Fig. 3. PCA plot comparing strains collected from blood (red) with strains collected from sputum (green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

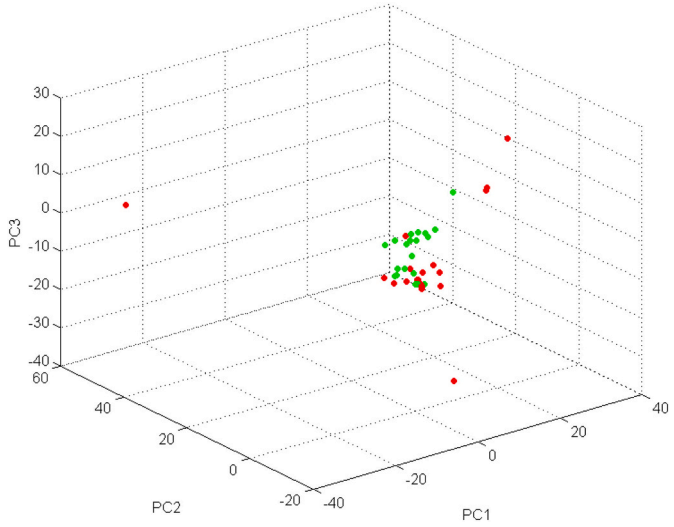


Fig. 5. PCA plot comparing strains collected from patients with CF (red) with patients without CF (green).

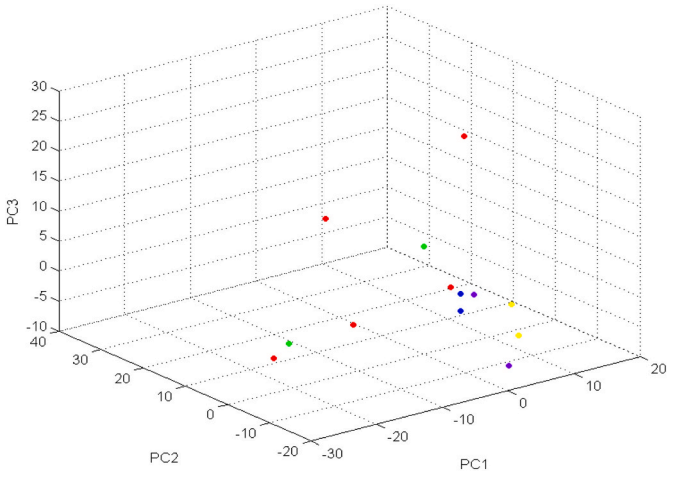


Fig. 4. PCA plot comparing strains where more than one strain was collected from a single patient. A (red) versus B (green) versus C (blue) versus D (yellow) versus E (purple). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

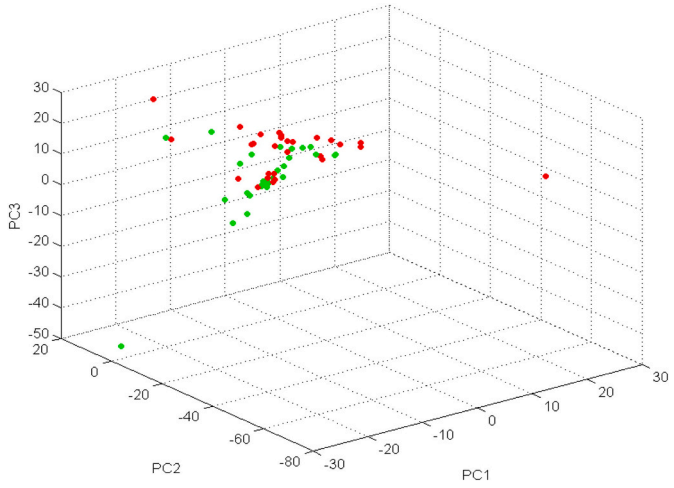


Fig. 6. PCA plot comparing strains collected from patients from Cordoba (red) with strains from all other locations (green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.4. Visual inspection

Spectra were examined to identify any relevant peaks in the initial set of isolates which may correlate with the ability or inability of a strain to be typed, and whether this correlation held for the external validation strains. One peak at *m/z* (mass to charge ratio) of approximately 6943 was found to be present in all non-typeable strains and absent in typeable strains (Fig. 8).

