
ADVANTAGES AND DISADVANTAGES OF WEB-BASED SYSTEMS TO IDENTIFY DNA PROFILES

Alexander Indra

AGES-Institut für medizinische Mikrobiologie und Hygiene Wien

LAB-based database Systems: *Clostridium difficile*

- one laboratory responsible for nomenclature and identification
 - Laboratory resources are limited
 - only few had/have access to typing information
- till now various „localised“ nomenclature systems
 - AI, SW,
- limited availability of a standard strain collection
 - till 2008 !!! no international strain collection available
 - strain collection sample size is limited
 - possibility for errors
 - strain collection is always months or years behind of the users epidemiological development

INTERNET vs LAB based Database Systems

•INTERNET

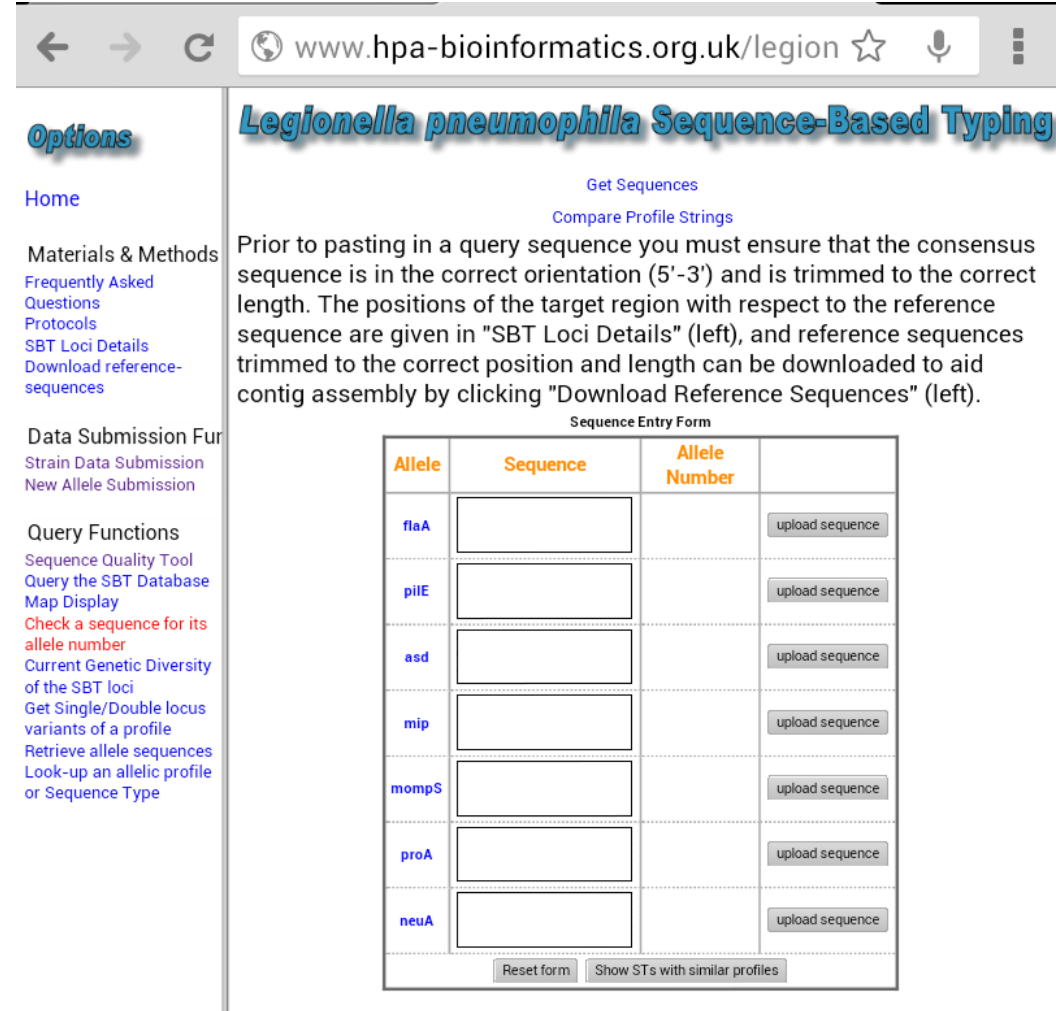
- **Up to date internet-database for all users**
- **Only** the sequencer **file is needed** to define a new type
- **No additional costs**
 - no soft- or hardware is needed
- **No restriction** for sequencer, gel, size-standard or primer-set
- **ALL users have access** to new Data
- **No shipping** costs

