P R O T E I N

R E A D I N G

Identification of microorganisms by mass spectrometry

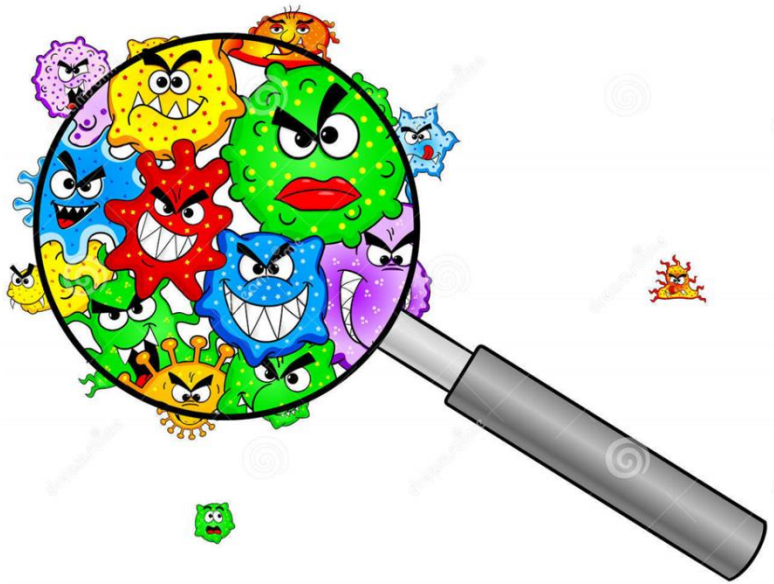
Mario Cindrić
Ruđer Bošković Institute, Zagreb, Croatia

email: mcindric@irb.hr

Identification of microorganisms by mass spectrometry

Dr. sc. Mario Cindrić
Proteomics and mass spectrometry
Institut Ruđer Bošković
Bijenička cesta 54
10 000 Zagreb
mcindric@irb.hr
Tel: +385 99 218 32 80

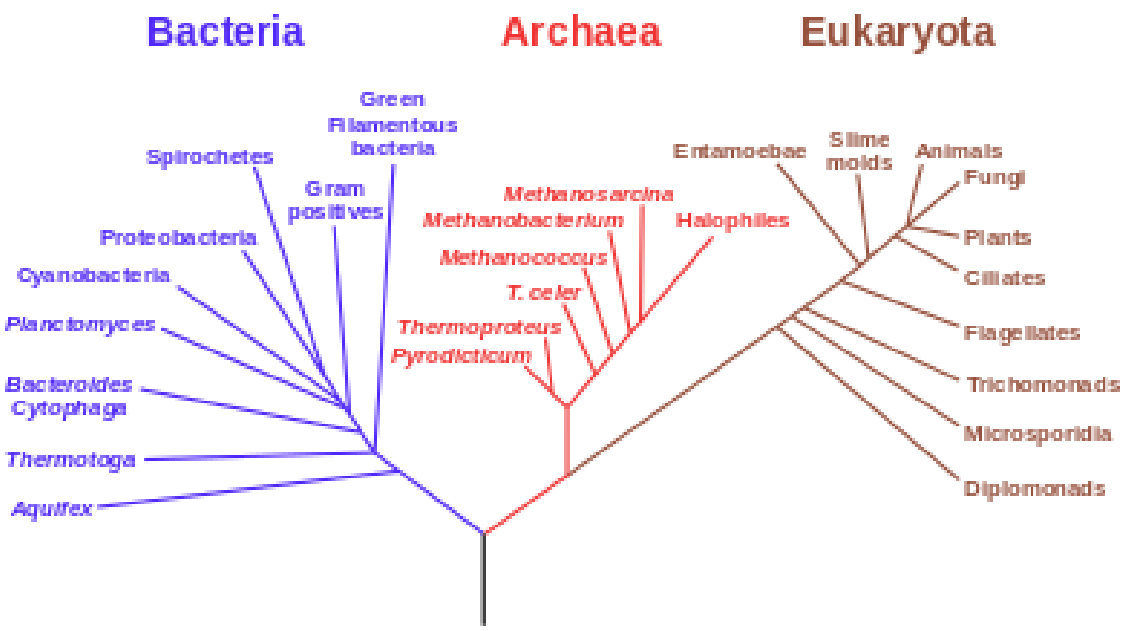
Currently used techniques for species identification are mostly relied on protein database matching or alternatively on immunological procedures. Although widely used, such an approach does not always provide satisfactory matching, sequence coverage or specific antigen-antibody reaction to unambiguously identify DNA, RNA, lipid, and sugar or peptide/protein of selected species. Mass spectrometry as a tool for species determination was introduced about ten years ago. Matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) identifies microbes using either intact cells or cell extracts. The most abounded proteins from the cell detected in the instrument are matched against commercially available databases that cover relatively limited number of microbes. On the other hand, high resolution mass spectrometers in contrast to MALDI-TOF MS could provide more accurate and precise results after *de novo* sequencing of analyzed proteins.



Biotypization means species identification

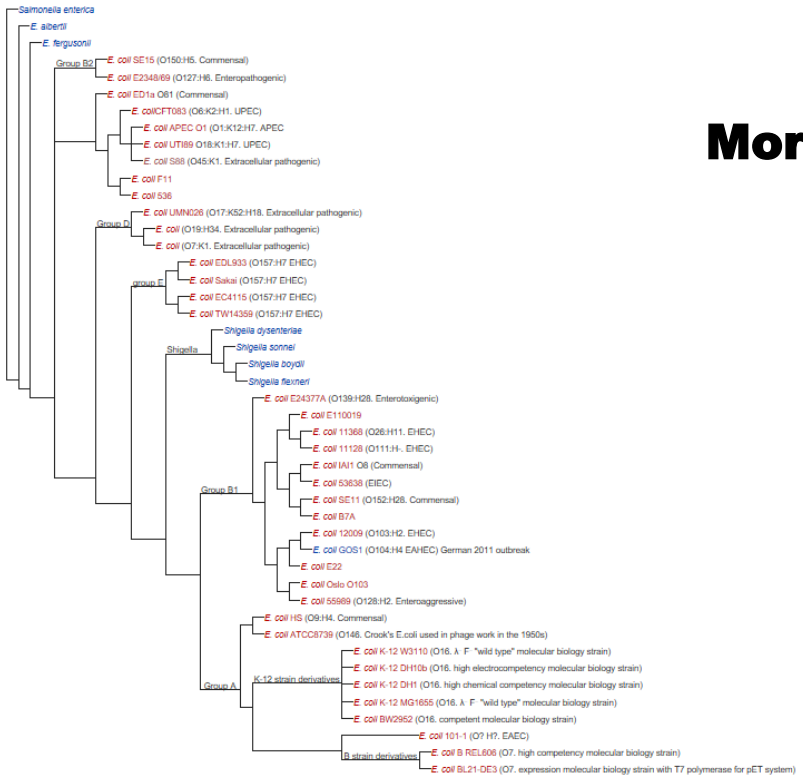
Subspecies identification?

What is the sub-species identification?



e.g. pathogenic Escherichia coli

Why species determination is not enough?

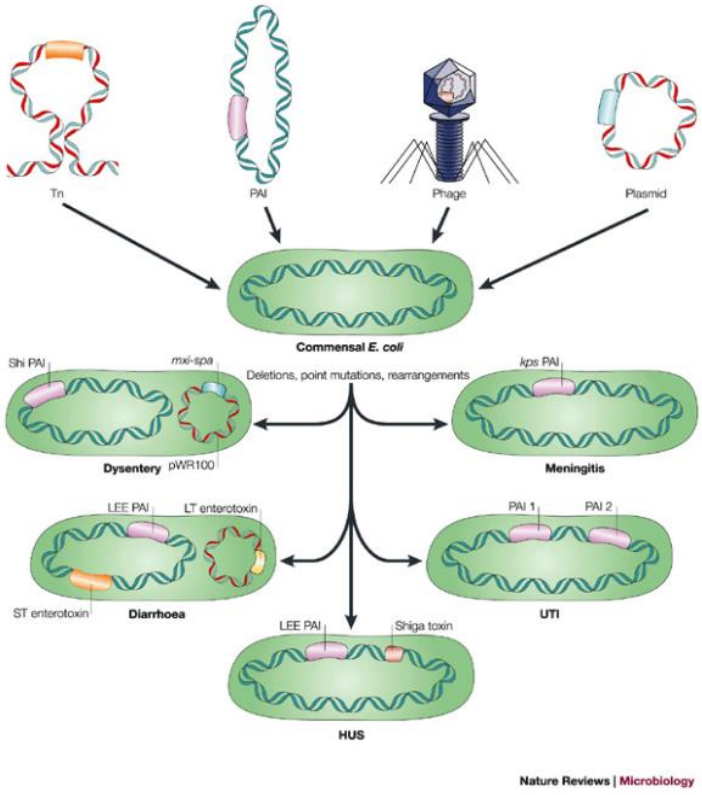


More than 600 known sub-types
50 pathogenic

FEMS Microbiology Reviews, 2016, S.M.Soto

The first complete DNA sequence of an *E. coli* genome(laboratory strain K-12 derivative MG1655) was published in 1997

General clinical syndromes can result from infection with one of these pathotypes: enteric/diarrhoeal disease, urinary tract infections (UTIs) and sepsis/meningitis. Among the intestinal pathogens there are six well-described categories:
enteropathogenic *E. coli* (EPEC),
enterohaemorrhagic *E. coli* (EHEC),
enterotoxigenic *E. coli* (ETEC),
enteroaggregative E. coli (EAEC),
enteroinvasive *E. coli* (EIEC)
and diffusely adherent *E. coli* (DAEC)



Science, 2017, Blattner F.R.

