Areas of uncertainties by using MALDI-TOF and problems to solve

Adrian Egli, MD PhD
Head of Division of Clinical Microbiology, University Hospital Basel
Research Group Leader, Department of Biomedicine, University of Basel

15.03.2017

Agenda

- Quality of spectra
- Database
- Harmonization of protocols
- Bioinformatics

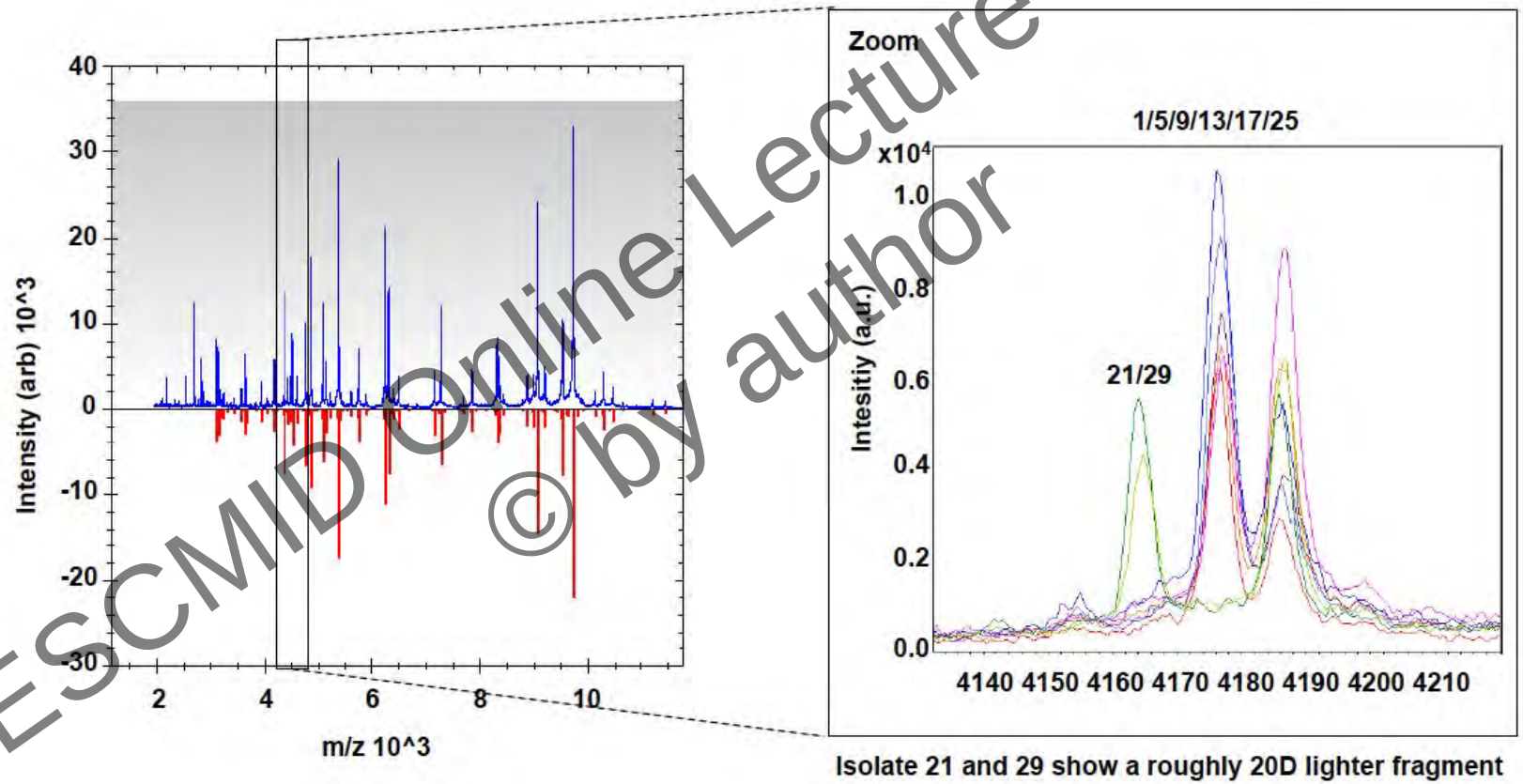
Quality of spectra

- High quality spectra allows
  - High reproducibility of data
  - Subtyping of species e.g. within a complex, sequence types
  - Eventually typing during outbreaks

- Dependent on
  - Colony age[1]
  - Culture conditions[2]
  - Protein extractions[3]
  - Colony density[1]

