

Evaluation of Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry and ClinPro Tools as a Rapid Tool for Typing *Streptococcus pyogenes*

Manal Tadros,^{1,2} Ana Cabrera,^{3,4} Larissa M. Matukas,^{1,2,5,6} and Matthew Muller^{5,6}

¹Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada, ²Division of Microbiology, Unity Health Toronto, Toronto, Ontario, Canada, ³Department of Pathology and Laboratory Medicine, Schulich School of Medicine & Dentistry, Western University, London, Ontario, Canada, ⁴Division of Microbiology, Pathology and Laboratory Medicine, London Health Sciences Centre, London, Ontario, Canada, ⁵Department of Medicine, University of Toronto, Toronto, Ontario, Canada, ⁶Division of Infectious Diseases, Unity Health Toronto, Toronto, Ontario, Canada

Background. Timely strain typing of group A *Streptococci* (GAS) is necessary to guide outbreak recognition and investigation. We evaluated the use of (matrix-assisted laser desorption ionization time-of-flight mass spectrometry) combined with cluster analysis software to rapidly distinguish between related and unrelated GAS isolates in real-time.

Methods. We developed and validated a typing model using 177 GAS isolates with known *emm* types. The typing model was created using 43 isolates, which included 8 different *emm* types, and then validated using 134 GAS isolates of known *emm* types that were not included in model generation.

Results. Twelve spectra were generated from each isolate during validation. The overall accuracy of the model was 74% at a cutoff value of 80%. The model performed well with *emm* types 4, 59, and 74 but showed poor accuracy for *emm* types 1, 3, 12, 28, and 101. To evaluate the ability of this tool to perform typing in an outbreak situation, we evaluated a virtual outbreak model using a “virtual outbreak strain; *emm*74” compared with a non-outbreak group or an “outgroup” of other *emm* types. External validation of this model showed an accuracy of 91.4%.

Conclusions. This approach has the potential to provide meaningful information that can be used in real time to identify and manage GAS outbreaks. Choosing isolates characterized by whole genome sequencing rather than *emm* typing for model generation should improve the accuracy of this approach in rapidly identifying related and unrelated GAS strains.

Keywords: GAS; MALDI-TOF; typing.

Group A *Streptococci* (GAS) causes a variety of diseases that range from uncomplicated pharyngitis to necrotizing fasciitis and streptococcal toxic shock syndrome and has frequently been associated with outbreaks in both healthcare and community settings [1, 2]. When an outbreak is suspected on epidemiological grounds, typing of the involved GAS isolates can confirm person-to-person transmission, help elucidate the chain of transmission, and guide public health and infection control interventions [3].

Several methods are available for GAS typing, including pulsed-field gel electrophoresis (PFGE) and *emm* typing [3, 4]. Both methods are labor intensive and not available in most

clinical microbiology laboratories. For example, in Canada, *emm* typing of GAS isolates is performed primarily at the National Microbiology Laboratory in Winnipeg, Manitoba [5]. For this reason, typing data are often available only after outbreak control measures are implemented. Whole genome sequencing (WGS) is the gold standard for molecular typing, but it is currently available primarily as a research tool for in depth cluster and phylogenetic analysis of *Streptococcus pyogenes* [6]. Although WGS has been used in the analysis of GAS outbreaks, this has been done retrospectively and not in a real-time manner to assist with outbreak identification and management

Timely characterization of GAS is necessary for rapid identification of outbreaks and transmission events and to guide outbreak investigation and control efforts, but our current methods cannot provide real-time data to public health or infection control providers. To address this issue, we evaluated the potential use of matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) combined with cluster analysis software to rapidly identify related and unrelated GAS isolates for real-time applications.

Received 18 July 2019; editorial decision 2 October 2019; accepted 3 October 2019.

Correspondence: Manal Tadros, MBBS, MSc, PhD, FRCPC, Unity Health Toronto, 30 Bond St., M5B 1W8, Room CCN 2-002 (manal.tadros@unityhealth.to).