4. Discussion

4.1. Epidemiological classification

Using unsupervised clustering methods, MALDI-TOF MS was unable to discriminate *B. contaminans* samples into epidemiological outbreak classes. However, using statistical modelling of the isolate spectra, the GA consistently produced models with 97%–100% recognition capability and over 88% accuracy when cross validated for classification by outbreak, location and PFGE typeability. This provides proof-of-concept for MALDI-TOF MS clonal, epidemiological recognition, with resultant impacts on public health outbreak monitoring, and in particular, tracing pathogen spread across regions.

4.2. Comparability of PFGE and MALDI-TOF MS methods for typeability of B. contaminans isolates

Externally validating the PFGE typeability model with other routinely collected strains showed promise in the distinction of typeable and non-typeable strains using MALDI-TOF MS, with 100% sensitivity and specificity in the QC model. It is unclear why the QC model surpassed the GA model in the external validation. This may be due to a difference in the model, or to so with the representativeness of the samples used for validation. However the success in both models merits further investigation with larger datasets.

The identification of a unique peak for PFGE non-typeable samples is consistent with reports of biomarkers of other phenotypic variants, including drug resistance and virulence, in a range of bacteria (Flores-Treviño et al., 2019; Hrabák et al., 2013; Oviaño and Bou, 2019).

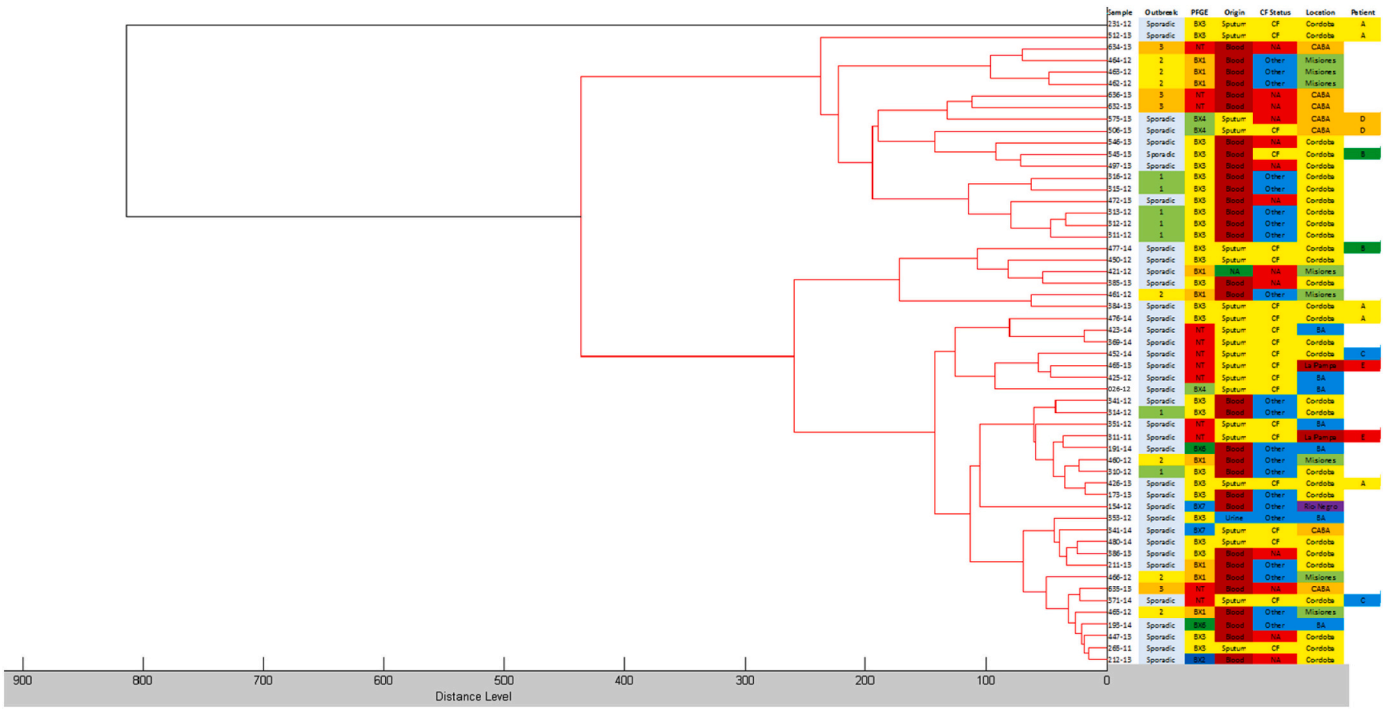


Fig. 7. Sample dendrogram demonstrating relationships between strains, using PCA residuals PFGE = Pulsed-Field Gel Electrophoresis, CF = Cystic Fibrosis, NT = Non-typeable, BA = Buenos Aires province CABA = Ciudad Autonoma de Buenos Aires.