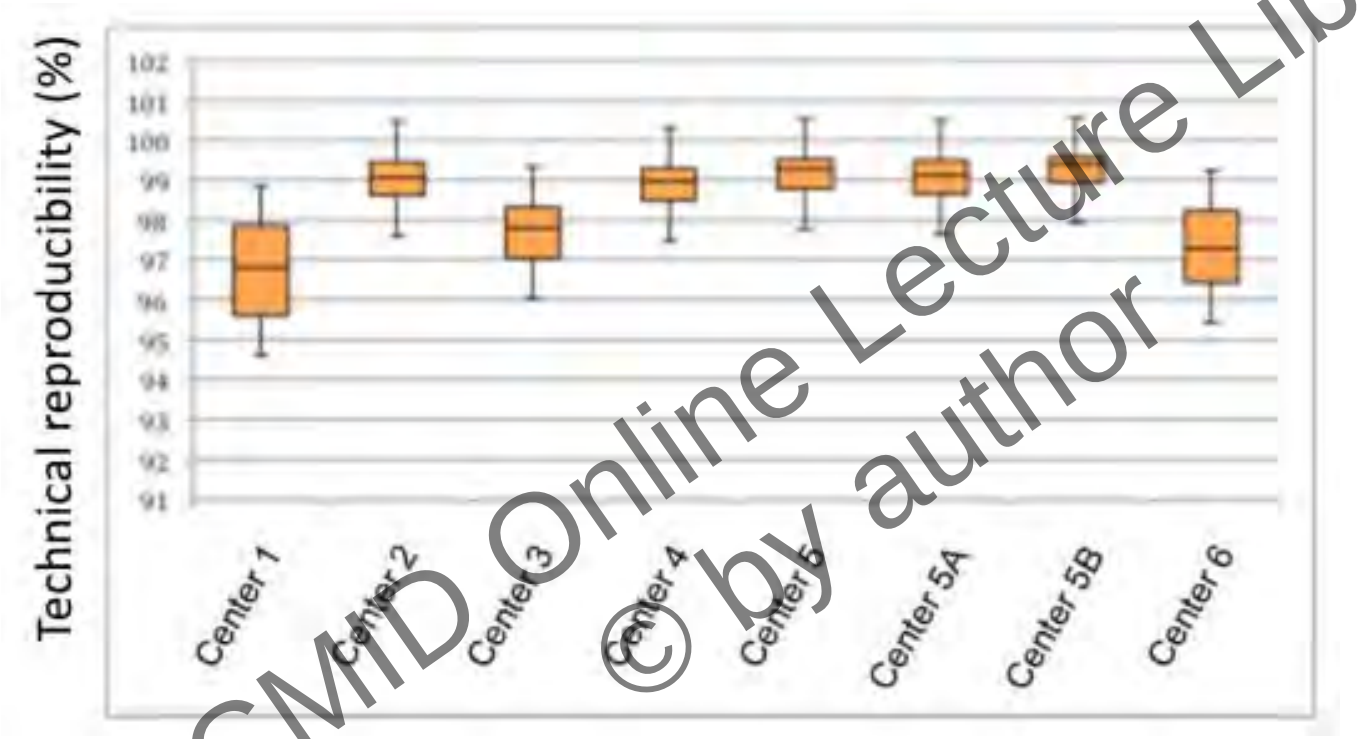
[1] Croxatto et al. FEMS Microbiology 2012; [2] Goldstein JE et al. Letters in App Micro 2013
[3] Cassagne C et al. Mycoses 2016

Even shift of single amino acids may be deteceted if there is a change in mass



Egli et al. Plos one 2014

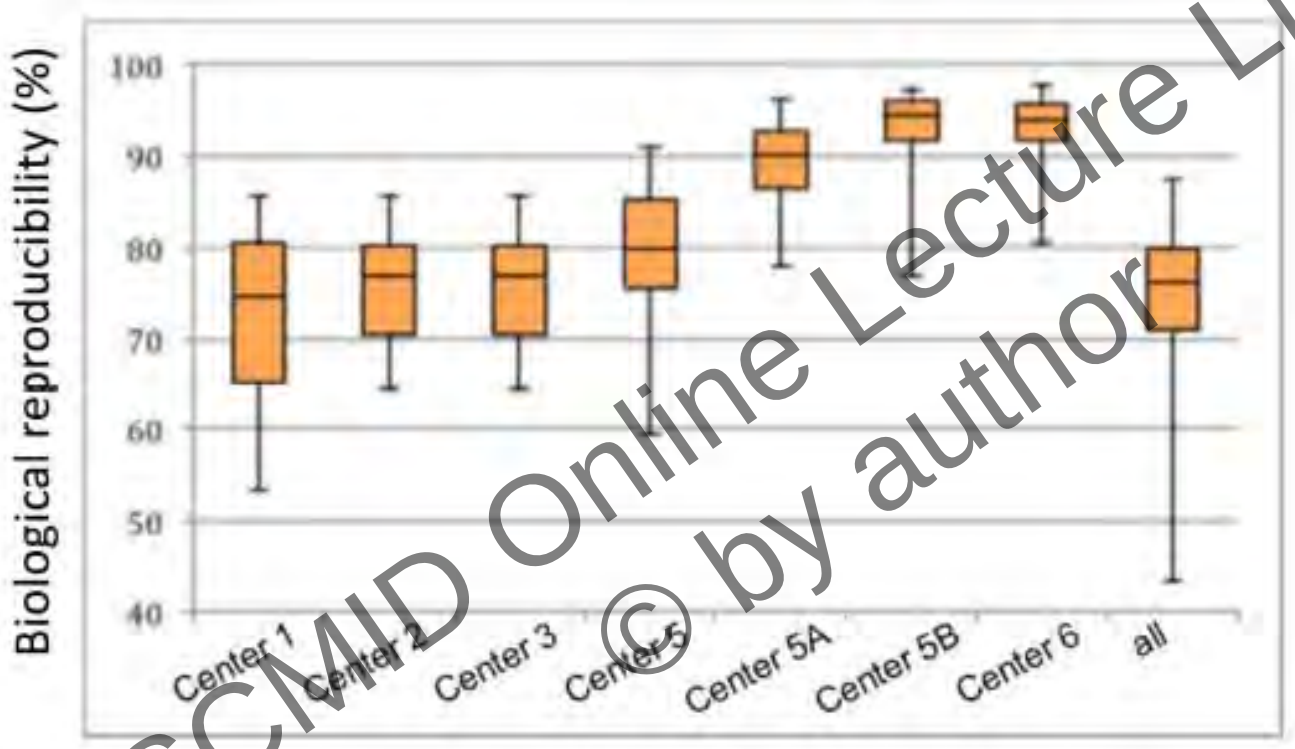
Technical reproducibility



12 ESBL-producing *E. coli* strains from 2 outbreaks and 2 non-related isolates.
Technical reproducibility: 4 spectra per isolate.

