Discrimination of *Enterobacteriaceae* and Non-fermenting Gram Negative Bacilli by MALDI-TOF Mass Spectrometry

Reiner Schaumann[1,†], Nicolas Knoop[1,†,*], Gelimer H. Genzel[1], Kevin Losensky[1], Christiane Rosenkranz[1], Catalina S. Stîngu[1], Wolfgang Schellenberger[2], Arne C. Rodloff[1] and Klaus Eschrich[2]

[1]*Institute for Medical Microbiology and Epidemiology of Infectious Diseases, University of Leipzig, Leipzig, Germany*

[2]*Institute of Biochemistry, School of Medicine, University of Leipzig, Leipzig, Germany*

Abstract: Discrimination of *Enterobacteriaceae* and Non-fermenting Gram Negative Bacilli by MALDI-TOF Mass Spectrometry

Matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) has proven to be an effective identification tool in medical microbiology. Discrimination to subspecies or serovar level has been found to be challenging using commercially available identification software. By forming our own reference database and using alternative analysis methods, we could reliably identify all implemented *Enterobacteriaceae* and non-fermenting gram negative bacilli by MALDI-TOF MS and even succeeded to distinguish *Shigella sonnei* from *Escherichia coli (E. coli)* and *Salmonella enterica* spp. *enterica* serovar Enteritidis from *Salmonella enterica* spp. *enterica* serovar Typhimurium. Furthermore, the method showed the ability to separate Enterohemorrhagic *E. coli* (EHEC) and Enteropathogenic *E. coli* (EPEC) from non-enteropathogenic *E. coli*.

Keywords: *E. coli*, EHEC, Enterobacteriaceae, EPEC, MALDI-TOF MS, *Salmonella spp.*, Shigella sonnei.

Since matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) has become a more and more effective tool for species identification in clinical microbiology [1-4], many studies tried to evaluate its performance compared to conventional phenotypic and genotypic identification systems [5-11]. Especially by using commercially available MALDI-TOF MS data analysis software, satisfying results in the discrimination of predominantly gram negative bacilli were achieved. Some authors already encountered difficulties in the discrimination of closely related species and began to develop alternative, not commercially available, analysis software.

In order to optimize reliable discrimination we evaluated the performance of MALDI-TOF MS in the identification of gram negative bacilli based on a database containing strains previously identified by the API systems (bioMérieux S.A., Marcy L'Etoile, France) and checked reproducibility in a blindfolded study with strains previously identified by the VITEK II system (bioMérieux S.A.). Both tools are widely spread standard identification systems for bacteria [12]. As already indicated above, the aim of our study was not only to discriminate between genetically distinct species, but also between closely related species like *E. coli* and *Shigella*

sonnei or *Salmonella enterica* spp. *enterica* serovars Enteritidi*s* and Typhimurium which so far need additional tests such as agglutination reactions for definite species identification. He *et al.* investigated the performance of the Biotyper software (Bruker Daltonics, Germany) in the identification of enteric bacterial pathogens and failed to distinguish *Shigella* spp. and Enterohemorrhagic *E. coli* (EHEC) from non-enteropathogenic *E. coli* [9]. Dieckmann *et al.* demonstrated a reliable method to distinguish strains of *Salmonella enterica* to the subspecies level, but not yet different serovars [5]. Karger *et al.* applied a method comparable to the one used by the aforementioned author to discriminate between shiga toxin producing serotypes of *E. coli* and encountered difficulties in finding optimal data analysis parameters allowing no reproducible and reliable discrimination [11].

Since we already achieved good results by using our own, alternative statistical analysis software in recent studies [13-15], we attempted to optimize identification rates for gram negative bacilli by MALDI-TOF MS by establishing our own database of gram negative bacilli with a total of 219 strains from 17 species and 12 genera. Fifteen species of *Enterobacteriaceae* and 2 species of non-fermenting gram negative bacilli were included (Table **1**). As proposed by Seng *et al.* [7], we established spectra of about 10 strains as a reference for each species in the database (for illustration see Fig. **1**). Those strains were clinical isolates from the routine diagnostics laboratory of our institute.

