MALDI-TOF mass spectrometry applied to *H. pylori* diagnosis and typing

E. Bessède, L. Bénéjat, E. Sifré,
M. Chapelle*, P. Mogabure*, F. Mégraud

CNR Campylobacters et Hélicobacters,
Hôpital Pellegrin et Université de Bordeaux, France
* Bruker, Marne la Vallée

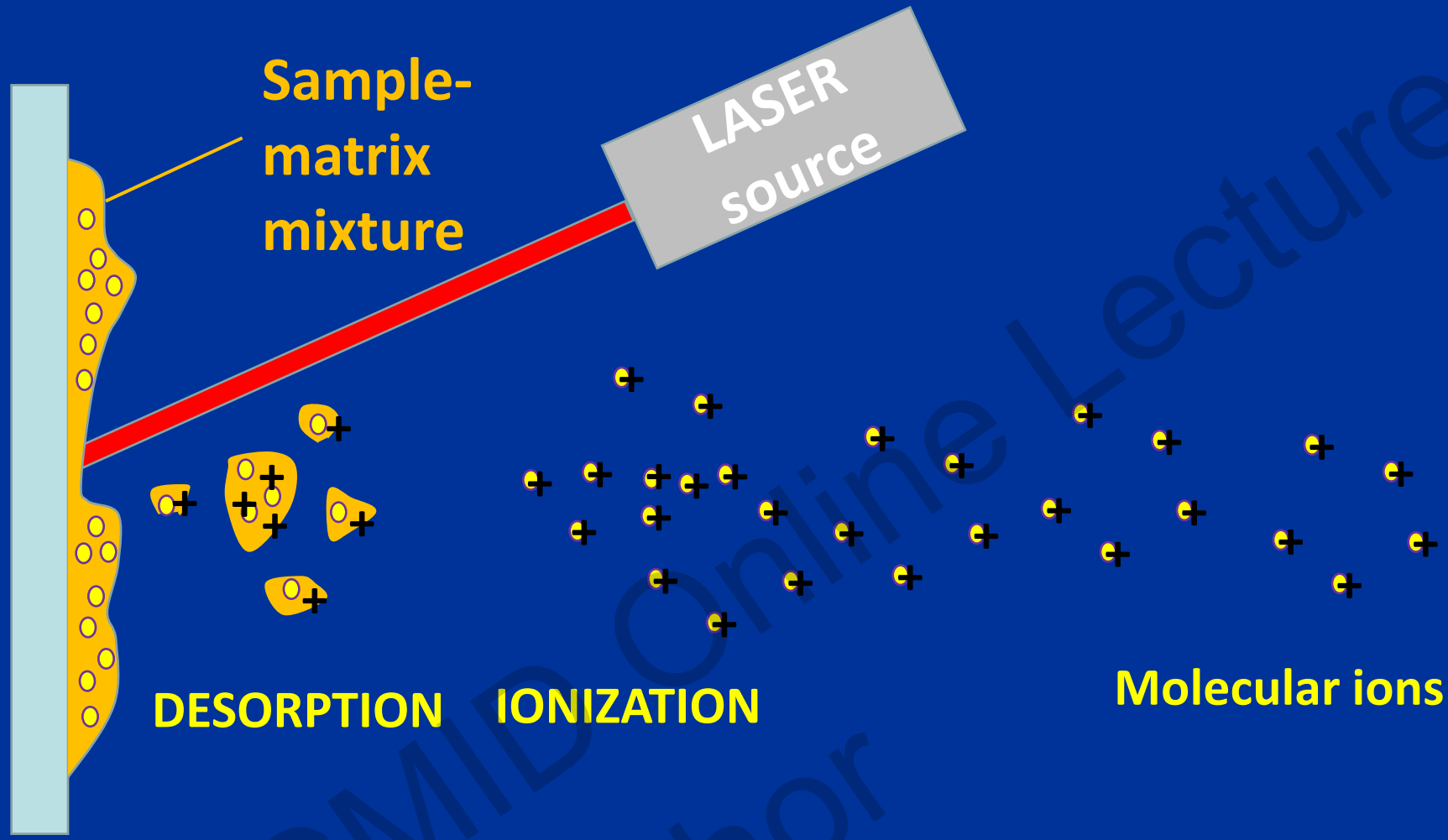
General principle of mass spectrometry

- To transform molecules from their natural state into ions in a gaseous state and get their molecular mass by analyzing the m/z ratio
- Different ionization technologies
  - Electrospray
  - MALDI: Matrix assisted laser desorption-ionization

  ➔Possible analysis of intact macromolecules and biomolecules

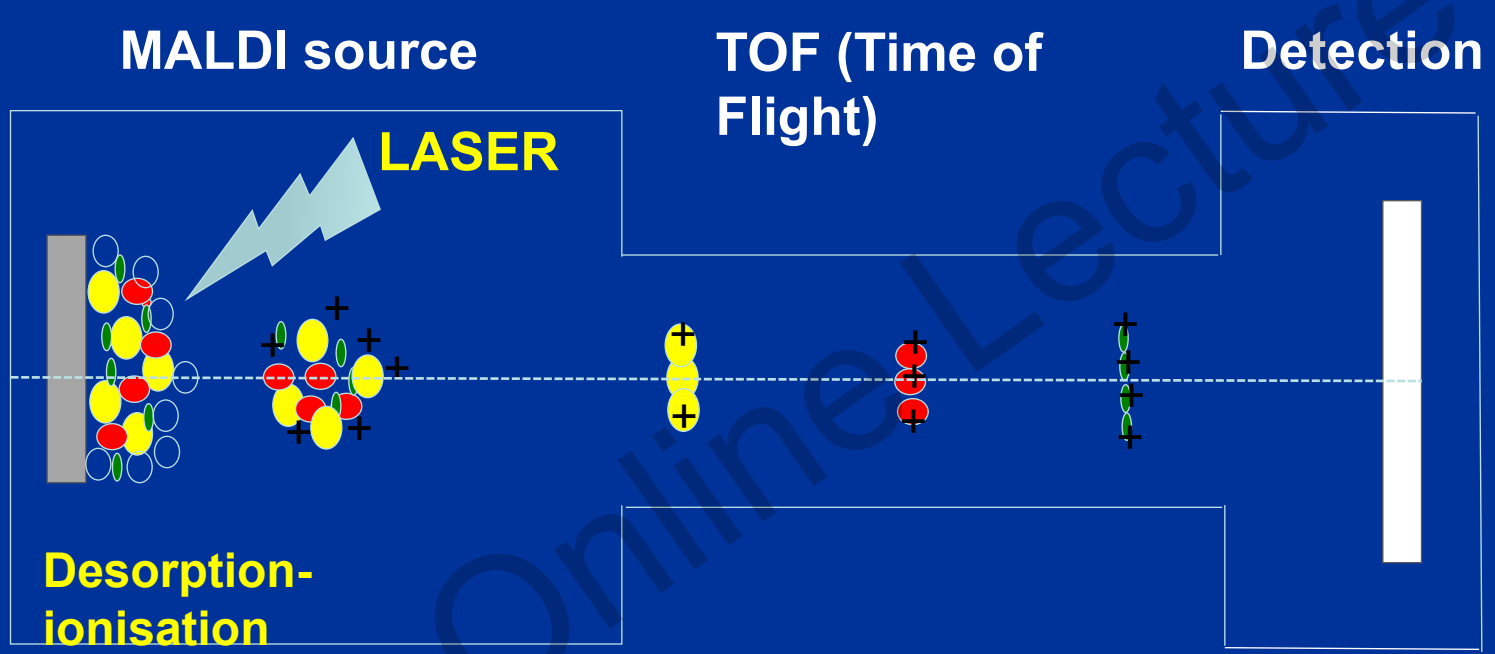
- 2000's : application to bacterial identification