The first complete DNA sequence of an *E. coli* genome(laboratory strain K-12 derivative MG1655) was published in 1997.

The genomic structure of the *E. coli* pathotypes that have been sequenced so far show a striking mosaic pattern, with 2,000 genes present in 247 islands in one pathotype that are not present in K-12. Up to 0.53 MB of DNA present in K-12 can also be absent from pathogenic *E. coli*.

K-12 only found in the lower intestineof warm-blooded organisms.

Nature Reviews Microbiology , 2004, Mobley H.T

Science, 2017, Blattner F.R.

POLIMERASE CHAIN REACTION (PCR) species determination

How to determine the difference?

Whole exome sequencing (**WES**)

- humans have about 180,000 exons, constituting about 1% of the human genome, or approximately 30 million base pairs.

- E. Coli strains have in average 4,500 gens, constituting about 90% of the bacteria genome, or approximately 4 million base pairs.

Sequencing the first human genome was a massive undertaking. The project cost $1 billion and lasted 13 years. But genetic technologies have progressed dramatically, and a team from the Rady Children's Institute for Genomic Medicine recently sequenced a genome in 19.5 hours, marking the feat with a Guinness World Record.

GenomeMag.com, 2018

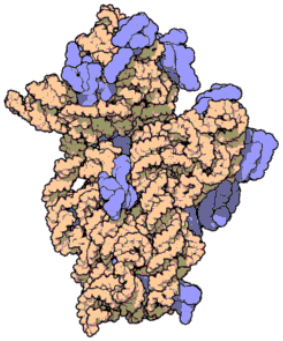
How to determine the difference?

Non-whole exome sequencing (non-**WES**)

- 16S ribosomal RNA (rRNA) sequencing is a common amplicon sequencing method used to identify and compare bacteria present within a given sample. 16S rRNA gene sequencing is a well-established method for studying phylogeny and taxonomy of samples from complex microbiomes or environments that are difficult or impossible to study.

- 16srRNA gene sequence based technique is the very commonly used techniques for the identification of bacterial species. During the Blast in the NCBI, it is the possibility of matching to two different species.

- Sometimes sequencing the entire 1,500-bp region is necessary to distinguish between particular taxa or strains.



Cell, 2000, Jonath A.

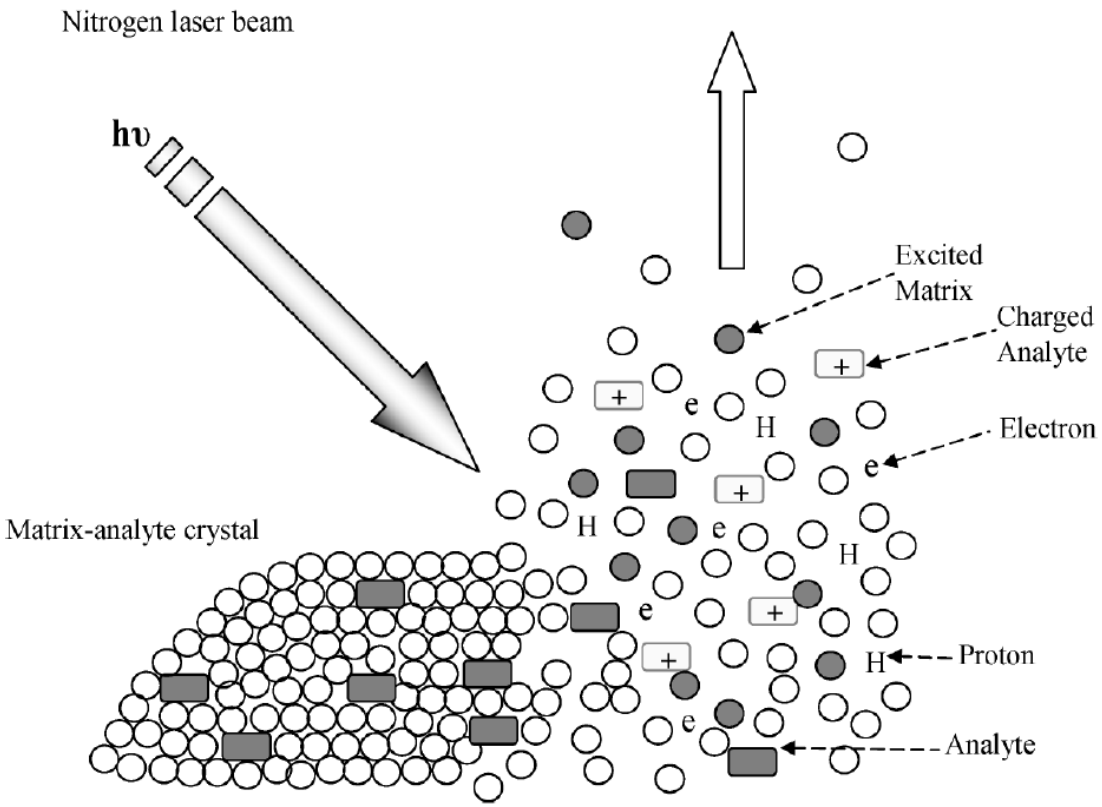
Clin Microbiol Rev, 2004, Clarridge, J. E. III.

Clinical practice- One day

One-day Workflow Scheme for Bacterial Pathogen Detection and Antimicrobial Resistance Testing from Blood Cultures

MALDI MASS SPECTROMETRY species determination

Matrix-Assisted Laser Desorption Ionization MALDI



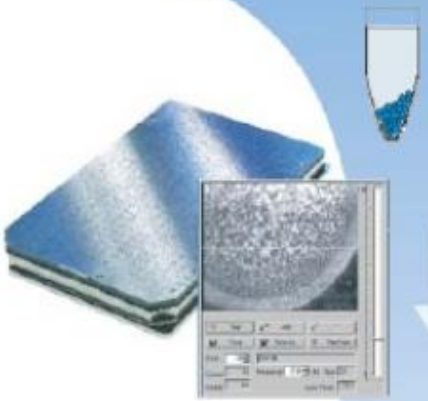
UNC SCHOOL OF MEDICINE

Bruker MALDI BioTyper Workflow

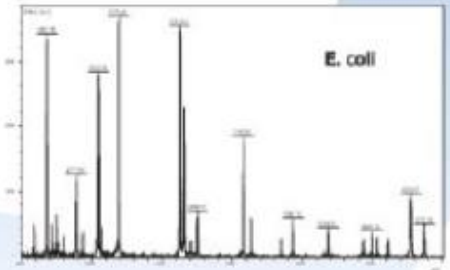
1. Select a Colony

Unknown Microorganism

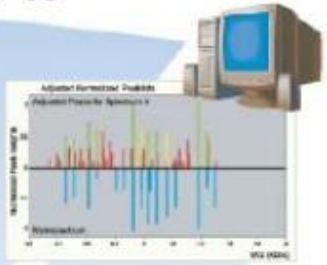
2. Smear a thin-layer onto Target Plate or perform rapid organic extraction & spot supernatant