Oberle et al. Plos one 2016

Biological reproducibility: high variability



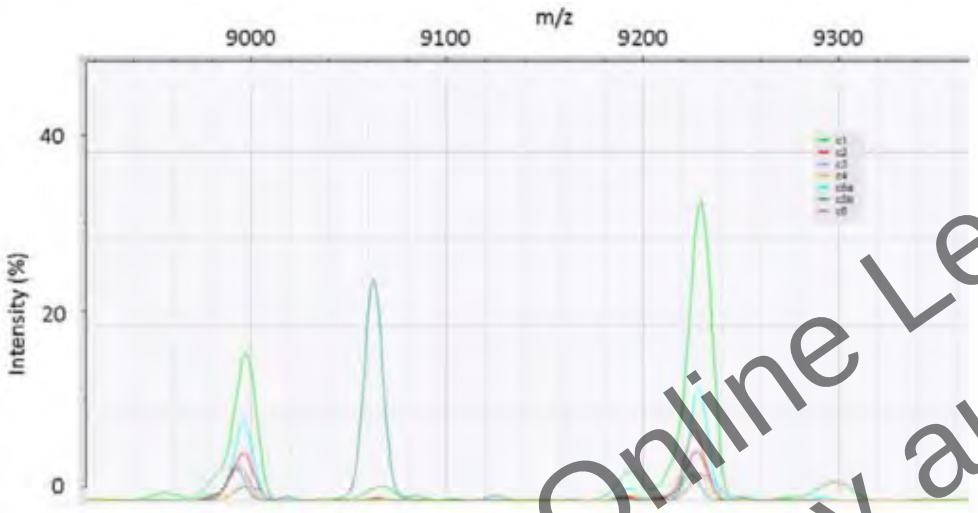
12 ESBL-producing *E. coli* strains from 2 outbreaks and 2 non-related isolates.
Biological reproducibility: 3 different days

Oberle et al. Plos one 2016

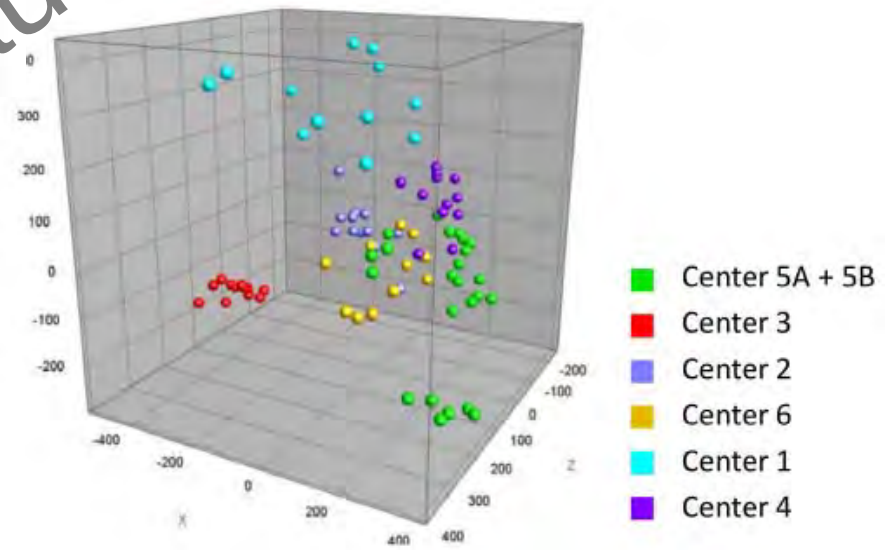
The consequence of variability between centers

- Identification of genus or species should be possible
- High resolution peak profiles allows more detailed analysis
  - Subspecies
  - Typing
- Low technical and biological reproducibility
  - Quality indicator for laboratory specific work-flows

Inter-center Reproducibility



Center dependent clustering of data



12 ESBL-producing *E. coli* strains from 2 outbreaks and 2 non-related isolates.

Oberle et al. Plos one 2016

Steps to improve data quality

- Standard operating procedures
  - Same workflow[1,2]
  - Same media[1,2]
- Same age of cultures if possible
  - Not more than 24h -> few exceptions slow growing bacteria
- More laser shots?
- Not from chromogenic agar
- If uncertain about a species – trust your instinct
  - Biochemical or sequence confirmation

[1] Egli A et al. Plos one 2015; [2] Oberle M et al. Plos one 2016

Commercial vs. homebrew database

- Most labs use commercial-available databases
  - Pros:
    - Standardized approach
    - Less labor intensive
    - Easy start
  - Cons:
    - Problems with limited spectra information
    - Dependence in case of rare or new pathogens -> wait for the update
- Curration of a high quality database is very expensive
  e.g. comparison to gold standard
  -> 16S sequencing or whole genome sequencing