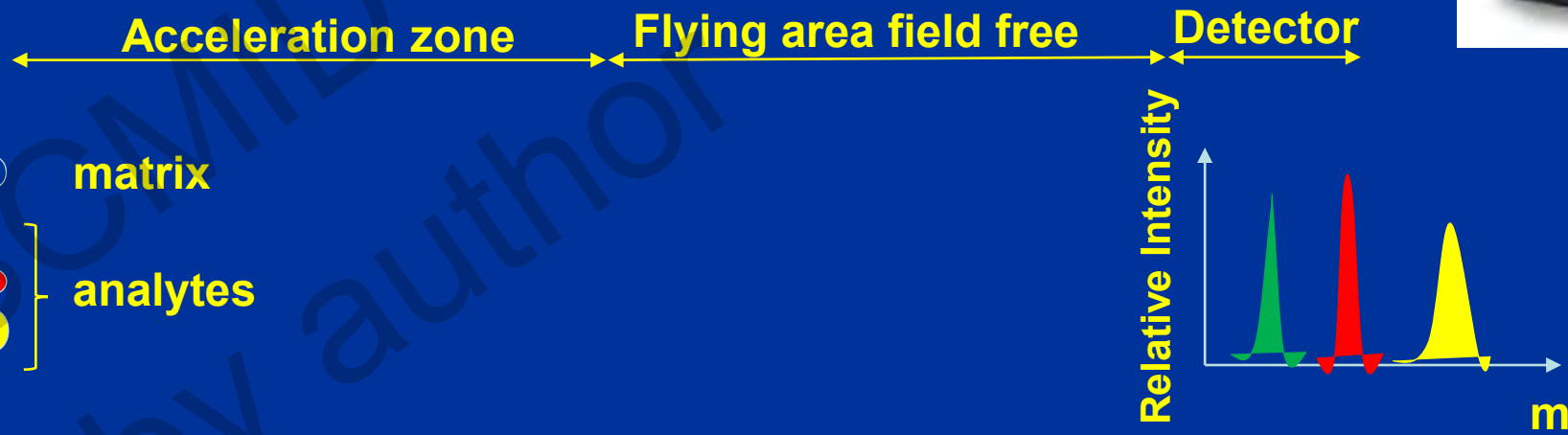
MALDI principle



Desorption and ionization of a matrix / sample mixture co-crystallized on a metal plate by a pulsed laser beam

Schema of a MALDI-TOF



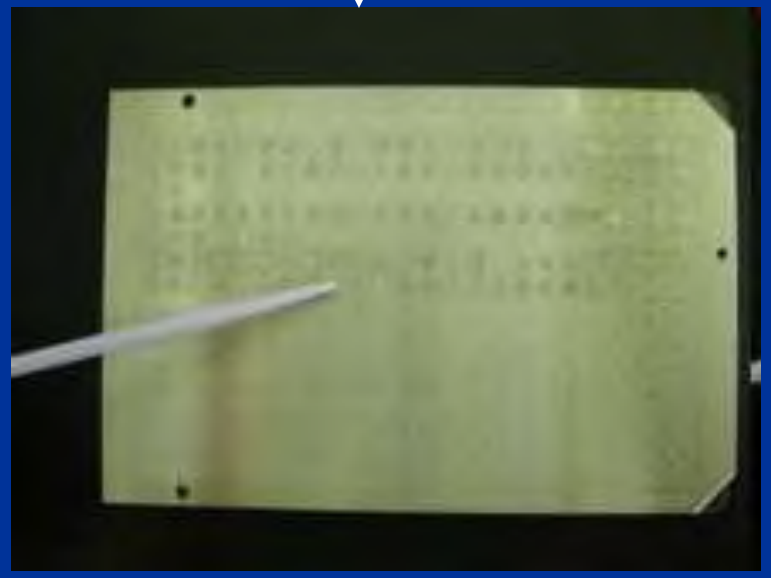


How to use the MALDI-TOF to identify bacteria?

From a fresh colony

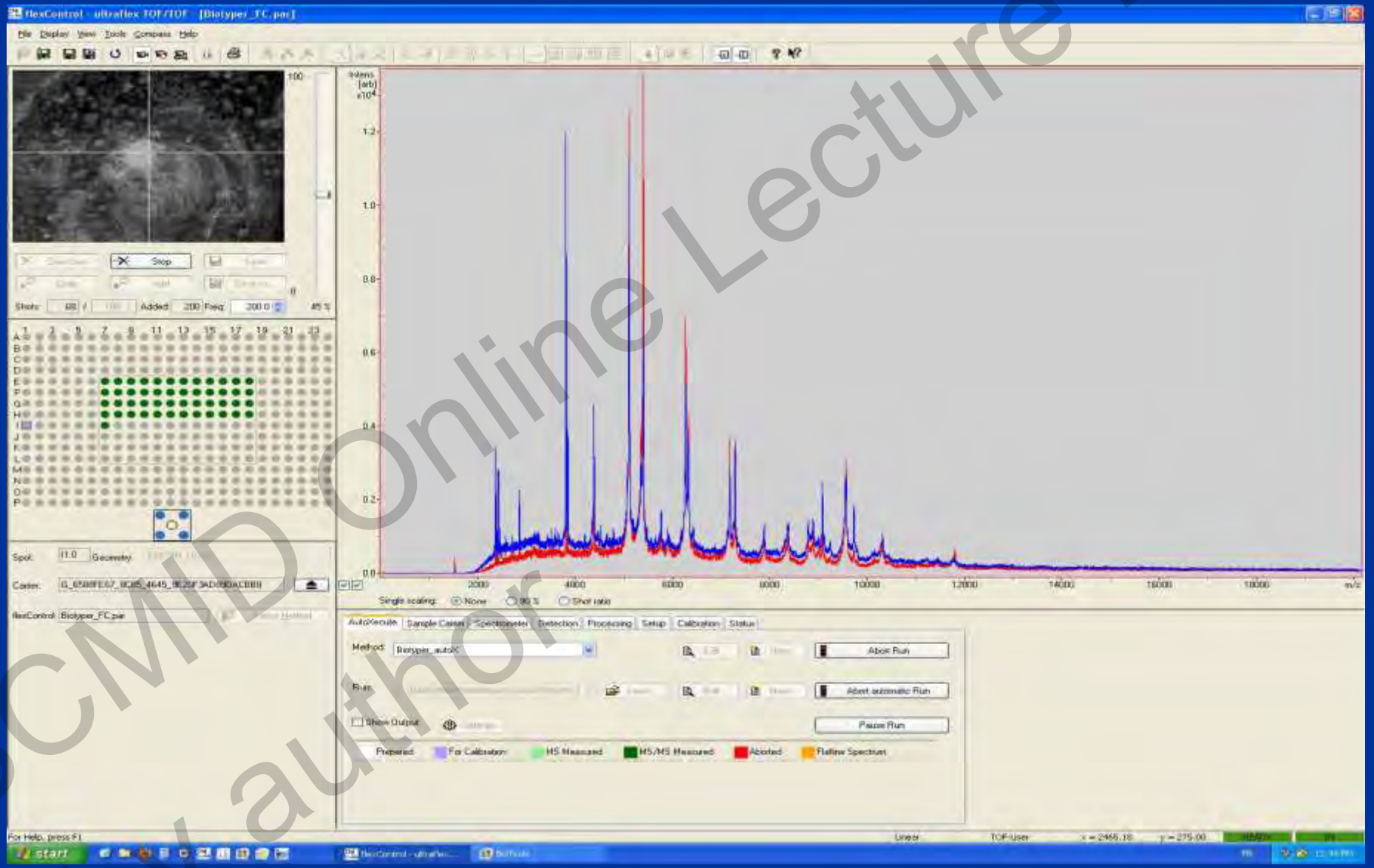
- Direct deposit on a steel MALDI target plate

or deposit after protein extraction



- Addition of 1 µL of matrix (HCCA, Bruker Daltonics)
- Dry at room temperature

Obtention of spectrum



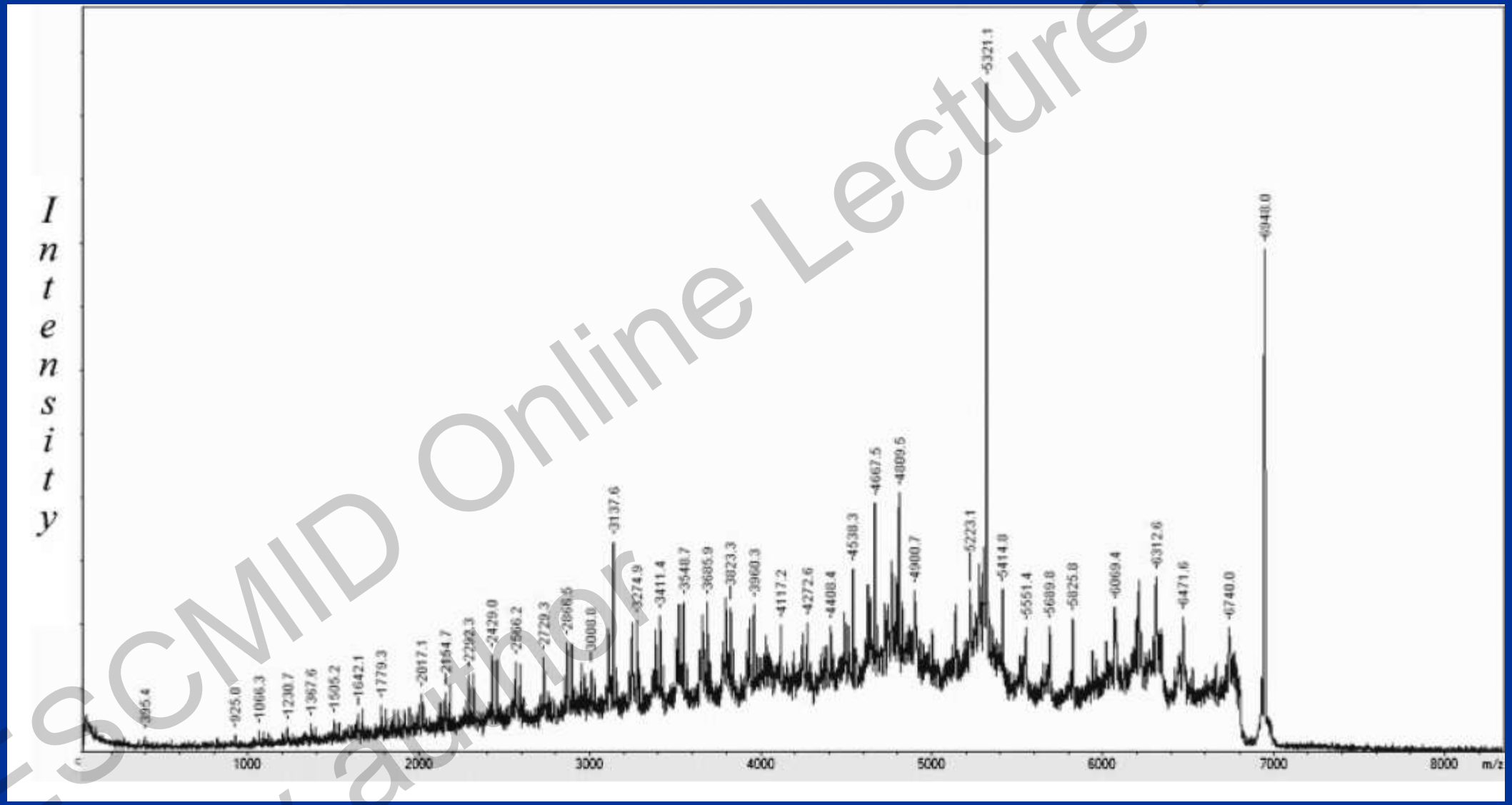
Concordance scores

Rank(Quality)	Matched Pattern	ScoreValue	NCBIIdentifier
1(++)	Campylobacter coli 11167_03 NVU	2158	195
2(++)	Campylobacter coli CCUG 11283 NVU	2149	195
3(+)	Campylobacter coli 10090_03 NVU	194	195
4(-)	Campylobacter jejuni MB_5195_05 THL	[illegible]	197
5(-)	Campylobacter jejuni ATCC 29428 THL	[illegible]	197
6(-)	Campylobacter jejuni MB_7240_05 THL	[illegible]	197
7(-)	Arthrobacter citreus DSM 20133 T_DSM	[illegible]	1670
8(-)	Campylobacter lari CCUG 23947 NVU	[illegible]	201
9(-)	Arthrobacter sp P1B386 UFL	[illegible]	1663
10(-)	Clostridium perfringens B1038_NCTC 4964_BOG	[illegible]	1502

Interpretation of the scores

- Score ≥ 2
  - Species validated
- 1.7 ≤ Score < 2
  - Genus validated
  - Species validated if ID matched several times
- Score < 1.7
  - No ID validated

Example of an *H. pylori* spectrum



Ilina *et al.*, 2010, Rapid Commun. Mass Spectrom. 2010; 24: 328–334

H. pylori identification with MALDI-TOF

Biotyper 2.0	N	%
Species	0	0
Genus	10	20.8
Not identified	38	79.2
total	48	

6 *H.pylori* / 3207 reference strains

Own database	N	%
Species	45	93.8
Genus	1	2.1
Not identified	2	4.1
total	48	

50 *H.pylori* reference strains

Bénéjat L., 2010-Abstract A 267-EHSG

H. pylori identification with MALDI-TOF

Strain	Detected species	log(Score)
Hp_13t	*Helicobacter pylori* 26695_ce PGM	**2.298**
	Helicobacter pylori J99_PGM	1.849
	Lactobacillus ingluviei 15946T_DSM	1.069
Hp_a63	*Helicobacter pylori* 26695_ce PGM	**2.040**
	Helicobacter pylori J99_PGM	1.432
	Clostridium tertium 1048_NCTC 541_BOG	1.142
Hp_57y	*Helicobacter pylori* 26695_ce PGM	1.880
	Helicobacter pylori J99_PGM	1.519
	Campylobacter coli 10090_03 NVU	1.217

Ilina *et al.*, 2010, Rapid Commun. Mass Spectrom. 2010; 24: 328–334

Comparison of 2 *H. pylori* reference strains by MALDI-TOF

- J99 & 26695
- 40 significant peaks for each mass spectrum (S/N > 10)
- Mass profiles ≠
- Correlation with macro and micro-heterogeneity of the *H. pylori* genome

M for J99	M for 26695	Description
4320	4320	RL36
5246	5246	RL34
5515		RL32
	5529	RL32
5541	5541	
6066	6066	RL33
6798*	6798*	RL28
6912	6912	
6946*	6946*	Hpn
7129*	7129*	RL35
7652	7652	RL31
	7683	RL29
7752		RL29
	7905	RL24
7915		RL24
8482	8482	RS21
	8657	
8971		RS16
	8985	RS16
	9114	
9129		
9278		
10065	10065	RS20
10260	10260	
	10384	
10414		
10448	10448	RS18
	10543	RS19
10557		

M – average value of experimental *m/z* for each peak (mass spectra collected from 10 independent experiments have been used); * – with N-terminal methionine loss.

Ilina *et al.*, 2010, Rapid Commun. Mass Spectrom. 2010; 24: 328–334

Comparison of 2 *H. pylori* reference strains by MALDI-TOF

- J99 & 26695
- 40 significant peaks for each mass spectrum (S/N > 10)
- Mass profiles ≠
- Correlation with macro and micro-heterogeneity of the *H. pylori* genome

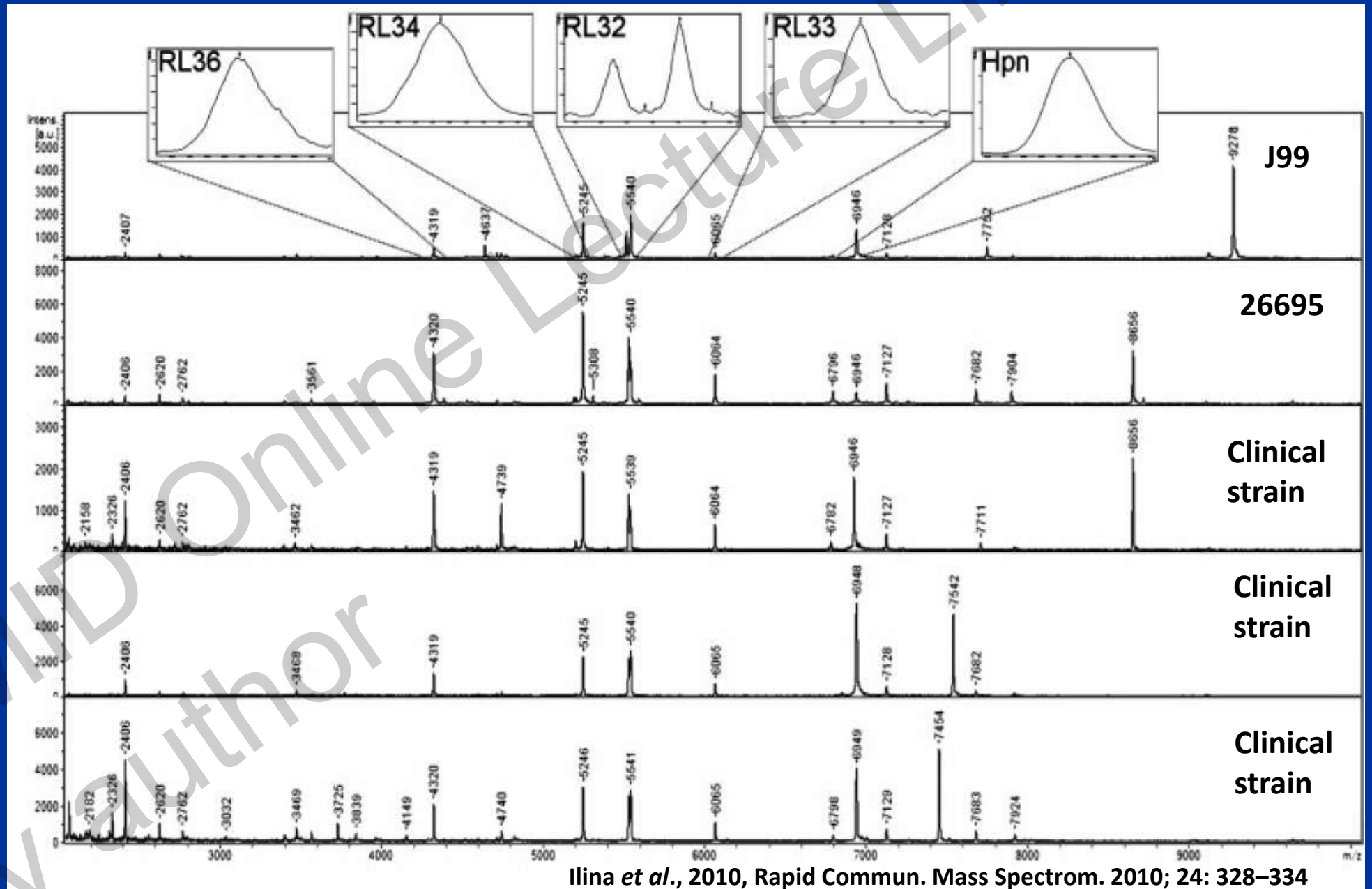
M for J99	M for 26695	Description
4320	4320	RL36
5246	5246	RL34
5515		RL32
	5529	RL32
5541	5541	
6066	6066	RL33
6798*	6798*	RL28
6912	6912	
6946*	6946*	Hpn
7129*	7129*	RL35
7652	7652	RL31
	7683	RL29
7752		RL29
	7905	RL24
7915		RL24
8482	8482	RS21
	8657	
8971		RS16
	8985	RS16
	9114	
9129		
9278		
10065	10065	RS20
10260	10260	
	10384	
10414		
10448	10448	RS18
	10543	RS19
10557		

M – average value of experimental *m/z* for each peak (mass spectra collected from 10 independent experiments have been used); * – with N-terminal methionine loss.

Ilina *et al.*, 2010, Rapid Commun. Mass Spectrom. 2010; 24: 328–334

Comparison with 17 clinical strains

Only 5 peaks were present in more than 70% of strains



Ilina *et al.*, 2010, Rapid Commun. Mass Spectrom. 2010; 24: 328–334

In summary

Because of the important heterogeneity between the strains, a specific database is required to correctly identify *H. pylori*

MALDI-TOF for *H. pylori* typing: is detection of the different pathovars possible?

Virulence factors of *H. pylori*

- **Outer membrane proteins**
  - adhésines :*bab*A2, *sab*A
  - *hop*Z, *hop*Q
- **Inflammation (IL-8 induction)**
  - *cag*PAI
  - *oip*A
  - *ice*A1
- **VacA**

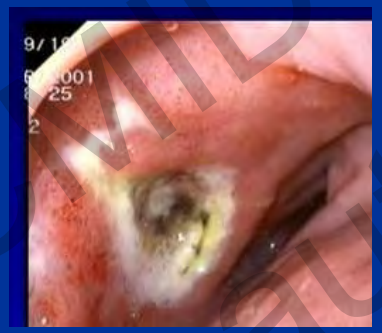
Helicobacter pylori infection and pathologies

Normal gastric mucosa

H. pylori → Chronic gastritis 100%

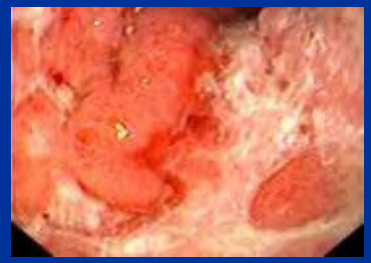
Asymptomatic

Duodenal ulcer



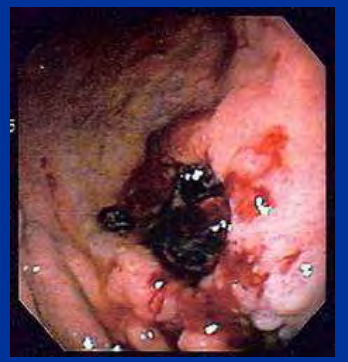
5 - 10 %

Gastric adenocarcinoma



1 %

LG - MALT



< 0,1%

AGE

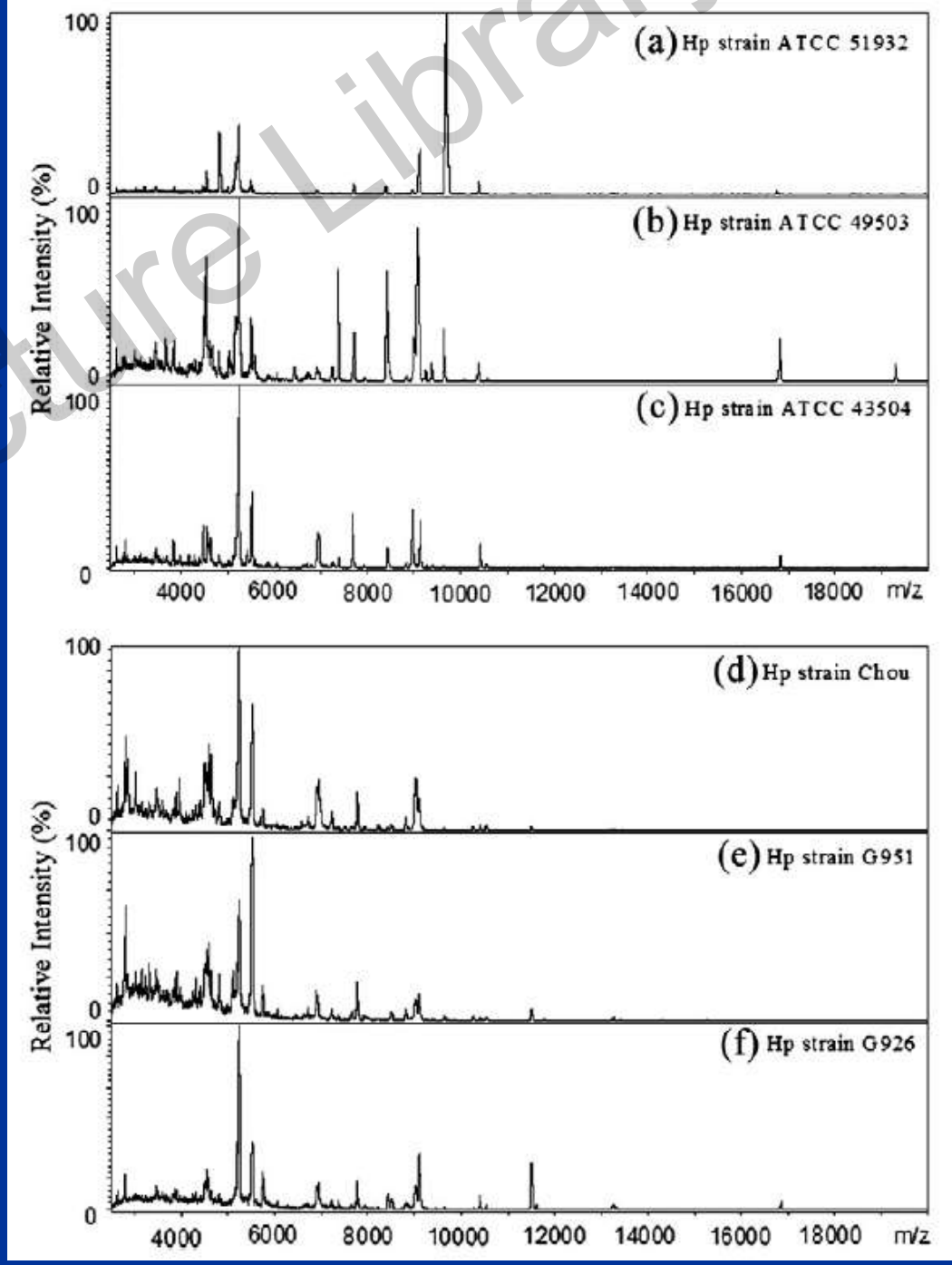
Differentiation of *H. pylori* based on pathogenicity factors

Comparison of 3 *H. pylori* reference strains and 3 clinical isolates

- 20 potential interesting peaks detected

- No peak corresponding to pathogenicity factors

→ whole protein profile analysis needed



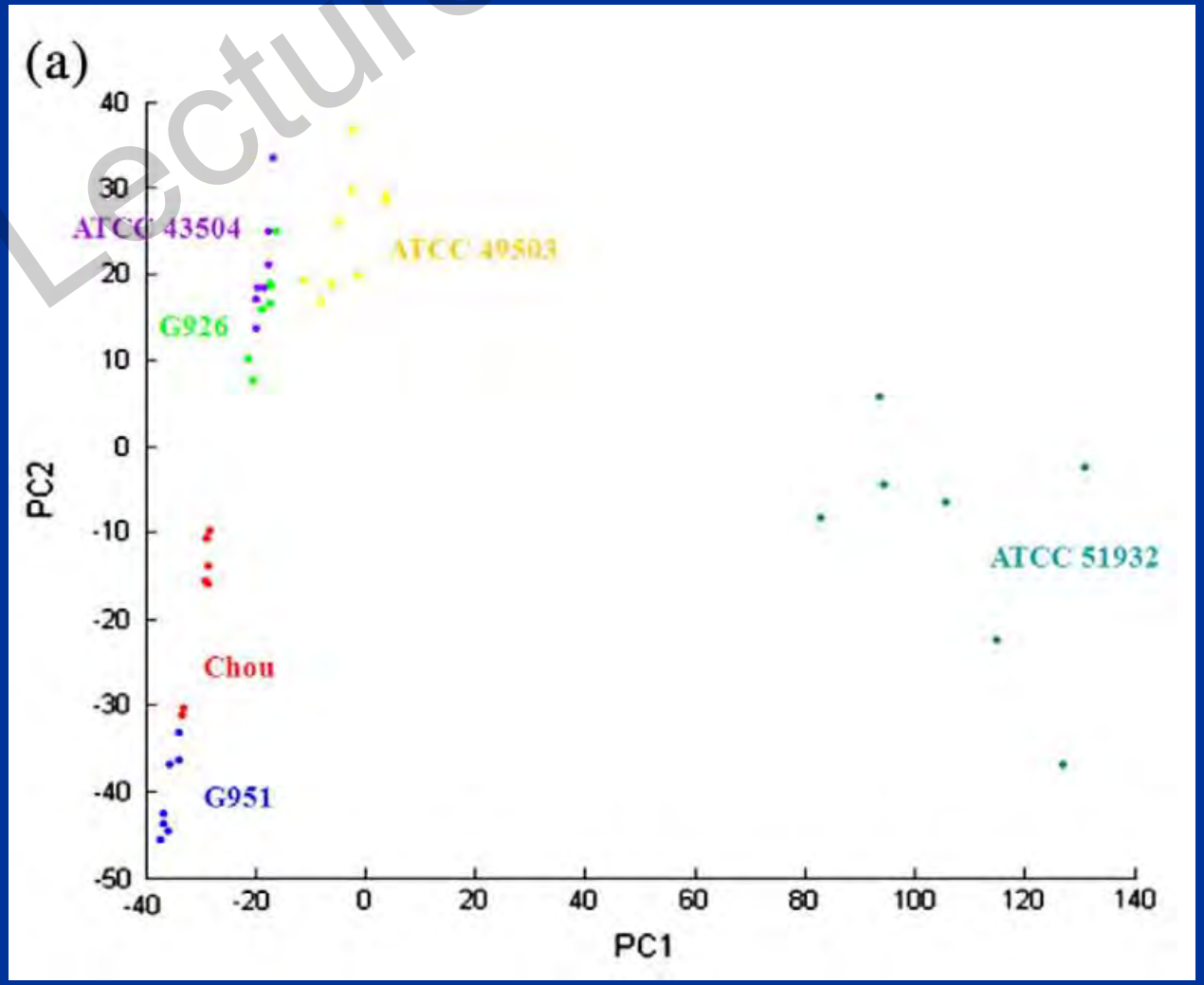
Cho *et al.*, Clinica Chimica Acta 424 (2013) 123-130

Differentiation of *H. pylori* based on pathogenicity factors

Principal Component Analysis = multivariate analysis

- Fingerprintings used to differentiate the strains

- Suitability of this approach for a rapid grouping



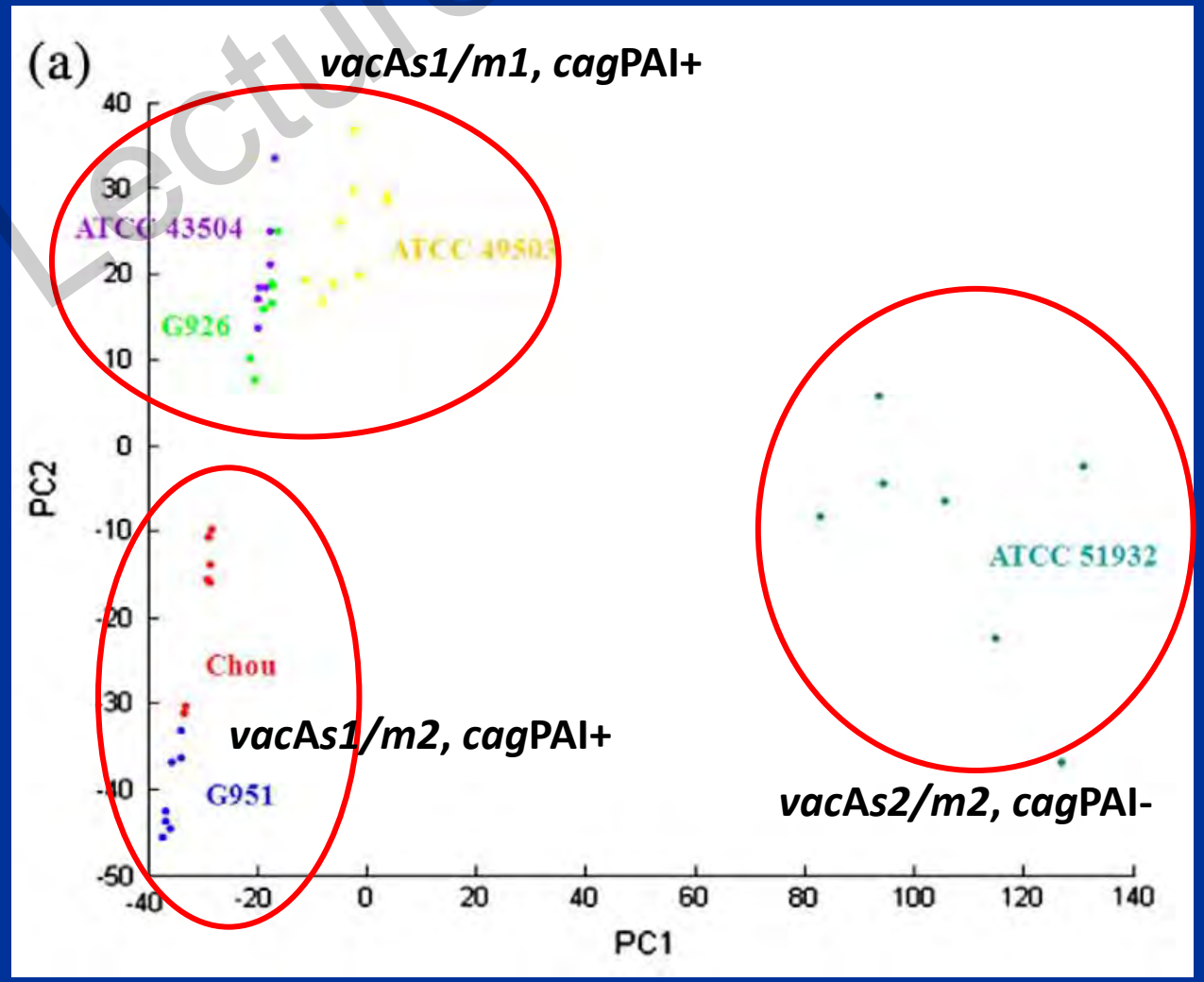
Cho *et al.*, Clinica Chimica Acta 424 (2013) 123-130

Differentiation of *H. pylori* based on pathogenicity factors

Principal Component Analysis = multivariate analysis

- Fingerprintings used to differentiate the strains

- Suitability of this approach for a rapid grouping



Cho *et al.*, Clinica Chimica Acta 424 (2013) 123-130

Typing *H. pylori*
A preliminary study

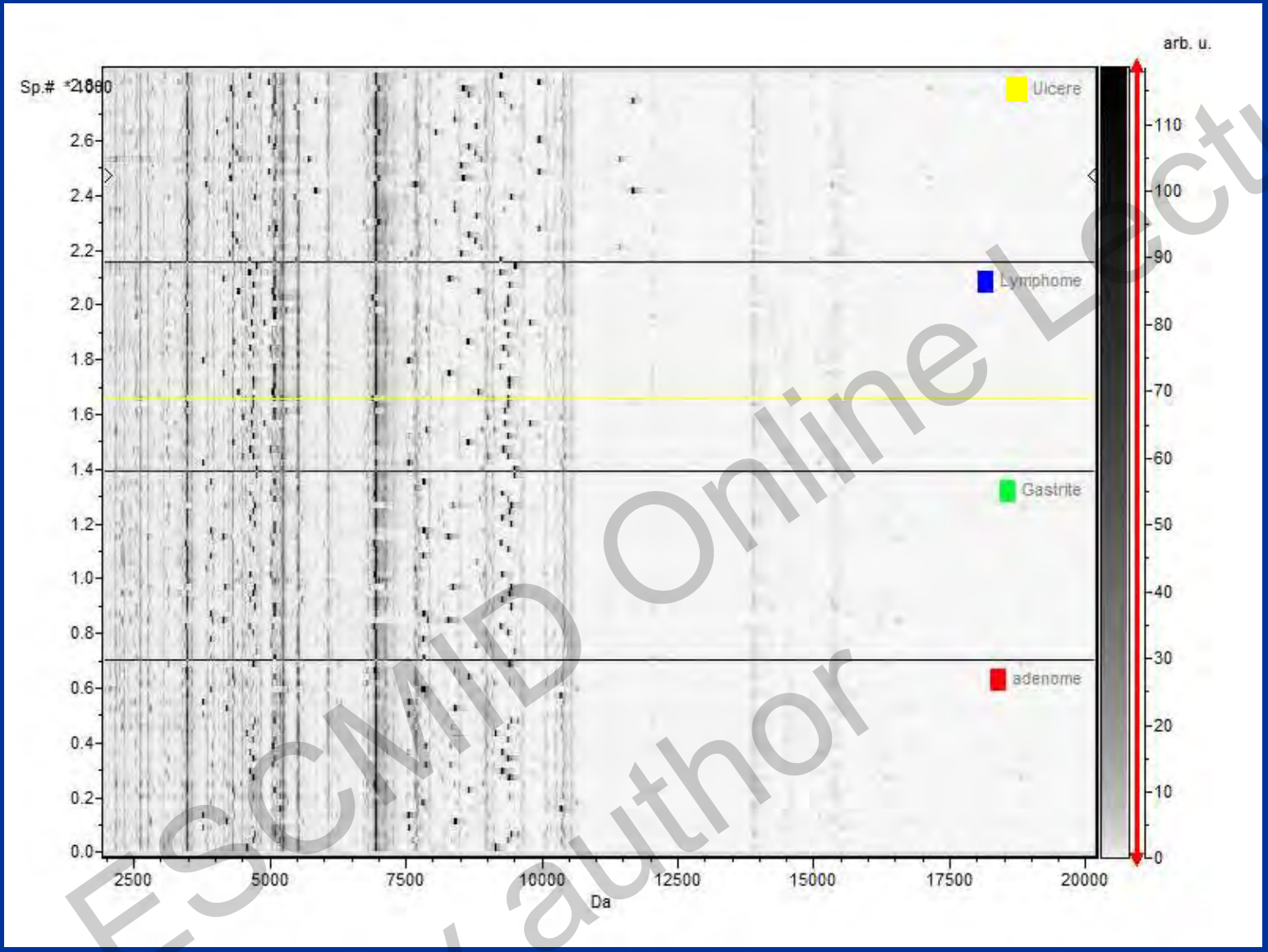
Material

62 strains composed of 4 pathovars:

- 15 Adenocarcinoma
- 16 Gastric MALT lymphoma
- 15 Gastritis
- 16 Peptic ulcer disease (PUD)

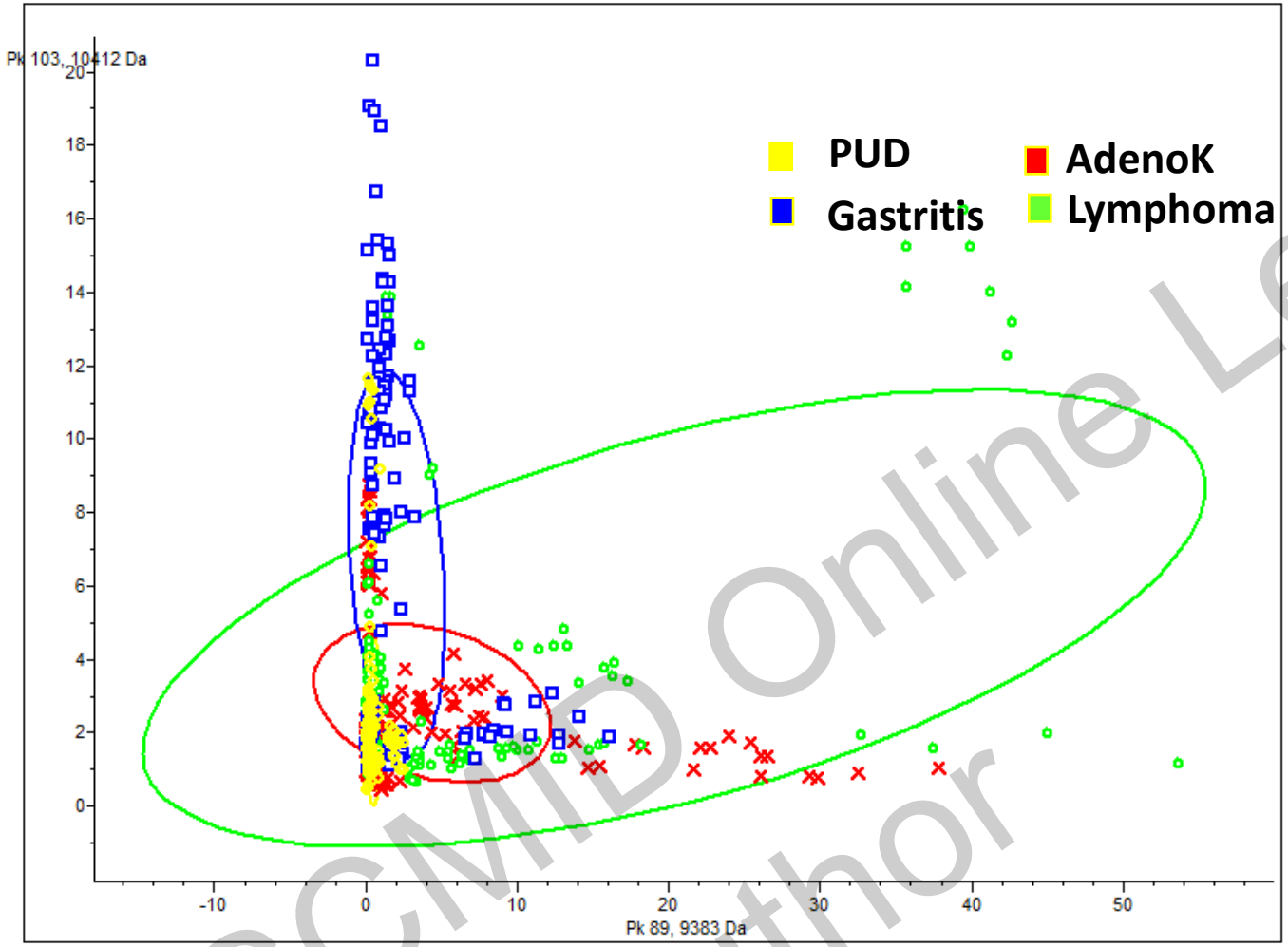
For each strain: 2 cultures (incubation of 48h)
For each culture: 24 deposits (obtained from a suspension of 3McF)
For each deposit: 1 measure
$\Rightarrow \approx$ 3000 spectra

Spectra obtained with Biotyper 3.0



Large heterogeneity into each pathovar (No common peaks)

Analysis of all the data by ANOVA-test



Intensities of the 2 peaks (10412Da *vs*. 9383Da) with the lowest t-test (ANOVA)

Superposition of the 4 groups → no possible use as a specific marker

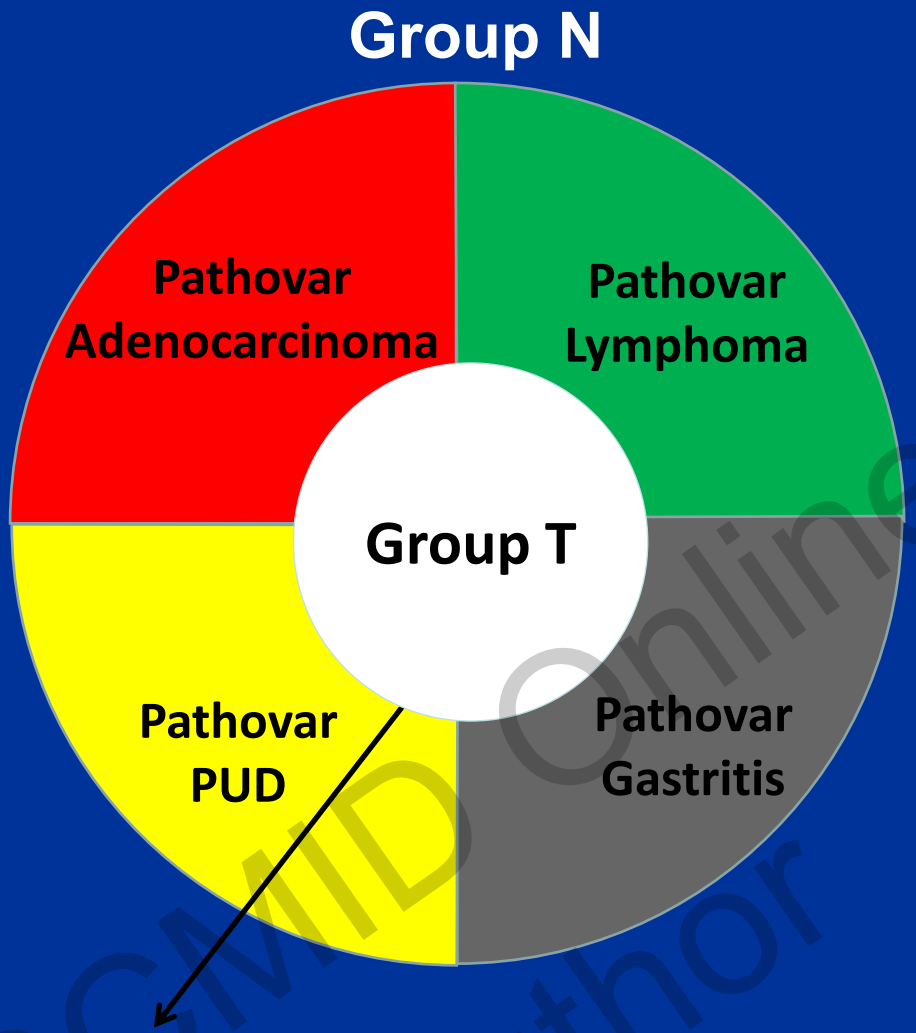
Idem for all of the peaks taken individually

No possibility to classify the strains when using only a few peaks → multiplex analysis is needed

Typing *H. pylori* with ClinProTools 3.0 rel 2.2 (Bruker)

- **Algorithms tested for classification**
  - Genetical Algorithm
  - Geometrical Algorithm
  - Informatical Algorithm
  - Statistical analysis Algorithm

Validation of the algorithms



Used for modelisation
(140 spectra per pathovar = 20% of the entire data)

- **2 steps for validation:**

- **Modelisation**

Training data are injected into a classification model
-> estimation of the recognition capability of the model

- **Validation** of the model used
-> calculate the recognition capability

Results: training data

Model name	Recognition capability
Genetical Algorithm	99.46%
Geometrical Algorithm	97.67%
Informatical Algoritm	67.93%
Statistical analysis Algorithm	41.26%

Results: validation data

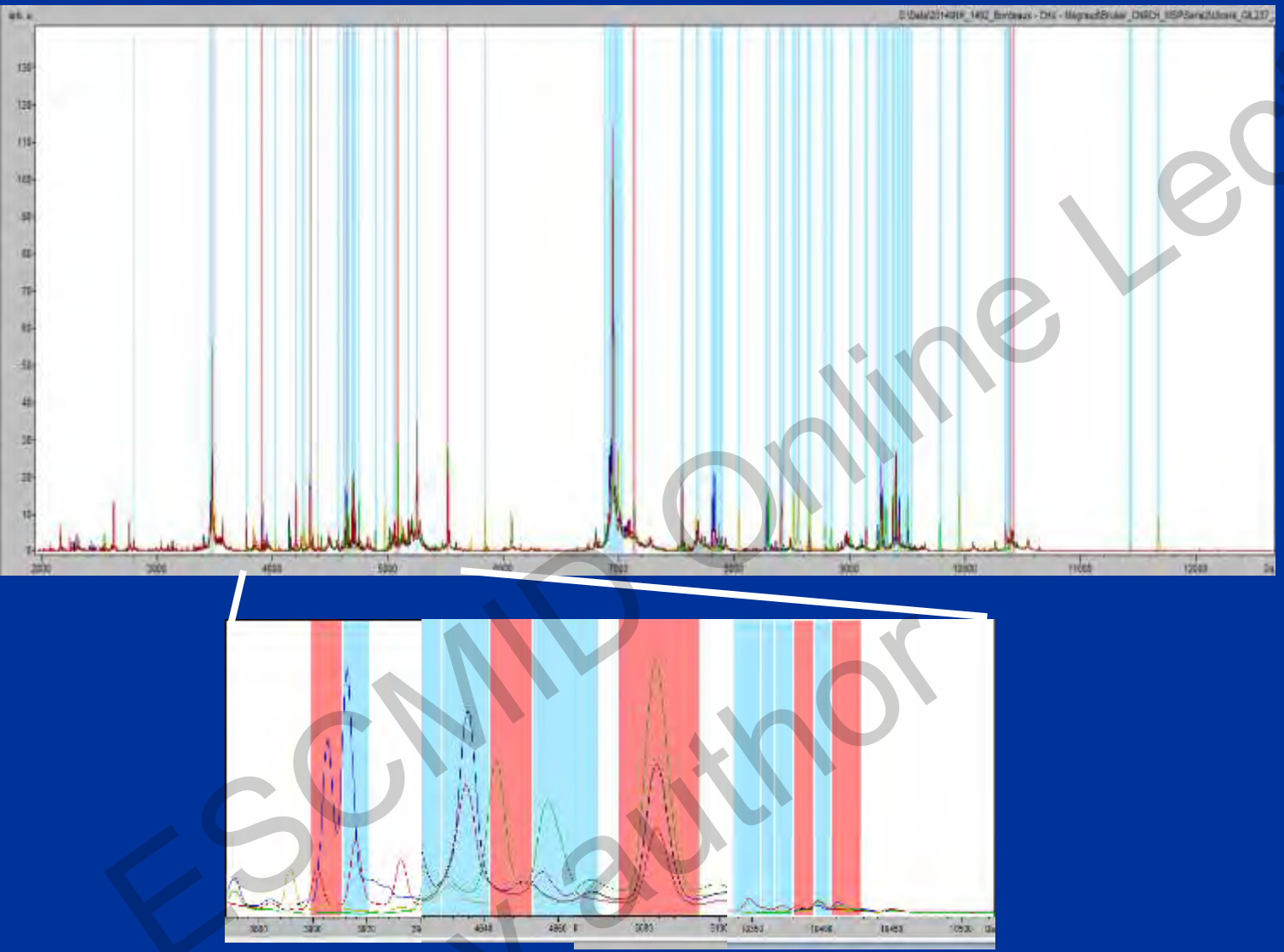
Validation performed on the 80% of the data which were not used for model construction

Recognition rates:

Adenocarcinoma	90.8%
MALT lymphoma	97.6%
Gastritis	97.6%
PUD	98.2%

ClinProt Model

Peaks used with the genetic model to separate the pathovars



Index	Masse
23	4644,4
38	5517,5
6	3473,9
101	10383,3
52	7131,2
103	10412,0
35	5083,2
10	3905,6
106	13891,1
17	4325,0

Comparison « Cancer » *vs.* « No cancer » strains

Non cancerous : gastritis + PUD

Cancer : adenocarcinoma + lymphoma

Model name	Recognition capability
Genetical Algorithm	100%
Geometrical Algorithm	96.72%
Informatical Algorithm	87.39%
Statistical analysis Algorithm	62.49%

Index	Mass
43	5517,59
42	5248,49
122	13893,47
117	10393,98
60	7685,27
91	8966,56
57	7131,19
37	5015,28
90	8959,56
53	6965,18

Perspectives

1. To consolidate results on a larger strains library

2. To optimize reproducibility

3. To identify the proteins detected as markers to check if they correspond to pathogenicity factors

Conclusion

H. pylori identification: difficult but possible

H. pylori typing: also possible but still needs to be optimized

Algorithmes de classification(analyse supervisée)

- **Algorithme génétique** (GA) : Chaque pathovar est une famille, chaque spectre est considéré comme un individu avec un seul chromosome et chaque pic est considéré comme un gène de ce chromosome. Une méthode itérative simule des croisements avec des taux de mutation et de cross-over aléatoires et tente de déterminer quels gènes transmis permettent d'identifier la famille d'appartenance des ancêtres, les individus de départ. La méthode est particulièrement adaptée à la modélisation de processus biologiques. Le résultat peut varier d'un calcul à l'autre en raison de l'appel à des méthodes aléatoires de génération des croisements

- Machine a vecteur support (SVM) : **algorithme géométrique** qui tente de trouver les surfaces de plus grand pouvoir séparateur entre les individus des chaque groupe (régressions hyperplanaires).

- Réseau de neurones supervisé (SNN) : **algorithme informatique** qui mime un réseau de neurones produisant un arbre de décision construit branche par branche avec chaque nouvel individu du jeu de données d'entrainement, conduisant à diriger l'individu vers son groupe d'origine

- Quick Classifier : algorithme basé uniquement sur l'analyse statistique du jeu de données : les pics de meilleur score de t-test sont intégrés dans le modèle.

Validation en deux étapes :
- Les données du jeu d'entrainement sont injectées dans le modèle pour classification -> calcul de la capacité de reconnaissance du modèle : donne une idée de la puissance du modèle
- Un jeu de données de taille supérieure au jeu d'entrainement (jusqu'à 10 fois plus), et dont le groupe d'appartenance des individus est connu est injecté dans le modèle et les taux de reconnaissance globale et par classe est calculé. Si la marge d'erreur est acceptable, le modèle peut ensuite être utilisé pour la classification d'individus inconnus.

Genetical algorithm

- Each pathovar is a family. Each spectrum is considered as a human with one chromosome and each peak is considered as a gene of this chromosom. An iterative method simulates crosses with mutation rate and random cross-over, trying to identify genes transmitted allowing to identify family membership ancestors. This method is adapted to biological processes modelisation.

H. pylori's identification with MALDI-TOF

Classement(Qualité)	Profil de référence	Score	IdentifiantNCBI
1(-)	Helicobacter pylori DSM 10242 DSM	1.678	210
2(-)	Helicobacter pylori DSM 9691 DSM	1.477	210
3(-)	Helicobacter pylori 151 RLT	1.405	210
4(-)	Helicobacter pylori J99 PGM	1.35	210
5(-)	Lactobacillus aviarius ssp aviarius DSM 20654 DSM	1.323	147810
6(-)	Helicobacter pylori DSM 7492 DSM	1.268	210
7(-)	Lactobacillus paracasei ssp paracasei DSM 20244 DSM	1.244	47714
8(-)	Enterococcus faecium 20218_1 CHB	1.213	1352
9(-)	Clostridium sordellii 1070_ATCC 9714T BOG	1.144	1505
10(-)	Brevundimonas diminuta DSM 7234T HAM	1.14	293

Scores < 2, but several matches

Possible without the extraction step

Low scores?