•LOCAL

- Up to date database for **one/few local users**
- **Strain must be send** to reference Lab
- **Additional costs**
 - Soft- or hardware **is needed**
- **Methods restricted to achieve minimal exchangeability**
- **only local User** can access data
- **shipping costs**

WEB-BASED SYSTEMS TO IDENTIFY DNA PROFILES

- available systems
 - **HPA-Legionella SBT Typing**



The screenshot shows the HPA-Legionella SBT Typing web interface. The browser address bar displays www.hpa-bioinformatics.org.uk/legion. The page title is **Legionella pneumophila Sequence-Based Typing**. The left sidebar contains navigation links under the heading **Options**, including [Home](#), [Materials & Methods](#), [Frequently Asked Questions](#), [Protocols](#), [SBT Loci Details](#), [Download reference-sequences](#), [Data Submission Form](#), [Strain Data Submission](#), [New Allele Submission](#), [Query Functions](#), [Sequence Quality Tool](#), [Query the SBT Database](#), [Map Display](#), [Check a sequence for its allele number](#), [Current Genetic Diversity of the SBT loci](#), [Get Single/Double locus variants of a profile](#), [Retrieve allele sequences](#), and [Look-up an allelic profile or Sequence Type](#). The main content area includes links for [Get Sequences](#) and [Compare Profile Strings](#). A text block explains that prior to pasting a query sequence, the user must ensure the consensus sequence is in the correct orientation (5'-3') and is trimmed to the correct length. The positions of the target region with respect to the reference sequence are given in "SBT Loci Details" (left), and reference sequences trimmed to the correct position and length can be downloaded to aid contig assembly by clicking "Download Reference Sequences" (left). Below this text is the **Sequence Entry Form**, which is a table with four columns: **Allele**, **Sequence**, **Allele Number**, and an empty column. The table contains seven rows for different alleles: **flaA**, **pilE**, **asd**, **mip**, **mompS**, **proA**, and **neuA**. Each row has a text input field for the sequence and a button labeled "upload sequence". At the bottom of the form are two buttons: "Reset form" and "Show STs with similar profiles".

Options

- [Home](#)
- [Materials & Methods](#)
- [Frequently Asked Questions](#)
- [Protocols](#)
- [SBT Loci Details](#)
- [Download reference-sequences](#)
- [Data Submission Form](#)
- [Strain Data Submission](#)
- [New Allele Submission](#)
- [Query Functions](#)
- [Sequence Quality Tool](#)
- [Query the SBT Database](#)
- [Map Display](#)
- [Check a sequence for its allele number](#)
- [Current Genetic Diversity of the SBT loci](#)
- [Get Single/Double locus variants of a profile](#)
- [Retrieve allele sequences](#)
- [Look-up an allelic profile or Sequence Type](#)

[Get Sequences](#)
[Compare Profile Strings](#)