Table 3
Strain recognition using Genetic Algorithm, Supervised Neural Network and Quick Classifier algorithms in five classification models.

Class	Model	Recognition Capability (%)	Cross Validation (%)
Typeable	*GA*	100	93
	SNN	50	51
	QC	90	88
Location	*GA*	97	90
	SNN	83	80
	QC	82	80
Outbreak 1	*GA*	100	95
	SNN	52	50
	QC	83	83
Outbreak 2	*GA*	100	88
	SNN	51	51
	QC	86	74
Outbreak 3	*GA*	100	93
	SNN	51	50
	QC	81	80

GA: genetic algorithm, SNN: supervised neural network, QC: quick classifier.

Table 4
External validation of PFGE typeability classification model using a further 25 non-typeable Argentinian *B. contaminans* strains.

	Sensitivity (%)	95% CI	Specificity (%)	95% CI
GA	83	52 to 98	100	75 to 100
QC	100	74 to 100	100	75 to 100
SNN	100	74 to 100	0	0 to 25

CI: confidence interval, GA: genetic algorithm, SNN: supervised neural network, QC: quick classifier.

Further examination of the peak weight identified at *m/z* of approximately 6943 may elucidate a peptide, or peptide mutation, which may be responsible for the non-typeability of certain bacterial strains when using *XbaI* PFGE. We hypothesise this may be due to an interaction between the peptide of interest and the *XbaI* restriction endonuclease. This supports a larger aim of identifying phenotypic correlates of MALDI-TOF MS-genetic relationships. MALDI-TOF MS spectra have been used to detect discriminatory peaks which could reliably identify outbreak strains of Shiga-toxigenic *E. coli*. Using liquid chromatography MS and publicly available databases, the proteins underlying these were identified (Christner et al., 2014). As higher resolution studies are being carried out, it is important to tease out the link between peak weights, proteins and the genome, especially at the lower and higher kDa ranges. High quality genetic analysis will allow correlation of e.g. single nucleotide polymorphisms to MALDI-TOF MS spectra. Other possibilities include the detection of lipids and carbohydrates by MALDI-TOF MS systems, rather than proteins only.

4.3. Clonality in outbreaks

The outbreaks in this study were classed as such due to their increased frequency at particular locations, however this does not guarantee clonality. Genetic analysis of the *recA* genes of 120 Argentinian *B. contaminans* strains from 2004 to 2010 using BOX-PCR revealed 10 patterns. During a particular outbreak, 55% of strains were found to be of the B8 subtype (implying a nosocomial clonal outbreak strain) whilst the other 45% of the same "outbreak" were thought to be circulating from other sources, such as environmental, but spread simultaneously (Martina et al., 2013), indicating the ambiguous nature of epidemiologically defined outbreak clusters.

There can also be significant genetic diversity within *B. contaminans* strains in the same patient chronically. In one study, over half of infected CF patients had a change in PCR-BOX subtype, and that multiple subtypes were likely coexistent in the same patient, indicating that at the time of infection or outbreak there is high genetic diversity (Martina et al., 2013). This implies that much of the aforementioned environmental diversity may in fact be due to co-colonization by multiple strains, none of which may be associated with outbreak transmission dynamics. A genetic study by Nunyar et al showed the presence of two concurrent phenotypes in the same patient in both blood and sputum, with different virulence factor expression and genomes, with 1433