From the beginning most common genus and species could be identified

TABLE 2. Retrospective analysis of identification of bacteria and yeasts by MALDI-TOF MS

Organisms	Sample data (no.)			MALDI-TOF MS analysis (no. [%])[a]					
	Isolates	Genera	Species	Genus correct	Species correct	Major error	Minor error	No ID	No uniform ID
Enterobacteriaceae	89	9	23	89 (100)	86 (96.6)	0 (0)	3 (3.4)	0 (0)	0 (0)
Nonfermentatitve Gram-negative bacteria	55	7	9	45 (81.8)	41 (74.5)	2 (3.6)	1 (1.8)	2 (3.6)	6 (10.9)
Gram-positive cocci	87	6	26	85 (97.7)	70 (80.5)	0 (0)	1 (1.1)	2 (2.3)	0 (0)
Miscellaneous bacteria	77	7	13	73 (94.8)	65 (84.4)	1 (1.3)	2 (2.6)	0 (0)	1 (1.3)
Yeasts	19	2	8	19 (100)	18 (94.7)	0 (0)	0 (0)	0 (0)	0 (0)
Total	327	29	80	311 (95.1)	280 (85.6)	3 (0.9%)	7 (2.1)	6 (1.8)	7 (2.1)

[a] MALDI-TOF MS identifications were compared to final identification (ID) based on biochemical phenotypic methods and/or 16S sequencing. Isolates were tested in duplicate for MALDI-TOF MS identification. A major error occurred when the genus was incorrect; a minor error occurred when the genus was correct but the species was incorrect. No uniform results between duplicate results were scored.

Van Veen SW et al. JCM 2010

From the beginning most common genus and species could be identified

TABLE 2. Retrospective analysis of identification of bacteria and yeasts by MALDI-TOF MS

Organisms	Sample data (no.)			MALDI-TOF MS analysis (no. [%])[a]					
	Isolates	Genera	Species	Genus correct	Species correct	Major error	Minor error	No ID	No uniform ID
Enterobacteriaceae	89	9	23	89 (100)	86 (96.6)	0 (0)	3 (3.4)	0 (0)	0 (0)
Nonfermentatitve Gram-negative bacteria	55	7	9	45 (81.8)	41 (74.5)	2 (3.6)	1 (1.8)	2 (3.6)	6 (10.9)
Gram-positive cocci	87	6	26	85 (97.7)	70 (80.5)	0 (0)	1 (1.1)	2 (2.3)	0 (0)
Miscellaneous bacteria	77	7	13	73 (94.8)	65 (84.4)	1 (1.3)	2 (2.6)	0 (0)	1 (1.3)
Yeasts	19	2	8	19 (100)	18 (94.7)	0 (0)	0 (0)	0 (0)	0 (0)
Total	327	29	80	311 (95.1)	280 (85.6)	3 (0.9%)	7 (2.1)	6 (1.8)	7 (2.1)

[a] MALDI-TOF MS identifications were compared to final identification (ID) based on biochemical phenotypic methods and/or 16S sequencing. Isolates were tested in duplicate for MALDI-TOF MS identification. A major error occurred when the genus was incorrect; a minor error occurred when the genus was correct but the species was incorrect. No uniform results between duplicate results were scored.

The problem are rare pathogens with few spectra in the database

Van Veen SW et al. JCM 2010

Commercial database: misidentifications

Table 3. Misidentification by both systems when the microorganisms is referenced in the database.

		Reference identification	Proposed ID (confidence value)[a]
Bruker Biotyper	**Misidentification to species level**	*Neisseria subflava* biovar *flava*	*Neisseria flavescens* (2.013)
		Neisseria subflava biovar *flava*	*Neisseria flavescens* (2.047)
		Neisseria subflava biovar *perflava*	*Neisseria flavescens* (2.036)
		Neisseria subflava biovar *subflava*	*Neisseria flavescens* (2.129
		Neisseria sicca	*Neisseria mucosa* (2.055))
		Aeromonas hydrophila	*Aeromonas caviae* (2.356)
		Burkholderia cepacia	*Burkholderia multivorans* (2.377)
		Pseudomonas pseudoalcaligenes	*Pseudomonas oleovorans* (2.2)
		Shewanella algae	*Shewanella putrefaciens* (2.143)
		Corynebacterium pseudodiphtheriticum	*Corynebacterium propinquum*(2.303)
		Staphylococcus carnosus	*Staphylococcus condimenti* (2.0)
		Streptococcus australis	*Streptococcus parasanguinis* (2.028)
		Streptococcus australis	*Streptococcus parasanguinis* (2.011)
		Streptococcus infantarius	*Streptococcus lutetiensis* (2.021)
		Streptococcus infantis	*Streptococcus peroris* (2.0)
	Misidentification to genus level	*Campylobacter hyointestinalis*	*Pandoraea sputorum* (2.213)
		Rahnella aquatilis	*Ewingella americana* (1.872)
		Actinomyces naeslundii	*Clostridium halophilum* (1.828)
		Bacillus sp.	*Corynebacterium tuberculostearicum* (2.02)
		Microbacterium lacticum	*Arthrobacter castelli* (2.154)
		Candida freyschussii	*Rhodotorula bogoriensis* (2.2)
VITEK MS (IVD)	**Misidentification to species level**	*Neisseria cinerea*	*Neisseria subflava* (99.9)
		Burkholderia cepacia	*Burkholderia multivorans* (99.9)
		Fusarium proliferatum	*Fusarium oxysporum* (97.6)
	Misidentification to genus level	*Campylobacter hyointestinalis*	*Comamonas testosteroni* (99.7)
		Psychrobacter sp.	*Eubacterium limosum* (99.8)
		Corynebacterium sp.	*Sphingobium xenophagum*(99.8)
		Microbacterium aurum	*Paenibacillus durus* (86.0)
		Staphylococcus hominis	*Kocuria kristinae* (99.9)
		Gordonia bronchialis	*Clavibacter michiganensis* (99.8)
		Mycobacterium avium	*Paenibacillus durus* (99.9)