Previous identification of the implemented strains from clinical specimens was conducted in a standard manner be-

*Address correspondence to these authors at the Department for Medical Microbiology and Epidemiology of Infectious Diseases, University Hospital Leipzig, University of Leipzig, Liebigstr. 20, D-04103 Leipzig, Germany; Tel: +49 341 97 17 947; Fax: +49 341 97 17 539; E-mail: nicolas.knoop@medizin.uni-leipzig.de
†Both authors contributed equally to this work

Table 1.   Formation of Our Database

Species	Strains (n)	Spectra (n) *
Escherichia coli	19	148
Klebsiella pneumoniae	19	136
Klebsiella oxytoca	10	80
Enterobacter cloacae	19	146
Enterobacter aerogenes	9	72
Proteus mirabilis	20	152
Proteus vulgaris	10	78
Salmonella Enteritidis	10	80
Salmonella Typhimurium	10	80
Serratia marcescens	10	80
Shigella sonnei	9	71
Yersinia enterocolitica	9	70
Citrobacter freundii	10	77
Morganella morganii	8	57
Acinetobacter baumanii	10	80
Pseudomonas aeruginosa	37	219
Total	**219**	**1626**

*Multiple spectra from a total of four specimens per each single strain were acquired at two distinct dates to check reproducibility and taking strain and recording specific variations into account.
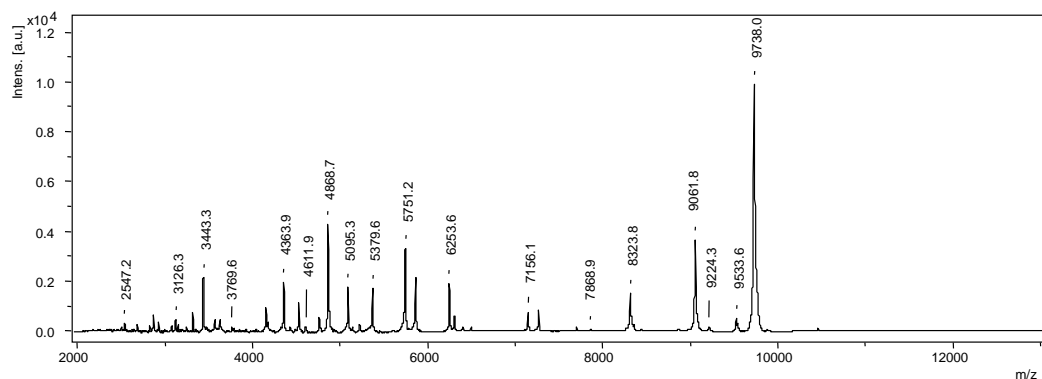
Fig. (1). Illustration of a typical MALDI-TOF MS spectrum for *E. coli*. Mass over charge ratios (m/z) are plotted with their corresponding relative intensities. By the parameters and materials we used, mass peaks were generally found up to a m/z of 10.000. Above 10.000 only few samples showed reproducible peaks.

ginning with microbiological culture which yielded pure growth of lactose non-fermenting and fermenting Gram negative rods (selective medium Endo-Agar (bioMérieux S.A.). The strains were further identified by Gram stain and key reactions, i.e. KOH, indole, oxidase as well as with commercially available API systems (API 20 E, ID 32 E, ID 32 GN and API 20 NE, bioMérieux S.A.). Organisms identified as *Salmonella* spp. (lactose negative) were subjected to subsequent serovar determination by serotyping agglutination according to the Kauffmann-White scheme using *Salmonella* antisera from Sifin (SIFIN Institut für Immun-

präparate und Nährmedien GmbH, Berlin, Germany). Any inconsistent specimens were identified genotypically by 16S rDNA amplification which was performed with the primers 609V and 699R as described by Ackermann *et al.* [16] resulting in a PCR product of approx. 300 bp. Amplification sequences were compared using the NCBI BLAST database (http://www.ncbi.nlm.nih.-gov/blast/Blast.cgi).