Open Forum Infectious Diseases®

© The Author(s) 2019. Published by Oxford University Press on behalf of Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
DOI: 10.1093/ofid/ofz441

MATERIALS AND METHODS

Selection of Isolates

This study was conducted using 177 GAS isolates, all of which had been previously characterized by *emm* typing using polymerase chain reaction. The isolates were identified from clinical or surveillance specimens. Surveillance specimens were throat and/or rectal cultures collected to identify asymptomatic carriers during GAS outbreaks. Isolates were either identified at the St. Michael's Hospital clinical microbiology laboratory (N = 66) or provided by Toronto Invasive Bacterial Disease Network (TIBDN) (N = 111). The TIBDN operates a surveillance program for invasive GAS (iGAS), which includes all hospitals and laboratories in metropolitan Toronto and Peel region (population 4 million) (<http://tibdn.ca/>) [7]. These isolates were obtained from patients presenting for clinical care in both ambulatory and inpatient settings. In addition, a group of isolates were identified that were part of a large and sustained clonal iGAS outbreak occurring at a local men's shelter in 2016 [6]. A variety of different *emm* types were identified within this sample, including isolates of emerging *emm* types such as *emm101* [8]. In addition, TIBDN provided isolates from the 5 GAS *emm* types most frequently isolated in metropolitan Toronto from 2012 to 2016 (*emm* type 1, 3, 4, 12, and 28) (TIBDN, written personal communication, December 5, 2016), 21 isolates that were *emm59*—a strain that caused a major outbreak in Ontario before 2010 [9]—and 30 isolates of *emm* types that are less frequently seen in Ontario. The sources of the isolates were as follows: blood, 113; wound, 35; screening, 11; joint fluid, 7; abscess, 2; peritoneal fluid, 2; pleural fluid, 1; tissue, 4; throat, 2.

Isolate Preparation for Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry

Stored isolates were cultured from frozen beads or Amies transport medium with charcoal on Columbia blood agar plates containing 5% sheep blood, (Oxoid, Nepean, Canada) and incubated at 35°C with 5% CO₂. Group A *Streptococci* species were confirmed by beta-hemolysis on sheep blood agar, grouping of carbohydrate antigen using MedicoDx Strep Grouping Kit (Intermedico, Markham, Canada), and MALDI-TOF MS. The tube extraction method was performed as previously described [10]. The details of the tube extraction method are provided in the [Supplementary Section](#).

Development of a Typing Model

MALDI-TOF Microflex LT 60 (Bruker Daltonics GmbH, Bremen, Germany) with a laser frequency of 60 Hz and a vacuum setting of -6 mbar was used. ClinProTools software, version 3.0 (Bruker Daltonics) was used to generate a model for cluster analysis of the GAS isolates according to the procedures provided in the user manual. Within ClinProTools, the support vector machine (SVM) algorithm was chosen for GAS

classification. The SVM is an artificial intelligence tool that analyzes a variety of peaks and ranks them based on their separation properties. It then uses those selected peaks to generate a pattern recognition for each "class" or cluster. This pattern recognition enables SVM to predict whether a new spectra from a test isolate is related to one of the classes within the model.

Model Generation

The spectra generated by MALDI-TOF for all isolates of a certain *emm* type were given a unique class designation. ClinProTools then "cleaned up" these spectra using spectra recalibration (1000 ppm maximal peak shift, 30% match to calibrant peaks, exclusion of spectra that could not be recalibrated), average spectra calculation (resolution, 800), average peak list calculation (signal-to-noise threshold, 5), and peak calculation in the individual spectra. The software then determined whether there were sufficient differences between classes based on differences in spectra between classes.

The typing model for our study included 8 different classes where each class comprised different GAS isolates that belonged to the same *emm* type. The following *emm* types were used to develop the typing model: *emm1*, *emm74*, *emm3*, *emm4*, *emm12*, *emm28*, *emm101*, and *emm59*. A total of 43 isolates were used for model generation. Each *emm* type was represented by at least 5 isolates, with a minimum of 60 spectra for each class.

Cross-Validation

The typing model was cross-validated and recognition capability values were calculated. Cross-validation tests the ability of the model to discriminate between the classes defined by the model. Cross-validation was done by dividing the data set into k equal parts. From each data set, 20% of the data were removed and the remaining 80% were used to develop a typing model. Each model was then tested on the 20% of data not used in model development. This entire process was repeated for 10 iterations.

