Originally published as:

Peter Lasch, Michal Drevinek, Herbert Nattermann, Roland Grunow, Maren Stämmler, Ralf Dieckmann, Torsten Schwecke, and Dieter Naumann.
Characterization of *Yersinia* Using MALDI-TOF Mass Spectrometry and Chemometrics.
(2010) Analytical Chemistry, 82 (20), pp. 8464–8475.

DOI: 10.1021/ac101036s

Characterization of *Yersinia* Using MALDI-TOF Mass Spectrometry and Chemometrics

Peter Lasch[*†], Michal Drevinek [‡], Herbert Nattermann [§], Roland Grunow [§], Maren Stämmler [†], Ralf Dieckmann [∥], Torsten Schwecke [†], and Dieter Naumann [†]

Biomedical Spectroscopy (P 25) and Centre for Biological Security (ZBS 2), Robert-Koch-Institut, Nordufer 20, D-13353 Berlin, Germany, National Institute for Nuclear, Biological and Chemical Protection, Kamenna 71, CZ-26231 Milin, Czech Republic, and Department of Biological Safety, Antibiotic Resistance and Resistance Determinants, Federal Institute for Risk Assessment (BfR), Diedersdorfer Weg 1, D-12277 Berlin, Germany

† Biomedical Spectroscopy, Robert-Koch-Institut.
‡ National Institute for Nuclear, Biological and Chemical Protection.
§ Centre for Biological Security, Robert-Koch-Institut.
∥ Federal Institute for Risk Assessment.

Abstract

Yersinia are Gram-negative, rod-shaped facultative anaerobes, and some of them, Yersinia enterocolitica, Yersinia pseudotuberculosis, and Yersinia pestis, are pathogenic in humans. Rapid and accurate identification of Yersinia strains is essential for appropriate therapeutic management and timely intervention for infection control. In the past decade matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) in combination with computer-aided pattern recognition has evolved as a rapid, objective, and reliable technique for microbial identification. In this comprehensive study a total of 146 strains of all currently known Yersinia species complemented by 35 strains of other relevant genera of the Enterobacteriaceae family were investigated by MALDI-TOF MS and chemometrics. Bacterial sample preparation included microbial inactivation according to a recently developed mass spectrometry compatible inactivation protocol. The mass spectral profiles were evaluated by supervised feature selection methods to identify family-, genus-, and species-specific biomarker proteins and—for classification purposes—by pattern recognition techniques. Unsupervised hierarchical cluster analysis revealed a high degree of correlation between bacterial taxonomy and subproteome-based MALDI-TOF MS classification. Furthermore, classification analysis by supervised artificial neural networks allowed identification of strains of Y. pestis with an accuracy of 100%. In-depth analysis of proteomic data demonstrated the existence of Yersinia-specific biomarkers at m/z 4350 and 6046. In addition, we could also identify species-specific biomarkers of Y. enterocolitica at m/z 7262, 9238, and 9608. For Y. pseudotuberculosis a combination of biomarkers at m/z 6474, 7274, and 9268 turned out to be specific, while a peak combination at m/z 3065, 6637, and 9659 was characteristic for strains of Y. pestis. Bioinformatic approaches and tandem mass spectrometry were employed to reveal the molecular identity of biomarker ions. In this way, the Y. pestis-specific biomarker at m/z 3065 could be identified as a fragment of the plasmid-encoded plasminogen activator, one of the major virulence factors in plague infections.

The genus *Yersinia* belongs to the family Enterobacteriaceae and presently comprises 13 validly described species.(1-3) Three of them, *Yersinia enterocolitica*, *Yersinia pseudotuberculosis*, and *Yersinia pestis*, are important pathogens with relevance to animal and human health which have been studied extensively by various methods. The other species, *Yersinia aldovae*, *Yersinia aleksicae*, *Yersinia bercovieri*, *Yersinia frederiksenii*, *Yersinia intermedia*, *Yersinia kristensenii*, *Yersinia massiliensis*, *Yersinia molaretii*, *Yersinia rohdei*, *Yersinia ruckeri*, and *Yersinia similis*, are less intensively characterized possibly because of their lower clinical importance.(4, 5) Pathogenic *Yersinia* are the cause of serious diseases in animals and humans; *Y. enterocolitica*, for example, is a food-borne enteropathogen which can produce severe gastroenteritis, terminal ileitis, and mesenteric lymphadenitis. *Y. pseudotuberculosis* is another enteropathogenic species which primarily causes diseases in animals. Symptoms in humans frequently mimic an infection with *Y. enterocolitica*, which

sometimes makes the resulting condition difficult to diagnose. Infections with *Y. pseudotuberculosis* may be accompanied by secondary complications such as erythema nodosum and reactive arthritis. *Y. pestis*, the causative agent of plague, is endemic throughout the world and primarily infects a wide range of rodents. Transmission to humans occurs primarily via the bites of fleas from infected rodents. Although the infectious disease is now rare, plague has been responsible for killing in three former waves of pandemic hundreds of million people. Three major clinical forms of an infection with *Y. pestis* have been described, bubonic, pneumonic, and septicaemic. The latter two forms are extraordinarily contagious and—if untreated—are associated with fatality rates close to 100%.(6) Intentional release of *Y. pestis* via aerosols would preferentially cause a primary pneumonic plague outbreak in the exposed population. Plague is thus considered to be one of the most serious bioterrorism threats.(6) The diagnosis of an infection by *Y. pestis* is based on clinical symptoms, microbiological tests (cultivation, Gram-staining), confirmation of the *Y. pestis* F1 antigen (immunostaining, serum titer), and polymerase chain reaction (PCR)-based techniques. These techniques are costly and time-consuming and require trained personnel. Furthermore, the differentiation of *Y. pseudotuberculosis* and *Y. pestis* is laborious, possibly because the genetic differences between these species are rather small.(7-10) In fact, there is convincing experimental evidence suggesting that *Y. pestis* is a highly uniform clone that diverged only recently (1500−20000 years ago) from *Y. pseudotuberculosis.*(8, 11)
There is an increasing need for new diagnostic technologies which allow rapid, nonsubjective, and accurate identification of microorganisms. These techniques should ideally complement traditional microbiological and PCR-based methods. Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) of whole microbial cells, or their extracts, is considered such a method. This technique was known in the past as intact cell mass spectrometry (ICMS) and has evolved in the past 15 years from a niche research procedure into a widely used practical application with the potential to revolutionize the way microorganisms are identified.(12-16) The technique is based on reproducible detection of microbial protein patterns and thus delivers complementary information to classical microbiological or genotyping methods. In a typical approach characteristic mass spectral fingerprints are obtained which subsequently can be compared against validated databases of bacterial reference spectra.(17, 18) Therefore, an extensive knowledge of biomarker identities is not required, which facilitates high-throughput routine analyses in clinical microbiology and the food industry or to assess public health hazards.
It has been shown by numerous studies that MALDI-TOF MS can be used under standardized conditions to discriminate bacteria at the genus, species, or even subspecies level by the identification of distinctive sets of protein mass peaks, referred to as taxon-specific biomarker ions. This fact as well as the availability of a mass spectrometry compatible inactivation protocol for highly pathogenic biosafety level 3 (BSL-3) microorganisms(19, 20) prompted us to systematically study the MALDI-TOF MS profiles of *Yersinia* and related genera. In the present paper, we describe the strategies to establish a database of MALDI-TOF reference spectra of *Yersinia* and present examples of genus- and species-specific biomarkers. Using unsupervised cluster and supervised artificial neural network (ANN) analysis, we will furthermore demonstrate how microbial mass spectra can be employed to establish an MS-based classification system of the genus *Yersinia*.

Experimental Section

Microbial Strains and Isolates

With the exception of selected *Yersinia* strains, all strains of the Enterobacteriaceae family originated from the strain collection at the Robert-Koch-Institut (RKI). The RKI collection was complemented by strains of *Y. pseudotuberculosis* and *Y. pestis* (strain collection at the National Institute for Nuclear, Biological and Chemical Protection, Milin, Czech Republic), reference strains from the German Collection of Microorganisms and Cell Cultures GmbH (DSMZ; Braunschweig, Germany), and *Y. enterocolitica* and *Y. pseudotuberculosis* strains from a clinical study carried out by AnagnosTec GmbH (Potsdam, Germany) or were kind gifts of Dr. T. M. Fuchs (*Y. enterocolitica*-like species; ZIEL, Technical University Munich, Germany), Dr. J. Rau (*Y. pseudotuberculosis*, *Y. enterocolitica*-like species; CVUA, Stuttgart, Germany), and Dr. N. Schürch (*Y. pestis*, *Y. pseudotuberculosis*; Labor Spiez, BABS, Spiez, Switzerland). Strains and isolates used in this study are listed by Table 1. Bacterial biomass was obtained by growing each strain for two passages under aerobic conditions on caso agar (Merck KGaA, Darmstadt, Germany) for 48 h at 28 °C. Cells were harvested from the

second passage by transferring an equivalent of three wire loops (dry weight of about 4 mg) from each agar plate to 20 µL of water.