All strains were isolated from clinical specimens, including blood cultures, wound swabs, urine, stool and specimens from the respiratory tract, i.e. sputum, tracheal aspirate, bronchoalveolar lavage (BAL) and tissue specimens.

The following reference strains were used to validate appropriate species affiliation of the clinical samples: *E. coli* ATCC 25922, *E. coli* GSM 2163, *Klebsiella pneumoniae (K. pneumoniae)* 911 (a strain from an interlaboratory test in Germany) and *Pseudomonas aeruginosa (P. aeruginosa)* ATCC 27853.

To check the accuracy of our database and to evaluate MALDI-TOF MS discrimination potential compared to another phenotypic detection system (VITEK II), 99 strains isolated from clinical specimens (blood cultures, wound swabs, urine, stool and specimens from the respiratory tract, i.e. sputum, tracheal aspirate, broncheoalveolar lavage and tissue specimens) and routinely identified by VITEK II in our lab were subjected to MALDI-TOF MS analysis in a blindfolded manner.

In order to analyze bacterial specimens by MALDI-TOF MS, multiple colonies from a pure culture were transferred to plastic tubes and cell lysis was conducted with trifluoroacetic acid (Merck, Darmstadt, Germany). The centrifuged cell lysates were brought to MALDI-TOF MS stainless steel target plates after vacuum centrifugation and α-cyano-4-hydroxycinnamic acid (Bruker Daltonics, Bremen, Germany) was used as a matrix [15].

MALDI-TOF MS spectra were acquired with a Bruker Autoflex II mass spectrometer (Bruker Daltonics, Germany) with a nitrogen laser (λ=337nm) operated in the positive linear mode (delay 350ns, voltage 20 kV; mass over charge ratio (m/z) ranging from 2,000 to 12,000) under control of FlexControl software (version 2.4; Bruker Daltonics). Each spectrum was obtained by averaging 500 laser shots acquired in automatic mode at the minimum laser power being necessary for ionization of the samples. The calibration strain for each MALDI-TOF MS optical measurement was *E. coli* GSM 2163 from New England BioLabs. Data files were transferred to FlexAnalysis software (version 2.4; Bruker Daltonics) for automated peak extraction. With this software, 5 0 peaks were automatically labeled in each spectrum according to their appearance above the background (threshold ratio 1.5). Correct labeling was controlled manually. Peak lists containing masses and intensities were exported as Excel-files [14, 15].

The data analysis was not performed by the Biotyper system provided by Bruker Daltonics, but by MatLab based algorithms (Version 7.10.0.499; The Math-Works Inc., Natick, MA) which were edited at the Institute of Biochemistry [14, 15]. Thus, peaks from all spectra belonging to one species were aligned to species-specific peak-clusters by forming mass-dependent windows over the peaks, taking the strain- and recording-specific variations of peak masses into account. Statistical similarity analysis was carried out by hierarchical clustering and dendrograms ware calculated on the basis of distance matrices by using a complete linkage function. For closely related species, we used classification algorithms, based on support-vector-machines (SVM) featuring an efficient two class classification. This was achieved by training the algorithm with a set of spectra of bacteria of known identity, according to our reference database. An error estimate of the class prediction was carried out by calculation of a 10-fold cross-validation error for the training group. For this purpose, the training set was first divided into

10 subsets of equal size. Sequentially, one subset was tested by using the classifier trained on the remaining nine subsets. Thus, each probe of the training set was predicted once. The cross-validation accuracy is the percentage of data which were correctly classified (correct rate) [14, 15]. While establishing the reference database we soon achieved excellent results in discriminating the species from one another by using statistical similarity analysis software. SVM algorithms were necessary for the separation of *E. coli* and *Shigella sonnei* and the two salmonella serovars. In some cases *Enterobacter cloacae (E. cloacae)* clustered together with *E. coli* and *Shigella sonnei* – in those cases SVM application could also resolve the discrepancies. In order to proof our results while establishing the database we subjected the 99 strains identified by VITEK II to MALDI-TOF MS identification.