External Validation

External validation was done using 134 GAS isolates of known *emm* types that were not included in model generation. The test isolates for external validation were prepared using the same ethanol formic acid extraction method and were each spotted 12 times on the target plate to generate 12 spectra for each isolate. External validation was performed using the validate function of ClinProTools 3.0, which provides a probability for assigning the tested isolate to each class. The percentage agreement between ClinProTools and *emm* typing was calculated at different cutoffs with respect to the proportion of spectra for a given isolate that must be within a class to designate that isolate as belonging to the class. Cutoffs of 90%, 80%, and 70% were evaluated to determine the cutoff that provided the best agreement between ClinProTools and *emm* typing. For example, a

Table 1. Cross-Validation of Individual *emm* Types (Classes) Within the Model

<i>emm</i> Type (Class)	Result of Cross-Validation
1 (1)	94.0%
74 (2)	100%
3 (3)	98.8%
4 (4)	93.3%
12 (5)	97.1%
28 (6)	95.3%
101 (7)	100%
59 (8)	98.1%

cutoff value of 90% meant that, for an isolate to be belonging to a certain class, more than 10 of 12 of its spectra were grouped within this particular class.

Peak Analysis

Data on all peaks used in model development were provided, along with their mass-to-charge ratio (m/z) and their average intensity in each class.

Virtual Outbreak Model

Another typing model was generated to evaluate the performance of MALDI-TOF/ClinPro Tools during a potential clonal outbreak of GAS. The model was generated using the same steps described above but with only 2 classes. Class 1

was the “virtual outbreak group”, which was composed of spectra generated from 5 different isolates of *emm74*. Class 2 was the non-outbreak group or the “outgroup” and was composed of spectra from 15 different isolates of other *emm* types (*emm1*, *emm3*, and *emm28*). Cross-validation of the model was performed as described above. External validation of the “virtual outbreak model” was performed using 70 isolates, which included 30 isolates that were “outbreak cases”, ie, belonged to *emm74* and *emm40* isolates that were of other *emm* types: 1, 3, 4, 6, 11, 12, 28, 48, 53, 59, 80, 83, 89, and 101.

RESULTS

Cross-Validation

The GAS SVM model passed cross-validation with an overall accuracy of 97.1%. Results of cross-validation of individual classes are provided in Table 1. Twenty-two peaks were used for classification, with the smallest mass equal to 2277.59 Da and the largest mass equal to 9807.13 Da. A list of the statistics for peaks included in this model is provided in the [Supplementary Material](#).

For visual evaluation of characteristic peaks, the gel view (Figure 1) was used within the GAS SVM model. The gel view is a display of the peaks of each class in a pattern similar to PFGE gel based on their masses. The depth of the color of each peak corresponds to the intensity of the peak.