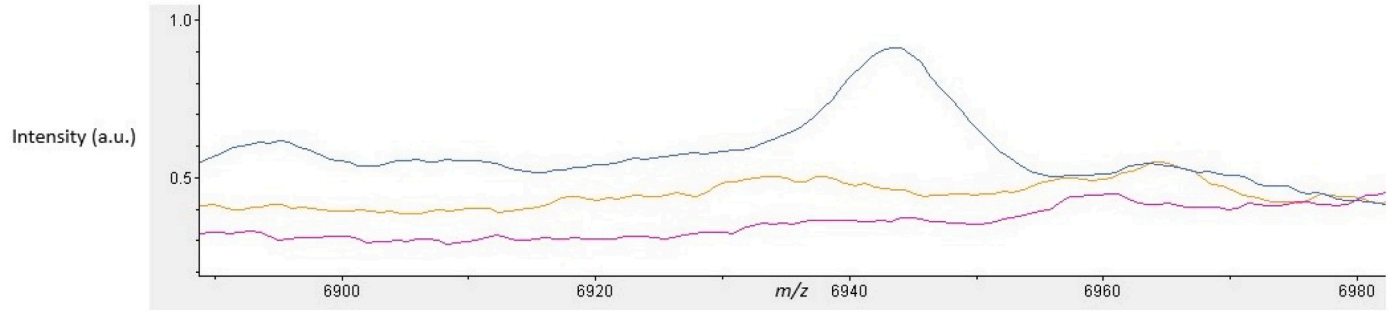


Fig. 8. Representative typeable and non-typeable spectra. Purple – sample 312–12, Outbreak 1 (PFGE BX3); sample 464–12, Yellow – Outbreak 2 (PFGE BX1); Green – sample 636–13, Outbreak 3 (PFGE non-typeable). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mutations detected (Nunvar et al., 2016). This correlates with the lack of obvious clustering of our samples when two samples from the same patient were analysed in unsupervised methods.

Further weakening the link between genetic clonality and outbreak is the possibility that outbreak strains are primarily propagated through methods independent to their genetics, such as patient characteristics or healthcare worker behaviours. This is interesting in the context of the non-typeable strains being correctly characterised by MALDI-TOF MS, as this likely does represent a biochemical or genetic trait affecting *XbaI* restrictase functioning. The performance of the GA outbreak classification models in our study also suggests that the outbreak strains may share elements such as epigenetic factors, environmental factors driving expression of particular phenotypes, or genotype fragments.

4.4. Limitations and recommendations

One limitation with the data set was the possible confounding of the CF vs non-CF analysis with the blood vs sputum analysis. All patients with CF had samples collected from sputum, whilst only one patient with a collection from blood samples had CF. Furthermore 11 of the 32 blood samples had an unknown disease aetiology. Therefore, excluding them from the analysis of CF vs non-CF may have also had the effect of disproportionately excluding strains taken from blood samples. A detailed analysis of MALDI-TOF MS discriminatory capacity by PFGE profile would also be useful. It would therefore be worthwhile to carry out a similar project using a larger sample size with strains with complete epidemiological data.

Due to time and resource limitations external validation was unable to be performed on all models. It is unclear how the location models would perform if externally validated using other samples from a similar time and place, and further study should include collecting strains form these regions in order to test this. External validation of outbreak strains is only feasible in real-time studies, where an outbreak has been detected and one wants to know whether a new strain is part of this clonal outbreak or an unrelated sporadic strain appearing at the same time.

5. Conclusion

MALDI-TOF MS successfully discriminates *B. contaminans* isolates into clonal, epidemiological clusters, and can recognise isolates non-typeable by PFGE. Further work should investigate this capability, including peptide studies and genomic sequencing to identify individual proteins or genes responsible for this non-typeability, which may be linked to the spectra peak identified.

Funding

This work was partially supported by the National Institutes of Health (NIH) [grant numbers R01GM103525 and U01CA207167], the Oxford University Clinical Academic Graduate School and the Walter Guinness Charitable Trust. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the funders, which had role in study design, data collection and analysis, or decision to publish, or preparation of the manuscript.

Contributions

Selali Fiamanya: Conceptualisation, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Writing - original draft. Lucía Cipolla: Conceptualisation, Investigation, Methodology, Resources, Supervision, Writing - review & editing. Mónica Prieto: Resources, Supervision, Writing - review & editing. John Stelling: Supervision, Funding Acquisition, Writing - review & editing.

Declaration of Competing Interest

The authors have no conflicting affiliations or financial or non-financial interests in the subject matter discussed in this manuscript.

References

Barco, L., Barrucci, F., Olsen, J.E., Ricci, A., 2013. Salmonella source attribution based on microbial subtyping. Int. J. Food Microbiol. 163 (2–3), 193–203. Elsevier B.V. https://doi.org/10.1016/j.ijfoodmicro.2013.03.005.

Bar-Meir, M., Berliner, E., Kashat, L., Zeevi, D.A., Assous, M.V., 2020. The utility of MALDI-TOF MS for outbreak investigation in the neonatal intensive care unit. Eur. J. Pediatr. 179 (12), 1843–1849. https://doi.org/10.1007/s00431-020-03696-3.