[a]Bruker Biotyper confidence value is up to 3.0, and VITEK MS is up to 100.

Lévesque S et al. Plos one 2015

Commercial database: misidentifications

Table 3. Misidentification by both systems when the microorganisms is referenced in the database.

		Reference identification	Proposed ID (confidence value)[a]
Bruker Biotyper	Misidentification to species level	*Neisseria subflava* biovar *flava*	*Neisseria flavescens* (2.013)
		Neisseria subflava biovar *flava*	*Neisseria flavescens* (2.047)
		Neisseria subflava biovar *perflava*	*Neisseria flavescens* (2.036)
		Neisseria subflava biovar *subflava*	*Neisseria flavescens* (2.129
		Neisseria sicca	*Neisseria mucosa* (2.055))
		Aeromonas hydrophila	*Aeromonas caviae* (2.356)
VITEK MS			
		Fusarium proliferatum	*Fusarium oxysporum* (97.6)
	Misidentification to genus level	*Campylobacter hyointestinalis*	*Comamonas testosteroni* (99.7)
		Psychrobacter sp.	*Eubacterium limosum* (99.8)
		Corynebacterium sp.	*Sphingobium xenophagum*(99.8)
		Microbacterium aurum	*Paenibacillus durus* (86.0)
		Staphylococcus hominis	*Kocuria kristinae* (99.9)
		Gordonia bronchialis	*Clavibacter michiganensis* (99.8)
		Mycobacterium avium	*Paenibacillus durus* (99.9)

[a]Bruker Biotyper confidence value is up to 3.0, and VITEK MS is up to 100.

Both commerical databases:

- Misidentification to species level
  e.g. *S. australis* -> *S. parasanguinis*

- Misidentification to genus level
  e.g. *Corynebacterium sp.* -> *Sphingobium xenophagum*

Improvement of the database is key for high quality diagnostics

Lévesque S et al. Plos one 2015

Database: comparison

Table 5. Identification results when the microorganism is not present in the Bruker Biotyper database.

Reference identification	Number of isolates	Bruker Biotyper				Proposed IDs (confidence value)
		Misidentification to the species	Correct ID to the level of		No ID	Mis ID	
			Genus	Complex/group			
Total number of strains (%)	70	2 (2.9)	10 (14.3)	2 (2.9)	52 (74.2)	4 (5.7)	

No ID = No identification obtained. Mis ID = Misidentification obtained

Lévesque S et al. Plos one 2015

Database: comparison

Table 5. Identification results when the microorganism is not present in the Bruker Biotyper database.

Reference identification	Number of isolates	Bruker Biotyper				Proposed IDs (confidence value)
		Misidentification to the species	Correct ID to the level of		No ID	Mis ID	
			Genus	Complex/group			
Total number of strains (%)	70	2 (2.9)	10 (14.3)	2 (2.9)	52 (74.2)	4 (5.7)	

No ID = No identification obtained. Mis ID = Misidentification obtained

No ID

Lévesque S et al. Plos one 2015

Database: comparison

Table 5. Identification results when the microorganism is not present in the Bruker Biotyper database.

Reference identification	Number of isolates	Bruker Biotyper				Proposed IDs (confidence value)
		Misidentification to the species	Correct ID to the level of		No ID	Mis ID	
			Genus	Complex/group			
Total number of strains (%)	70	2 (2.9)	10 (14.3)	2 (2.9)	52 (74.2)	4 (5.7)	

No ID = No identification obtained. Mis ID = Misidentification obtained

No ID

Table 6. Identification results when the microorganism is not present in the VITEK MS IVD database.

Reference identification	Number of isolates	VITEK MS (IVD)				Proposed IDs (confidence value [%])
		Misidentification to the species	Correct ID to the level of		No ID	Mis ID	
			Genus	Complex/group			
Total number of strains (%)	178	29 (16.2)	6 (3.4)	6 (3.4)	97 (54.5)	40 (22.5)	