The blindfolded study showed that MALDI-TOF MS could correctly identify 95% of the strains according to the previous VITEK II identification. Four strains were misidentified: one *K. pneumoniae* as *E. coli* (S36), one *Enterobacter aerogenes (E. aerogenes)* as *Serratia marcescens (S. marcescens)* (S37), one *E. cloacae* as *E. coli* (S89), and one *Shigella sonnei* as *E. coli* (S67). The four strains were subsequently subjected to API identification and 16s rDNA sequencing to resolve the discrepancies.

By this re-testing procedure isolate S37 was identified as *S. marcescens* and isolate S67 as *Shigella sonnei*. The two remaining misidentified specimen could not be identified unequivocally and were excluded from our study. In conclusion, a correct identification rate of 99% was achieved. Even *Shigella sonnei* was correctly distinguished from *E. coli* in 4 of 5 cases and the closely related *S. enterica* spp. *enterica* serovar Enteritidis and *S. enterica* spp. *enterica* serovar Typhimurium were properly identified (Table **2**).

Table **2** illustrates our results in the identification by MALDI-TOF MS compared to the VITEK II machine. As inconsistencies have arisen, those specimens were identified by another biochemical method (API) or even by serological and genotypic methods as mentioned above. The third column illustrates the results after the second identification (re-testing). The correct identification rate is calculated on the match after the re-testing.

In addition, we evaluated the performance of our analysis software in the discrimination of non-enteropathogenic *E. coli* from its pathotypes Enteropathogenic *E. coli* (EPEC) and Enterohemorrhagic *E. coli* (EHEC), respectively.

Spectra of 12 EPEC and 7 EHEC strains, previously identified by genotypic systems as mentioned above [16] and according to Hauswaldt *et al.* [17] (also including Endo-Agar selective growth and beta-lactamase screening by Etest (AB bioMérieux, Solna, Sweden)), were acquired. Further analysis was conducted as described for the aforementioned study. Definite discrimination was achieved by using SVM algorithms, where one class was defined by the known spectra of non-enteropathogenic *E. colis* from our database and the other by those of one of the pathotypes, respectively. The classification resulted in a correct rate of 96% for EHEC and a correct rate of 97% for EPEC indicating a reliable discrimination of the two pathotypes from non-enteropathogenic *E. coli*.

Table 2.   Results of the Blindfolded Study

Species	Strains Identified by MALDI-TOF MS (n) / Strains Routinely Identified by VITEK II (n)		Strains Correctly Identified by MALDI-TOF MS After Re-testing (n)	Strains Correctly Identified by MALDI-TOF MS (%)
Escherichia coli	6/6			100
Klebsiella pneumoniae	5/6		5/5[1]	100
Klebsiella oxytoca	6/6			100
Enterobacter cloacae	5/6		5/5[1]	100
Enterobacter aerogenes	5/6		5/5[1]	100
Proteus mirabilis	6/6			100
Proteus vulgaris	7/7			100
Salmonella Enteritidis	5/5			100
Salmonella Typhimurium	6/6			100
Serratia marcescens	7/7		8/8[2]	100
Shigella sonnei	4/5		4/5	80
Yersinia enterocolitica	6/6			100
Citrobacter freundii	7/7			100
Morganella morganii	7/7			100
Acinetobacter baumanii	7/7			100
Pseudomonas aeruginosa	6/6			100
Total	95/99		27/28	99

[1]One strain was excluded due to ambiguous identification
[2]One strain was previously misidentified by VITEK II as *E. aerogenes*. The strain was identified by re-testing as *S. marcescens*, implicating now a total amount of eight *S. marcescens* in the study.

CONCLUSION

Our results show that MALDI-TOF MS has a definite potential to distinguish closely related species and subspecies of *Enterobacteriaceae* from one another. As we could already indicate for the future detection of antibiotic resistance in *Enterobacteriaceae* and *P. aeruginosa*, powerful statistical classification algorithms even manage to distinguish between closely related serovars and serotypes of *Enterobacteriaceae*.