3. Add MALDI Matrix



4. Generate MALDI-TOF Profile Spectrum

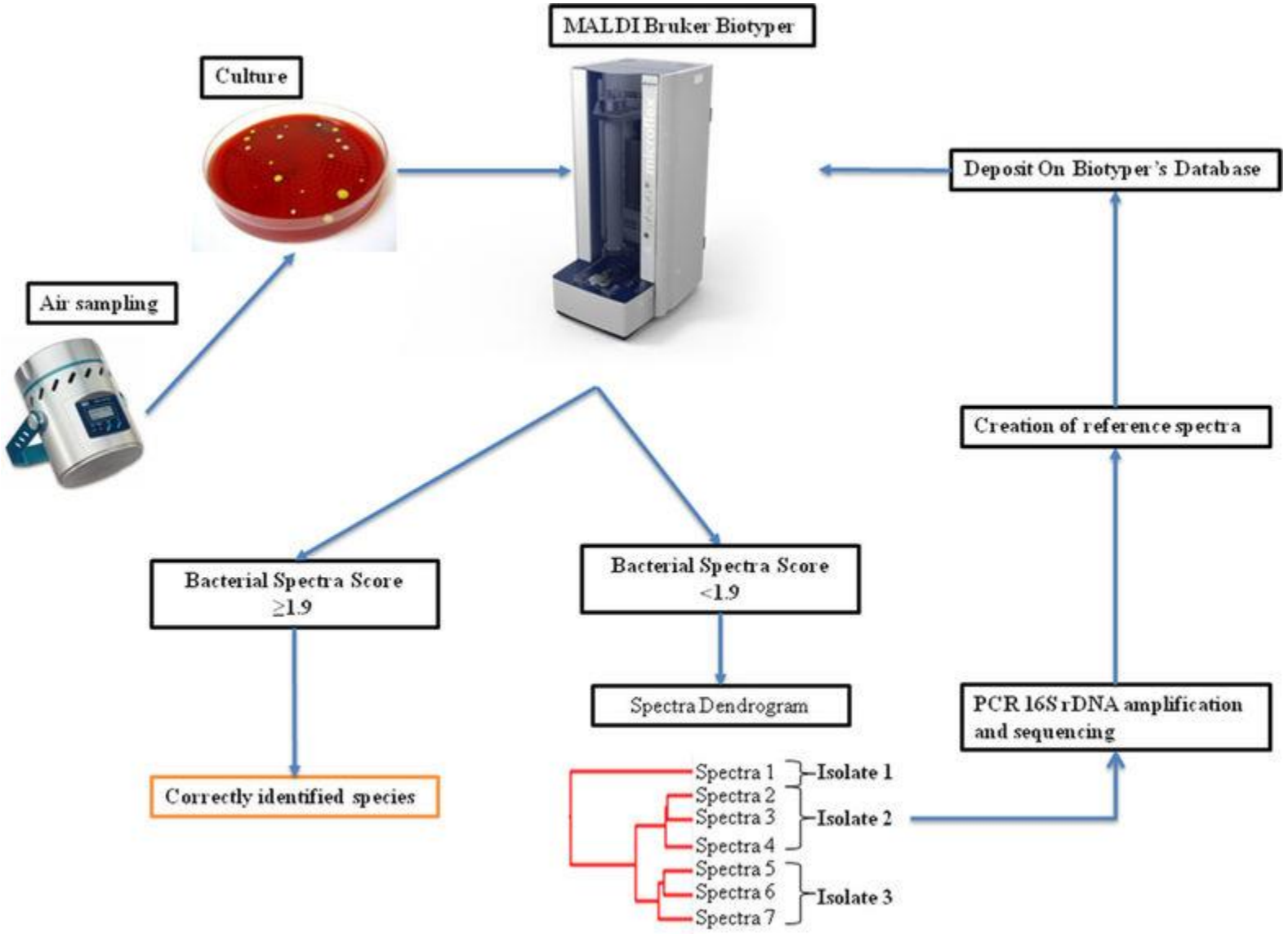


5. Data Interpretation

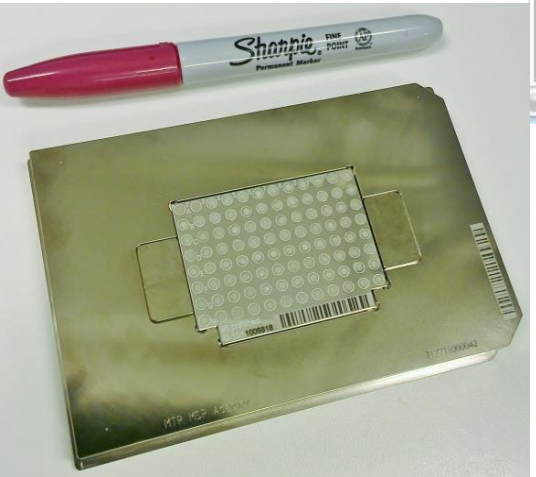
6. Match patterns to database to identify species

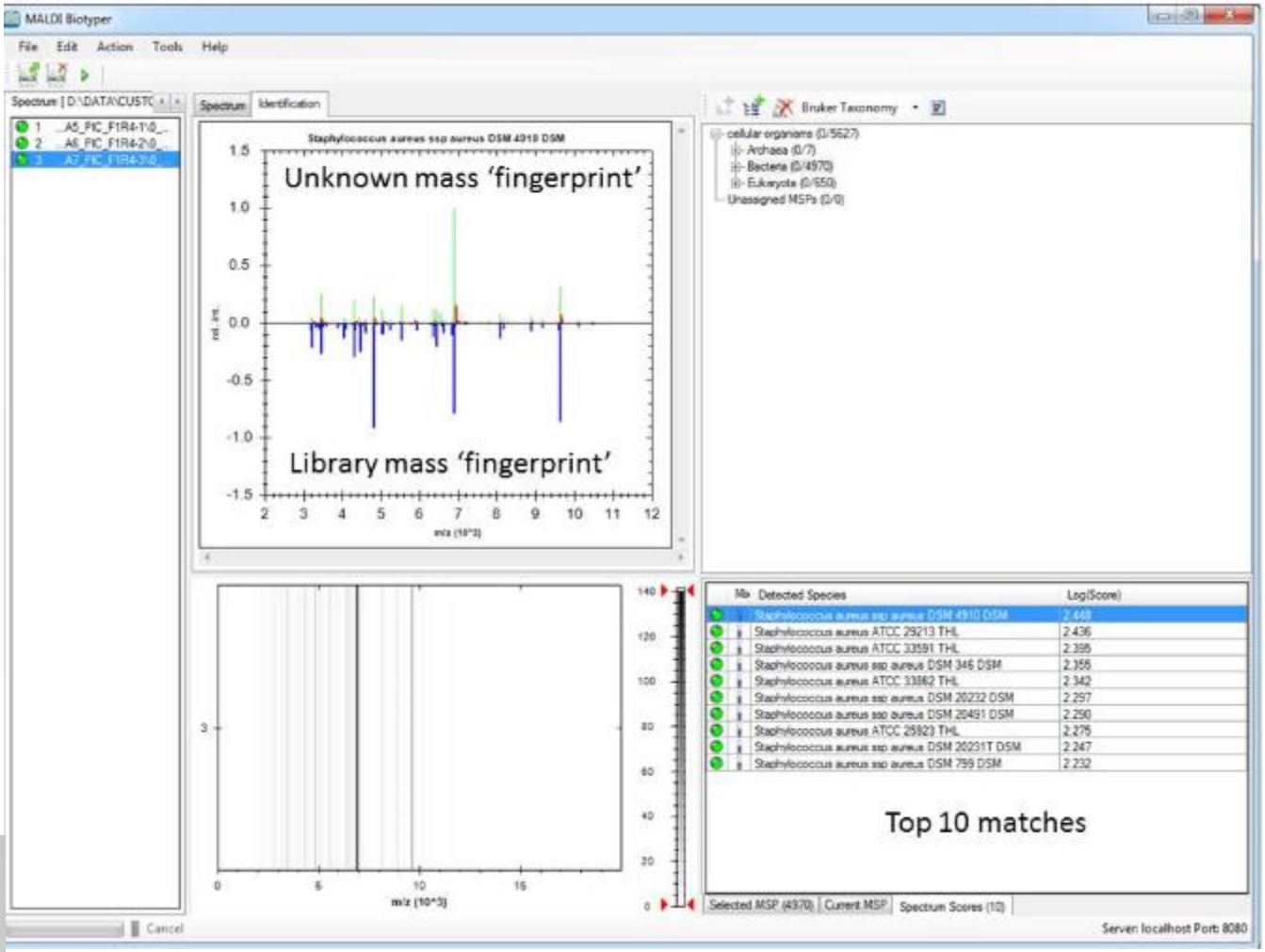
http://pharmaceuticalmicrobiologi.blogspot.hr/2017/09/

LOW RESOLUTION MALDI MASS SPECTROMETRY species determination with commercially or home-made database



BMC Research Notes, 2014, Raoult D.





The reference library now includes more than 2,700 species of 458 microorganism genera.