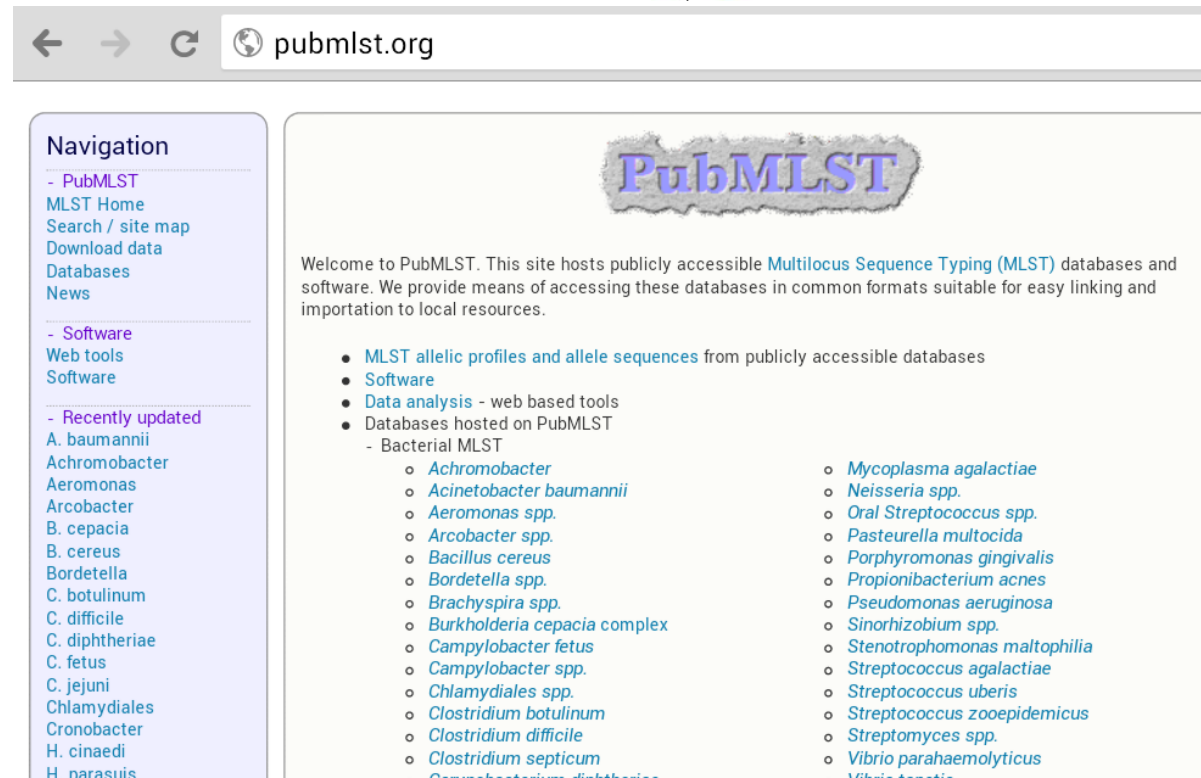
Prior to pasting in a query sequence you must ensure that the consensus sequence is in the correct orientation (5'-3') and is trimmed to the correct length. The positions of the target region with respect to the reference sequence are given in "SBT Loci Details" (left), and reference sequences trimmed to the correct position and length can be downloaded to aid contig assembly by clicking "Download Reference Sequences" (left).

Sequence Entry Form

Allele	Sequence	Allele Number	
flaA			upload sequence
pilE			upload sequence
asd			upload sequence
mip			upload sequence
mompS			upload sequence
proA			upload sequence
neuA			upload sequence

WEB-BASED SYSTEMS TO IDENTIFY DNA PROFILES

- available systems
 - HPA-Legionella SBT Typing
 - various MLST-Databases



The screenshot shows the PubMLST website interface. At the top, there is a browser address bar with the URL 'pubmlst.org'. Below the address bar, the 'Navigation' menu is visible on the left, listing links for 'PubMLST', 'MLST Home', 'Search / site map', 'Download data', 'Databases', 'News', 'Software', 'Web tools', and 'Software'. The main content area features the 'PubMLST' logo and a welcome message: 'Welcome to PubMLST. This site hosts publicly accessible Multilocus Sequence Typing (MLST) databases and software. We provide means of accessing these databases in common formats suitable for easy linking and importation to local resources.' Below this, there are three main categories of resources: 'MLST allelic profiles and allele sequences from publicly accessible databases', 'Software', and 'Data analysis - web based tools'. The 'Databases hosted on PubMLST' section is expanded, showing a list of bacterial MLST databases, including 'Achromobacter', 'Acinetobacter baumannii', 'Aeromonas spp.', 'Arcobacter spp.', 'Bacillus cereus', 'Bordetella spp.', 'Brachyspira spp.', 'Burkholderia cepacia complex', 'Campylobacter fetus', 'Campylobacter spp.', 'Chlamydiales spp.', 'Clostridium botulinum', 'Clostridium difficile', 'Clostridium septicum', 'Mycoplasmatales spp.', 'Neisseria spp.', 'Oral Streptococcus spp.', 'Pasteurella multocida', 'Porphyromonas gingivalis', 'Propionibacterium acnes', 'Pseudomonas aeruginosa', 'Sinorhizobium spp.', 'Stenotrophomonas maltophilia', 'Streptococcus agalactiae', 'Streptococcus uberis', 'Streptococcus zooepidemicus', 'Streptomyces spp.', and 'Vibrio parahaemolyticus'.

WEB-BASED SYSTEMS TO IDENTIFY DNA PROFILES

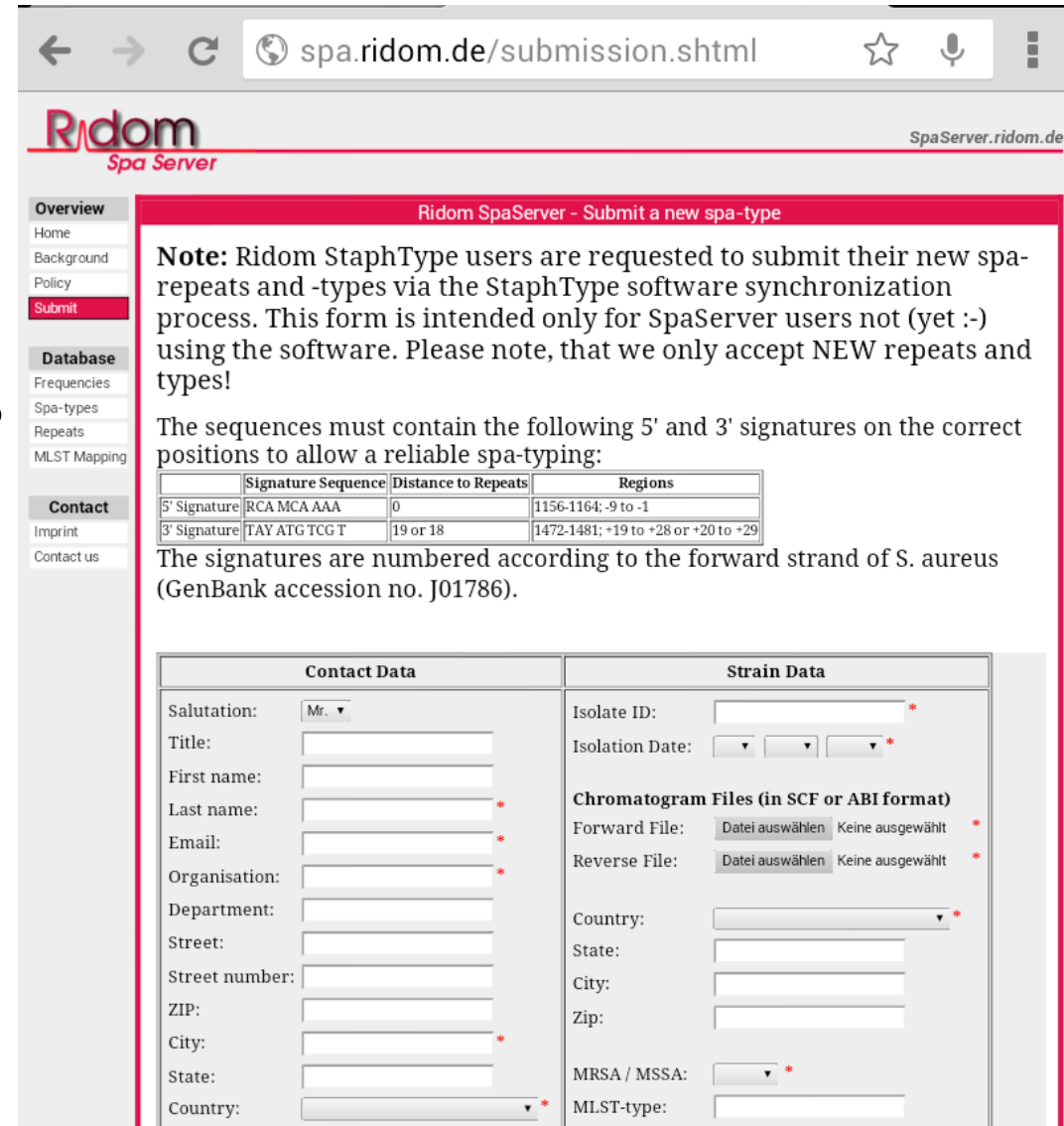
- available systems
 - HPA-Legionella SBT Typing
 - various MLST-Databases
 - **MIRU-VNTRplus**