Therefore, further improvement of data analysis software and the continuous acquisition of new reference spectra are of outstanding importance. The lack of specific reference spectra, especially of clinical isolates and genetically closely related specise, depicts an issue of the Biotyper database that other authors [5, 9] have encountered before. The variation of parameters in the statistical analysis also constitutes a burden for the implementation of MALDI-TOF MS in routine diagnostics for closely related serotypes or serovars of one species as indicated by Karger *et al*. for shiga toxin producing *E. coli* serotypes [11]. Thus it is preferable to conjoin multi-central acquired reference spectra and experience in the statistical analysis of the latter to form a reliable and powerful identification tool for routine microbiological diagnostics.

CONFLICT OF INTEREST

The authors have no conflicts of interest to report.

REFERENCES

[1]   Holland RD, Wilkes JG, Rafii F, *et al*. Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 1996; 10: 1227-32.
[2]   Krishnamurthy T, Ross PL, Rajamani U. Detection of pathogenic and non-pathogenic bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 1996; 10: 883-8.
[3]   Demirev PA, Ho YP, Ryzhov V, Fenselau C. Microorganism identification by mass spectrometry and protein database searches. Anal Chem 1999; 71: 2732-8.
[4]   Fenselau C, Demirev PA. Characterization of intact microorganisms by MALDI mass spectrometry. Mass Spectrom Rev 2001; 20: 157-71.
[5]   Dieckmann R, Helmuth R, Erhard M, Malorny B. 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 2008; 74: 7767-78.

[6]   Mellmann A, Cloud J, Maier T, *et al.* Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass spectrometry in comparison to 16S rRNA gene sequencing for species identification of nonfermenting bacteria. J Clin Microbiol 2008; 46: 1946-54.

[7]   Seng P, Drancourt M, Gouriet F, *et al.* Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 2009; 49: 543-51.

[8]   Cherkaoui A, Hibbs J, Emonet S, *et al.* Comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level. J Clin Microbiol 2010; 48: 1169-75.

[9]   He Y, Li H, Lu X, Stratton CW, Tang Y. Mass spectrometry biotyper system identifies enteric bacterial pathogens directly from colonies grown on selective stool culture media. J Clin Microbiol 2010; 48: 3888-92.

[10]   Benagli C, Rossi V, Dolina M, Tonolla M, Petrini O, Ahmed N. Matrix-assisted laser desorption ionization-time of flight mass spectrometry for the identification of clinically relevant bacteria. PLoS ONE 2011; 6: e16424.

[11]   Karger A, Ziller M, Bettin B, Mintel B, Schares S, Geue L. Determination of serotypes of Shiga toxin-producing *Escherichia coli*

isolates by intact cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microbiol 2011; 77: 896-905.

[12]   O'Hara CM. Manual and automated instrumentation for identification of *Enterobacteriaceae* and other aerobic gram-negative bacilli. Clin Microbiol Rev 2005; 18: 147-62.

[13]   Rupf S, Breitung K, Schellenberger W, Merte K, Kneist S, Eschrich K. Differentiation of mutans streptococci by intact cell matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Oral Microbiol Immunol 2005; 20: 267-73.

[14]   Friedrichs C, Rodloff AC, Chhatwal GS, Schellenberger W, Eschrich K. Rapid identification of viridans streptococci by mass spectrometric discrimination. J Clin Microbiol 2007; 45: 2392-97.

[15]   Schaumann R, Knoop N, Genzel GH, *et al.* A step towards the discrimination of beta-lactamase-producing clinical isolates of *Enterobacteriaceae* and *Pseudomonas aeruginosa* by MALDI-TOF mass spectrometry. Med Sci Monit 2012; 18: MT71-7.

[16]   Ackermann G, Tang YJ, Kueper R, *et al.* Resistance to moxifloxacin in toxigenic *Clostridium difficile* isolates is associated with mutations in gyrA. Antimicrob Agents Chemother 2001; 45: 2348-53.

[17]   Hauswaldt SI, Fickweiler K, Rodloff AC. Enteropathogene *E. coli*: Epidemiologie in Sachsen und Evaluation diagnostischer Methoden. Mikrobiologie 2009; 16: 206-16.