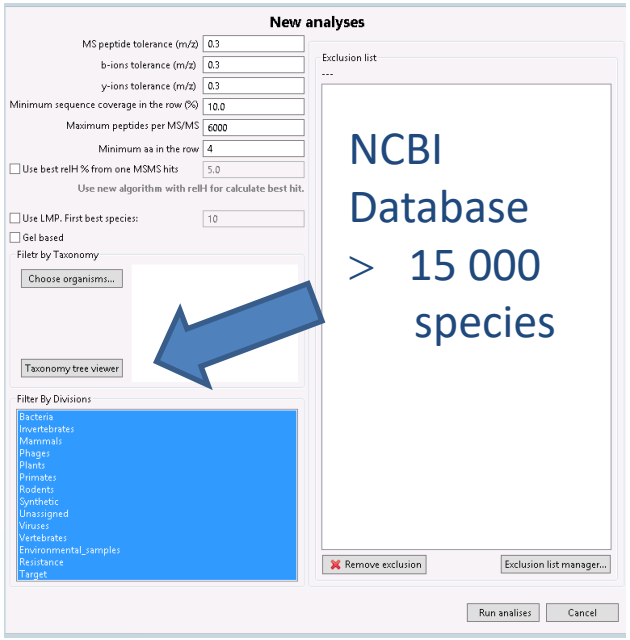
http://sites.psu.edu/msproteomics/tag/bruker-biotyper-3-1/
2018

These were general aspects of POLIMERASE CHAIN REACTION (PCR) and LOW RESOLUTION MALDI-MS spectra matching concept species determination

Is there any other concept?

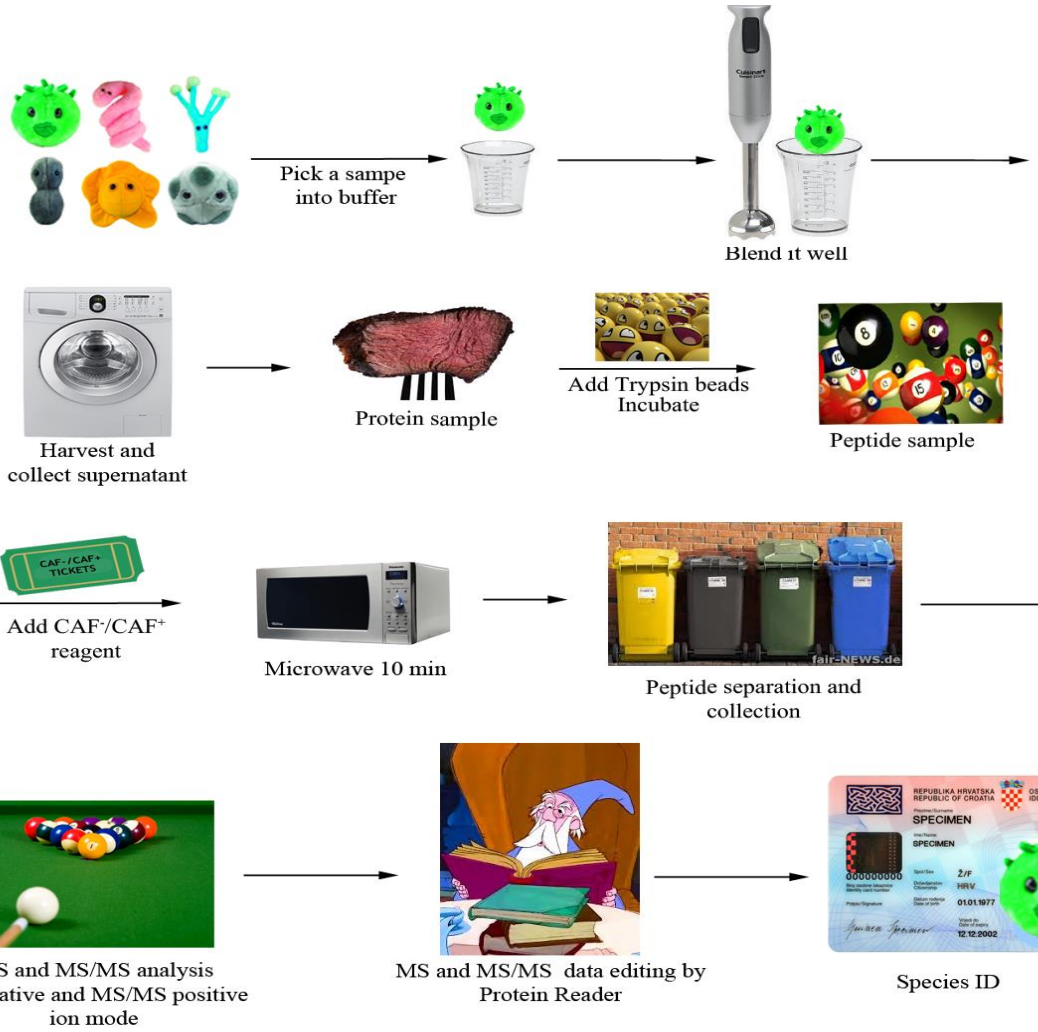
Protein reading concept

P R O T E I N
R E A D I N G



NCBI Database > 15 000 species

Sample preparation workflow



Sample preparation workflow

Pick a sampe into 100 µL of 25mM NH_4HCO_3, 0.1% v/v Triton X

Add app. 50 µL glass beads (d=0,1mm)

sample
glass beads

Vortex 30 sec
Put on ice 1 min
Repeat 5 times

OR

TissueRupter

Grinding cycle:
45 seconds ON / 45 seconds OFF
Total grinding times: 5-6 cycles

Harvest at 10000 rpm for 5 min
Collect supernatant
Store at -20 °C before analysis

Protein sample

Add Mag-Trypsin
(trypsin immobilized
on magnetic beads)

Incubate 1hour at 37 °C

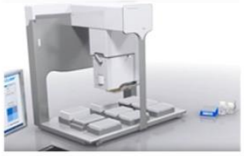
Peptide sample

Add 30 µL of CAF^-/CAF^+ reagent
onto 20 dried sample

Microwave in household oven at
90W for 10 minutes

Labeled peptide sample

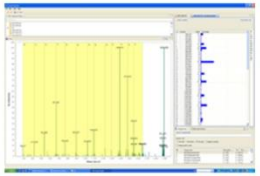
US 8,647,880 B2



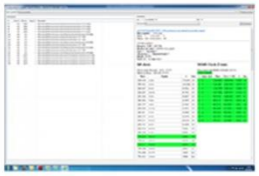
2-D LC peptide separation

Peptides mixed with CHCA
matrix and
spoted onto MALDI plate

MALDI TOF/TOF
MS and MS/MS analysis
in negative and MS/MS
positive ion mode



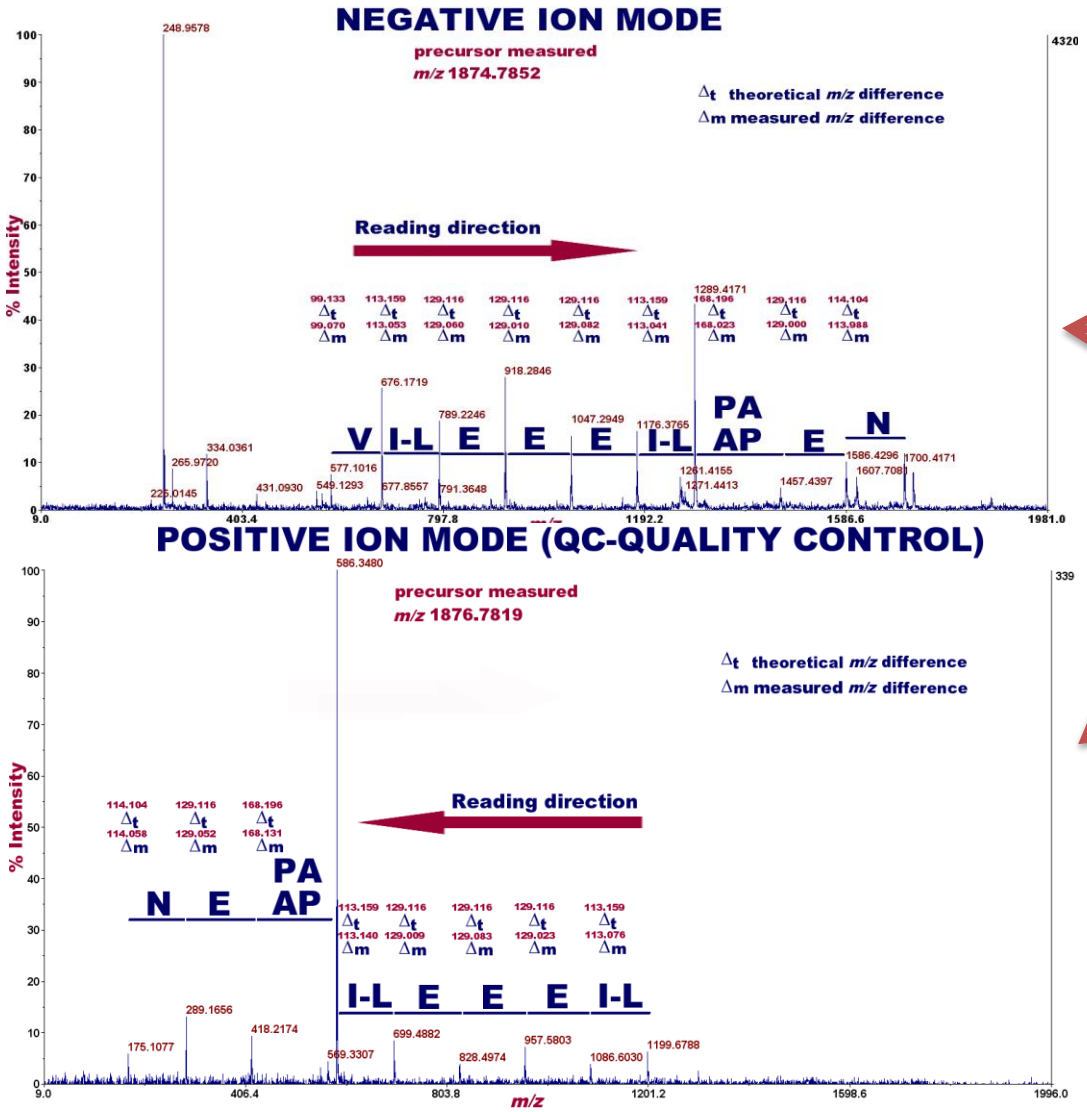
MS and MS/MS data editing by
Protein Reader