The screenshot shows the MIRU-VNTRplus web application interface. At the top, there is a navigation bar with links: Home, Browse Database, Identification by Similarity Search, Nomenclature, Background, Policy, Help, About us, Contact us, and Imprint. The main content area welcomes users to the application and provides instructions on how to use the Home button and the Navigate menu. It also displays the number of user strains uploaded (0) and offers options to enter a single user strain or import multiple strains from a file or clipboard. Below this, there are settings options: Change VNTR loci set (24 loci), Change VNTR loci order, and Reset default parameters. A section titled 'Molecular typing of bacteria from the Mycobacterium tuberculosis complex (MTBC)' describes the importance of MIRU typing for epidemiological purposes and mentions the use of a web server for implementing a web application. The text states that a collection of 186 strains representing the major MTBC lineages was used for implementing the web server, MIRU-VNTRplus (http://www.miru-vntrplus.org/). For each strain species, lineage, and epidemiologic information was stored together with copy numbers of 24 MIRU loci, spoligotyping patterns, regions of difference (RD) profiles, single nucleotide polymorphisms (SNPs), susceptibility data, and IS6110 RFLP fingerprint images.

WEB-BASED SYSTEMS TO IDENTIFY DNA PROFILES

- available systems
 - HPA-Legionella SBT Typing
 - various MLST-Databases
 - MIRU-VNTRplus
 - **RIDOM-StaphType**



The screenshot shows the 'Ridom SpaServer - Submit a new spa-type' page. It includes a navigation menu on the left with sections: Overview (Home, Background, Policy, Submit), Database (Frequencies, Spa-types, Repeats, MLST Mapping), and Contact (Imprint, Contact us). The main content area contains a note about submitting new spa-repeats and -types via the StaphType software synchronization process. Below the note is a table showing the required 5' and 3' signatures for S. aureus. The bottom section contains two columns of form fields: 'Contact Data' (Salutation, Title, First name, Last name, Email, Organisation, Department, Street, Street number, ZIP, City, State, Country) and 'Strain Data' (Isolate ID, Isolation Date, Chromatogram Files (Forward and Reverse), Country, State, City, Zip, MRSA / MSSA, MLST-type).

← → ↻ 🌐 spa.ridom.de/submission.shtml ☆ 🔊

Ridom
Spa Server

SpaServer.ridom.de

Ridom SpaServer - Submit a new spa-type

Note: Ridom StaphType users are requested to submit their new spa-repeats and -types via the StaphType software synchronization process. This form is intended only for SpaServer users not (yet :) using the software. Please note, that we only accept NEW repeats and types!