Sample Preparation/Safety Considerations

The preparation of microbial samples was carried out by using the trifluoroacetic acid (TFA) inactivation protocol for highly pathogenic microorganisms.(19) Briefly, 80 µL of pure TFA (Uvasol, Merck) was added to 20 µL of the bacterial suspensions. After 30 min of gentle shaking, the solutions were diluted 10-fold with HPLC grade water (Mallinckrodt Baker B.V., Deventer, The Netherlands). The sample dilutions of *Y. pestis* were additionally checked for sterility. For MALDI-TOF MS, 2 µL of the microbial dilutions was then mixed with 2 µL of a 12 mg/mL α-cyano-4-hydroxycinnamic acid (HCCA) solution (Bruker Daltonics, Bremen, Germany). HCCA solutions were prepared by dissolving HCCA in a 2:1 (v/v) mixture of 100% acetonitrile and 0.3% TFA (TA_2). A 1 µL volume of the sample/HCCA mixture was then spotted onto ground steel sample targets from Bruker Daltonics.

MALDI-TOF Mass Spectrometry

Mass spectra of microbial TFA extracts were collected by an Autoflex I mass spectrometer from Bruker Daltonics. The instrument was controlled by Bruker's FlexControl 3.0 data collection software and was equipped with a UV nitrogen laser (λ = 337 nm), which operated in a slightly defocused mode. MS measurements were carried out in linear mode using acceleration voltages of 20.00 and 18.45 kV (ion sources 1 and 2, respectively). The lens voltage was 6.70 kV. Spectra were collected in the mass range between *m*/*z* 2000 and *m*/*z* 20000. For external calibration Bruker's protein calibration standard I was employed (for details see ref 20). To achieve a high signal-to-noise ratio (SNR), each spectrum represents the integration of at least 600 individual laser shots.

Liquid Chromatography−MALDI Tandem Mass Spectrometry

For off-line analysis by MALDI-TOF MS, the 10-fold-diluted TFA extracts (see the sample preparation of microbial extracts and ref 19) were first pelleted by centrifugation. The supernatant was discarded and the pellet then dissolved in 6 M urea and 1% acetic acid. Peptide separation was performed on an Agilent (Palo Alto, CA) 1200 series binary HPLC instrument fitted with a 4.6 × 50 mm mRP-C_{18} reversed-phase column (Agilent). Peptides were eluted at a flow rate of 0.75 mL/min employing a gradient of 3%−30% B in 5 min and 30%−50% B in 33 min, where solvent A consisted of 0.1% TFA in water and solvent B of 0.08% TFA in acetonitrile. The eluate was monitored at λ = 210 nm, and 0.75 mL fractions were collected. Following evaporation to dryness, peptides of each fraction were dissolved in 10 µL of TA_2 (see above). A 1 µL volume of each fraction was mixed with 1 µL of HCCA solution (6 mg/mL in TA_2) and the resulting solution air-dried on a 384-well polished steel target plate (Bruker Daltonics). MALDI-TOF MS measurements were carried out under the control of FlexControl software, version 3.0 (Bruker Daltonics), using an Ultraflex II MALDI-TOF/TOF (TOF/TOF = tandem time-of-flight) mass spectrometer (Bruker Daltonics) equipped with a frequency-tripled solid-state Smartbeam Nd:YAG laser (λ = 355 nm) which operated at 100 Hz. Fractions containing target peptides were identified by recording spectra in linear positive mode with external calibration using a standard mixture of peptides. To sequence peptides, an exploratory scan from *m*/*z* 2000 to *m*/*z* 5000 was performed in the reflectron mode to assign a mass window for fragmentation and peptide sequencing in the "LIFT" tandem mass spectrometry (MS/MS) mode. The spectra were obtained by averaging up to 3000 laser shots acquired at a fixed laser power, which was set to the minimum laser power necessary for ionization of selected samples before the analyses were started. The mass spectra were visualized and processed using FlexAnalysis software, and sequence tag hints were obtained by analyzing tandem MS spectra using the Biotools 3.0 software (Bruker Daltonics). A BLAST search restricted to the *Yersinia* taxonomic identifier was performed, and finally the MS/MS spectrum was reinvestigated by alignment to the sequence obtained from the database search.

Data Analysis

The analysis of mass spectra was carried out using a collection of dedicated Matlab routines (The Mathworks Inc., Natick, MA) developed by the principal author of this study. These routines are available as Matlab p-code on request (free of charge). The Matlab functions provide direct access to

the raw spectral data and allow spectral preprocessing (e.g., smoothing, baseline correction, intensity normalization, internal calibration) as well as the generation of peak lists. Spectra were internally calibrated using a set of ribosomal biomarker proteins common to Enterobacteriaceae. We have used the Matlab routines also for producing gel views from preprocessed spectra and for generating bar-coded spectra. Furthermore, the routines were employed for univariate feature selection (independent *t* test) and for cluster analyses and producing dendrograms(21, 22) and as a software interface to the NeuroDeveloper program, an artificial neural network simulator from Synthon (Synthon Analytics GmbH, Heidelberg, Germany(23)). Tentative molecular assignments of experimental biomarkers were identified according to a strategy pioneered by Demirev and co-workers.(24-26) For this purpose the complete set of protein sequence data of a species of interest was obtained from the UniProtKB/SwissProt database. A custom-designed Matlab function was employed to calculate protein masses and to find potential matches with experimental mass peaks. A more detailed description of the functionality of the Matlab functions, the general strategy of supervised ANN classification, and the software interface between Matlab functions and the NeuroDeveloper can be found elsewhere.(20, 27)

Results

To study the intra- and interspecies relationships of strains from the genus *Yersinia*, MALDI-TOF MS measurements were carried out on a total of 146 strains which originated from the 13 currently known species of this genus. This database of *Yersinia* spectra was complemented by MS data acquired from 35 strains of other Enterobacteriaceae (see Table 1 for an overview). If possible, mass spectra were acquired in triplicate, i.e. from three independent bacterial cultivations. In this study microbial samples were prepared as described,(19) which allowed MALDI-TOF MS measurements of BSL-3 microorganisms outside of BSL-3 facilities.

MALDI-TOF Spectra of Clinically Relevant *Yersinia*

Figure 1 shows three examples of MALDI-TOF mass spectra from type strains of *Y. enterocolitica* (Figure 1A, DSM 11503), *Y. pseudotuberculosis* (Figure 1B, DSM 8992), and *Y. pestis* (Figure 1C, NCTC 5923). The spectral SNR typically allowed detection of 50−100 discrete ions per mass spectrum. Mass spectral profiles obtained from microbial TFA extracts characteristically represent molecular fingerprints that are easily distinguishable even by visual inspection. For example, an ion triplet at *m*/*z* 7262, 7288, and 7318 was found to be characteristic for spectra from *Y. enterocolitica* strains. Other examples of mass signals typical for *Y. enterocolitica* are the peaks at *m*/*z* 9608 and 9651 (cf. Figure 1A). A detailed inspection of spectra of the phylogenetically closely related species *Y. pseudotuberculosis* and *Y. pestis* exhibited a number of signals present in both species, for example, at *m*/*z* 4830, 6241, 7274, and 9659 (see Figure 1B,C).

Pseudo “gel views” are popular means for visual inspection of larger MS data sets. In gel view representations spectral peak intensities are usually converted to gray scales which can be subsequently plotted as a function of the *m*/*z* values. The pseudo gel view of Figure 2 displays the complete MS database in the mass range of *m*/*z* 3500−10500. Spectra of Enterobacteriaceae strains other than *Yersinia* are shown in the upper part of Figure 2 (74 spectra, lines 1−74), while all other 414 spectra were generated from *Yersinia* strains (lines 75−489). In general, we found that mass spectral profiles of *Yersinia* strains exhibited consistent and relatively homogeneous peak patterns which contained a number of distinct peaks commonly not found in spectra of other Enterobacteriaceae members. As illustrated by the gel view of Figure 2, candidates of such *Yersinia*-specific (genus level) biomarkers can be found at *m*/*z* 4350, 5427, 6046, and 6241.

Examples of species-specific signals are mass peaks at *m*/*z* 7149, 7262, 7318, 9238, 9608, and 9651 (all specific for *Y. enterocolitica*) and *m*/*z* 6637, 7274, 7783, 9268, and 9659 (specific for *Y. pseudotuberculosis* and *Y. pestis*) (see Figure 2).

Unsupervised Hierarchical Cluster Analysis

The presence of genus- and species-specific peak patterns in the mass spectra of strains from *Yersinia* was confirmed by unsupervised hierarchical cluster analysis (UHCA) (see Figure 3). This particular type of data-driven multivariate classification technique was carried out on the basis of bar-

coded mass spectra in the molecular mass range of *m*/*z* 3500−10500. Logical distance values were used as similarity measures between pairs of individual spectra, and Ward's algorithm was used as the clustering method. The dendrogram given by Figure 3 demonstrates that spectra of the MS database fall into three distinct clusters: the first cluster (I) contains two well-separated subclusters which contain spectra of *Y. enterocolitica* (Ia) and *Y. enterocolitica*-like species (Ib). Cluster II is formed by spectra of the non-*Yersinia* Enterobacteriaceae family members, i.e., from strains of the genera *Citrobacter*, *Edwardsiella*, *Enterobacter*, *Escherichia*, *Hafnia*, *Klebsiella*, *Proteus*, *Salmonella*, and *Serratia* and four outliers from *Y. enterocolitica* and *Y. pestis*. The third cluster (III) contains two subclusters formed exclusively by mass spectra of four *Yersinia* species. While spectra from *Y. pseudotuberculosis*, *Y. pestis*, and *Y. similis* constitute subcluster IIIa, cluster IIIb contains only spectra of *Y. ruckeri.*

Taxon-Specific Biomarker Ions

The spectral distinctness of strains of *Y. enterocolitica* (57 strains) on the one hand and *Y. pseudotuberculosis* and *Y. pestis* (26 and 21 strains, respectively) on the other is illustrated in greater detail by the pseudo gel view of Figure 4. This figure displays the diagnostically important mass regions between *m*/*z* 7080 and *m*/*z* 7390 (left) and *m*/*z* 9200 and *m*/*z* 9830 (right). While spectra from *Y. enterocolitica* show a characteristic triple peak with signals at *m*/*z* 7262, 7288, and 7318 (see above), the mass spectral profiles of *Y. pseudotuberculosis* and *Y. pestis* exhibit characteristic ions at *m*/*z* 7188, 7274, and 7359. The enlarged spectral region in the right panel of Figure 4 shows the presence of additional biomarkers for *Y. enterocolitica* at *m*/*z* 9238, 9608, and 9651. In the same spectral region *Y. pseudotuberculosis* and *Y. pestis* reproducibly display two other characteristic signals at *m*/*z* 9268 and 9659, which could additionally serve as biomarkers for these species. The gel view of Figure 5 illustrates the presence of biomarker ions that allow discrimination between *Y. pestis* and *Y. pseudotuberculosis*. A candidate of a *Y. pestis*-specific biomarker is a peak at *m*/*z* 3065 (Figure 5, left panel), while spectra of *Y. pseudotuberculosis* could be identified by the presence of a peak at *m*/*z* 6474 (right panel).

An overview of the biomarkers identified in the present study is given in Table 2. The table also gives an idea of the possible molecular identity of some of these biomarkers. Peak assignment was carried out by comparing the experimental MALDI-TOF data with molecular masses obtained from protein sequences of *Y. pseudotuberculosis*, *Y. pestis*, and *Y. enterocolitica* contained in the UniProtKB/Swiss-Prot database.

No possible molecular identity of the potential *Y. pestis*-specific biomarker at *m*/*z* 3065 could be deduced from sequence database searches based on the mass of the intact molecule alone. We therefore performed liquid chromatography (LC)−MALDI MS/MS to obtain amino acid sequence information. Following separation of the TFA-extracted proteins by liquid chromatography on a reversed-phase C_{18} column, a fraction containing the respective polypeptide was identified using MALDI-TOF MS. A prominent mass peak with an average mass of *m*/*z* 3065 (linear mode) and a monoisotopic mass of 3062.6 (reflectron mode) was detected in a fraction eluting at 32% solvent B. Using MALDI-TOF tandem MS, a contiguous sequence tag, YTVTAG, was obtained. A BLAST similarity search (http://blast.ncbi.nlm.nih.gov) of this tag restricted to the *Yersinia* taxonomic identifier for nonredundant protein sequences returned with 100% sequence coverage protein entries related to the plasminogen activation factor (Pla) of *Y. pestis*. The sequence tag is located in a C-terminal region of the 312 amino acid Pla precursor, NSGDSVSIGGDAAGISNKNYTVTAGLQYRF, which displays a calculated molecular weight of the singly protonated molecular species $[M + H]^+$ of *m*/*z* 3062.5 corresponding to the molecular mass observed by MALDI-TOF MS of the intact precursor ion. Careful re-examination of the empirical fragmentation profile using Biotools software revealed that the observed fragment ion masses were highly consistent with the predicted cleavages calculated from the amino acid sequence, generating an almost complete y-ion series (data not shown).

Within the scope of the present study we also employed a statistical technique for more objective identification of biomarker ions. One of these methods is based on univariate *t* tests carried out on labeled subsets of mass spectra. Typically, the microbial database is divided into two classes, for example, into a class "*Yersinia*" (class a) and a class "Enterobacteriaceae other than *Yersinia*" (class b). Then the discriminative power of each spectral feature—typically an intensity at a given *m*/*z* value—is determined by a univariate independent *t* test. Each *t* test provides a *p* value which is a statistical measure of the distinctness between the mass spectral features of classes a and b at the given *m*/*z* value. A *p* value of 1 confirms the null hypothesis (equal class means), while *p* values close

to 0 cast doubt on this hypothesis. One can plot the *p* values as a function of the *m*/*z* values. The smaller the *p* value, the higher the discriminative potential of the spectral feature under investigation. In Figure 6 three examples of *p* value curves (note the inverse logarithmic scaling with −log *p*) are given. Figure 6A shows discriminative biomarkers that allow differentiation between MALDI-TOF mass spectra of Enterobacteriaceae other than *Yersinia* and of the genus *Yersinia*. The three most discriminating biomarkers, at *m*/*z* 6241, 4350, and 6046, are all found in mass spectra of strains from *Yersinia* (class a). Two other peaks at *m*/*z* 4365 and 5382 are less discriminative and typical for a biomarker ion of class b (Enterobacteriaceae other than *Yersinia*).

Discriminative mass peaks within the genus *Yersinia* were statistically verified in a similar way. Species-specific biomarkers of *Y. enterocolitica* (class a in Figure 6B) were identified at *m*/*z* 9608, 3632, and 9238, which confirmed the results of the visual inspection of the gel views. Note that the signal at *m*/*z* 4805 represents the doubly charged ion $(M + 2H)^{2+}$ of the parent ion at *m*/*z* 9608. Univariate *t* tests also revealed marker peaks in MS data of *Y. pestis* and *Y. pseudotuberculosis* (see Figure 6C). Again, the most differentiating biomarkers at *m*/*z* 6474, typical for *Y. pseudotuberculosis*, and at *m*/*z* 3065 for *Y. pestis* were found in good agreement with the visual analysis of mass spectra (Figure 1) and gel views (Figures 2 and 5).

ANN Analysis

Supervised classification of strains from members of the Enterobacteriaceae family was carried out on the basis of a modular hierarchical ANN. Bar code mass spectra which contained 30 peaks per spectrum taken from the mass range of *m*/*z* 2200−12000 served as ANN inputs, and none of the feature selection methods supplied by the NeuroDeveloper software package were employed. The hierarchical ANN approach allowed separate optimization of three distinct ANNs: An ANN top-level net was created which allowed differentiation between spectra of strains of the genus *Yersinia* and (i) other Enterobacteriaceae family members. Spectra identified as *Yersinia* were further analyzed by a second ANN, called “sublevel 1”, which was trained to differentiate between the classes (ii) *Y. enterocolitica*-like and (iii) *Y. enterocolitica* and *Y. pestis*/*Y. pseudotuberculosis*. To discriminate between the species of the latter category, (iv) *Y. pseudotuberculosis* and (v) *Y. pestis*, a third ANN, “sublevel 2”, was employed. All individual subnets consisted of three-layer feed-forward multilayer perceptron (MLP) ANNs with connected layers of input, hidden, and output neurons. Teaching was achieved by using the rprop learning rule. The NeuroDeveloper software package from Synthon(23) allowed combination of the neural libraries into a single hierarchically organized neural network. A schematic overview of the hierarchy of the modularly organized neural network along with the predefined classes is illustrated by Figure 7.

The general strategy of ANN analysis routinely includes procedures of training and internal and external validation of the individual networks. Training and internal validation are accomplished on the basis of labeled mass spectra, i.e., spectra with known class assignments. During training, the performance of classification is determined on the basis of internal validation spectra not used for teaching. At the training stage ANN performance can be optimized by modifying the method of preprocessing, adding or eliminating spectral features, or changing the network’s architecture. When training is finished, the classifier can be challenged by a second (external) set of data which contains spectra kept totally separate.

To train the toplevel ANN, 335 MALDI-TOF mass spectra from strains of *Yersinia* and 54 spectra from Enterobacteriaceae strains not belonging to *Yersinia* (class i) were combined into a teaching/internal validation subset which included 20% internal validation spectra selected randomly. The external validation subset for top-level classification comprised 20 spectra of class i and 80 spectra of *Yersinia* strains (classes ii−v). Using this data set for external validation, perfect classification could be obtained with 100% accuracy. Similar results were obtained with the sublevel 1 ANN. To train this subnet, the teaching/internal validation subset consisted of 69 spectra of class ii, 114 spectra of class iii, and 88 spectra from class iv + class v. Note that this data set also contained 20% randomly selected internal validation spectra. External validation by 20/58/66 spectra (class ii/class iii/class iv + class v) by the sublevel 1 ANN resulted in 100% classification accuracy. Final identification of spectra from strains of (iv) *Y. pseudotuberculosis* and (v) *Y. pestis* was tested by the sublevel 2 ANN (see Table 3 for details). Although the number of spectra, specifically of the external validation subset, is comparably small, the results of classification by this specific ANN demonstrate separability of MALDI-TOF mass patterns of *Y. pseudotuberculosis* and *Y. pestis.*

Discussion

Family-Level Peaks

MALDI-TOF mass spectra of strains from the Enterobacteriaceae family suggest the existence of family-specific biomarkers at *m*/*z* 4185 and 8370. These peaks could be reproducibly detected in the mass spectra of almost all bacterial strains characterized within the context of the present study (see Figures 1 and 2). Using the UniProtKB/SwissProt database, both mass peaks could be related to the 30S ribosomal protein S21 either as the singly charged ($[M + H]^+$) or the doubly charged ($[M + 2H]^{2+}$) ion (cf. Table 2). Our preliminary peak assignment is additionally supported by the results of a BLAST search. For example, the program NCBI BLASTP 2.2.22 revealed for the 30S ribosomal protein S21 of *Y. pestis* (SwissProt entry B0H6A7), with a calculated MW of 8500 (terminal Met is cleaved!), a total of 129 database entries with 100% amino acid sequence identity. Interestingly, all of these entries originated from altogether 16 genera of the Enterobacteriaceae family such as *Salmonella*, *Escherichia*, *Erwinia*, *Yersinia*, and others. The family-specific mass signal at *m*/*z* 8370 was reported also by others.(28, 29) These authors report for various strains of *Escherichia coli* a peak at *m*/*z* 8370 which could be identified as the S21 protein of the small (30S) ribosomal subunit. Furthermore, Dieckmann et al.(16) categorized the signal of the 30S ribosomal protein S21 as a “specificity category I peak”, meaning that it is specific for a genus (*Salmonella*) or a higher level taxon. These findings are supported also by a very recent top-down bioinformatics study in which experimental MALDI-TOF peaks of a strain of *Y. rohdei* were annotated and subsequently compared in silico (using MW obtained from the SwissProt protein database) with ribosomal proteins of various species of the Enterobacteriaceae family. On the basis of bioinformatics, the authors hypothesized that the 30S ribosomal protein S21 has the same sequence in 27 of 28 strains of Enterobacteriaceae contained in the protein database(30) (spring 2010), which is consistent with our experimental data.

Genus-Level Peaks

All of the analyzed *Yersinia* species displayed a genus-specific peak at *m*/*z* 6241 (see Figures 1, 2, and 6). This specifically identifying mass signal could be assigned to the 50S L33 ribosomal subunit protein (terminal Met cleavage + methylation), which is confirmed by a recent literature assignment of *Y. rohdei*.(30) Interestingly, a BLAST search for the 50S L33 protein of *Y. pestis* revealed 100% sequence identity to strains of *Y. pseudotuberculosis* (and additionally to strains of *Erwinia carotovora*, *Pectobacterium wasabiae*, and *Serratia odorifera*). In other *Yersinia* such as *Y. enterocolitica* and seven *Y. enterocolitica*-like species, a Leu−Ile exchange at position 47 is found. Since this modification is not associated with a change of a protein’s MW, the latter finding would explain the occurrence of the peak at *m*/*z* 6241 in all spectra of *Yersinia*.

Another BLAST search of the second genus-specific signal at *m*/*z* 6046 of *Y. pestis* (50S ribosomal protein L32) returned sequence identity to strains of the clinically relevant *Yersinia* but also to strains of *Y. ruckeri*, *Y. intermedia*, and *Y. fredericksenii*. Database entries for the 50S L32 subunit of the remaining *Yersinia* species either could not be found or—in the case of strains of other Enterobacteriaceae—showed modified amino acid sequences. The third peak identified as genus-specific was the signal at *m*/*z* 4350 (50S L36, Table 2). Although this peak was experimentally confirmed in almost all MALDI-TOF spectra of *Yersinia* (see Figure 2), the (incomplete?) protein database contained no entries for this protein in *Y. enterocolitica*-like species. Furthermore, as the BLAST search for the 50S L36 protein of *Y. pestis* revealed 100% sequence identity with the 50S L36 subunit of selected strains of *Erwinia*, *Sodalis*, *Pectobacterium*, and *Serratia*, the mass peak at *m*/*z* 4350 may not be considered as a genus-specific peak on its own.

Species-Level Peaks

While we found for many of the recognized *Yersinia* species spectral biomarkers that would allow precise typing at the species level, the following discussion will be focused on the biomarker ions of *Y. pseudotuberculosis* and *Y. pestis*.
Y. pestis is highly related to *Y. pseudotuberculosis*.(31) The 16S rRNAs of both species are identical,(7) and they share an exceptionally high number of identical proteins. To give an example, a systematic comparison of all ribosomal subunit proteins in the UniProtKB/SwissProt database did not reveal even a single amino acid exchange (data not shown). In light of these considerations, it was remarkable to detect a clearly differentiating peak (*m*/*z* 3065) in the mass spectra of *Y. pestis*. As stated above the analysis of MALDI tandem MS measurements helped to reveal the molecular identity of this ion as a fragment of one of the virulence factors of *Y. pestis*, Pla. Pla is a surface protease which is known to play a key role in primary pneumonic plague infections by promoting the invasion of *Y. pestis* from subcutaneous sites of inoculation into the lymphatic system.(32) Upon flea-borne transmissions the Pla virulence factor is also required to develop bubonic plague.(33) Pla is encoded on a 9.5 kB plasmid (pPCP1) of *Y. pestis* and is thus a potential *Y. pestis*-specific biomarker. Sodeinde and Goguen showed that Pla is expressed as a transient 34.6 kDa primary *pla* product, which is processed by N-terminal cleavage upon insertion into the outer membrane to yield 32.6 kDa α-Pla.(34) This product (292 amino acids) is then autoprocessed at the C-terminus to yield a slightly smaller derivative, termed β-Pla (262 amino acids).(35) α-Pla and β-Pla are both present in the outer membrane of *Y. pestis*, and no significant functional differences between both products could be detected.(35) Interestingly, the sequence of the cleavage product (30 amino acids) was found to be identical to the sequence of the *m*/*z* 3065 ion as derived by MALDI tandem MS. Although the functional significance of the Pla fragment is hitherto unknown, it is apparent that the observation of the *Y. pestis*-specific biomarker at *m*/*z* 3065 is the result of fragmentation of α-Pla resulting in the formation of β-Pla.

Figure 5 demonstrates that the *m*/*z* 3065 ion is absent in mass spectra of the strains O3-01501-08 (see the arrow in Figure 5). Since these strains have lost the plasmid pPCP1, the absence of the peak at 3065 *m*/*z* further supports our assignment of this biomarker.
Another biomarker candidate which allows differentiation of *Y. pestis* and *Y. pseudotuberculosis* is a peak at *m*/*z* 6474. This peak was found in spectra of *Y. pseudotuberculosis* and some strains of *Y. enterocolitica* (which considerably reduces its diagnostic value). As sequence database searches were not helpful to unravel the protein identity of the *m*/*z* 6474 signal, we are currently investigating HPLC fractions of *Y. pseudotuberculosis* extracts by MALDI-TOF tandem MS.

Classification Analysis by UHCA

Aside from biomarker characterization, the protein expression profiles as revealed by MALDI-TOF MS can also be analyzed by multivariate classification techniques. For example, the entirely data-driven technique of UHCA was employed to investigate whether there were homogeneous groups in the database of microbial mass spectra.(21, 22) The results of UHCA clearly demonstrated the presence of species-specific mass patterns of *Y. enterocolitica* (see species cluster Ia of Figure 3) and also certain phenotypic relatedness between *Y. enterocolitica* and *Y. enterocolitica*-like bacteria. Interestingly, there is also some evidence that even a typing at the subspecies level of *Y. enterocolitica* is possible. This assumption can be taken from the diversity of the mass spectral patterns of *Y. enterocolitica* in the mass region of *m*/*z* 7500−9500 (see the arrows in Figure 2) where some of the signals appear to be correlated with serogroups of *Y. enterocolitica*. However, systematic analyses and perhaps further measurements of more strains are required to prove this hypothesis. Another conclusion that can be drawn from the dendrogram of Figure 3 is that *Y. ruckerii* forms a distinct cluster that is well-separated from all other *Yersinia*. This is particularly interesting since the taxonomic status of *Y. ruckerii* is still a subject of controversial discussions.(4, 5) UHCA also reveals a close relationship of *Y. pestis*, *Y. pseudotuberculosis*, and *Y. similis* (cf. cluster IIIa of Figure 3). Again, this finding is in good agreement with literature reports in which the intra- and interspecies genetic relationships of *Yersinia* were studied. For example, *Y. pestis*, *Y. pseudotuberculosis*, and *Y. similis* are found to be tightly clustered in phylogenetic trees obtained by 16S rRNA gene sequence analysis.(3) In this study *Y. similis*, whose name stems from “similar” (to *Y. pseudotuberculosis*), formed one specific subcluster with the *Y. pseudotuberculosis*/*Y. pestis* complex within the *Yersinia* root cluster. While the question of whether *Y. similis* can be reliably differentiated from *Y. pseudotuberculosis*/*Y. pestis* cannot be finally answered using MS data of only one strain of *Y. similis* (see Table 1), the dendrogram of Figure 3 suggests an even closer proteomic relationship between *Y.*

pseudotuberculosis and *Y. pestis*. In fact, we found that the latter two species were indistinguishable on the basis of plain UHCA, which is obviously a consequence of a relatively small interspecies variance compared with the intraspecies variance, reproducibility/repeatability errors, and other factors. Obviously, existing systematic differences such as the *Y. pestis*-specific signal at *m*/*z* 3065 (see above) or the signal at *m*/*z* 6474 (*Y. pseudotuberculosis*) are too small to be of much consequence for the protein pattern based UHCA results.
The experimental observation of a close relationship of strains from *Y. pestis* and *Y. pseudotuberculosis* is backed by methods studying the genetic relatedness among *Yersinia*, for example, by 16S rDNA(31) and 16S rRNA analyses or multilocus sequence typing (MLST) of concatenated sequences from various housekeeping gene loci.(3) Data of these techniques consistently suggested an unusually high degree of genetic relatedness between strains of *Y. pseudotuberculosis* and *Y. pestis*,(5) and some authors even stated that *Y. pestis* is a clone which only recently (1500−20000 years ago) has evolved from *Y. pseudotuberculosis.*(8)

ANN Classification Analysis

When MLP ANNs are trained for classification analysis, complex multivariate discriminant functions are established. These functions take advantage of the specific intra- and interclass variance at each spectral feature. In this way, the ANN classifier not only “knows” average microbial mass patterns but at the same time considers also the complex feature-specific intraclass variance. Compared to classification analysis by straightforward peak matching algorithms, this results in an improved classification accuracy and makes ANN classification of microbial mass spectra exceptionally robust. However, the advantage of improved classification accuracy is gained at the cost of higher requirements in terms of sample/spectrum numbers and standardization. Furthermore, classification in microbiology with sometimes hundreds or even thousands of classes would prove unwieldy even for hierarchical ANN classifiers, mainly because the minimum number of training samples per class is not always available and the required number of subnets cannot be trained in a realistic time frame. For the clinically relevant *Yersinia* species we could nonetheless systematically study a sufficient number of bacterial strains and, owing to the hierarchical ANN architecture, reduce the complexity of the individual subnets. With the present study we could thus demonstrate that a combination of MALDI-TOF mass spectrometry and classification analysis by MLP ANN can be successfully employed to unambiguously differentiate between strains of *Yersinia* and strains from other Enterobacteriaceae. With classification accuracies of close to 100% we were furthermore able to reliably identify the three important pathogens of the genus *Yersinia*, *Y. enterocolitica*, *Y. pseudotuberculosis*, and *Y. pestis*.

Conclusions

In this paper we report on the rapid and reliable identification of bacteria from the genus *Yersinia* using a combination of MALDI-TOF mass spectrometry and advanced chemometric methods. Statistical methods for biomarker identification, unsupervised hierarchical clustering, and modular artificial neural networks turned out to be optimal to identify genus- and species-specific biomarkers and to establish classification models for rapid, reliable, and objective identification of highly pathogenic microorganisms. We propose this approach as an additional technique not only for scientific research purposes but also as a routine method to rapidly diagnose the BSL-3 microorganism *Y. pestis* from animal or patient materials.

Tables and Figures

Table 1. Overview of the Microbial Strains and Isolates Used in This Study[a]

genus	species	strain
Citrobacter	*amalonaticus*	04/08695; ATCC 25405
	diversus	ATCC 25408
	freundii	DSM 30039; 07/07764
Edwardsiella	*tarda*	DSM 30052
Enterobacter	*aerogenes*	DSM 30053; NM 20
	(Pantoea) agglomerans	ATCC 27988
	cloacae	DSM 30054
	gergoviae	ATCC 33426
	sakazakii	04/01242
Escherichia	*coli*	08/01585; K12 DSM 3871; NCTC 104118; Nm I, RKI A139
Hafnia	*alvei*	DSM 30163
Klebsiella	*oxytoca*	ATCC 13182
	pneumoniae spp.ozeanae	ATCC 25926; DSM 681
Proteus	*mirabilis*	DSM 788; MN 12
	(Morganelle) morganii	DSM 30117
	stuartii	ATCC 25827
	vulgaris	718/91; ATCC 33420
Salmonella	*enterica serovar minnesota*	SF 1111
	hadar hadar	08/01744
	enteritidis	LT21/Ib 08/01796
	typhimurium	SH 9178; DT 104 08/01627
Serratia	*grimesii*	DSM 30063
	marcesens	DSM 30121
Shigella	*boydii*	II 06/06820
	flexneri	3A 07/02503
Yersinia	*aldovae*	DSM 18303; CIP 3488
	aleksiciae	DSM 14987; IMB 4355
	bercovieri	DSM 18528
	enterocolitica	ssp. *enterocolitica* DSM 11503; ssp. *enterocolitica* DSM 4780; DSM 9676; O:10 32; O:13, 7 6; O:15 31098; O:30 3; O:3 116; O:3; 13169; O:3 211; O:3; 21820; O:3; 25202; O:3 26817; O:3 26972; O:3; 29211; O:3 29213; O:3 29460 II; O:3 29918; O:3 31100; O:3 63/1; O:3 63/2; O:3 68; O:3 79; O:3 82/1; O:3 82/2; O:3 85; O:3 93; O:3 K 201; O:3 K 61; O:3 K 71; O:3 K 79; O:3 K 81; O:3 K 92; O:3 K Y9; O:3 RK/111; O:41,43 31075; O:5, 27 966/89; O:5, 27, Biotyp2; O:5, 27 IP 885; O:50,51 1; O:5 14/91; O:5 29961; O:5 29987; O:5 29988; O:6, 30 29835; O:7, 8 30344;O:8 ATCC 9610; O:9 28144; O:9; 31077; O:9 31079; O:9; 31080; O:9; 31084; O:9;

genus	species	strain
		383; O:9; 7192; O:9 7191; O:9 H 705/86; O:9 H 739/87
	frederiksenii	27; 29; 30; O:44,45 21; O:44,45 7211; CIP 3489; DSM 18490
	intermedia	28; 67; 71; CIP 3490; DSM 18517; O:37 or O:13,7* 16; O:42,54 2885
	kristensenii	DSM 18543; 259/87; CIP 3491; 571/87; 572/87; 573/87; O:46 7230
	mollaretii	DSM 18520; 29211; 50; 51; 55; 60
	pestis	NCTC 10029; NCTC 10030; NCTC 10329; NCTC 10330; NCTC 2028; NCTC 2868; NCTC 570; NCTC 5923; CCUG EV 76; O3-01500; O3-01501; O3-01502; O3-01503; O3-01504; O3-01505; O3-01506; O3-01506; O3−01507; O3-01508; RV 3; vaccine ICM 1/41
	pseudotuberculosis	DSM 8992; 04PA01423; 04WDK36747; 04WI40834; 05WDK14320; 06PW40285; 06WI22829; 07PW12234; 07WI00989; 25743; 25858; 27705; 27707; 28819; 28928; 29490; 29827; BV 1; BV 2; BV 4; G525; I; J9; Nr. 1.5; Typ3 P- INV+; VI
	rohdei	DSM 18270
	ruckeri	04FGD15769; 04FGD26364; 04FGD31174; 04FGD38762; 06FGD29560; 07FGD30558; CIP 3492; DSM 18506; J6
	similis	IMB 4354

a Abbreviations: DSM, Deutsche Sammlung von Mikroorganismen; ATCC, American Type Culture Collection; NCTC, National Collection of Type Cultures; CCUG, Culture Collection, University of Gö̈teborg, Sweden; FGD; CIP, Collection Institute Pasteur, Paris, France; IMB, Institut fü̈r Mikroorganismen der Bundeswehr, Mü̈nchen, Germany.

Table 2. Selected Biomarkers Found at the Family, Genus, and Species Levels

	obsd av mass [*m*/*z*]	tentative protein identity	predicted mass [*m*/*z*]
Enterobacteriaceae (family level)	4185	$[M + 2H]^{2+}$; 30S ribosomal protein S21	4185.3[a]
	8370	$[M + H]^{+}$; 30S ribosomal protein S21	8369.6[a]
Yersinia (genus level)	4350	$[M + H]^{+}$; 50S ribosomal protein L36 1	4350.3
	6046	$[M + H]^{+}$; 50S ribosomal protein L32	6045.8[a]
	6241	$[M + H]^{+}$, 50S ribosomal protein L33	6241.4[a,b]
Y. enterocolitica (species level)	3632	$[M + 2H]^{2+}$; 50S ribosomal protein L29	3631.7
	4805	$[M + 2H]^{2+}$; DNA-binding protein HU-α	4804.5
	7262	$[M + H]^{+}$; 50S ribosomal protein L29	7262.3
	7318	$[M + H]^{+}$; putative Yop protein translocation protein E	7318.4[a]
	9238	$[M + H]^{+}$; DNA-binding protein HU-β	9238.4
	9608	$[M + H]^{+}$; DNA-binding protein HU-α	9607.9
Y. pseudotuberculosis and *Y. pestis* (complex)	6637	$[M + 2H]^{2+}$; ribosomal subunit interface protein	6636.9[a]
	7274	$[M + H]^{+}$; 50S ribosomal protein L29	7274.4
	9268	$[M + H]^{+}$; DNA-binding protein HU-β	9268.5
	9659	$[M + H]^{+}$; 30S ribosomal protein S20	9659.2[a]
Y. pseudotuberculosis (species level)	6474	unassigned [c]	
Y. pestis (species level)	3065	$[M + H]^{+}$; Pla (fragment) NSGDSVSIGGDAAGISNKNYTVTAGLQYRF	3064.3

a Listed masses without N-terminal Met.
b Methylated.
c Not present in *Y. pestis*, but found also in many strains of *Y. enterocolitica*. The protein identity was determined using a protein database search engine (http://www.uniprot.org/) which allows the UniProtKB/Swiss-Prot database to be accessed.

Table 3. Differentiation of MALDI-TOF Mass Spectra from Strains of *Y. pseudotuberculosis* and *Y. pestis* by an ANN[a]

		classification based on ANN analysis of MS spectra	
		Y. pseudotuberculosis	***Y. pestis***
gold standard	*Y. pseudotuberculosis*	20 (63)	0 (0)
	Y. pestis	0 (1)	9 (61)

a In this approach, spectra of *Y. pestis* and *Y. pseudotuberculosis* were used to train and internally validate the *so-called* sublevel 2 ANN (see Figure 7). The confusion matrix shows classification results of the external validation data set and the combined training and internal validation subsets (numbers in parentheses).

Figures

Figure 1. Typical MALDI-TOF mass spectra between *m*/*z* 2500 and *m*/*z* 12000 from three different *Yersinia* species: *Y. enterocolitica* (A), *Y. pseudotuberculosis* (B), and *Y. pestis* (C). Preparation/inactivation of the microbial samples was carried out according to a recently published protocol.(19)

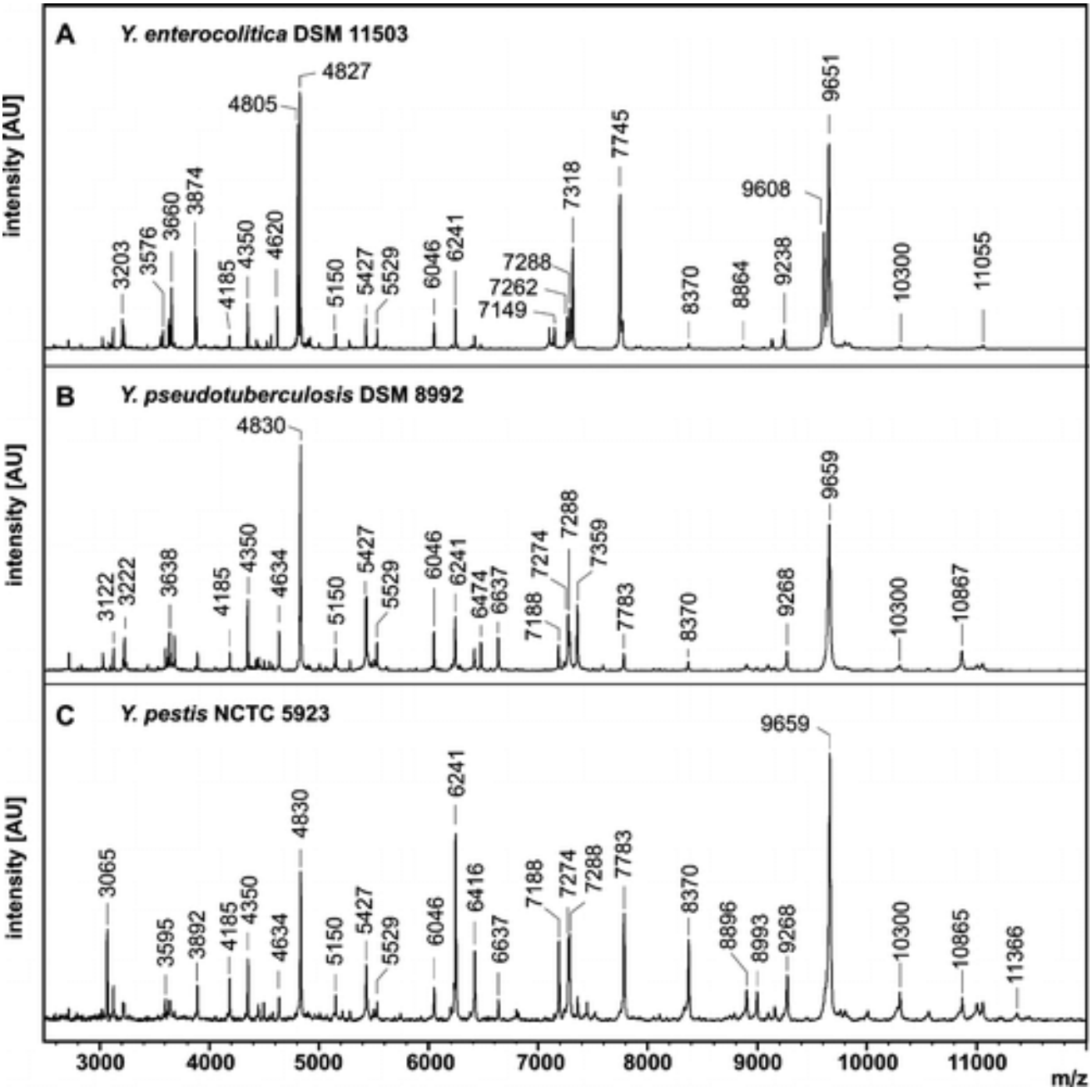

Figure 2. Pseudo gel view in the mass range of *m*/*z* 3500−10500 showing 489 MALDI-TOF spectra from a variety of Enterobacteriaceae, among them 415 spectra from 13 *Yersinia* species. For constructing the gel view representations, mass spectra have been baseline-corrected, smoothed, and vector-normalized. *Yersinia*-specific biomarkers (genus level) are given in italics, *Y. enterocolitica*-specific biomarkers (species level) are underlined, and *Y. pseudotuberculosis* and *Y. pestis* biomarkers are given in bold (also see the text for details). Peaks at *m*/*z* 4185 $(M + 2H)^{2+}$ and *m*/*z* 8370 $(M + H)^{+}$ are found in all mass spectra of Enterobacteriaceae family strains.

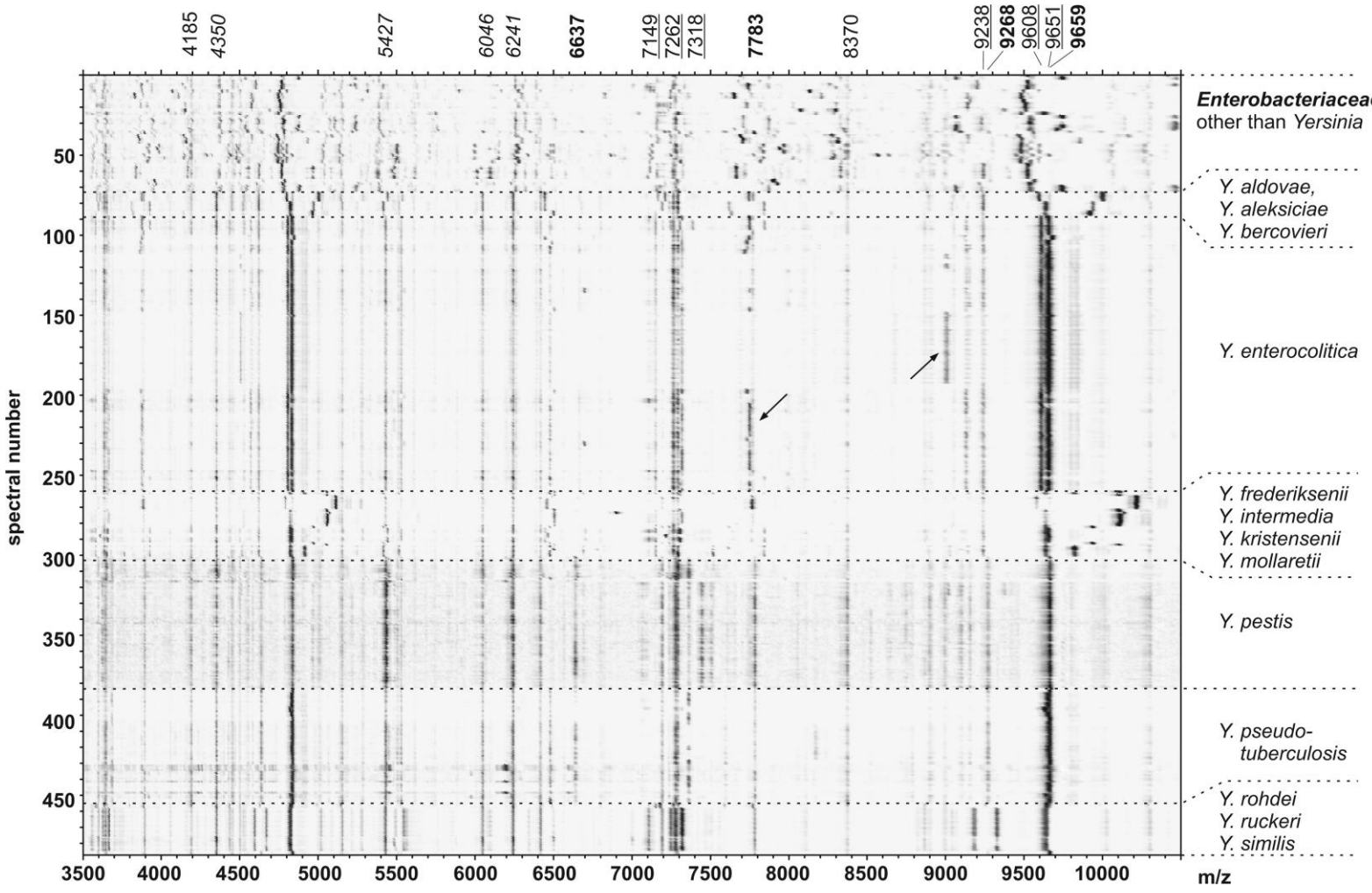

Figure 3. Unsupervised hierarchical cluster analysis of microbial MALDI-TOF mass spectra. In this analysis, a database of spectra from 182 strains of the Enterobacteriaceae family (11 genera, 38 species) was systematically investigated. The dendrogram shows the existence of three main clusters. Cluster I is formed by spectra of *Y. enterocolitica* (cluster Ia) and *Y. enterocolitica*-like species (Ib). Cluster II contains spectra from Enterobacteriaceae family members other than *Yersinia*. Cluster III is composed of MS data from strains of *Y. pseudotuberculosis*, *Y. pestis*, and *Y. similis* (IIIa) and *Y. ruckerii* (IIIb). Note that cluster II contains four outliers from strains of *Y. enterocolitica* (&) and *Y. pestis* (*).

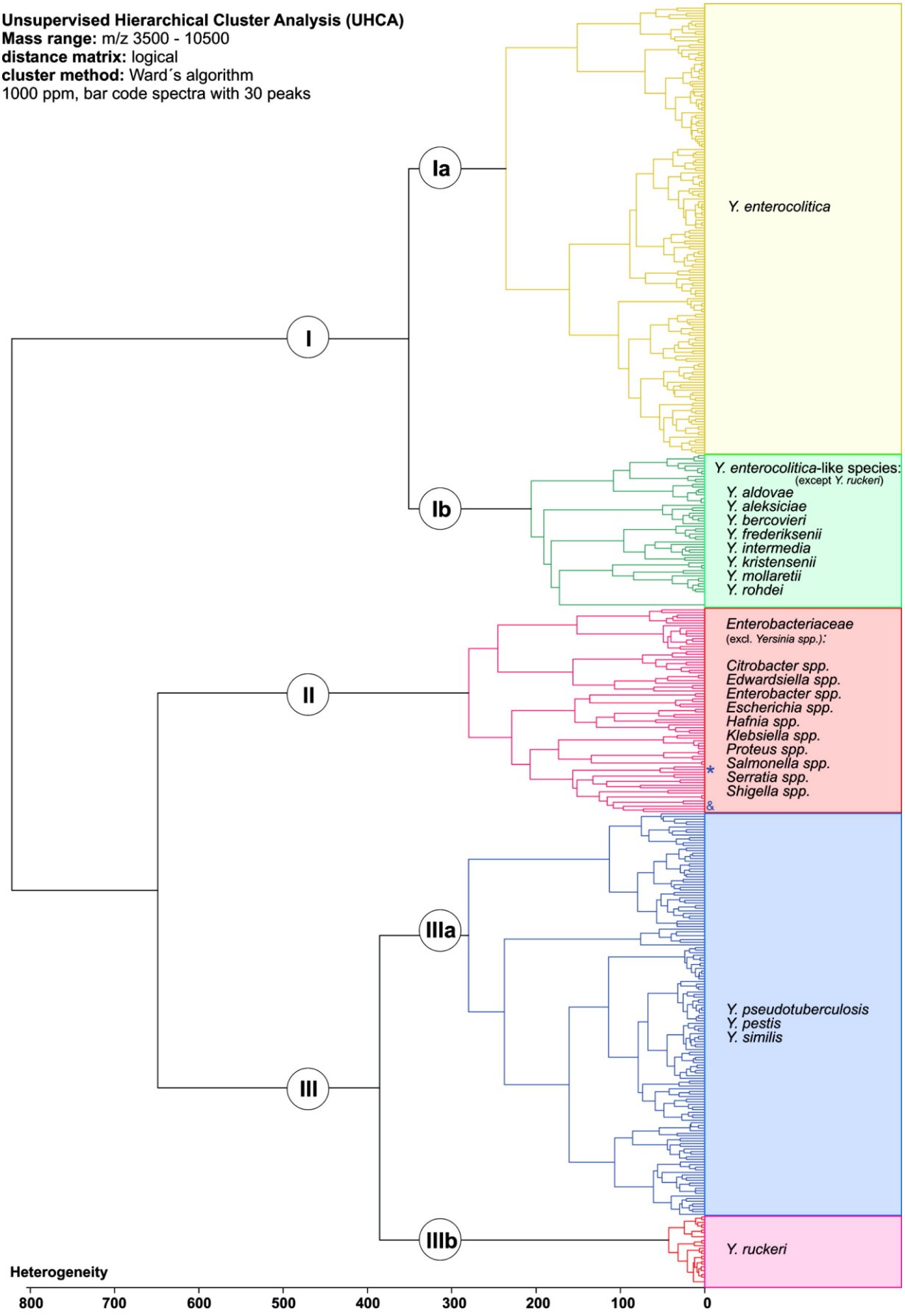

Figure 4. Pseudo gel view of mass spectra from *Y. enterocolitica* (57 strains), *Y. pestis* (21 strains), and *Y. pseudotuberculosis* (26 strains) in the diagnostically relevant mass ranges of *m*/*z* 7080−7390 (left panel) and *m*/*z* 9200−9830 (right panel). The illustration shows selected MALDI-TOF MS biomarkers of *Y. enterocolitica* (underlined) and *Y. pestis* and *Y. pseudotuberculosis* (italics).

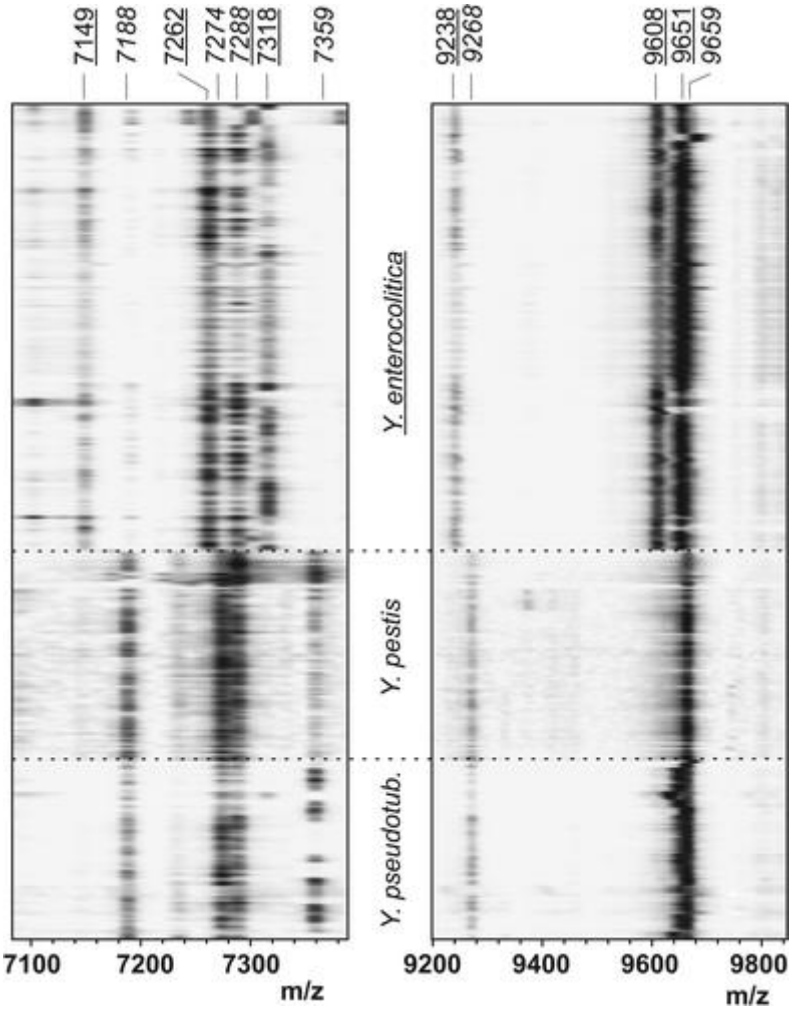

Figure 5. Pseudo gel view with biomarkers of *Y. pestis* (21 strains) and *Y. pseudotuberculosis* (26 strains) in the mass ranges of *m*/*z* 2930−3300 (left panel) and *m*/*z* 6200−6700 (right panel). The gel view demonstrates species-specific biomarkers of *Y. pestis* (underlined) at *m*/*z* 3065 and *Y. pseudotuberculosis* (italics) at *m*/*z* 6474.

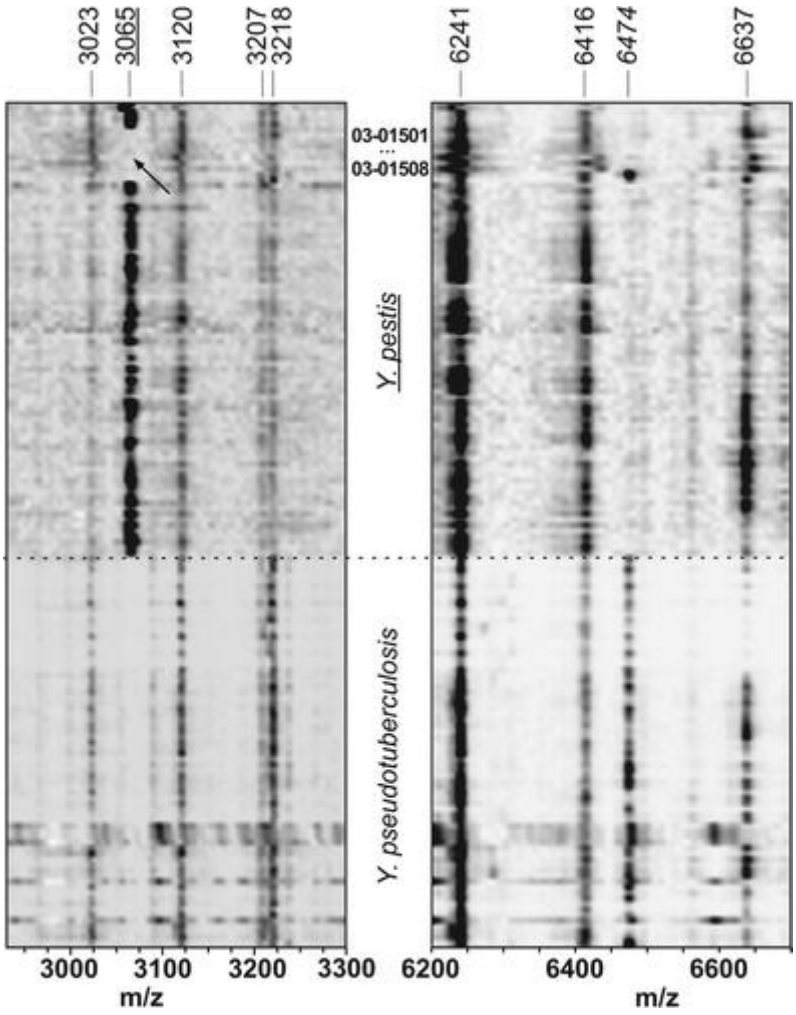

Figure 6. Identification of biomarker ions by a series of independent *t* tests. Univariate *t* tests were carried out in the mass range of *m*/*z* 3500−10500 on the basis of peak tables with 30 peaks per individual mass spectrum. In these analyses, small *p* values cast doubt on the null hypothesis of equal class means. Note the inverse logarithmic scaling. (A) Identification of biomarkers specific for the genus *Yersinia*. Class a contains 415 mass spectra of 13 species of the genus *Yersinia*. Class b contains 74 spectra from 25 Enterobacteriaceae species other than *Yersinia*. (B) Identification of species-specific mass signals of *Y. enterocolitica*: 57 strains (172 spectra) of (class a) *Y. enterocolitica* vs 243 MALDI-TOF spectra of (class b) 12 *Yersinia* species other than *Y. enterocolitica* (88 strains). (C) Differentiation of (class a) *Y. pestis* (21 strains, 82 spectra) and (class b) *Y. pseudotuberculosis* (26 strains, 72 spectra). Key: a, biomarker ions of class a; b, biomarker ions of class b.

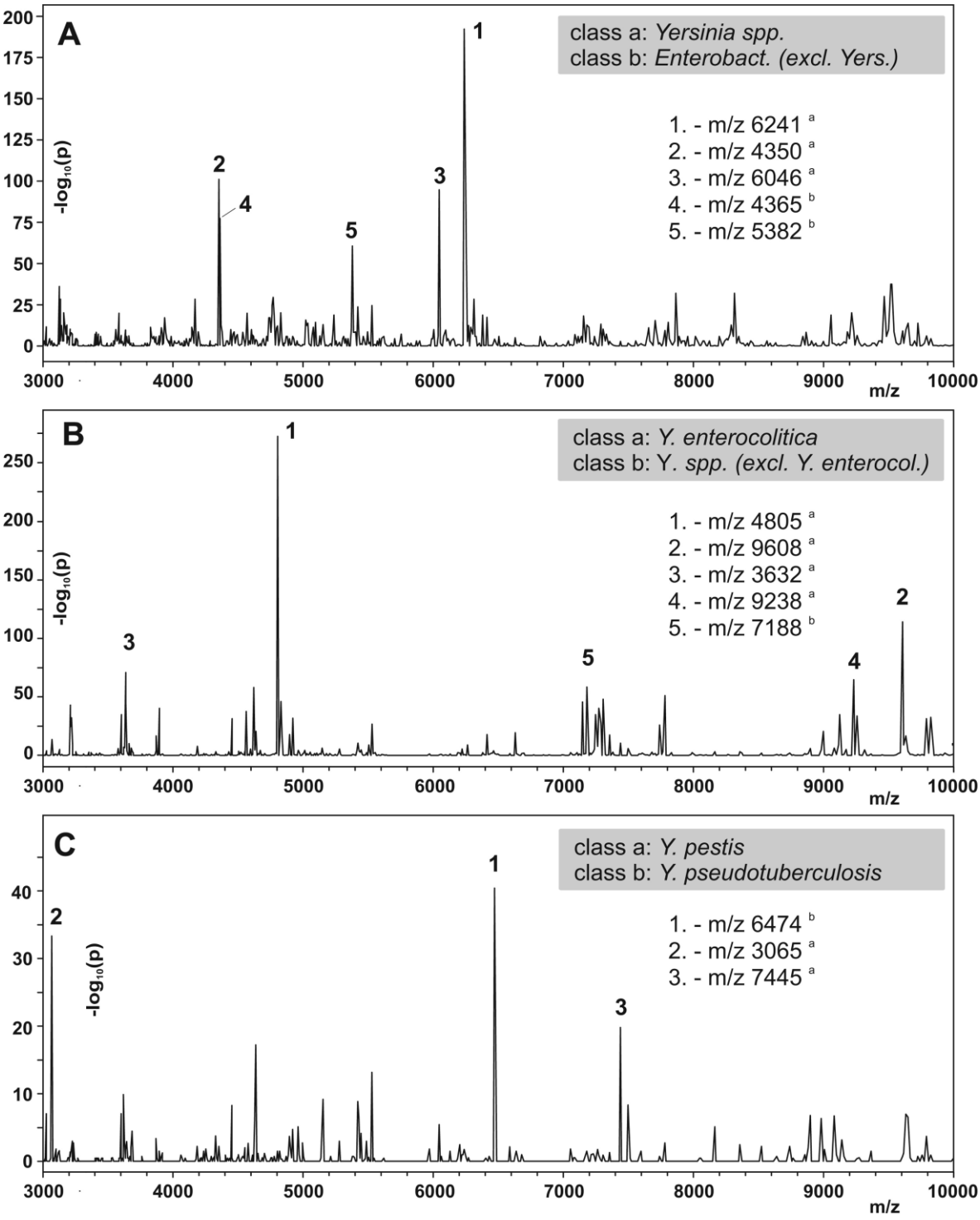

Figure 7. Identification of strains from the Enterobacteriaceae family by MALDI-TOF MS using a hierarchically organized ANN. The modular ANN consists of a top-level ANN which differentiates between *Yersinia* and (i) the other genera of the Enterobacteriaceae family. Two sublevel ANNs allow classification of strains from the genus *Yersinia* with the classes (ii) *Y. enterocolitica*-like species, (iii) *Y. enterocolitica*, (iv) *Y. pseudotuberculosis*, and (v) *Y. pestis*.