Wrong ID

Lévesque S et al. Plos one 2015

Homebrew fungal database outcompetes commerical database

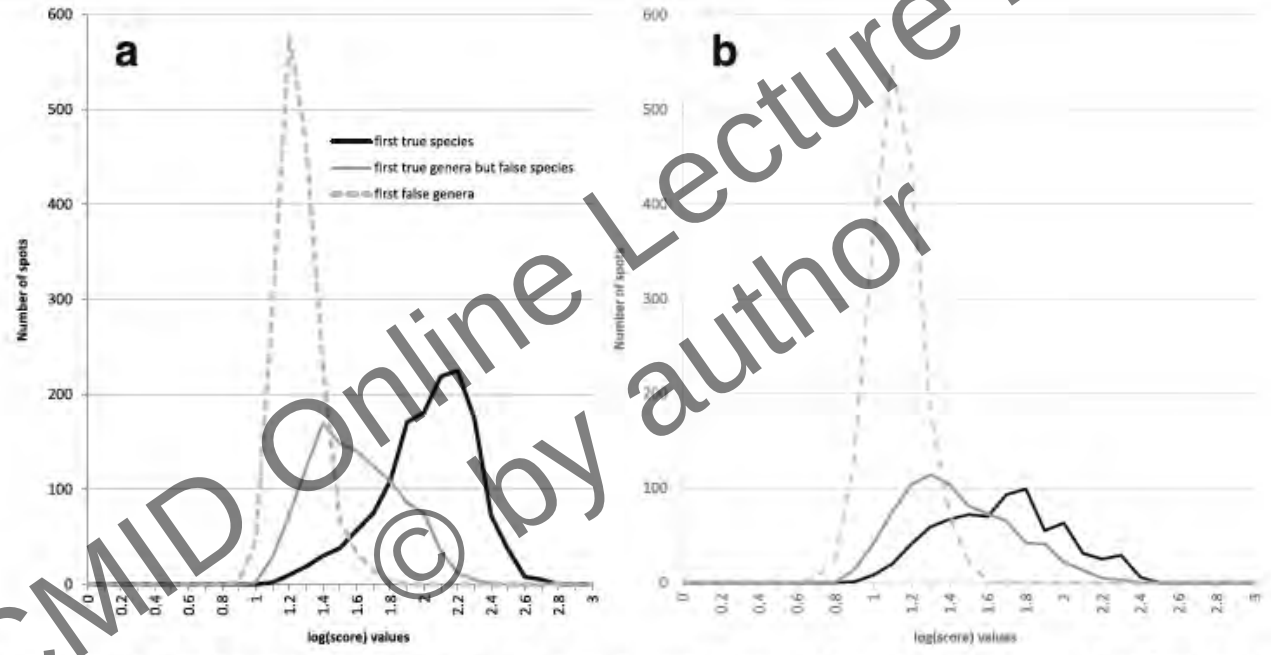


Fig. 1 Highest log(score) value distribution of the 1688 tested spectra. **a** considering the first accurate species identification (*dark line*), the first accurate genus identification (*gray line*), and the first false genus identification (*dotted line*) using the in-house reference database. **b** considering the first accurate species identification (*dark line*), the first accurate genus identification (*gray line*), and the first false genus identification (*dotted line*) using the Bruker reference database

Normand AC et al. BMC Microbiology 2017

Specific analytic algorithm for fungal species identification

Table 9 Comparison of the identification efficiency of the two databases per species, using the D1 algorithm of identification with a 1.7 LS threshold, for 180 strains that are represented by three or more strains in both reference databases

		In-House database		Bruker database	
	nb of isolates	% accepted	species PPV	% accepted	species PPV
Absidia corymbifera	2	100.00	1.00	50.00	1.00
Alternaria alternata	2	100.00	0.50		
Aspergillus flavus	15	100.00	1.00		
Aspergillus fumigatus	20	95.00	1.00		
Aspergillus nidulans	8	100.00	1.00		
Aspergillus niger	4	100.00	1.00	50.00	1.00
Aspergillus parasiticus	2	100.00	1.00	0.00	Not Relevant
Aspergillus terreus	10	100.00	1.00	30.00	1.00
Aspergillus versicolor	3	66.67	1.00	33.33	1.00
Exophiala dermatitidis	2	50.00	1.00	50.00	1.00
Fusarium oxysporum	21	80.95	1.00	19.05	1.00
Fusarium proliferatum	10	100.00	1.00	60.00	0.33
Fusarium solani	7	57.14	0.25	28.57	1.00
Microsporum canis	2	100.00	1.00	50.00	1.00
Purpureocillium lilacinus	3	66.67	1.00	66.67	1.00
Paecilomyces variotii	3	66.67	1.00	66.67	1.00
Penicillium chrysogenum	27	100.00	0.96	33.33	1.00
Scedosporium apiospermum	11	100.00	0.82	81.82	1.00
Schizophyllum commune	4	100.00	1.00	100.00	1.00
Scopulariopsis brevicaulis	15	86.67	1.00	46.67	0.71
Trichophyton interdigitale	1	100.00	1.00	0.00	Not Relevant
Trichophyton rubrum	8	87.50	0.86	37.50	0.67
Total Isolates	180	91.67	0.95	47.22	0.92

Nb of isolates : number of isolates from the panel of strain that belong to each species. % accepted: percentage of submitted strains that fulfilled the criteria for identification at either the genus or species level; species PPV: positive predictive value at the species level

D_1: Only the identification corresponding to the highest of the four scores is taken into account, plotted and categorized.

91.7% vs. 47.2%

Reliability of cut-off values

- What is the best cut-off to use for sensitive and specific identification?
  - Meta-analysis with 11 studies from blood culture[1]
    - MALDI-TOF Score ≥2.0: Sensitivity 74.6%, Specificity: 88.0%
    - MALDI-TOF Score ≥1.7: Sensitivity 92.8%, Specificity: 81.2%

Table 1 Performance of MALDI-TOF MS identification of 149 monomicrobial blood cultures by in-house method A and in-house method B compared with definitive identification

Definitive identification (with number of isolates)		In-house method A with recommended and reduced cut-off score values								In-house method B with recommended and reduced cut-off score values							
		>2.0	>1.7 < 2.0	<1.7	Concordant id	>1.7	>1.5 < 1.7	<1.5	Concordant id	>2.0	>1.7 < 2.0	<1.7	Concordant id	>1.7	>1.5 < 1.7	<1.5	Concordant id
Total G + and G - bacteria	149	75	47	7	122	122	12	15	133	39	59	52	98	98	18	33	116
		50.3 %	31.5 %	18.1 %	81.9 %	81.9 %	8.05 %	10 %	89.3 %	26.2 %	39.6 %	34.9 %	65.8 %	65.8 %	12 %	22.1 %	77.8 %

>2.0 = 50.3% >1.7 = 81.9%

[1] Scott JS et al. Infectious Dis 2016; [2] Jakovljev A et Bergh K, BMC Microbiologogy 2015

Reliability of cut-off values: Experiences from Basel – a few examples

- Every “new” from MALDI-TOF identification has to be confirmed with an additional independent method – most commonly 16S sequencing

- Critical IDs – few examples:
  - *Pantoea ananatis/calida/dispersa/septica* only genus
  - *Serratia ureilytica* -> report as *S. marcescens* as only 1 entry
  - *Salmonella sp.* -> only genus, serotyping required
  - *Burkholderia cepacia* complex members -> report as complex
  - *Pseudomonas fluorescens* and *P. putida* group -> report as group
  - *Staphylococcus intermedius* -> no reporting on species level
  - *Streptococcus mitis, S. mutans, S. bovis* group -> report as group
  - *S. pneumoniae* -> use additonal biochemical testing!
  - *B. cereus* group -> culture morphology!
  - … many more!

Improvement of Databases e.g. *Mycobacteria spp*.

TABLE 3 Impact of database update on identification of clinical isolates

Mycobacterium species	No. of isolates	No. of references in:		Mean (range) log(score) value(s) with:	
		v2.0	v3.0	v2.0	v3.0
M. abscessus	13	11	24	2.036 (1.815–2.183)	2.059 (1.830–2.209)
M. arupense	1	1	8	1.918	2.067
M. avium	17	10	39	1.872 (1.587–2.222)	1.963 (1.738–2.298)
M. bohemicum	1	3	10	1.903	2.048
M. chelonae	6	10	19	2.058 (1.989–2.208)	2.084 (2.017–2.208)
M. europaeum	1	1	2	1.847	1.847
M. fortuitum	17	9	20	2.052 (1.648–2.393)	2.160 (1.806–2.393)
M. gordonae	8	10	23	1.824 (1.543–2.123)	1.904 (1.737–2.226)
M. haemophilum	1	1	6	1.891	2.058
M. insubricum	1	3	3	2.018	2.018
M. chimaera-M. intracellulare group	6	9	32	1.971 (1.705–2.137)	2.105 (1.939–2.170)
M. kansasii	8	14	26	1.927 (1.636–2.243)	1.992 (1.636–2.243)
M. lentiflavum	5	3	13	2.063 (2.014–2.132)	2.175 (2.043–2.308)
M. mageritense	2	2	4	2.146 (2.091–2.207)	2.249 (2.214–2.310)
M. malmoense	1	5	14	2.233	2.233
M. marinum	1	8	18	2.345	2.345
M. mucogenicum-M. phocaicum group	2	3	21	2.218/2.101[a]	2.362/2.127[a]
M. palustre	1	2	2	2.086	2.086
M. peregrinum	2	1	15	2.349/2.155[b]	2.349/2.299[b]
M. porcinum	1	2	8	2.150	2.361
M. shimoidei	1	1	6	2.225	2.225
M. simiae	1	3	13	2.166	2.166
M. smegmatis	1	5	9	2.034	2.034
M. szulgai	1	5	15	2.125	2.199
M. triplex	1	1	2	1.933	2.019
M. xenopi	4	5	18	1.852 (1.485–2.099)	2.093 (1.947–2.237)
Total	104	128	370		

[a] Score values for the two representative *M. mucogenicum-M. phocaicum* group strains.
[b] Score values for the two representative *M. peregrinum* strains.

No of references:
128 -> 370

Increase of the MALDI-TOF score
v2.0 -> 2.044
v3.0 -> 2.085

Rodriguez-Sanchez B et al. JCM 2016

Updates of commerical databases… What is important to ask?

- Which species were updated?
- How many new species?
- Reclassification?
- How many isolates per species?
- You better do some reading – the infectious diseases specialist will call you within hours… e.g. *Klebsiella variicola*

What else is on the (MALDI) spot?