The sequences must contain the following 5' and 3' signatures on the correct positions to allow a reliable spa-typing:

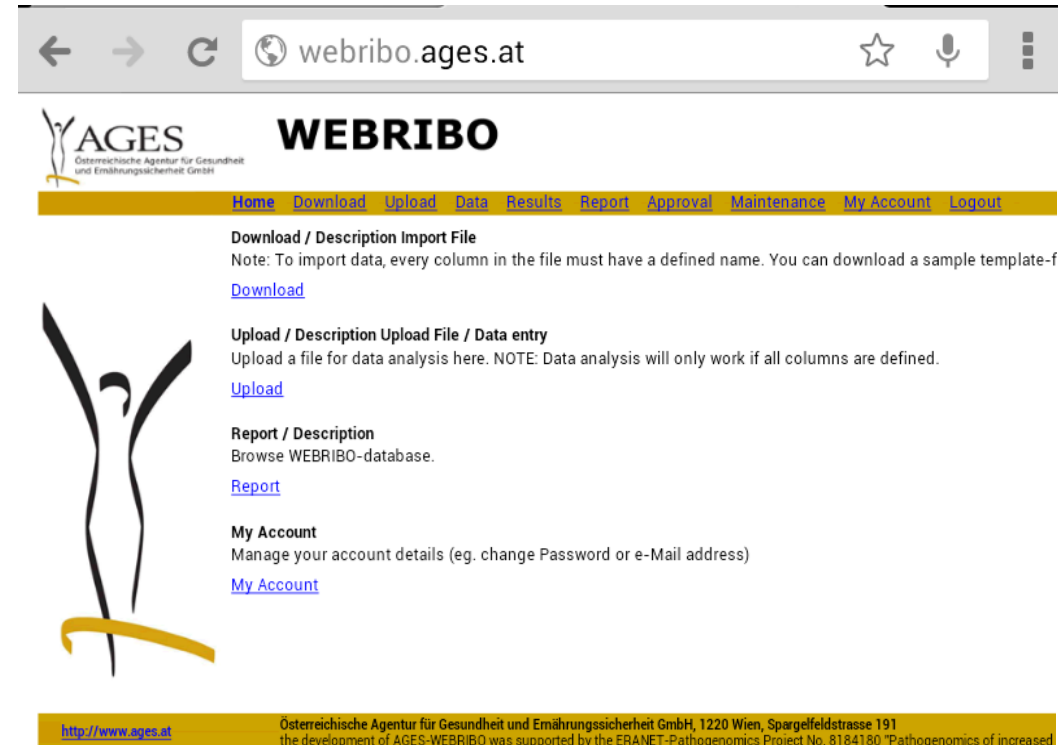
	Signature Sequence	Distance to Repeats	Regions
5' Signature	RCA MCA AAA	0	1156-1164; -9 to -1
3' Signature	TAY ATG TCG T	19 or 18	1472-1481; +19 to +28 or +20 to +29

The signatures are numbered according to the forward strand of *S. aureus* (GenBank accession no. J01786).

Contact Data		Strain Data	
Salutation:	Mr. ▼	Isolate ID:	
Title:		Isolation Date:	▼ ▼ ▼
First name:		Chromatogram Files (in SCF or ABI format)	
Last name:		Forward File:	Datei auswählen Keine ausgewählt
Email:		Reverse File:	Datei auswählen Keine ausgewählt
Organisation:		Country:	▼
Department:		State:	
Street:		City:	
Street number:		Zip:	
ZIP:		MRSA / MSSA:	▼
City:		MLST-type:	
State:			
Country:	▼		

WEB-BASED SYSTEMS TO IDENTIFY DNA PROFILES

- available systems
 - HPA-Legionella SBT Typing
 - various MLST-Databases
 - MIRU-VNTRplus
 - RIDOM-StaphType
 - **AGES-WEBRIBO**



the only internet database doing
automated fragment analysis

WEB-BASED SYSTEMS TO IDENTIFY DNA PROFILES: **Cons**

- analysis algorithm **quite sensitive**
 - HPA-Legionella
- **RAW-data** import for **Beckman or Licor-Systems** not always available
 - most Databases missing RAW-Data Import
- **nomenclature can be divergent**
 - overdiscrimination
 - Kreiswirth vs. RIDOM, AGES-WEBRIBO - Cardiff
- **Analysis can take some time**