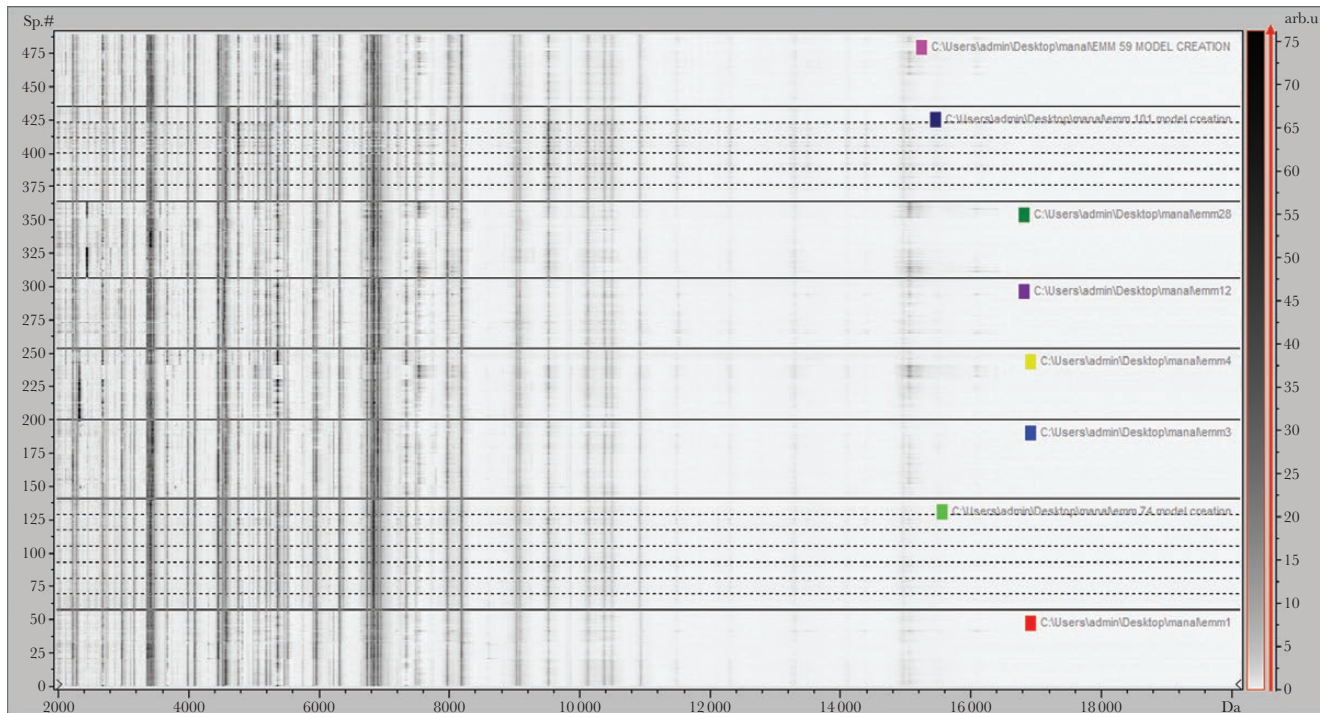


Figure 1. Gel view showing the spectra of 8 model generation classes, each depicted with a different colored square. The x-axis records the mass (m/z value) of the selected spectra. The left y-axis displays the running spectrum number originating from subsequent spectra loading. The peak intensity is expressed by a color code. The right y-axis indicates the peak intensity in arbitrary units.

The 2D Peak Distribution View (Figure 2) displays the distribution of 2 selected peaks in the included spectra of the loaded model generation classes. This view is a pictorial demonstration of the ability of the 2 peaks shown in it to discriminate between the different classes within the model. The peak numbers and mass (m/z) values are indicated on the x- and y-axes, respectively.

External Validation

The accuracy of the GAS SVM model at the 70%, 80%, and 90% cutoff was 73%, 74%, and 70%, respectively. The model demonstrated the best accuracy in external validation for isolates belonging to the following *emm* types: *emm74*, *emm4*, and *emm59* with accuracies of 88%, 100%, and 100% using the 80% cutoff. In addition, the model was successful in classifying *emm* types 6, 11, 31, 48, 53, 60, 73, 77, 83, 87, 89, and 114 as unrelated with an accuracy of 83% at the same cutoff. The model showed poor accuracy with other *emm* types (1, 3, 12, 28, and 101). Table 2 provides details of the number of isolates used for external validation

from each class as well as the accuracy of the GAS SVM model with individual classes.

Results of the Virtual Outbreak Model

The Virtual Outbreak Model passed cross-validation with an accuracy of 100% for each of class 1 and class 2. Ten peaks were used for classification. Receiver operating curve characteristics (ROC) was used to demonstrate the discriminatory capability of different peaks and presented as the area under the curve (AUC). A peak with an AUC of 1 has the best accuracy in discriminating class 1 from class 2. Supplementary Figure S1 shows ROC of peak 35; 3447.8 Da that had an AUC of 0.97. External validation showed an accuracy of 91.4% at a cutoff of 80%. Two isolates that were not “outbreak related” (*emm 4*) were misclassified as outbreak cases (assigned to class 1), and 4 outbreak related isolates (*emm74*) were misclassified as non-outbreak strains (assigned to class 2). Supplementary Figures S2 and S3 show correct classifications of outbreak related and “non-outbreak” strains, respectively. Supplementary Figure S4 shows a suggested workflow algorithm during a potential GAS outbreak.