Böhme, K., Morandi, S., Cremonesi, P., Fernández No, I.C., Barros-Velázquez, J., Castiglioni, B., Brasca, M., Cañas, B., Calo-Mata, P., 2012. Characterization of *Staphylococcus aureus* strains isolated from Italian dairy products by MALDI-TOF mass fingerprinting. Electrophoresis 33 (15), 2355–2364. https://doi.org/10.1002/elps.201100480.

Bruker, 2020. MALDI Biotyper® Training Movies - Service. Bruker. https://www.bruker.com/service/education-training/training-courses/microbiology-diagnostics-training-courses/maldi-biotyper-training-movies.html.

Christner, M., Trusch, M., Rohde, H., Kwiatkowski, M., Schlüter, H., Wolters, M., Aepfelbacher, M., Hentschke, M., 2014. Rapid MALDI-TOF mass spectrometry strain typing during a large outbreak of Shiga-toxigenic *Escherichia coli*. PLoS One 9 (7). https://doi.org/10.1371/journal.pone.0101924.

Cipolla, L., Rocca, F., Armitano, R.I., Martinez, C., Almuzara, M., Faccone, D., Vay, C., Prieto, M., 2019. Development and evaluation of an in-house database for quick identification of Burkholderia contaminans by MALDI-TOF MS. Rev. Argentina Microbiol. 51 (3), 255–258. https://doi.org/10.1016/j.ram.2018.09.001.

Clark, A.E., Kaleta, E.J., Arora, A., Wolk, D.M., 2013a. Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. Clin. Microbiol. Rev. 26 (3), 547–603. https://doi.org/10.1128/CMR.00072-12.

Clark, C.G., Kruczkiewicz, P., Guan, C., McCorrister, S.J., Chong, P., Wylie, J., van Caeseele, P., Tabor, H.A., Snarr, P., Gilmour, M.W., Taboada, E.N., Westmacott, G.R., 2013b. Evaluation of MALDI-TOF mass spectroscopy methods for determination of *Escherichia coli* pathotypes. J. Microbiol. Methods 94 (3), 180–191. https://doi.org/10.1016/j.mimet.2013.06.020.

Degand, N., Carbonnelle, E., Dauphin, B., Beretti, J.L., Le Bourgeois, M., Sermet-Gaudelus, I., Segonds, C., Berche, P., Nassif, X., Ferroni, A., 2008. Matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of nonfermenting gram-negative bacilli isolated from cystic fibrosis patients. J. Clin. Microbiol. 46 (10), 3361–3367. https://doi.org/10.1128/JCM.00569-08.

Dieckmann, R., Malorny, B., 2011. Rapid screening of epidemiologically important *Salmonella enterica subsp. enterica* serovars by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl. Environ. Microbiol. 77 (12), 4136–4146. https://doi.org/10.1128/AEM.02418-10.

Dieckmann, R., Helmuth, R., Erhard, M., Malorny, B., 2008. Rapid classification and identification of salmonellae at the species and subspecies levels by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl. Environ. Microbiol. 74 (24), 7767–7778. https://doi.org/10.1128/AEM.01402-08.

Fehlberg, L.C.C., Andrade, L.H.S., Assis, D.M., Pereira, R.H.V., Gales, A.C., Marques, E.A., 2013. Performance of MALDI-ToF MS for species identification of *Burkholderia cepacia* complex clinical isolates. Diagn. Microbiol. Infect. Dis. 77 (2), 126–128. https://doi.org/10.1016/j.diagmicrobio.2013.06.011.

Flores-Treviño, S., Garza-González, E., Mendoza-Olazarán, S., Morfín-Otero, R., Camacho-Ortiz, A., Rodríguez-Noriega, E., Martínez-Meléndez, A., Bocanegra-Ibarias, P., 2019. Screening of biomarkers of drug resistance or virulence in ESCAPE pathogens by MALDI-TOF mass spectrometry. Sci. Rep. 9 (1) https://doi.org/10.1038/s41598-019-55430-1.

Florio, W., Tavanti, A., Barnini, S., Ghelardi, E., Lupetti, A., 2018. Recent advances and ongoing challenges in the diagnosis of microbial infections by MALDI-TOF mass spectrometry. Front. Microbiol. 9 (May) https://doi.org/10.3389/fmicb.2018.01097. Frontiers Media S.