WEB-BASED SYSTEMS TO IDENTIFY DNA PROFILES: **Wishlist**

- **Freely accessible constantly updated internet-database**
- **ABI gained files** are **analyzed** automatically
- **Reliable results** regardless of any primer, gel, size-standard or sequencing machine used
- **No commercial software is needed**
(Bionumerics, Genemapper, ...)
- **no manual data manipulation** needed

Capillary Gel Electrophoresis based PCR Ribotyping Analysis using WEBRIBO

Labwork

- Direct or culture samples
- DNA-extraction
- PCR
- Capillary sequencer PCR-ribotyping

Analysis

- results in 4 steps using
WEBRIBO (webribo.ages.at)



WEBRIBO:

Equipment requirements

- **PCR-Cycler**

- **Primers**

- any labelled PCR-Ribotyping primer available (e.g. Bidet, Stubbs or Janežič)

- **Sequencer**

- Applied Biosystems (ABI) Systems tested (310, 3100, 3130(xl), 3730)
 - any POP-Gel
 - any capillary length
 - Beckman Coulter Sequencer

- **Size-Standard**

- any Size-Standard like
 - Geneflo 625 ROX/Tamra
 - LIZ 600/LIZ 120
 - Genemapper 1000
 - Beckman GenomeLab 600

- **Capable of** differentiating **ribotypes 014 and 020**
 - additional 014 subtypes can be separated
- **Capable of** differentiating **various 001 subtypes** first described by Northey et al.

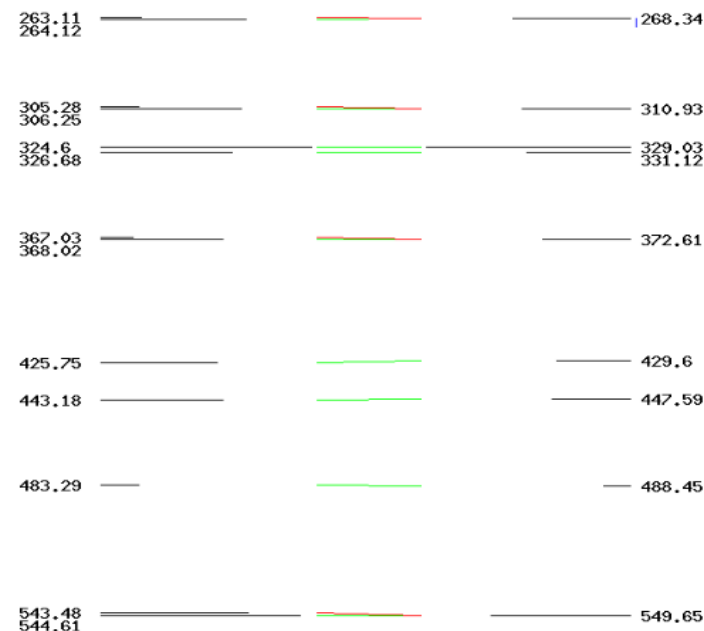


WEBRIBO: Cons

- .fsa data analysis algorithm **quite sensitive**
- **No RAW-data** import for **Beckman Systems**
- PCR-Ribotype **nomenclature not in total accordance**
 - **Subtype** nomenclature is sometimes **confusing**
- **Assignment algorithm has to be improved**
- **Analysis can take some time**

- Up to 99% correct **identification**
of samples analysed with

- **different** size-standards
- **different** sequencing machines
- **different** primers
 - Stubbs, Bidet, Janežič
- **a supervised learning algorithm**
copes with double-peaks due to PCR-artefacts



QUESTIONS?

webribo.ages.at

alexander.indra@ages.at