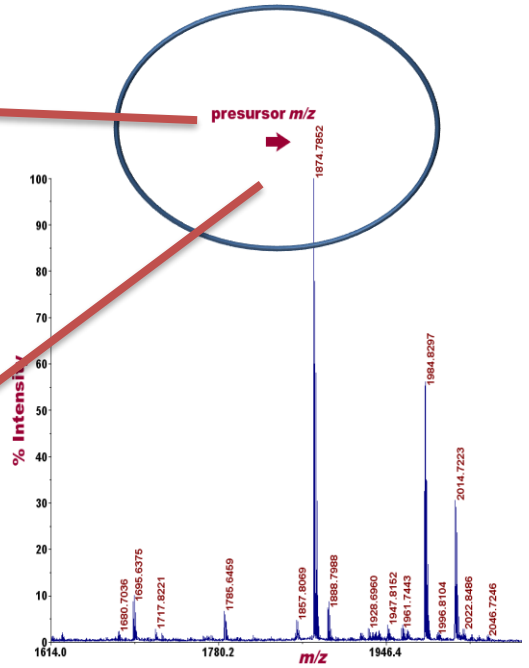
Species ID

http://rapidcell.proteinreader.com/

P R O T E I N
R E A D I N G



MS negative ion mode



CAF-/CAF+ or shorter CAF/CAF

PROTEIN READING

Protein reading concept:

1. Peptides MS negative ionization
2. Peptide sequencing in neg. MS/MS
3. Peptide sequencing in pos. MS/MS
4. Looking for overlapped readings
5. ID of the microorganisms by overlapped readings

e.g.

N → C

neg. ion mode

V I/L E E E I/L P A E N

N E A P I/L E E E I/L

pos. ion mode

N ← C

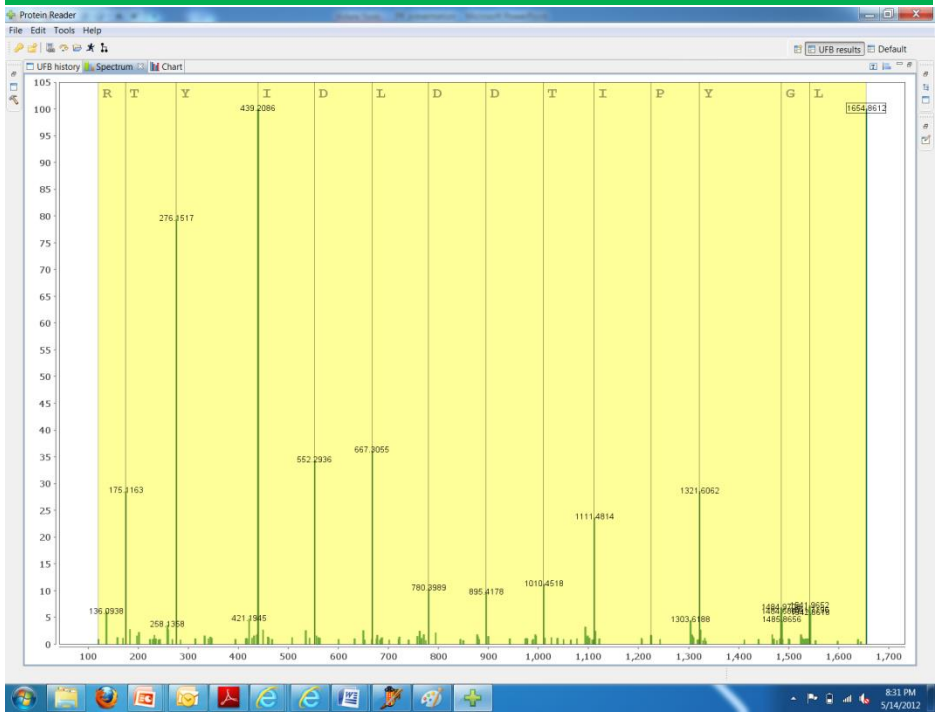
P R O T E I N

R E A D I N G

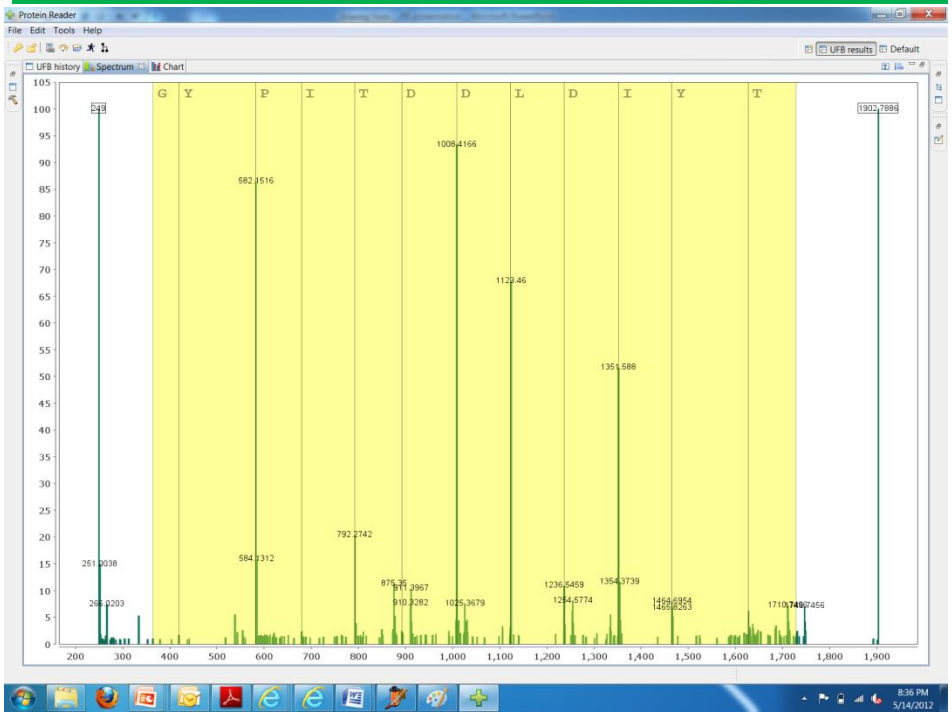
Protein reader features, sure shot concept (SS)