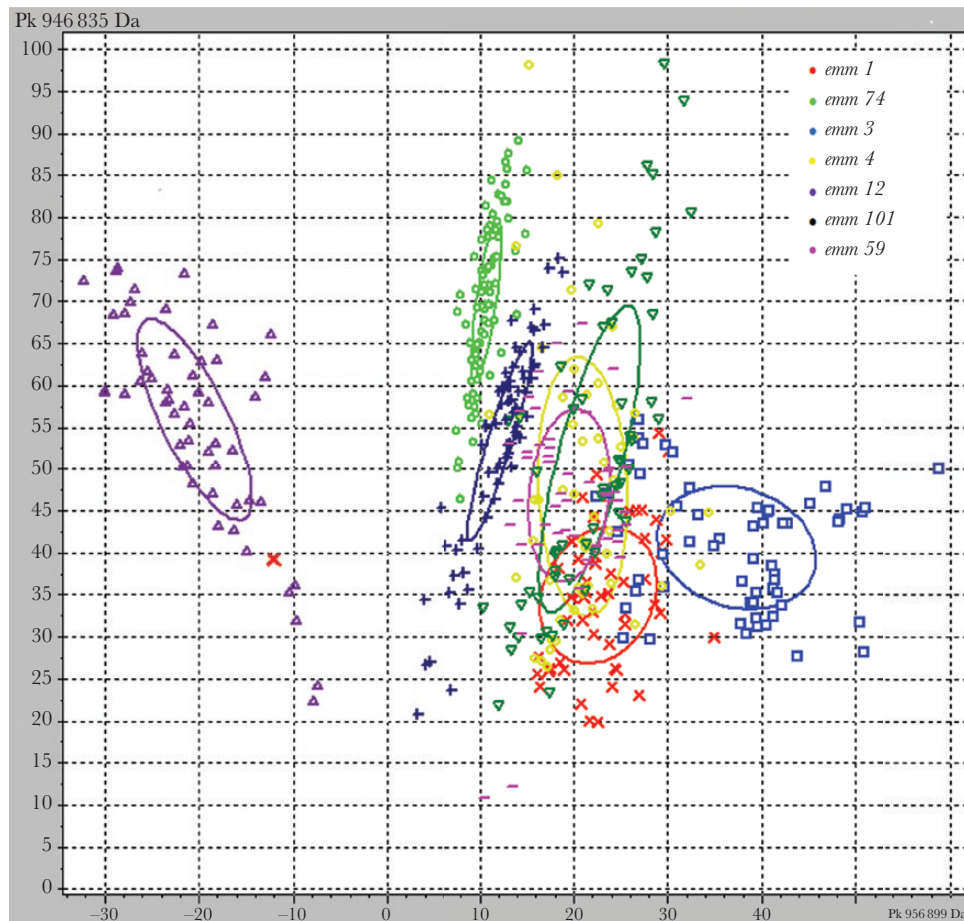


Figure 2. Two-dimensional peak distribution view. The x-axis shows the peak intensity values for peak 95 (6899 Da) within the different classes of the model, and the y-axis shows the peak intensity values for peak 94 (6835 Da) within the different classes of the model. The ellipses represent spectra with greater distinction within the model classes.

Table 2. External Validation Results of Individual *emm* Types at Different Cutoff Values Within the Model

<i>emm</i> Type (Class)	Number of Isolates for External Validation	Accuracy at Cutoff 90%	Accuracy at Cutoff 80%	Accuracy at Cutoff 70%
1 (1)	10	30%	50%	60%
74 (2)	32	78%	88%	91%
3 (3)	10	30%	30%	50%
4 (4)	10	100%	100%	100%
12 (5)	7	57%	57%	57%
28 (6)	7	0%	14%	14%
101 (7)	6	33%	33%	50%
59 (8)	16	94%	100%	100%
Other <i>emm</i> types (<i>emm</i> 6, 11, 31, 48, 53, 60, 73, 77, 83, 87, 89, and 114 not included in model)	36	89%	83%	67%

