# Analysis of MS Data in Clinical Microbiology

## Polymicrobial Identification

## Accessing and Analyzing Clinical Spectral Data

*Capacity-building Workshop*
*Applications of MALDI-TOF Mass-Spectrometry (MS) in Clinical Microbiology*

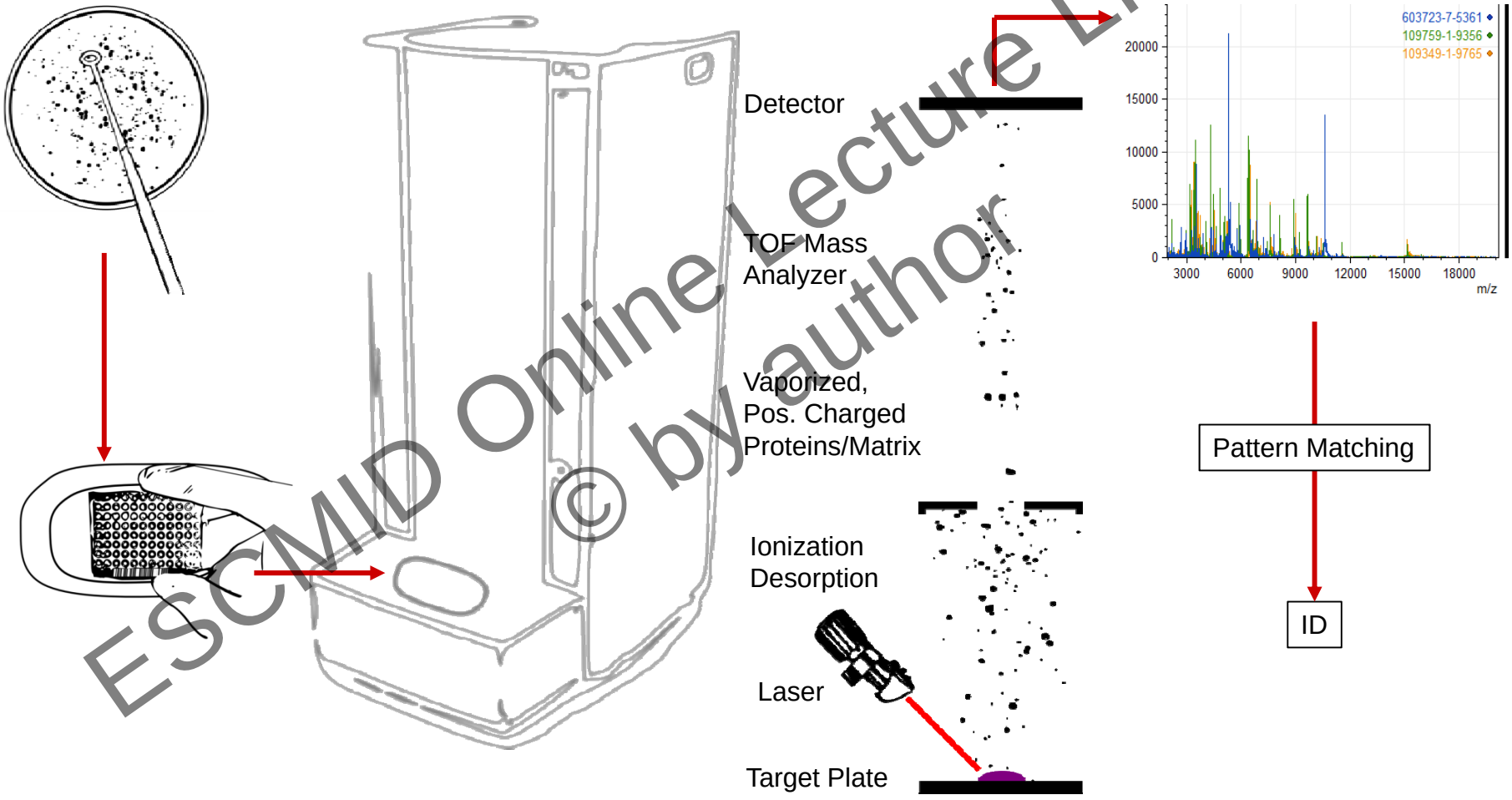
Christian Gehringer

MD, M.Sc. (ETH), Dr. rer. medic
Internal Medicine / Clinical Microbiology
University Hospital Basel
Petersgraben 4
CH-4031 Basel
+41 61 32 85430
christian.gehringer@usb.ch

# Overview

- Microbial identification by MALDI-TOF MS in Clinical Microbiology
- Polymicrobial identification of blood cultures with the Sepsityper module
- How to access your spectral data from MBT
- Tools for spectral data analysis
- Examples of clusters and groups

# Workflow: MALDI-TOF MS in Clinical Microbiology



# MALDI-TOF MS in a Clinical Laboratory

- Mass spectrometry is fast, cost-efficient[1] and specific for microbial identification

| | MBT MSP Library[2] | VITEK MS™V3 DB[3] |
|---|---|---|
| Bacteria | 2,226 | 882 |
| Yeast and molds | 235 | 164 |
| MSP (MBT) or spectra (VITEK) | 6,903 | 38,428 |
| MSP/spectra per species | 2.8 | 36.7 |

## Some current challenges

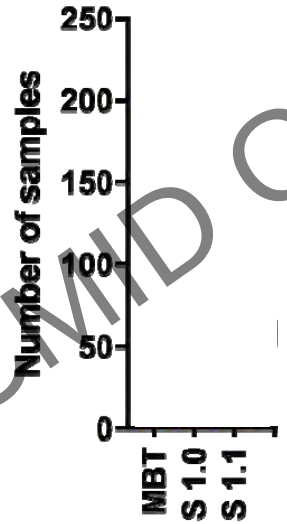
- Discrimination between phylogenetically related species not always meets clinical needs[4]
- Source material from microbial cultures only

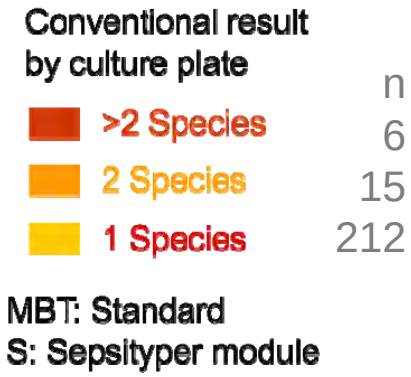


1) Dixon, P., et al. (2015). *Eur J Clin Microbiol Infect Dis* **34, 863-876**.
2) MBT 6903 MSP Library, Apr 2016, Bruker
3) Vitek MS ™ V3 Database, 2017, accessed biomerieux-diagnostics.com/vitek-ms, bioMérieux
4) Croxatto, A.*, et al. (2012). FEMS Microbiol Rev* ***36, 380-407.***

# Identification of Mixed Infections

- Mixed infections are present in approx. 10% of blood culture samples
- MALDI-TOF MS directly from source material or blood culture allows identification of mixed infections
- Biotyper® Sepsityper module promises mixed infection algorithm






27th ECCMID, Poster:1303, Manuscript in preparation

| Conventional microbiological result from culture plates | | Identifications by direct MALDI-TOF MS from positive blood cultures Correct / incomplete[1] / incorrect[2] / no identification[3] | | |
|---|---|---|---|---|
| | | Standard algorithm | Mixed 1.0 | Mixed 1.1 |
| **1 Species** | 212 | 193 / 0 / 0 / 19 | 154 / 0 / 42 / 16 | 195 / 0 / 1 / 16 |
| **2 Species** | 15 | 0 / 15 / 0 / 0 | 7 / 4 / 4 / 0 | 2 / 13 / 0 / 0 |
| **>2 Species** | 6 | 0 / 5 / 0 / 1 | 0 / 5 / 1 / 0 | 0 / 6 / 0 / 0 |

1: Reporting of correct identification not covering all species in a polymicrobial sample.
2: Reporting of species that are not present in the sample (additionally to correct identifications).
3: Defined as scores < 1.7 independently of the identified organism.

# Conclusion: Identification from Mixed Infections

- Intact cell MALDI-TOF MS works best for microbial identification on the current databases from cultured monobacterial samples

- Identification directly from blood cultures is possible (in >80-90%), even from polymicrobial samples (leading to monomicrobial results with standard MBT)

- Mixed infection algorithm (might) result in…
  - … slightly higher identification rates (less “no identification possible”)
  - … higher wrong identifications (version 1.0)
  - … missed mixed infections (➔ incomplete reporting persists, both versions)

- Risk of over-/under-treatment
- Complete dependence on manufacturer’s decisions

# Increasing Use of MALDI-TOF MS

# Access MS Data from CM Routine Identifications

- The microflex LT (Biotyper system) is optimized for routine analysis
- Investigation of individual spectra is possible but larger (retrospective) analysis is not encouraged
- All spectra are stored but sample identifiers are not readily available

Why should you care?

- Dataset shift by sample selection bias possible

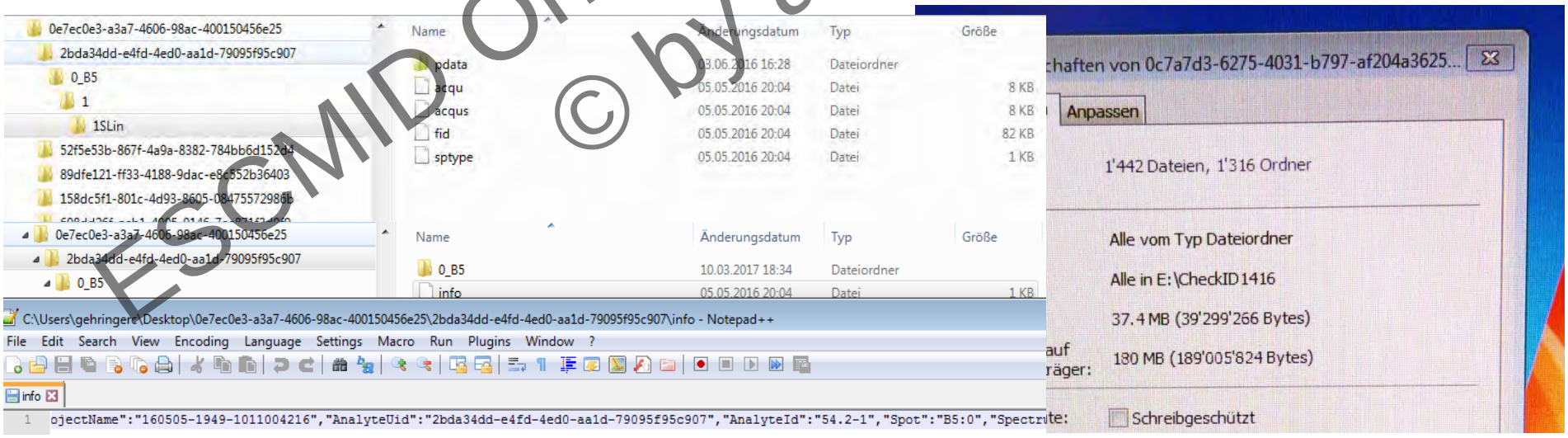
on dataset shift: Moreno-Torres, J. G., et al. (2012). Pattern Recognition 45, 521-530.

# Potentials of Analyzing Routine Mass Spectra

- Years' worth of spectral data stored
- Usually possible to access further (clinical, antibiotic susceptability,…) details
- Retrospective analyses when new findings are evident
- Mass spectra are of varying quality (different personnel and conditions) but represent a realistic set of data
- Includes regional variants of specific species
- Independence, research potentials etc…

# Structure of MBT Spectral Data Allows no Direct Bulk Analyses

- Summary spectra stored in *.fid-files (8 files in 5(!) folders/sample)
- Random alphanumeric strings
- Sample identifier is stored in a different location (info file / folder name)
- Command line exporting tool (CompassXport) is offered but the sample identifier is not contained (mzML file positions empty).
- Data/folder structure changed over time, including the location of the sample identifier (approx. mid-2015, and for “archived” spectra)
- Data/folder structure has to be preserved for MBT analysis



## Aim

- Structure spectral information in an accessible format
- Retrieve respective sample identifier
- Preserve original files
- Allow the connection of the sample identifier to information from LIS



LIS: Laboratory information (management) system

# Workflow to Make your Data Accessible

- Copy all raw data to external device
- Convert to *.mzML format using CompassXport and custom win bat-scripts ("convert_MBT1.bat" and/or "convert_MBT2.bat")[1]
- Automatically store all metadata in a *.csv file ("metadata.bat")[1]
- For MBT analysis, create list of sample identifiers (automatically copied by "Selektorenliste.bat")[1]
- Result
  - 1 file/spectrum
  - 1 spectra database containing all spectra ever measured
  - Connectable to clinical data using (unique) identifiers
  - Access to fid file and structure possible (to be re-analyzed by MBT-SW)



1) download https://github.com/gehringer/MALDI (available approx. 5/2017) or e-mail christian.gehringer@usb.ch

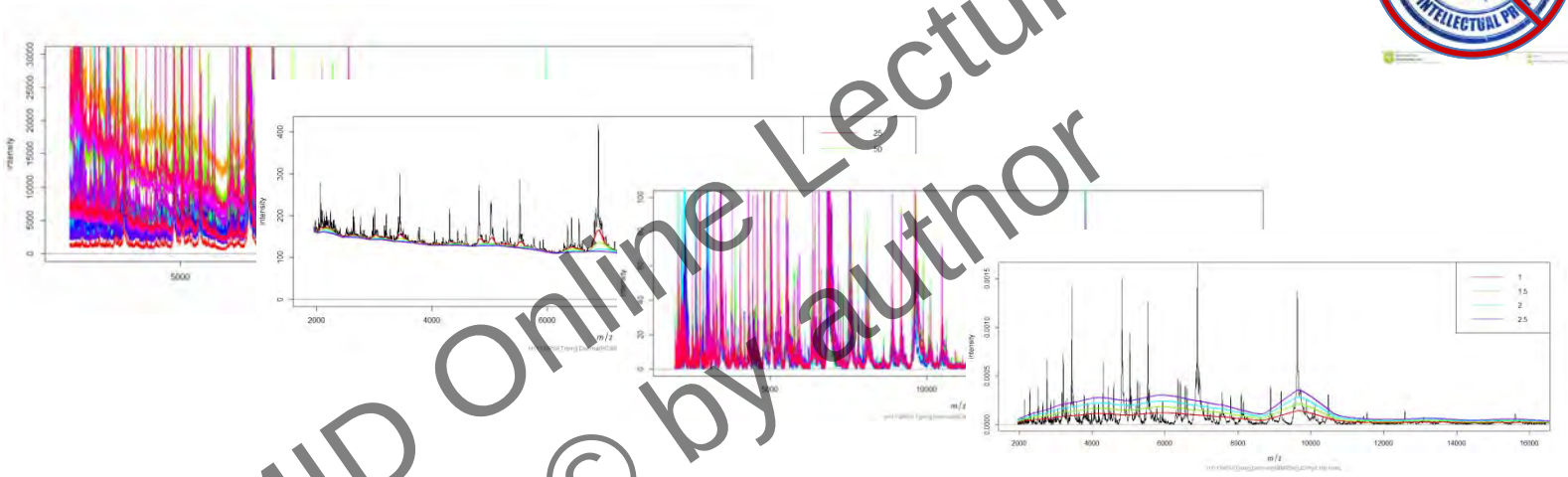
# Analysis Software for Spectral Data

- R-packages: MALDIquant/MALDIquantForeign, Peak Probability Contrasts
- Viewer: mMass (http://www.mmass.org/) and many others…
- Multiple sources, e.g. http://strimmerlab.org/notes/mass-spectrometry.html

# Example 1: Analysis Using MALDIquant Package[1] (R)

- Import mzML

  spectra<-laply(list.files(path = importfolder, pattern = ".mzML$"), function(x)importMzML(x))



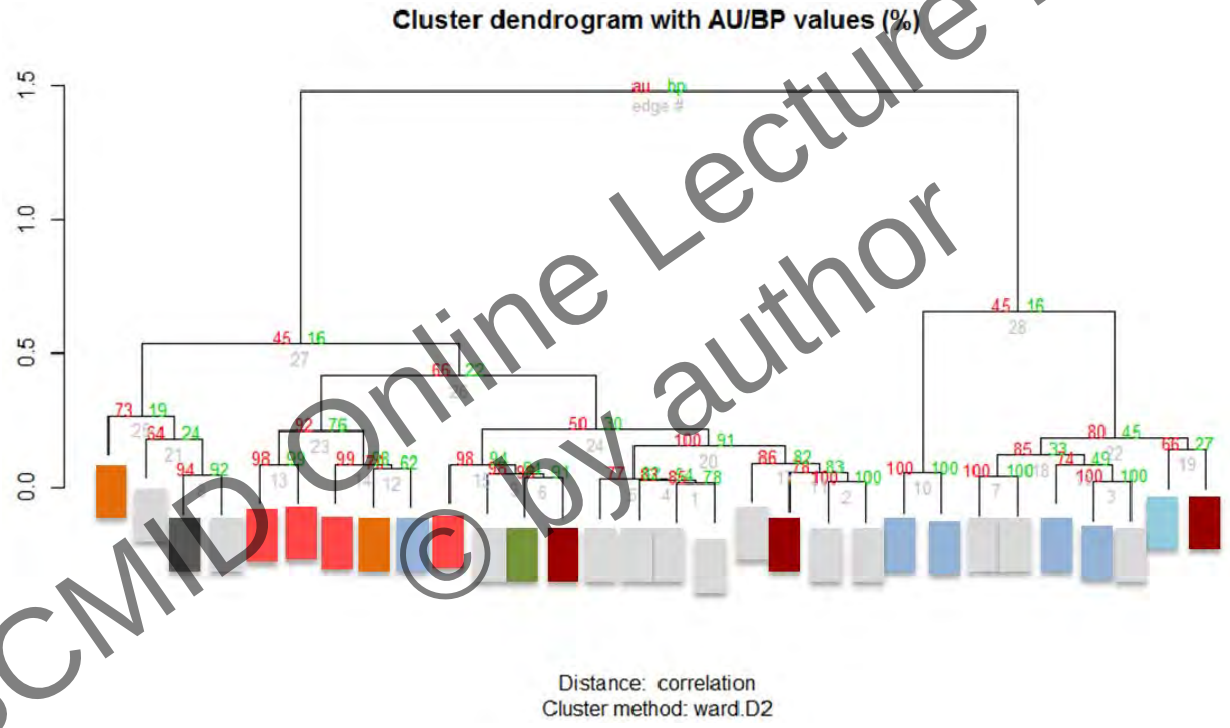

- Make spectra comparable:
  trimming, transforming, smoothing, baseline subtraction, normalization, noise and shift, warping (see package documentation and examples)
- Peak detection (supersmoother, MAD…)



Gibb, S. & Strimmer, K. (2012). *Bioinformatics* ***28, 2270-2271.***
https://github.com/sgibb/MALDIquantExamples/

# Analysis with R
(continued)

- Simple cluster analyses (MRSA, groups according to WGS)





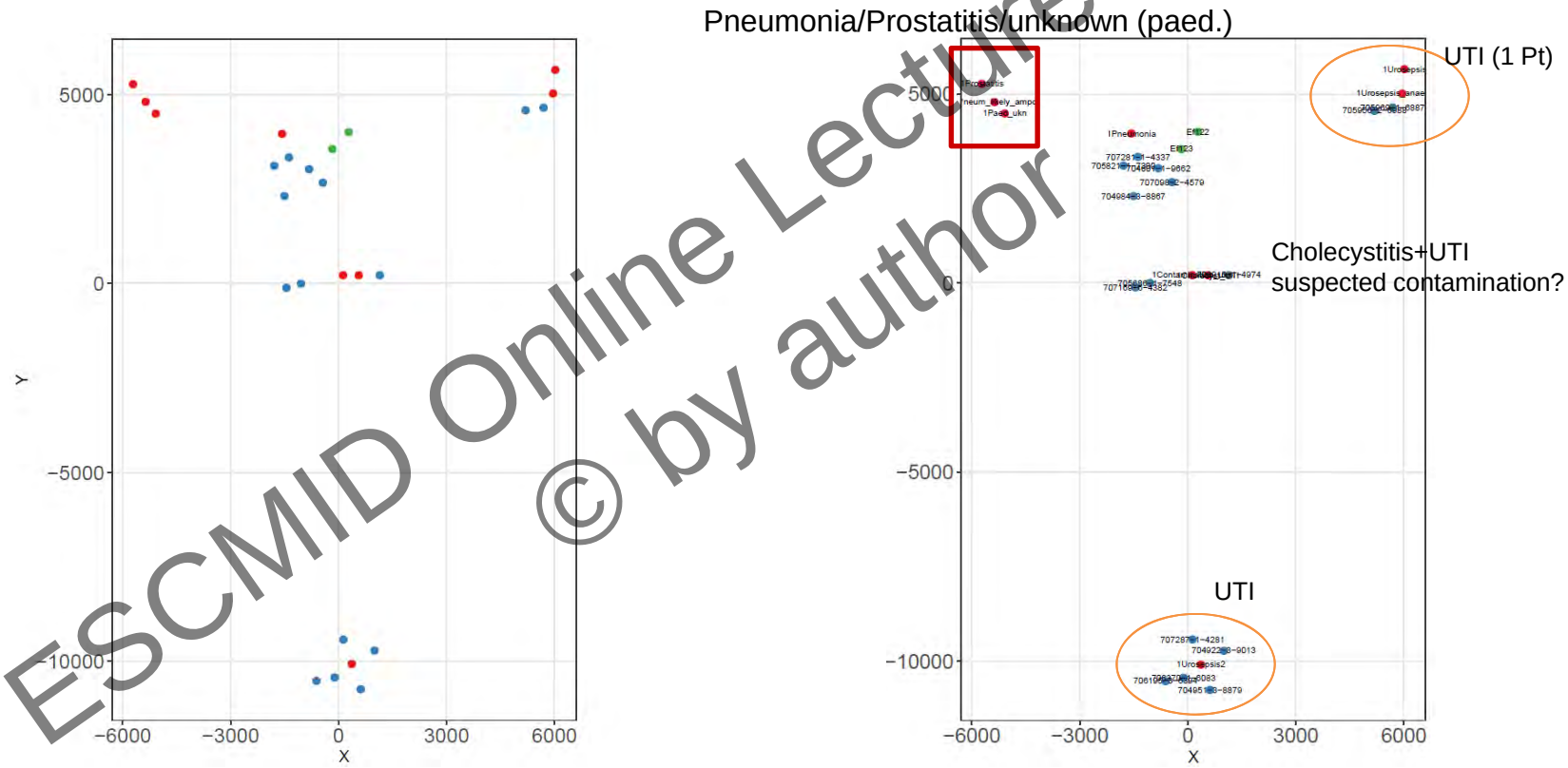

- Insufficient separation (sensitive to peak intensity, spectrum quality, growth conditions..)
- e.g. Mass Spectrometry-based PhyloProteomics (MSPP) more reliable[1]

1) Zautner, A. E., et al. (2015). Sci Rep 5, 13431.

# t-SNE

t-distributed stochastic neighbor embedding[1]

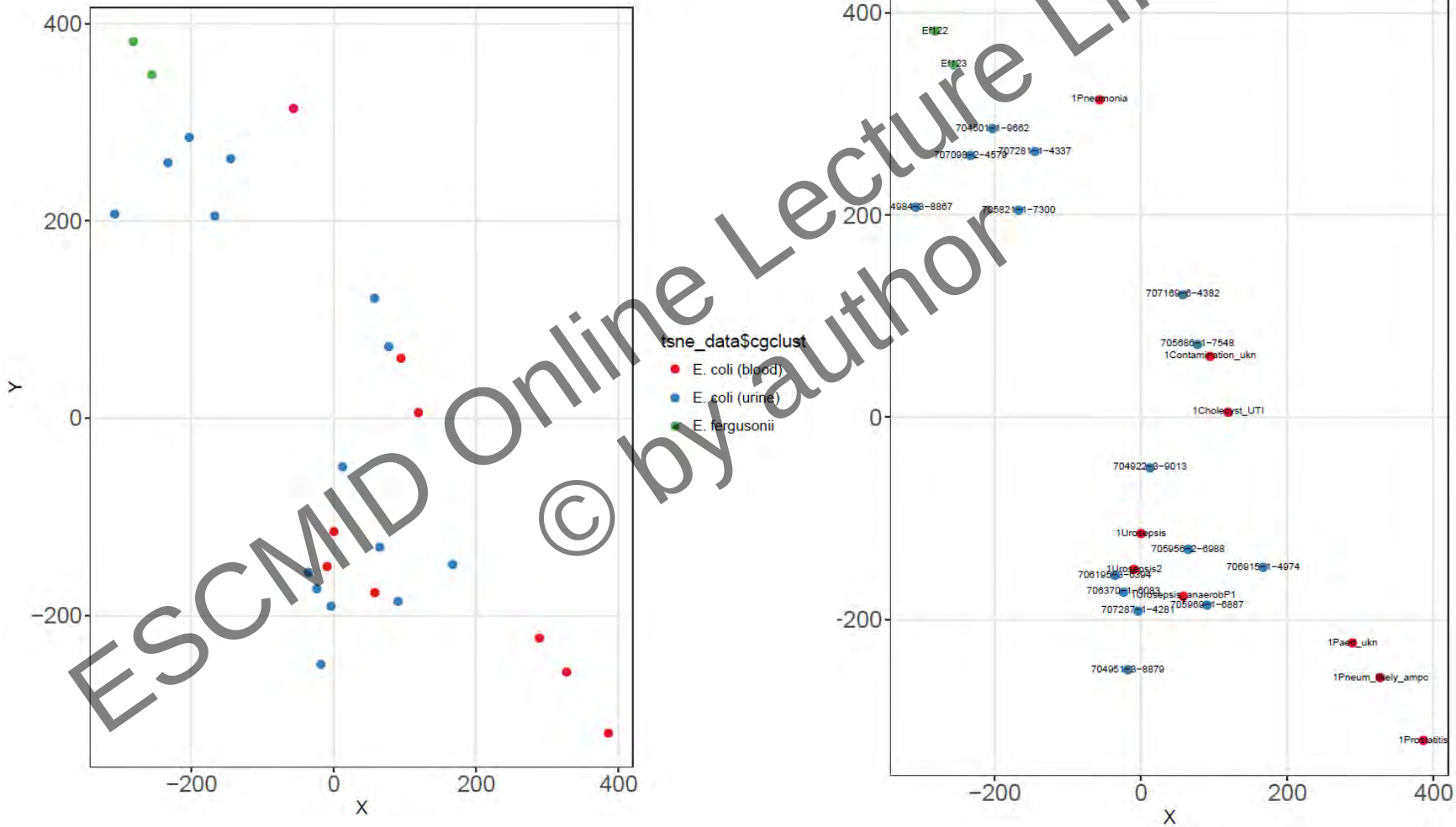
- Dimensionality reduction (flexible but complex for interpretation)[2]






1) *van der Maaten, L.J.P.; Hinton, G.E. (Nov 2008). Journal of Machine Learning Research.* ***9****: 2579–2605.*
2) http://distill.pub/2016/misread-tsne/

# Flexibility and Dynamics of t-SNE



# Conclusion

- There's more than microbial identification MALDI-TOF MS (in CM)
- Polymicrobial identification might become standard but needs more refinement (and extension to other materials)
- Access to primary data should be facilitated further since reanalyzing data offers several research opportunities
  - As new data becomes available
  - In exploratory projects
- Powerful, free, state of the art analysis tools are available
- Collaborative sharing of annotated spectral data might result in **improved** and **extended** predictions based on actual clinical samples

# Thank you

# for your attention

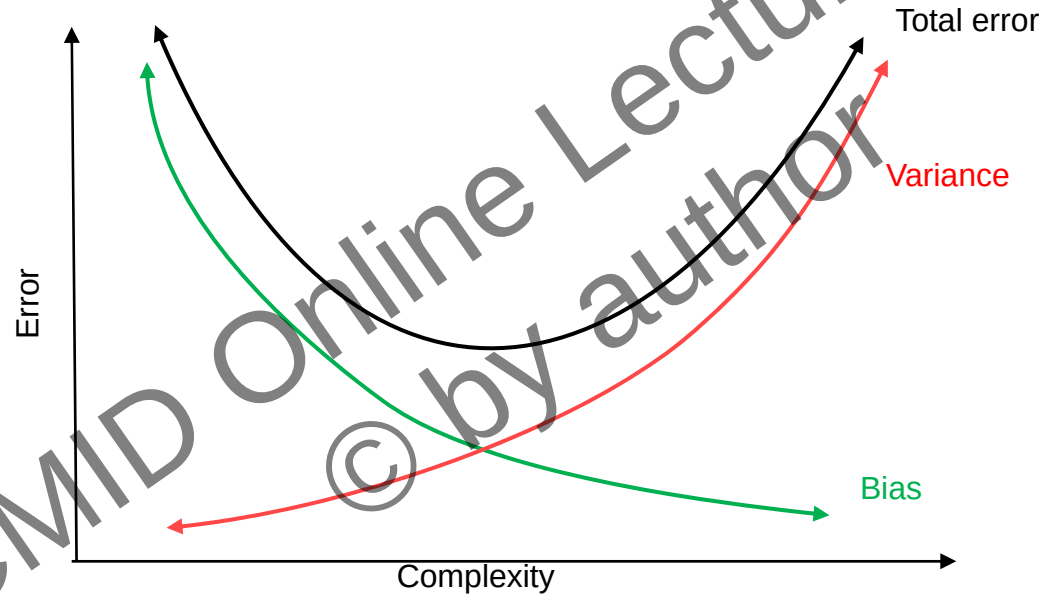
*For questions and comments please feel free to contact me at*
christian.gehringer@usb.ch

Acknowledgements

*Clinic Internal Medicine*
**Prof. Stefano Bassetti**

*Clinical Microbiology*
**PD Dr. Adrian Egli**
Dr. Vladimira Hinic
Dr. Dominik Meinel
Dr. Helena Seth-Smith
PD Dr. Daniel Goldenberger
Clarisse Straub
Doris Hohler
Birke Mebold

*Applied Microbiology Research Lab, Department Biomedizin*
Jana Linnik
Dominik Vogt
Yaseen Syedbasha
Christina Homberger






Bias vs. variance: http://scott.fortmann-roe.com/docs/BiasVariance.html and in general: http://www.kdnuggets.com/