- peptide sequence confirmed by reading in pos. and neg. MS/MS

Pos. MS/MS reading LGYPITDDLDIYTR



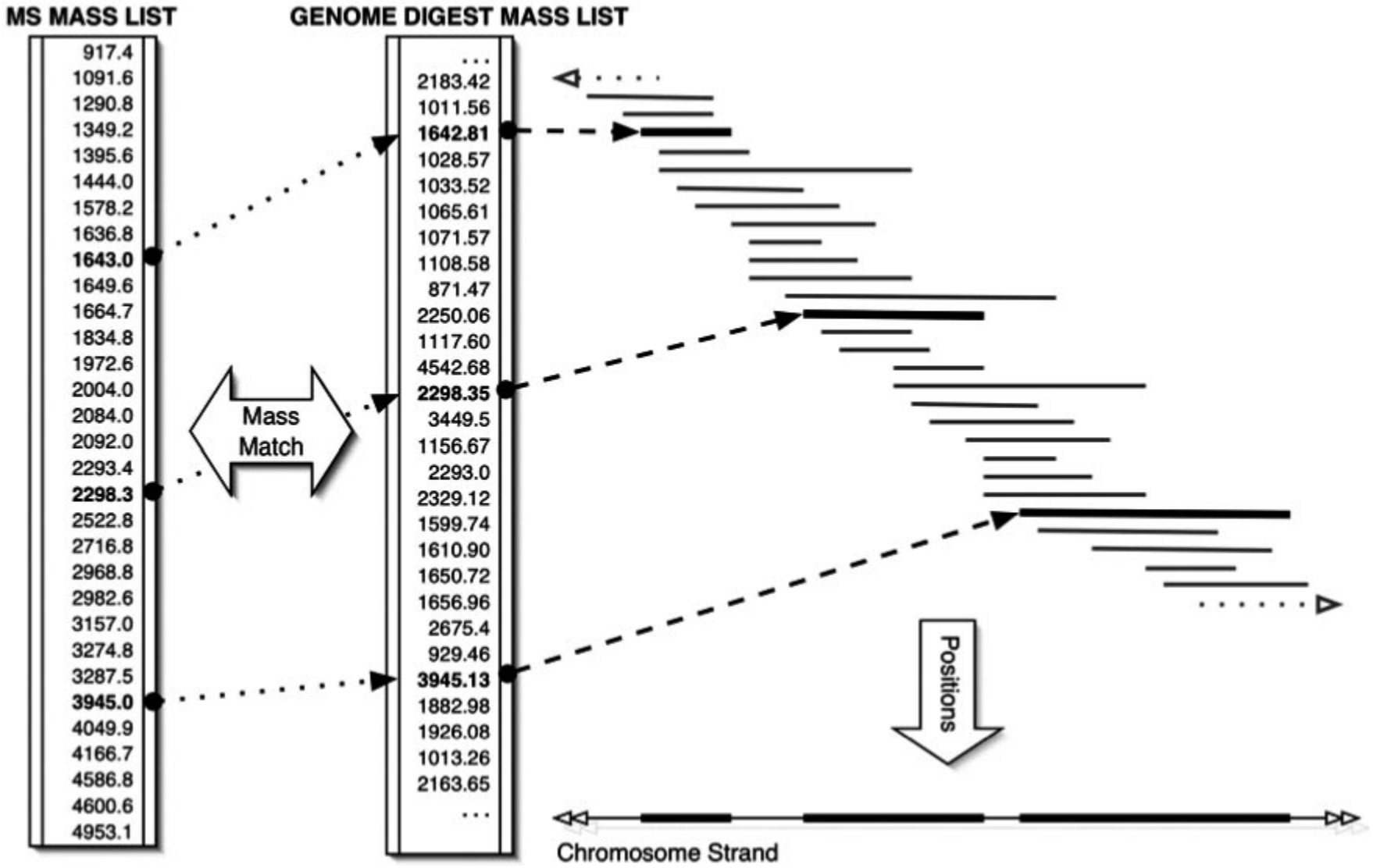
Neg. MS/MS reading LGYPITDDLDIYTR



Pos. MS/MS Outer membrane protein A precursor (*Escherichia coli* WV_060327)

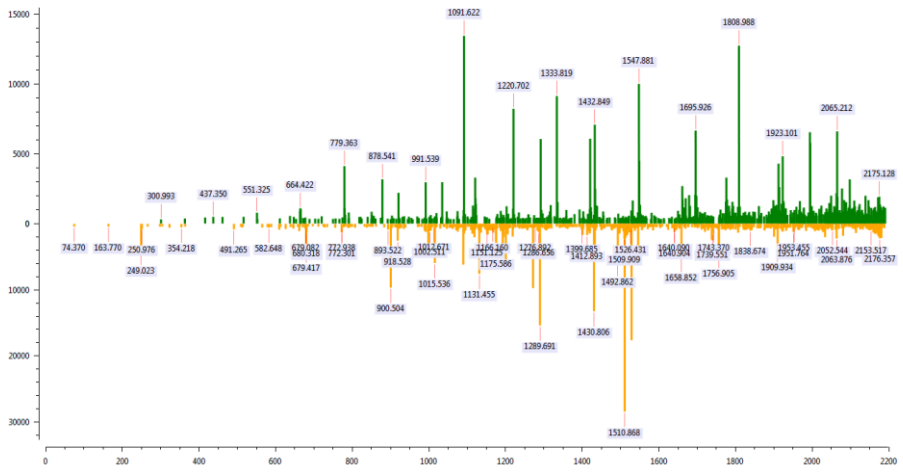
Neg. MS/MS Outer membrane protein A precursor (*Escherichia coli* WV_060327)

HIGH RESOLUTION MALDI MASS SPECTROMETRY species determination

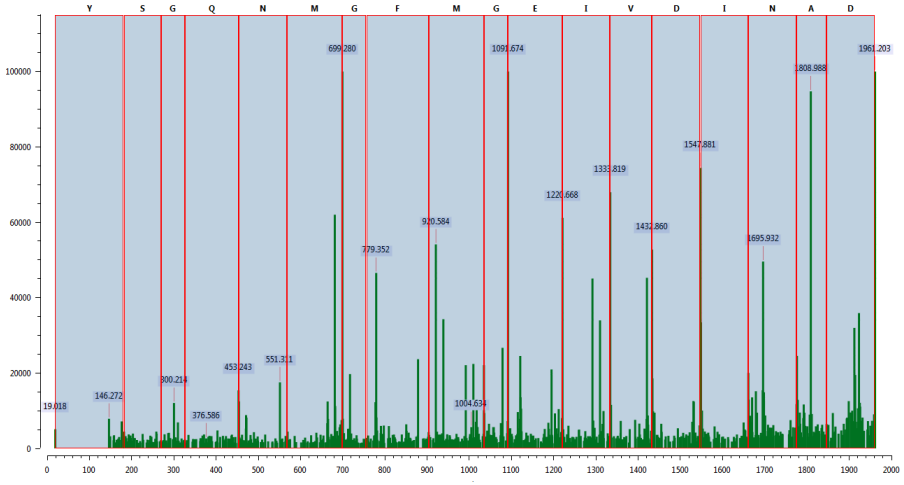


Giddings, PNAS 2003

PROTEIN READING



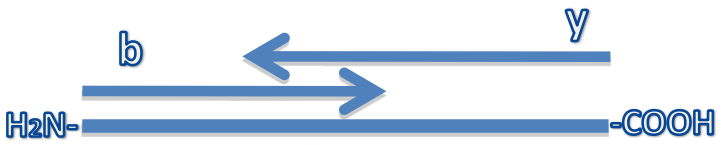
Positive and negative mass spectra



Fused positive and negative mass spectra

CAF/CAF MALDI analysis

Analysis in positive and negative ion mode of peptides

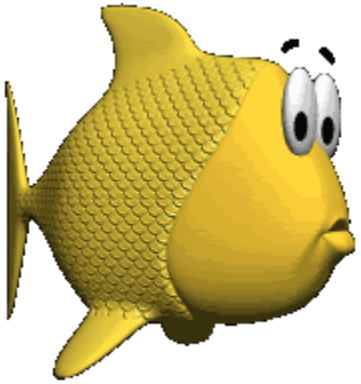


BENEFITS!!!