DISCUSSION

The optimal approach to outbreak assessment and management is through the combined use of epidemiological and molecular data [11]. In most cases, however, the molecular data are not available in time to assist with initial outbreak management interventions. Furthermore, although molecular methods remain the gold standard for cluster analysis of bacteria, lack of access to these tests at most facilities and the complexity and cost associated with them has led microbiologists to explore alternative typing methods [12]. The utility of MALDI-TOF and ClinPro Tools has been assessed for the typing of a variety of organisms including vancomycin-resistant *Enterococcus faecium* [12, 13], certain sequence types of extended-spectrum beta lactamases, *Escherichia coli* [14], *Mycobacterium abscessus* [15], *Streptococcus pneumoniae* [16], *Haemophilus influenzae* [17], carbapenem-resistant *Klebsiella pneumoniae* [18], *Mycoplasma pneumoniae* [19], *Lactobacillus casei* [20], and *Propionibacterium acnes* [21, 22]. In this study, we evaluated the utility of MALDI-TOF and Clin ProTools for cluster analysis of *S pyogenes* compared with *emm* typing. At a cutoff of 80%, the GAS SVM model had an overall accuracy of 74%. There was a wide variation in the successful assignment of tested isolates to the correct class for different *emm* types. Isolates that belonged to *emm* types 74 and 59 used in our study were derived from outbreaks and were closely related temporally. For these isolates, our model was quite successful in assigning them to the correct class with an accuracy of 88% for class 2 (*emm*74) and 100% for class 8 (*emm*59). We created a virtual outbreak model that included an outbreak class (*emm*74) and an outgroup (isolates from 3 other *emm* types). This model showed a good performance with 91.4% accuracy. This preliminary work suggests that MALDI and ClinPro Tools can be of utility at an early stage of an ongoing GAS outbreak once the first 1–3 isolates have been identified as clonal by molecular methods. Some reports have noted extensive allelic variation within the variable N terminal of the M protein in many *emm* types such as *emm*3, *emm*1, and *emm*28 [22]. This might explain the poor performance

of the GAS SVM model with these *emm* types. Moura et al [23] reported that GAS isolates from cases of necrotizing fasciitis were clustered together using MALDI-TOF MS and were distinct from isolates associated with noninvasive infections, despite their sharing the same *emm* type. This suggests that *emm* typing is not be the ideal “gold standard” to be used as a comparator to MALDI-TOF typing. Refining the model by choosing isolates characterized by WGS for *emm* types that demonstrate decreased clonality will improve its accuracy in cluster analysis of these *emm* types.

The GAS SVM model generation time was less than 5 minutes. For each new isolate, the hands-on time was approximately 40 minutes, which included approximately 30 minutes for tube extraction, target plate spotting, and matrix overlay and 10 minutes for spectral preparation for testing and external validation.

CONCLUSIONS

Our study has several limitations. The analysis was done retrospectively from archived isolates and its impact in an actual outbreak situation was not assessed. The gold standard comparator that we used (*emm* typing) might not be a true gold standard for some *emm* types, and the use of isolates characterized by WGS for model construction would be preferred in *emm* 1, 3, and 28 and perhaps others. It is also likely that performance would be higher if a larger number and variety of GAS isolates had been used to create the model. Despite these limitations, our results add to an increasing body of literature supporting the potential for applying proteomics to bacterial typing for a variety of organisms. Given its favorable accuracy with clonal isolates, the model has the potential of providing real-time meaningful information that can be used for early detection and management of GAS outbreaks. The low operational cost and the quick turnaround time makes it a practical tool for this purpose.

Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of

the authors, so questions or comments should be addressed to the corresponding author.

Acknowledgments

We thank Dr. Allison McGeer and the Toronto Invasive Bacterial Disease Network for providing most of the isolates used in this study.

Author contributions. M. T. contributed to developing and validating the typing method and manuscript creation. A. C. contributed to developing the typing method and manuscript review. L. M. M. and M. M. contributed to manuscript review and assistance with acquiring software for typing.

Financial support. This study was supported by the Innovation Fund of the Alternative Funding Plan for the Academic Health Sciences Centres of Ontario.

Potential conflicts of interest. All authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

References

1. Walker MJ, Barnett TC, McArthur JD, et al. Disease manifestations and pathogenic mechanisms of group A *Streptococcus*. *Clin Microbiol Rev* **2014**; 27:264–301.
2. Cunningham MW. Pathogenesis of group A streptococcal infections. *Clin Microbiol Rev* **2000**; 13:470–511.
3. Deutscher M, Schillie S, Gould C, et al. Investigation of a group A streptococcal outbreak among residents of a long-term acute care hospital. *Clin Infect Dis* **2011**; 52:988–94.
4. Facklam R, Beall B, Efstratiou A, et al. emm typing and validation of provisional M types for group A streptococci. *Emerg Infect Dis* **1999**; 5:247–53.
5. Pilon PA, Savard N, Aho J, et al. Invasive group A streptococcal infection outbreaks of type emm118 in a long-term care facility, and of type emm74 in the homeless population, Montréal, Quebec. *Can Commun Dis Rep* **2019**; 45:26–31.
6. Teatero S, McGeer A, Tyrrell GJ, et al. Canada-wide epidemic of emm74 group A *Streptococcus* invasive disease. *Open Forum Infect Dis* **2018**; 5:ofy085.
7. Page AV, Kotb M, McGeer A, et al. Systemic dysregulation of angiotensin-1/2 in streptococcal toxic shock syndrome. *Clin Infect Dis* **2011**; 52:e157–61.
8. Tyrrell G, Fathima S, Kakulphimp J, Bell C. Increasing rates of invasive group A streptococcal disease in Alberta, Canada; 2003–2017. *Open Forum Infect Dis* **2018**; 5:ofy177.
9. Athey TB, Teatero S, Sieswerda LE, et al. High incidence of invasive group A *Streptococcus* disease caused by strains of uncommon emm types in thunder bay, Ontario, Canada. *J Clin Microbiol* **2016**; 54:83–92.
10. CLSI. Methods for the Identification of Cultured Microorganisms Using Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. 1st ed. CLSI guideline M58. Wayne, PA: Clinical and Laboratory Standards Institute; 2017.
11. Nutman A, Marchaim D. How to: molecular investigation of a hospital outbreak. *Clin Microbiol Infect* **2019**; 25:688–695.
12. Griffin PM, Price GR, Schooneveldt JM, et al. Use of matrix-assisted laser desorption/ionization-time of flight mass spectrometry to identify vancomycin-resistant enterococci and investigate the epidemiology of an outbreak. *J Clin Microbiol* **2012**; 50:2918–31.
13. Schlebusch S, Price GR, Gallagher RL, et al. MALDI-TOF MS meets WGS in a VRE outbreak investigation. *Eur J Clin Microbiol Infect Dis* **2017**; 36:495–9.
14. Matsumura Y, Yamamoto M, Nagao M, et al. Detection of extended-spectrum- β -lactamase-producing *Escherichia coli* ST131 and ST405 clonal groups by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol* **2014**; 52:1034–40.
15. Fangous MS, Mougari F, Gouriou S, et al. Classification algorithm for subspecies identification within the *Mycobacterium abscessus* species, based on matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol* **2014**; 52:3362–9.
16. Nakano S, Matsumura Y, Ito Y, et al. Development and evaluation of MALDI-TOF MS-based serotyping for *Streptococcus pneumoniae*. *Eur J Clin Microbiol Infect Dis* **2015**; 34:2191–8.
17. Chen JHK, Cheng VCC, Wong CP, et al. Rapid differentiation of *Haemophilus influenzae* and *Haemophilus haemolyticus* by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry with ClinProTools mass spectrum analysis. *J Clin Microbiol* **2017**; 55:2679–85.
18. Angeletti S, Dicuonzo G, Lo Presti A, et al. MALDI-TOF mass spectrometry and blakpc gene phylogenetic analysis of an outbreak of carbapenem-resistant *K. pneumoniae* strains. *New Microbiol* **2015**; 38:541–50.
19. Xiao D, Zhao F, Zhang H, et al. Novel strategy for typing *Mycoplasma pneumoniae* isolates by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry coupled with ClinProTools. *J Clin Microbiol* **2014**; 52:3038–43.
20. Huang CH, Li SW, Huang L, Watanabe K. Identification and classification for the *Lactobacillus casei* group. *Front Microbiol* **2018**; 9:1974.
21. Nagy E, Urbán E, Becker S, et al. MALDI-TOF MS fingerprinting facilitates rapid discrimination of phylotypes I, II and III of *Propionibacterium acnes*. *Anaerobe* **2013**; 20:20–6.
22. Wang J, Zhou N, Xu B, et al. Identification and cluster analysis of *Streptococcus pyogenes* by MALDI-TOF mass spectrometry. *PLoS One* **2012**; 7:e47152.
23. Moura H, Woolfitt AR, Carvalho MG, et al. MALDI-TOF mass spectrometry as a tool for differentiation of invasive and noninvasive *Streptococcus pyogenes* isolates. *FEMS Immunol Med Microbiol* **2008**; 53:333–42.