- **Antibiotic resistance testing**
  - Huge development in recent years[1,2]
  - Direct testing of enzyme activity -> fast turnaround time
  - Chance for laboratory automation

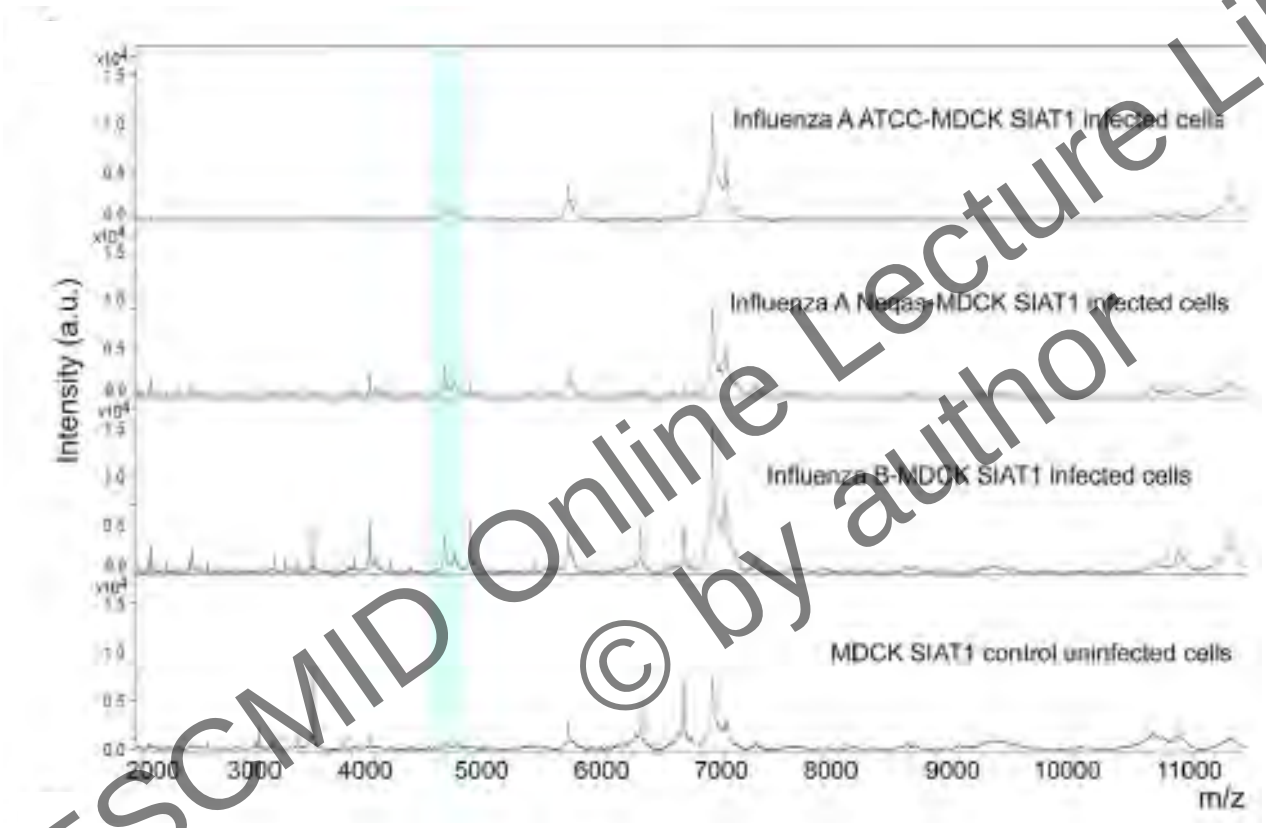
  - Gap 1: no epidemiological data on the type of carabapenemases
  - Gap 2: Oxa48 could be challenging
  - Gap 3: direct from specimen -> low sensitivity

  - Induction of resistance mechanisms with specific media
    - e.g. oxacillin -> MRSA
  - More validation studies needed with multiple clinical isolates

- What about other pathogens?

[1] Papagiannitsis CC et al JCM 2015; [2] Hrabak J et al, Methods Mol Biol 2015

Virus detection using MALDI-TOF



Cell culture infected with viruses – measurement of specific peaks

Calderaro A et al. Scientific report 2016

Where the companies should invest… (my personal opinion)

- High quality database
  - High quality spectra, with clinical and geographical diversity
  - More isolate per strain spectra
  - Sequence types associated with e.g. resistance

- Bioinformatic analysis -> this is where the money is…
  - Mixed infection algorithm
  - Automatic detection of similar peak profiles
  - Tool boxes for user driven research

- Data mining tools for research
  - Better access to data

- Improved kits for MALDI-TOF directly from patient material e.g. urine, CSF, joint puncture

- Fungal database for dermatophytes

Where the companies should not invest… (my personal opinion)

- (Troublesome) antibiotic resistance
  - Special isotope labeled media
  - Too labor intensive to detect, multiple controls
  - Phenotypic vs. genotypic assays should be critically evaluated

  - Rather go for peak detection e.g. *B. fragilis* with Carbapenem resistance, implementation into lab flows e.g. automation with cleavage protocols

Where the community should invest

- Networking and sharing experiences
- Sharing database updates
- Protocols
- A common platform
- …

Summary and conclusion

- MALDI-TOF revolutionized the identification workflow, however, its full potential has still to be explored.

- Improved quality of spectra will help to
  - Improve identification down to subspecies levels
  - May even lead to typing applications

- Database content is key for routine and research!

Thank you for your attention!

Adrian Egli, MD PhD

Clinical Microbiology

University Hospital Basel

Email: adrian.egli@usb.ch

Phone: +41 61 556 57 49

ESGMD ESCMID STUDY GROUP FOR GENOMIC AND MOLECULAR DIAGNOSTICS

European Society of Clinical Microbiology and Infectious Diseases

Become a member now!