Gatlin-Bunai, C.L., Cazares, L.H., Cooke, W.E., Semmes, O.J., Malyarenko, D.I., 2007. Optimization of MALDI-TOF MS detection for enhanced sensitivity of affinity-captured proteins spanning a 100 kDa mass range. J. Proteome Res. 6 (11), 4517–4524. https://doi.org/10.1021/pr0703526.

Grundmann, H., Klugman, K.P., Walsh, T., Ramon-Pardo, P., Sigauque, B., Khan, W., Laxminarayan, R., Heddini, A., Stelling, J., 2011. A framework for global surveillance of antibiotic resistance. Drug Resist. Updat. 14 (2), 79–87. https://doi.org/10.1016/j.drup.2011.02.007.

Hou, T.Y., Chiang-Ni, C., Teng, S.H., 2019. Current status of MALDI-TOF mass spectrometry in clinical microbiology. J. Food Drug Anal. 27 (2), 404–414. Elsevier Taiwan LLC. https://doi.org/10.1016/j.jfda.2019.01.001.

Hrabák, J., Chudácková, E., Walková, R., 2013. Matrix-assisted laser desorption ionization-time of flight (MALDITOF) mass spectrometry for detection of antibiotic resistance mechanisms: from research to routine diagnosis. Clin. Microbiol. Rev. 26 (1), 103–114. https://doi.org/10.1128/CMR.00058-12.

Josten, M., Reif, M., Szekat, C., Al-Sabti, N., Roemer, T., Sparbier, K., Kostrzewa, M., Rohde, H., Sahl, H.G., Bierbaum, G., 2013. Analysis of the matrix-assisted laser desorption ionization-time of flight mass spectrum of Staphylococcus aureus identifies mutations that allow differentiation of the main clonal lineages. J. Clin. Microbiol. 51 (6), 1809–1817. https://doi.org/10.1128/JCM.00518-13.

Khot, P.D., Fisher, M.A., 2013. Novel approach for differentiating shigella species and escherichia coli by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J. Clin. Microbiol. 51 (11), 3711–3716. https://doi.org/10.1128/JCM.01526-13.

Kuhns, M., Zautner, A.E., Rabsch, W., Zimmermann, O., Weig, M., Bader, O., Groß, U., 2012. Rapid discrimination of *Salmonella enterica* Serovar Typhi from other serovars by MALDI-TOF mass spectrometry. PLoS One 7 (6), e40004. https://doi.org/10.1371/journal.pone.0040004.

Lau, A.F., Wang, H., Weingarten, R.A., Drake, S.K., Suffredini, A.F., Garfield, M.K., Chen, Y., Gucek, M., Youn, J.H., Stock, F., Tso, H., DeLeo, J., Cimino, J.J., Frank, K. M., Dekker, J.P., 2014. A rapid matrix-assisted laser desorption ionization-time of flight mass spectrometry-based method for single-plasmid tracking in an outbreak of carbapenem-resistant Enterobacteriaceae. J. Clin. Microbiol. 52 (8), 2804–2812. https://doi.org/10.1128/JCM.00694-14.

Martina, P., Bettiol, M., Vescina, C., Montanaro, P., Mannino, M.C., Prieto, C.I., Vay, C., Naumann, D., Schmitt, J., Yantorno, O., Lagares, A., Bosch, A., 2013. Genetic diversity of Burkholderia contaminans isolates from cystic fibrosis patients in Argentina. J. Clin. Microbiol. 51 (1), 339–344. https://doi.org/10.1128/JCM.02500-12.

Medina-Pascual, M.J., Valdezate, S., Carrasco, G., Villalón, P., Garrido, N., Saéz-Nieto, J. A., 2015. Increase in isolation of Burkholderia contaminans from Spanish patients with cystic fibrosis. Clin. Microbiol. Infect. 21 (2), 150–156. https://doi.org/10.1016/j.cmi.2014.07.014.

Miñán, A., Bosch, A., Lasch, P., Stämmler, M., Serra, D.O., Degrossi, J., Gatti, B., Vay, C., D'Aquino, M., Yantorno, O., Naumann, D., 2009. Rapid identification of Burkholderia cepacia complex species including strains of the novel Taxon K, recovered from cystic fibrosis patients by intact cell MALDI-ToF mass spectrometry. Analyst 134 (6), 1138–1148. https://doi.org/10.1039/b822669e.