N-term → C-term

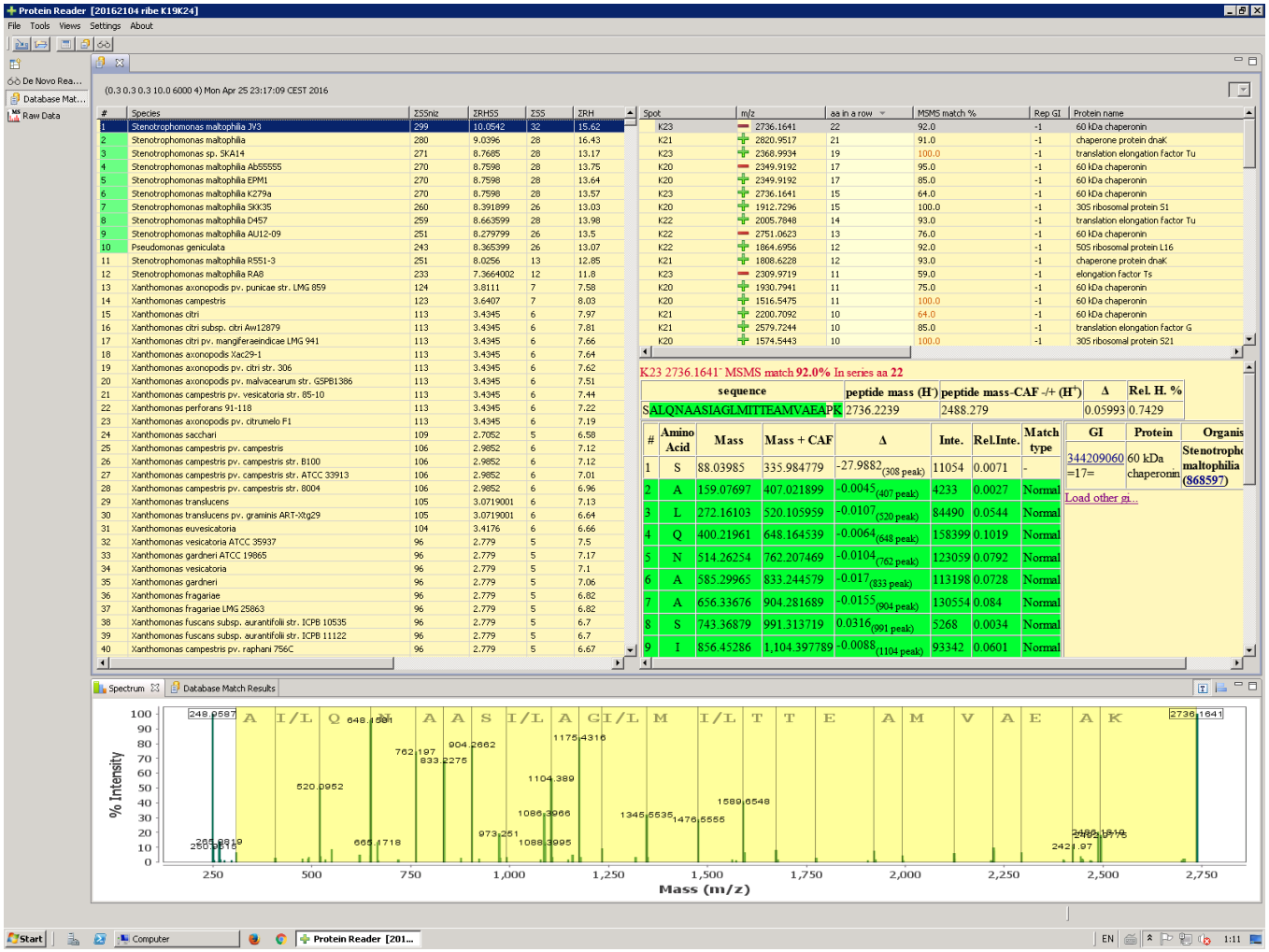
C-term → N-term

1-D chromatography on bacterial peptides from fish (45 min separation)



Proteins found

1. 60 kDa chaperonin (17)
2. translation elongation factor Tu (4)
3. chaperone protein dnaK (2)
4. translation elongation factor G (5)
5. 30S ribosomal protein S1 (2)
6. elongation factor Ts (2)
7. malate dehydrogenase (2)
8. 50S ribosomal protein L1 (2)
9. ATP synthase subunit beta (2)
10. 50S ribosomal protein L16 (1)
11. 30S ribosomal protein S21 (1)
12. histone family protein DNA-binding protein (1)
13. protein grpE (2)
14. 10 kDa chaperonin (1)
15. 2-polyprenylphenol 6-hydroxylase (1)
16. citrate synthase I (1)
17. two component winged helix family transcriptional regulator (1)
18. multi-sensor hybrid histidine kinase (1)
19. integral membrane sensor signal transduction histidine kinase (1)



19 SS proteins

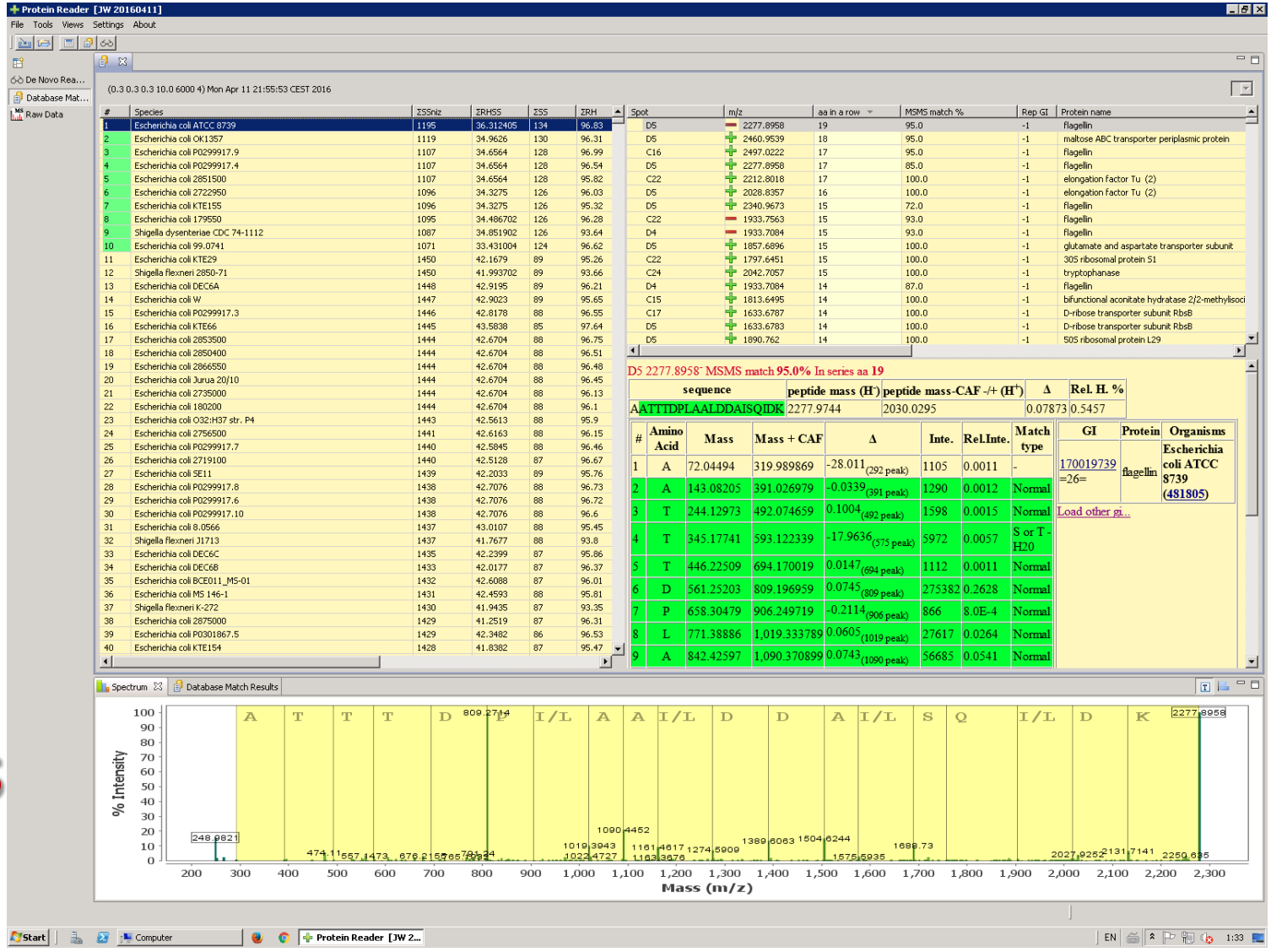
Common and rare clinical strains

Proteus mirabilis str. 33	*Acitenobacter baumanni* str. 3	*Trichophyton rubrum* CBS 118892	*Bacillus anthracis*
Proteus mirabilis str. 1312R	*Enterococcus faecium* str. E124	*Aspergillus fumigatus* Af293	*Aspergillus flavus*
Staphylococcus aureus strain 59	*Pseudomonas aurigenosa* strains: P1, P3, P104, P91, P33, P123	*Aureobasidium pullulans*	*Candida albicans*
Staphylococcus aureus MRSA	*Escherichia coli* CFT073	*Fusarium dimerum*	*Candida glabrata*
Yersinia enterocolica str. 94	*Helicobacter pylori* P12	*Fusarium delphinoides*	*Cryptococcus neoformans*