Nagy, E., Becker, S., Sóki, J., Urbán, E., Kostrzewa, M., 2011. Differentiation of division I (cfiA-negative) and division II (cfiA-positive) *Bacteroides fragilis* strains by matrix-assisted laser desorption/ionization time of-flight mass spectrometry. J. Med. Microbiol. 60 (11), 1584–1590. https://doi.org/10.1099/jmm.0.031336-0.

Novais, Â., Sousa, C., De Dios Caballero, J., Fernandez-Olmos, A., Lopes, J., Ramos, H., Coque, T.M., Cantón, R., Peixe, L., 2014. MALDI-TOF mass spectrometry as a tool for the discrimination of high-risk *Escherichia coli* clones from phylogenetic groups B2 (ST131) and D (ST69, ST405, ST393). Eur. J. Clin. Microbiol. Infect. Dis. 33 (8), 1391–1399. https://doi.org/10.1007/s10096-014-2071-5.

Nunvar, J., Kalferstova, L., Bloodworth, R.A.M., Kolar, M., Degrossi, J., Lubovich, S., Cardona, S.T., Drevinek, P., 2016. Understanding the pathogenicity of Burkholderia contaminans, an emerging pathogen in cystic fibrosis. PLoS One 11 (8). https://doi.org/10.1371/journal.pone.0160975.

Oviaño, M., Bou, G., 2019. Matrix-assisted laser desorption ionization–time of flight mass spectrometry for the rapid detection of antimicrobial resistance mechanisms and beyond. Clin. Microbiol. Rev. 32 (1) https://doi.org/10.1128/CMR.00037-18. American Society for Microbiology.

Sandrin, T.R., Goldstein, J.E., Schumaker, S., 2013. MALDI TOF MS profiling of bacteria at the strain level: a review. Mass Spectromet. Rev. 32 (3), 188–217. Mass Spectrom Rev. https://doi.org/10.1002/mas.21359.

Singhal, N., Kumar, M., Kanaujia, P.K., Virdi, J.S., 2015. MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. Front. Microbiol. 6 (Aug) https://doi.org/10.3389/fmicb.2015.00791. Frontiers Research Foundation.

Van Belkum, A., Welker, M., Erhard, M., Chatellier, S., 2012. Biomedical mass spectrometry in today's and tomorrow's clinical microbiology laboratories. J. Clin. Microbiol. 50 (5), 1513–1517. American Society for Microbiology Journals. https://doi.org/10.1128/JCM.00420-12.

Vanlaere, E., Sergeant, K., Dawyndt, P., Kallow, W., Erhard, M., Sutton, H., Dare, D., Devreese, B., Samyn, B., Vandamme, P., 2008. Matrix-assisted laser desorption ionisation-time-of of-flight mass spectrometry of intact cells allows rapid identification of *Burkholderia cepacia* complex. J. Microbiol. Methods 75 (2), 279–286. https://doi.org/10.1016/j.mimet.2008.06.016.

Vanlaere, E., Baldwin, A., Gevers, D., Henry, D., De Brandt, E., LiPuma, J.J., Mahenthiralingam, E., Speert, D.P., Dowson, C., Vandamme, P., 2009. Taxon K, a complex within the *Burkholderia cepacia* complex, comprises at least two novel species, Burkholderia contaminans sp. nov. and Burkholderia lata sp. nov. Int. J. Syst. Evol. Microbiol. 59 (1), 102–111. https://doi.org/10.1099/ijs.0.001123-0.

Wieser, A., Schneider, L., Jung, J., Schubert, S., 2012. MALDI-TOF MS in microbiological diagnostics-identification of microorganisms and beyond (mini review). Appl. Microbiol. Biotechnol. 93 (3), 965–974. Appl Microbiol Biotechnol. https://doi.org/10.1007/s00253-011-3783-4.

Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. Chemom. Intell. Lab. Syst. 2 (1), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9.

Wolters, M., Rohde, H., Maier, T., Belmar-Campos, C., Franke, G., Scherpe, S., Aepfelbacher, M., Christner, M., 2011. MALDI-TOF MS fingerprinting allows for discrimination of major methicillin-resistant *Staphylococcus aureus* lineages. Int. J. Med. Microbiol. 301 (1), 64–68. https://doi.org/10.1016/j.ijmm.2010.06.002.