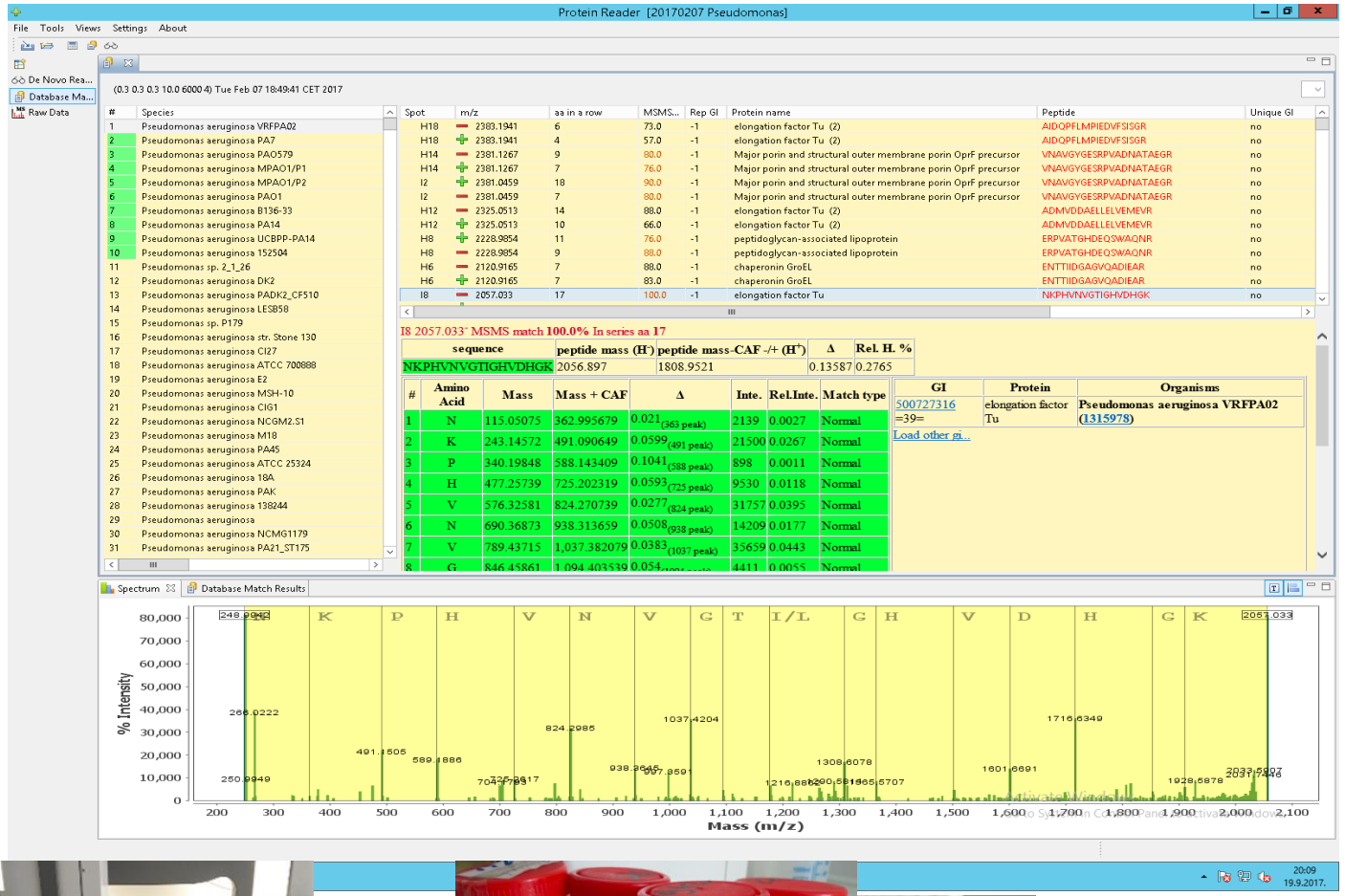
Separated in 45 min or identified by 3 h

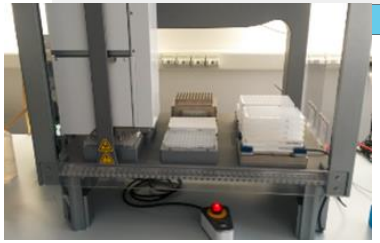
2-D chromatography on bacterial peptides up to the subspecies level

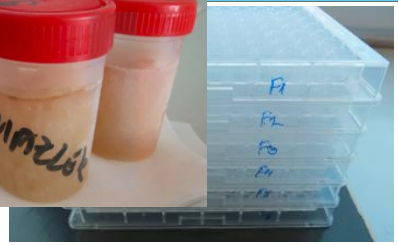
Over 100 SS proteins



2-D chromatography on bacterial proteins in urine sample







P R O T E I N

R E A D I N G

COMPARISON WITH OTHER MS BIOTYPIZATION APPROACHES

Bruker's Biotyper results

specie	strain	result	best match	score	secound best match	score	positive identification	
Lactobacillus plantarum	strain 7.1	(+++)(A)	Lactobacillus plantarum	2.438	Lactobacillus plantarum	2.43	+	
Enterococcus faecium	strain 7.2	(++)(A)	Enterococcus faecium	2.224	Enterococcus faecium	2.221	+	
Pediococcus pentosaceus	strain 2.2	(++)(A)	Pediococcus pentosaceus	2.134	Pediococcus pentosaceus	2.067	+	
Lactobacillus casei	strain 16.1	(++)(A)	Lactobacillus paracasei	2.257	Lactobacillus paracasei	2.119	+	False positive match!!!
Lactobacillus fermentum	strain 8.2	(++)(A)	Lactobacillus fermentum	2.05	Lactobacillus fermentum	1.794	+	
Lactobacillus fermentum	strain 10.2	(+)(B)	Lactobacillus fermentum	1.85	Lactobacillus fermentum	1.701	+/-	
Lactobacillus brevis	strain 15.2	(+)(B)	Lactobacillus brevis	1.938	Lactobacillus brevis	1.892	+/-	
Lactobacillus buchneri	strain 12.3	(+)(B)	Lactobacillus buchneri	1.895	Lactobacillus kefiri	1.725	+/-	
Lactobacillus plantarum	strain 16.3	(+)(B)	Lactobacillus plantarum	1.97	Lactobacillus plantarum	1.849	+/-	
Lactobacillus fermentum	strain 9.1	(+)(B)	Lactobacillus fermentum	1.885	Lactobacillus fermentum	1.719	+/-	
Lactoccocus lactis	strain 1.1	(-)(C)	7 not reliable identification	1.383	7 not reliable identification	1.355	-	
Lactococcus lactis	strain 6.8	(-)(C)	16 Not reliable identification	1.272	16 Not reliable identification	1.234	-	

THE PROTEIN READER

legend	score
A - positive	>2
B - maybe	1.7 to 2
negative resulti	< 1,7

A - positive identification
B - maybe positive identification
C - not reliable identification

CAF+ / CAF- results	positive identification
Lactobacillus plantarum	+
Enterococcus faecium	+
Pediococcus pentosaceus	+
Lactobacillus casei	+
Lactobacillus fermentum	+
Lactobacillus fermentum	+
Lactobacillus brevis	+
Lactobacillus buchneri	+
Lactobacillus plantarum	+
Lactobacillus fermentum	+
Lactoccocus lactis	+
Lactococcus lactis	+

Pediococcus pentosaceus	8 spots	5 precursors	15 read sequences
Lactobacillus brevis	8 spots	5 precursors	20 read sequences
Lactobacillus casei	8 spots	5 precursors	34 read sequences
Lactobacillus plantarum	8 spots	5 precursors	17 read sequences
Enterococcus faecium	8 spots	5 precursors	12 read sequences
Lactobacillus fermentum	8 spots	5 precursors	14 read sequences
Lactoccocus lactis	8 spots	5 precursors	36 read sequences
Lactobacillus buchneri	8 spots	5 precursors	33 read sequences

Final result: 12 out of 12

Examples of sequences reading

MALDI TOF/TOF
MS and MS/MS analysis
in negative and MS/MS
positive ion mode

Low resolution MS vs. High resolution MS

PROTEIN READING

Conclusions

- For the first time high resolution mass spectrometry can exploit both positive and negative ion mode for protein and species identification to the subspecies level
- Sequence reading includes only b-ions (neg. ion mode) and y-ions (pos. ion mode) making sequence reading easy and unambiguous
- Clinical application requires quality control gained by two orthogonal techniques (MS/MS neg. and pos.)
- Sample preparation is finished in 3 h and it can be completely automatized
- Species ID is relied on genetic sequence not on the “home made” database
- Developed software named PROTEIN READER exploits the benefits accomplished by CAF/CAF chemistry for fast, accurate, robust and reliable protein and species identification

Welcome!

14th International School of Biophysics

Since 1981, the School has been introducing the participants to various vantage points and aspects of the physics, chemistry and structure of biological and soft matter systems. The participants (mostly Ph.D. students) and lecturers came from comparably diverse scientific communities.

Dates and deadlines

Check out the previous schools: 2016, 2014, 2012

DO NOT MISS: A Poem about our School

Organized by:

Co-organized by:

BM1403 NATIVE MS
CM1306 MOLECULAR MACHINES
CA15126 MOBIEU

Supported by:

Endorsed by:

Sponsored by: