Video Article

Subtyping of *Campylobacter jejuni* ssp. *doylei* Isolates Using Mass Spectrometry-based PhyloProteomics (MSPP)

Andreas E. Zautner[1], Raimond Lugert[1], Wycliffe O. Masanta[1], Michael Weig[1], Uwe Groß[1], Oliver Bader[1]

[1]Institut für Medizinische Mikrobiologie, Universitätsmedizin Göttingen

Correspondence to: Andreas E. Zautner at azautne@gwdg.de

URL: http://www.jove.com/video/54165
DOI: doi:10.3791/54165

Keywords: Genetics, Issue 116, MALDI-TOF, below-species differentiation, phyloproteomics, ICMS, *Campylobacter jejuni* ssp. *doylei*, MSPP, microbiology

Date Published: 10/30/2016

Citation: Zautner, A.E., Lugert, R., Masanta, W.O., Weig, M., Groß, U., Bader, O. Subtyping of *Campylobacter jejuni* ssp. *doylei* Isolates Using Mass Spectrometry-based PhyloProteomics (MSPP). *J. Vis. Exp.* (116), e54165, doi:10.3791/54165 (2016).

Abstract

MALDI-TOF MS offers the possibility to differentiate some bacteria not only at the species and subspecies level but even below, at the strain level. Allelic isoforms of the detectable biomarker ions result in isolate-specific mass shifts. Mass spectrometry-based phyloproteomics (MSPP) is a novel technique that combines the mass spectrometric detectable biomarker masses in a scheme that allows deduction of phyloproteomic relations from isolate specific mass shifts compared to a genome sequenced reference strain. The deduced amino acid sequences are then used to calculate MSPP-based dendrograms.

Here we describe the workflow of MSPP by typing a *Campylobacter jejuni* ssp. *doylei* isolate collection of seven strains. All seven strains were of human origin and multilocus sequence typing (MLST) demonstrated their genetic diversity. MSPP-typing resulted in seven different MSPP sequence types, sufficiently reflecting their phylogenetic relations.

The *C. jejuni* ssp. *doylei* MSPP scheme includes 14 different biomarker ions, mostly ribosomal proteins in the mass range of 2 to 11 kDa. MSPP can in principle, be adapted to other mass spectrometric platforms with an extended mass range. Therefore, this technique has the potential to become a useful tool for strain level microbial typing.

Video Link

The video component of this article can be found at http://www.jove.com/video/54165/

Introduction

During the last decade, matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) has advanced to be a highly valued standard method for microbial genus and species identification in clinical microbiology[1,2]. Species identification is based on the recording of small protein fingerprints of intact cells or cell lysates. The typical mass range for a mass spectrometer used in routine clinical microbiology is 2-20 kDa. Additionally, the resulting spectra can be used to discriminate strains at the below-species and below-subspecies level[3]. Early pioneering studies have identified specific biomarker ions for a particular subgroup of strains in *Campylobacter jejuni*[4], *Clostridium difficile*[5], *Salmonella enterica* ssp. *enterica* serovar *Typhi*[6], *Staphylococcus aureus*[7-9], and *Escherichia coli*[10-12].

The combination of several variable biomarker masses corresponding to allelic isoforms offers the option for deeper subtyping. Previously, we successfully implemented a method to convert these variations in mass profiles into meaningful and reproducible phyloproteomic relations called mass spectrometry based phyloproteomics (MSPP) on a *C. jejuni* ssp. *jejuni* isolate collection[13]. MSPP can be used a mass spectrometric equivalent to DNA sequence based subtyping techniques like multilocus sequence typing (MLST).

Campylobacter species are the leading cause of bacterial gastroenteritis worldwide[14,15]. As a consequence of Campylobacteriosis post-infectious sequela, namely, Guillain Barré Syndrome, reactive arthritis and inflammatory bowel disease can arise[16]. The main sources of infection are contaminated livestock meat from chicken, turkey, swine, cattle, sheep and ducks, milk and surface water[15,17]. Therefore, regular epidemiological surveillance studies in the context of food safety are necessary. MLST is the "gold standard" in molecular typing for *Campylobacter* species[18]. Because the Sanger-sequencing based MLST method is labor intensive, time consuming and relatively expensive, MLST typing is restricted to relatively small isolate cohorts. Therefore, there is a need for cheaper and faster subtyping methods. This need could be met by mass spectrometric methods like MSPP.

This paper presents a detailed protocol for MSPP-typing using a collection of *Campylobacter jejuni* ssp. *doylei* isolates and comparison of its potential with MLST.

Protocol

1. Prepare a Safe Workplace by Considering Biosafety Conditions

1. Become familiar with the laboratory and safety regulations that are of relevance for working with microorganisms. Most human pathogenic microorganisms must be handled at biosafety level 2 conditions but some, such as *Salmonella enterica* serovar Typhi, require biosafety level 3. Information on level of handling each pathogen can be accessed at www.cdc.gov/biosafety.
2. Regardless of the biohazard classification of the specific microorganism, regard all materials that came in contact with the infectious agent as infectious waste that must be autoclaved before disposal. Respect regional safety guidelines for hazardous materials and biological substances. Ensure that suitable containers for immediate and proper disposal of potentially contaminated materials (biohazards) are available.
3. Ensure that sterile instruments (inoculation loops), solutions and culture media (agar plates) are available before commencing bacterial culture.
4. Wash hands with antiseptic soap and warm water immediately after handling infectious microorganisms.

2. Select Reference and Collection Isolates

1. Select and obtain one standard genome-sequenced reference isolate along with the sequences of the encoded proteome, ideally in FASTA format. If more genome-sequenced strains are available, include these in the analysis.
   Note: This isolate/these isolates will later on be used to predict the identity of the mass peaks observed in mass spectrometry (see section 7).
2. Select and obtain a variety of potentially diverse isolates in such a way that they cover the phylogeny of the species or subspecies of interest.
   Note: These isolates will later on be used to demonstrate the variability of biomarkers in the population (see section 8).
3. Ensure that the entire collection and reference isolate(s) are properly typed by the respective gold standard for this particular organism[18-20].
   Note: This may include a variety of (sub)-typing methods, but will likely resort to MLST, which still is the standard method to demonstrate genetic diversity of most microbial species.
4. To infer the phylogeny within the collection, calculate a phylogram from the typing data, *e.g.*, using the unweighted pair group method with arithmetic mean (UPGMA) in MEGA6 software for MLST data[21]. For MLST data, also consult a MLST database and assign sequence types and respective clonal complexes[22].
   Note: This will later on be used to analyze the congruency of MSPP with the earlier gold standard typing method (see section 9).

3. Prepare a MALDI Target Plate

CAUTION: TFA is a strong acid. Improper use of TFA bears the risk of severe skin burns, eye damage and severe irritation of the upper respiratory tract if inhaled. Therefore, stringent safety measures must be respected and proper personal protective equipment (PPE) including safety goggles, face shields, appropriate gloves, boots, or even a full protective suit is needed, while handling TFA. Possible exposure to TFA must be controlled by handling the substance under adequate ventilation with an effective exhaust ventilation system. In case of insufficient ventilation, a respirator with approved filter must be used. Additionally, TFA is harmful to aquatic life with long lasting effects. Any release of TFA in waste water to the environment must be avoided.

Note: Before spotting the samples onto a MALDI target, clean the target plate thoroughly if the plate was used previously.

1. Prepare 100 ml 70% aqueous ethanol solution using 30 ml deionized water and 70 ml pure ethanol.
2. Prepare 250 µl of an 80% aqueous trifluoroacetic acid (TFA) solution by mixing 50 µl of deionized water and 200 µl 100% TFA in a reaction tube and vortexing the tube for 1 min.
3. Clean the MALDI target by putting it into a glass dish and submersing it in 70% aqueous ethanol for about 5 min at room temperature.
4. Rinse the target under hot water.
5. Using a paper tissue, wipe the target plate intensively with 70% aqueous ethanol solution to remove all previous samples and other potential debris.
6. If further cleaning is required, rinse under hot water while wiping with a paper tissue.
7. Remove residual, and potentially invisible, contaminants, by covering the target surface with a thin layer of 80% aqueous TFA (~100 µl per 96 spots) and wiping all target positions clean with a paper tissue.
8. Finally, rinse the target to remove acid, wipe it dry using a paper tissue, and leave it for at least 15 min at room temperature to evaporate residual liquid.

4. Preparation of an α-Cyano-4-hydroxy-cinnamic Acid Matrix Solution Containing an Internal Calibrant

1. Prepare a saturated matrix solution by dissolving 10 mg α-cyano-4-hydroxy-cinnamic acid (HCCA) in 1 ml of a mixture of 50% acetonitrile, 47.5% water, and 2.5% TFA. Residual undissolved HCCA will remain if the solution is saturated.
2. Add recombinant human insulin as an internal calibrant. For this, prepare a stock solution to a final concentration of 10 pg/µl in 50% aqueous acetonitrile, aliquot and store at -20 °C for further use.

5. MALDI-TOF Mass Spectrometry

Note: Culture conditions specific for the organisms of interest must be used. Samples for MALDI-TOF MS can be prepared either by smear preparation or extraction, depending on the organism (see section 8.4.1). While the ethanol-formic acid extraction method provides sufficient inactivation of pathogens, smear preparation has to be performed under sufficient biosafety conditions as required (see section 1). Usually, there is no risk of infection after the application of the matrix, but for specific pathogens specific inactivation protocols are required. Thus, for example MALDI-TOF MS of *Nocardia* species requires previous lysis of the bacteria in boiling water, following by ethanol precipitation of proteins[23].
El Khéchine *et al*. developed a procedure for inactivation of *Mycobacteria*, heating the bacterial colonies at 95 °C for 1 hr in screw-cap tubes containing water and 0.5% Tween 20[24].

1. Smear Preparation
    1. Spread a pinhead-sized amount of a bacterial colony directly onto a MALDI target plate position ('spot').
    2. Overlay each spot with 1 µl of HCCA regular matrix or matrix containing the internal calibrant and leave to crystalize at room temperature. For determination of the exact mass of the calibration peak, overlay a control spot with 1 µl Test Standard and 1 µl of HCCA matrix containing the internal calibrant.
    Note: Here, as Test Standard an extract of *Escherichia coli* DH5 alpha is used that demonstrates a characteristic protein fingerprint in MALDI-TOF MS. It is spiked it with two proteins that extend the upper limit of the detectable mass range.

2. Extraction Method
    1. Harvest approximately five colonies from an agar plate culture with an inoculation loop and thoroughly suspend in 300 µl double-distilled water in a 1.5 ml reaction tube. Add 900 µl absolute ethanol and mix thoroughly by repeated pipetting until the bacterial colonies are completely suspended.
    Note: At this step it is possible and well established to store the samples at -20 °C. Additionally, inactivation of pathogens can be tested by streaking 1-10 µl of the extract onto a suitable agar plate following incubation at optimal growth conditions. Successful inactivation is indicated by the absence of microbial growth.
    2. Centrifuge the sample at 13,000 x g for 1 min, discard the supernatant, and dry the pellet at room temperature for 10 minutes. Resuspend the pellet thoroughly by pipetting up-and-down in 50 µl of 70% formic acid.
    3. Add 50 µl of acetonitrile and mix. Remove debris by centrifugation at 13,000 x g for 2 min. Transfer 1 µl of supernatant onto a sample position on a MALDI target plate and leave to dry for 5 min at room temperature.
    4. Overlay each spot with 1 µl of HCCA matrix containing the internal calibrant and leave to crystalize at room temperature.

3. Recording of Mass Spectra
    Note: Peak-picking from mass spectra is done using the standard procedures recommended (Centroid algorithm; S/N ratio: 2; rel. Intensity threshold: 2%; peak width 3 m/z, baseline subtraction: TopHat)
    1. Calibrate the instrument according to the manufacturers' protocol.
    2. For each spot, gather 600 spectra in 100-shots steps.
        1. Go to the "AutoXecute" tab of the configuration software of the mass spectrometer. Open the "Method" by left-clicking onto the "Method" button and choosing the method *e.g.,* "MBT_AutoX" from the pulldown menu.
        2. Left-click the "Edit…" button right of the "Method" menu to open the "AutoXecute Method Editor". Go to the "Accumulation" tab. Set the "Sum up" value to "600" and the "satisfactory shots in" _x_ "shot steps" value to "100".

4. Internal Spectrum Calibration Procedure
    Note: Minute measurement errors are inherent to mass spectrometry. Depending on intermittent instrument use, instrument temperature and re-calibration, the obtained measurement values may vary between experiments. Following pre-measurement instrument calibration and post-measurement spectrum calibration to an internal calibrant is the most precise way to ensure inter-spectrum comparability.
    1. Perform the following procedures for each calibration peak list:
        1. Start spectrum browser (*e.g.*, flexAnalysis and open spectrum: menu "File"→"Open…".)
        2. Create mass control list with calibrant peak: menu "Method"→"Open…".
        3. Choose method: MBT_Standard.FAMSMethod →"Open".
        4. Edit Mass Control List: Uncheck all Calibrants.
        5. Add Calibrant peak at bottom: Peak Label: "Insulin_HIStag[M+H]+_avg"; m_z: "5808.29"; Tolerance[ppm]: "50"; Check Calibrant checkbox.
        6. Save as, *e.g.*, "MSPP calibrant list".
    2. For each spectrum, choose the calibrant peak from the list, click "Automatic Assign" and press "Ok".

6. Verify the Internal Calibration Procedure

1. Experimentally Determine the Exact Mass of the Calibration Peak.
    1. Prepare two spots with 1 µl Test Standard each (step 5.1.1). Overlay the first with 1 µl regular HCCA matrix, the second with 1 µl calibrant-spiked matrix.
    2. Obtain mass spectra from each spot (section 5.3), and internally calibrate to the Test Standard peaks (section 5.4).
    3. Overlay both spectra by opening them with the spectrum browser (*e.g.,* flexAnalysis and open spectrum: menu "File"→"Open…") and finding the peak at the expected mass (insulin m/z = 5,808.29), which should be present in the calibrant-spiked spectrum, but not in the spectrum obtained with the regular matrix.

2. Check that the Calibrant Peak is not Obscured by Any Other Biomarker of the Organism of Interest.

1. Prepare two spots (section 5.1) with the reference strain and overlay the first with 1 µl regular matrix, the second with 1 µl calibrant-spiked matrix.
2. Acquire mass spectra from both spots (section 5.3) and overlay the resulting spectra by opening them with the spectrum browser (*e.g.,* flexAnalysis and open spectrum: menu "File"→"Open…"). Ensure that the calibrant peak is clearly visible in the spectrum obtained with the spiked matrix and not obscured by another adjacent signal. If this is not the case, choose another calibrant for this particular organism.
   Note: Using a spiked internal calibrant significantly increases the precision to determine variations of biomarker masses. Using this method, mass differences down to 1 Da can be detected. Alternatively, also invariant masses originating from the organisms may be used as calibrants. However, by definition, all organism-derived masses must be considered potentially variable, unless proven otherwise.

7. Identify Biomarker Ions in the Reference Strain

1. Measure the Mass Spectrum of the Reference Strain, Using Matrix Spiked with the Internal Calibrant.
   1. Spread a pinhead-sized amount of a bacterial colony (section 5.1) or 1 µl of bacterial protein extract (section 5.2) directly onto a MALDI target plate position ('spot').
   2. Overlay each spot with 1 µl of HCCA matrix containing the internal calibrant and leave the target plate to crystalize at room temperature (section 5.1.2/5.2.4).
   3. Record mass spectra of the reference strain (section 5.3).
2. Internally calibrate the reference spectrum to the calibrant mass (here: insulin at m/z = 5,806.29), and subsequently pre-process by baseline subtraction (TopHat) and smoothing (parameters: SavitzkyGolay; width: 2 m/z, 10 cycles).
   1. Start spectrum browser (*e.g.,* flexAnalysis and open spectrum: menu "File"→"Open…".)
   2. Choose method: pulldown menu "Method" →"Open…", Left-click the method of choice, *e.g.,* "MBT_Standard.FAMSMethod"→ "Open".
   3. Calibrate spectrum by choosing "Internal…" from the pulldown menu "Calibrate". A window opens listing the calibrant peak(s) (section 5.4). Left-click the calibrant peak (Insulin) →Left-click "OK". Choose "Process spectra" from the "Process" pulldown menu.
   4. For baseline subtraction activate the spectrum in the spectrum list at the right side. Choose "Subtract Mass Spectrum Baseline" from the "Process" pulldown menu.
   5. To smooth the spectrum, activate the spectrum in the spectrum list at the right side. Choose "Smooth Mass Spectrum Baseline" from the "Process" pulldown menu.
3. From the genome sequencing data of the reference strain, calculate the theoretical monoisotopic molecular weight of each of the encoded proteins by translating the DNA sequence into the corresponding amino acid sequence using a sequence alignment editor. Copy-paste this protein sequence into the input box at the ExPASy Bioinformatics Resource Portal (http://web.expasy.org/compute_pi/). and press "Click here" to compute pI/Mw. In the case of *C. jejuni* ssp. *doylei* calculate the mass of 14 detectable biomarkers.
4. Copy the results into a spreadsheet, with one column containing the gene identifier and the next the molecular weight. Sort the rows by calculated molecular weight to facilitate easier lookup of the masses. Note: Other columns are optional; functional annotation may be especially useful later for interpretation.
5. Insert a second column into the spreadsheet for the molecular weight of the de-methioninated form, subtracting 135 Da from the monoisotopic molecular weight. Note: This is because some proteins undergo posttranslational modification by proteolytic removal of the *N*-terminal methionine.
6. Assign each major measured biomarker mass to the calculated masses from the reference strain by looking up the measured mass from the genome table prepared above (**Table 1**).
7. If the biomarker ions cannot be assigned to predicted gene products, consider other posttranslational modifications (methylation, acetylation, prenylation, *etc.*; see http://www.abrf.org/delta-mass for a compilation of modifications and the associated mass changes). For any other known posttranslational modification that is frequently observed in the organisms of interest add another column to the table and recalculate the molecular weight analogous to the process for the de-methioninated form.
8. Set up another spreadsheet tab, and record for each biomarker ion the mass and identifier in a separate table column.

8. Assess Biomarker Variability in the Population

1. Calibrate mass spectra obtained from collection isolates, as done for the reference strain(s) (section 5.4.2).
2. Identify variant biomarkers in the mass spectra. A variant mass is characterized by the absence of a mass known from the reference spectrum and appearance of a novel mass not present in the reference. The mass difference must conform to a single amino acid exchange (**Table 2**), or a combination thereof.
3. At one isolate per row, record the measured mass for each biomarker and the predicted isoform in the respective table columns.
   Note: Rarely, some mass shifts may be attributable to several different amino acid exchanges, *e.g.,* both, N exchanged by D and Q exchanged by E, and vice versa, result in a mass shift of 0.985 Da (**Table 2**). This intrinsic problem cannot be resolved by mass spectrometry alone. Therefore, different isoforms on the protein sequence level having the same mass must be treated as a single MSPP type. Parallel Sanger sequencing of the particular biomarker ion genes confirmed that this problem did not occur while MSPP-typing the *C. jejuni* ssp. *doylei* isolate collection used in this study.
4. Confirm novel MSPP types by PCR-amplifying and sequencing the respective biomarker genes. In turn, this also serves as confirmation that the biomarker identity has been assigned correctly.
   1. Culture bacterial isolates under optimal growth conditions. Culture *C. jejuni* ssp. *doylei* strains on Columbia agar supplemented with 5% sheep blood at 37 °C under microaerophilic conditions (5% O_2, 10% CO_2, 85% N_2). Incubate for ca. 48 hr. Use a separate agar plate for each isolate to avoid cross-contamination.
   2. Extract genomic DNA of the bacterial isolates using an appropriate DNA extraction kit/ automated machinery according to manufacturer's instructions.

3. Amplify the respective biomarker genes using the primers listed in **Table 3**. Perform all PCR reactions under the following conditions: denaturation at 94 °C for 30 sec; annealing at 55 °C for 30 sec; elongation at 72 °C for 30 sec.
4. Determine the DNA sequence of each amplicon by Sanger sequencing using an appropriate amount of genomic DNA (usually 600-700 ng of DNA is sufficient at a concentration of ca. 100 ng/µl) and one of the amplification primers (usually this is done by use of a service provider).

5. In a separate table (see **Table 4**), record the deduced protein sequence for each novel isoform by using an appropriate translation tool (*e.g.*, Transseq: http://www.ebi.ac.uk/Tools/st/) [25].

9. Calculate a MSPP-based Phylogeny and Compare to the Gold Standard

1. Concatenate the particular biomarker amino acid sequences belonging to the MSPP type of the isolates into one continuous sequence using a sequence alignment editor, such as BioEdit (http://www.mbio.ncsu.edu/bioedit/bioedit.html)[26] or the biomarker spreadsheet.
2. Calculate phylogeny by clustering as done for the gold standard typing data, *e.g.*, UPGMA clustering [21].
3. Compare the MSPP-based phylogeny to the one obtained with the gold standard[4,13].

Representative Results

Previously, we successfully established a MSPP scheme for *C. jejuni* ssp. *jejuni*[13]. Here, we aimed to extend the method to the sibling subspecies *C. jejuni* ssp. *doylei*. In this specific setting, seven *C. jejuni* ssp. *doylei* isolates were acquired from the Belgian collection of microorganisms/Laboratory of Microbiology UGent BCCM/LMG Ghent, Belgium. All seven isolates used for our analyses were of human origin. The genome-sequenced strain ATCC 49349 (LMG 8843), originally isolated from feces of a 2-year-old Australian child with diarrhea served as the reference. From the collection, six additional strains were available: LMG 9143, LMG 9243, and LMG 9255 isolated from feces of children suffering diarrhea and living in Brussels, Belgium; LMG 7790 (ATCC 49350) isolated from a gastric biopsy sample of an individual from Germany; and LMG 8870 (NCTC A613/87) as well as LMG 8871 (NCTC A603/87), both isolated from different blood samples drawn from South African children.

All *C. jejuni* ssp. *doylei* isolates were stored at -80 °C as cryobank stocks, freshly thawed for each analysis and cultured by using Columbia agar base supplemented with 5% sheep blood and incubation for 48 hr at 37 °C under microaerophilic conditions (5% O_2, 10% CO_2, 85% N_2).

In contrast to the establishment of the MSPP technique for *C. jejuni* ssp. *jejuni*[13] it was not possible to base the method on an isoform database derived from a larger sequence collection. Only a single *C. jejuni* ssp. *doylei* entry was present in the rMLST database, and this corresponded to the reference strain *C. jejuni* ssp. *doylei* ATCC 49349[27]. Therefore, each biomarker mass shift detected in comparison to strain *C. jejuni* ssp. *doylei* ATCC 49349 was a new entry in the isoform database and needed to be reconfirmed by Sanger sequencing.

To analyze the diversity of the obtained *C. jejuni* ssp. *doylei* isolates, MLST was performed and a MLST-based UPGMA-dendrogram including the seven examined *C. jejuni* ssp. *doylei* isolates and *C. jejuni* ssp. *jejuni* strain 81-176 as the outgroup calculated (**Figure 1**). Each *C. jejuni* ssp. *doylei* isolate belonged to a different MLST-sequence type, falling into two subclusters. Strain LMG7790 had the longest phylogenetic distance as compared to the remaining *C. jejuni* ssp. *doylei* isolates.

The recordable MALDI-TOF reference mass spectrum of *C. jejuni* ssp. *doylei* ATCC 49349 contained 14 singularly charged biomarker ions that could be identified by comparison of the calculated molecular masses with the reference spectrum (**Figure 2**).

Within the collection, varying isoforms were detected for L32-M, L33, the hypothetical protein encoded by *gi|152939117*, S20-M, and S15-M (**Figure 3**). The remaining masses assigned to biomarker ions were invariable in the tested *C. jejuni* ssp. *doylei* isolates. The amino acid substitution corresponding to the mass shifts have been identified by PCR-amplification and Sanger sequencing of the particular gene using the primers listed in **Table 3**. **Table 4** lists the amino acid sequences of all biomarker ions included in the MSPP scheme as well as the detected allelic isoforms.

A MSPP-based phyloproteomic UPGMA-tree (**Figure 4**) was calculated for the same seven *C. jejuni* ssp. *doylei* isolates and *C. jejuni* ssp. *jejuni* strain 81-176 as the outgroup using the concatenated amino acid sequences of all 14 biomarker ions in order of their molecular weight in the reference strain. Each isolate represented an individual MSPP-sequence type. Although the obtained phyloproteomic relations were not fully identical to the ones obtained by MLST, the *C. jejuni* ssp. *doylei* isolates again arranged in two subclusters. The first subcluster was formed by LMG 8843, LMG 8870, and LMG 9243, the second by LMG 9255, LMG 9143, LMG 8871 and LMG 7790. As seen with the MLST-analysis, isolate LMG7790 showed the longest phyloproteomic distance in the MSPP-analysis.

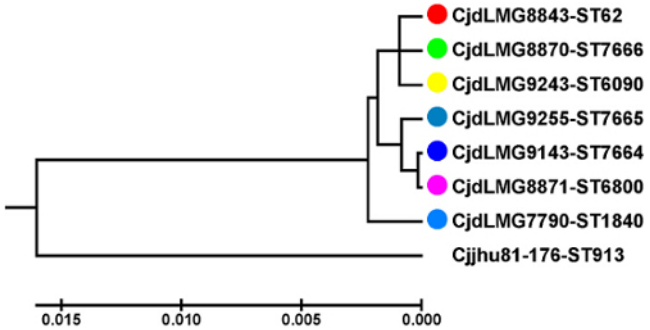

Figure 1: MLST-based phylogenetic UPGMA-tree. Balanced MLST-based UPGMA-dendrogram constructed from seven *C. jejuni* ssp. *doylei* isolates and *C. jejuni* ssp. *jejuni* strain 81-176. Strain color code: LMG 8843 (black), LMG 8870 (grey), LMG 9243 (blue), LMG 9255 (turquoise), LMG 9143 (green), and LMG 8871 (pink), LMG 7790 (yellow), and 81-176 (white). The MLST-sequence type is given behind the name of each isolate. Every *C. jejuni* ssp. *doylei* isolate belongs to a different MLST-sequence type. X-axis indicates the linkage distances. Strain LMG 7790 shows the longest phylogenetic distance compared to the remaining six isolates. Strains LMG 8843, LMG 8870, and LMG 9243 as well as LMG 9255, LMG 9143, and LMG 8871 form two different subclusters within the *C. jejuni* ssp. *doylei* cluster. Please click here to view a larger version of this figure.

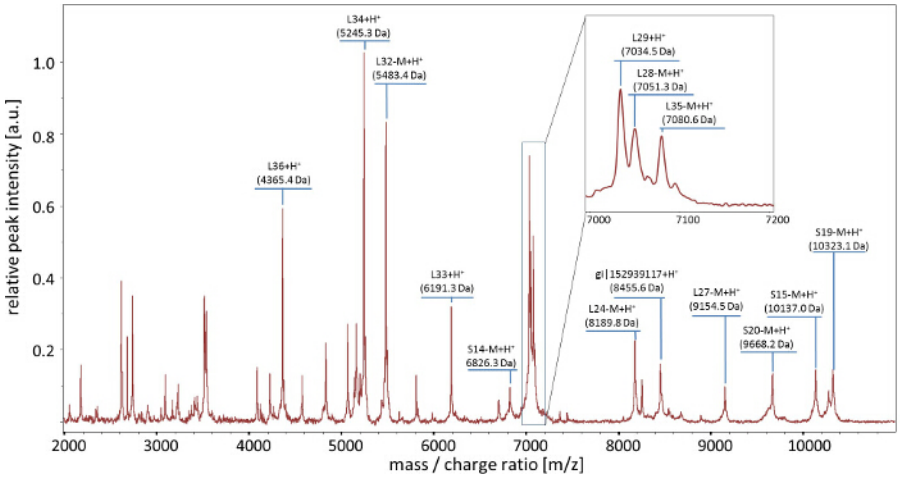

Figure 2: Tentative assignment of genomic correlates of *C. jejuni* ssp. *doylei* ATCC 49349 (LMG 8843) to observed biomarker masses. Based on the calculated masses (average isotopic composition) of predicted ORFs from the whole genome sequence of *C. jejuni* ssp. *doylei* strain ATCC 49349, biomarker masses were assigned to the corresponding protein coding sequences. However, within measurement range, the biomarker masses for the ribosomal subunits L31 (MW = 7,315 Da), S17 (MW = 9,549 Da), and S18 (MW = 10,285 Da) as well as their de-methioninated isoforms (inset m/z = 7,000-7,200 Da) could not be unambiguously assigned. Therefore, L31, S17, and S18 were not included in the *C. jejuni* ssp. *doylei* MSPP scheme. Please click here to view a larger version of this figure.

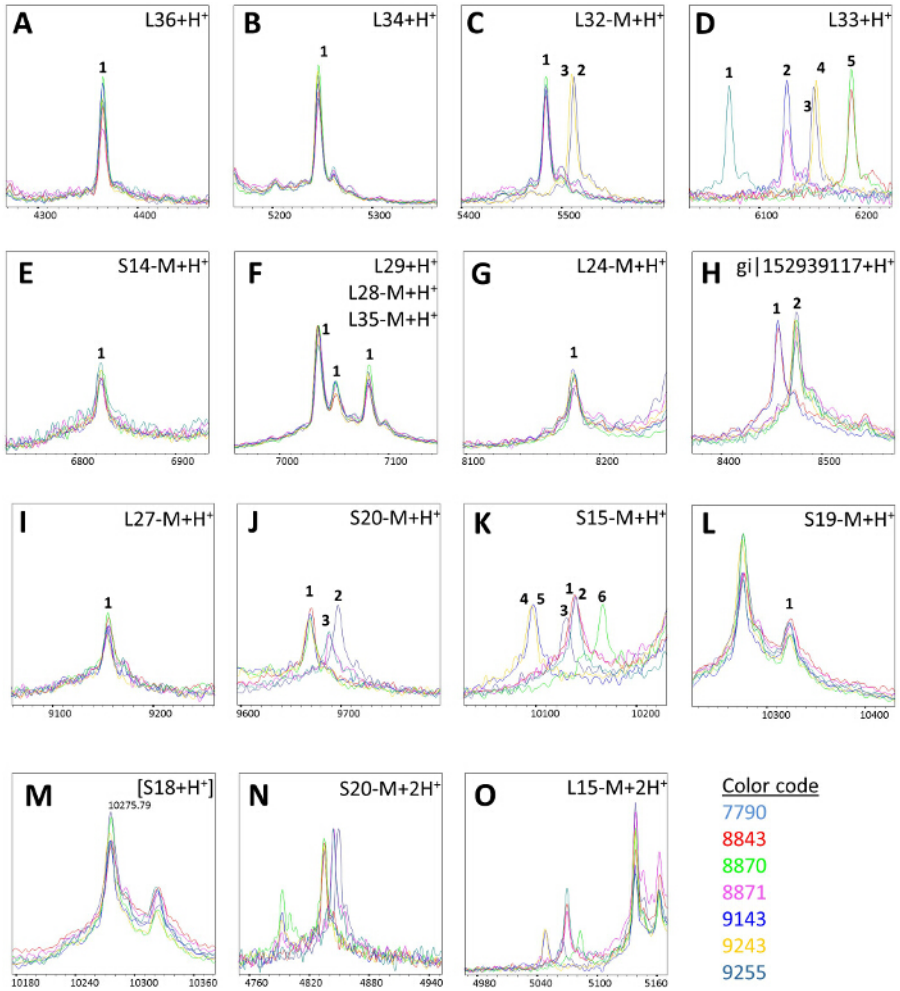

Figure 3: Specific biomarker mass peaks of the *C. jejuni* ssp. *doylei* MSPP scheme. In each panel the mass spectra of all seven tested *C. jejuni* ssp. *doylei* isolates (color code as in **Figure 1**) corresponding to the particular MSPP types have been overlaid to indicate biomarker mass shifts due to allelic isoforms: (**A**) L36+H$^+$; (**B**) L34+H$^+$; (**C**) L32-M+H$^+$; (**D**) L33+H$^+$; (**E**) S14-M+H$^+$; (**F**) L29+H$^+$, L28-M+H$^+$, and L35-M+H$^+$; (**G**) L24-M+H$^+$; (**H**) gi|152939117+H$^+$; (**I**) L27-M+H$^+$; (**J**) S20-M+H$^+$; (**K**) S15-M+H$^+$; (**L**) S19-M+H$^+$; (**M**) an intense mass obscuring variable potential biomarker peak of ribosomal protein S18; (**N**) S20-M+2H$^+$; (**O**) L15-M+2H$^+$; X-Axes: mass [Da]·charge-1 ratio, scale 200 Da. Y-Axes: intensity [arbitrary units], spectra were individually adjusted to a comparable noise level for better visualization of low-intensity peaks. "-M" after the name of a ribosomal subunit indicates the de-methioninated isoform. Please click here to view a larger version of this figure.

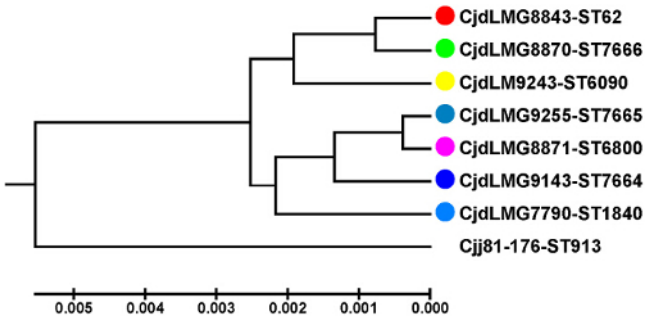

Figure 4: MSPP-based phyloproteomic UPGMA-tree. The depicted phyloproteomic UPGMA-tree includes the same seven *C. jejuni* ssp. *doylei* isolates and *C. jejuni* ssp. *jejuni* strain 81-176 as in **Figure 1**. A colored spot (color code as in **Figure 1**) indicates each strain. For better comparison the isolates are arranged in input order, which was the order obtained from the MLST-based tree. The axis below the dendrogram indicates the linkage distances. Although the obtained phyloproteomic relations are not completely identical to the phylogenetic MLST-based tree, the global picture is comparable. The population splits into two groups: the three isolates LMG 8843, LMG 8870, and LMG 9243 form one cluster, while LMG 9255, LMG 9143, LMG 8871 and LMG 7790 form a second cluster. Within this cluster LMG 7790 shows the longest phyloproteomic distance compared to the subcluster formed by LMG 9255, LMG 9143, and LMG 8871. Within this subcluster LMG 9143 switches position in relation to LMG 9255. However, even using the MSPP-method each isolate represents an individual MSSP-sequence type. Please click here to view a larger version of this figure.

Table 1: Calculated masses of all annotated ORFs of the *C. jejuni* ssp. *doylei* ATCC 49349 strain. List of all calculated monoisotopic molecular weights of each protein encoded in the *C. jejuni* ssp. *doylei* ATCC 49349 genome. The ExPASy Bioinformatics Resource Portal was used for calculation of the particular molecular masses. Column B lists the methioninated isoforms whereas column C lists the de-methioninated isoforms. Biomarker masses included in the *C. jejuni* ssp. *doylei* MSPP scheme are highlighted in red. Note: The biomarker protein L36 is not annotated in the genome sequence of *C. jejuni* ssp. *doylei* ATCC49349. Please click here to download this file.

Table 2: Mass changes induced by amino acid changes due to single SNPs. All potential single nucleotide polymorphisms were checked for the resulting amino acid exchange using the standard genetic code. All silent mutations were discarded, and nonsynonymous mutations compiled together with the resulting mass change. Please click here to download this file.

Table 3: Oligonucleotide primers used for sequencing of the *C. jejuni* ssp. *doylei* genes included in the MSPP-scheme Please click here to download this file.

Table 4: Overview of all isoforms included in the *C. jejuni* ssp. *doylei* MSPP-scheme. This table lists all biomarker ions included in the *C. jejuni* ssp. *doylei* MSPP scheme. The amino acid sequence of the reference strain ATCC 49349 is given completely. For further detectable isoforms, only specific amino acid substitutions are listed. The amino acid numbering always includes the start-methionine; if mass spectrometry indicates its absence, it is written in brackets (M). For each isoform molecular mass, mass difference to reference strain ATCC 49349 isoforms and frequency within the isoform dataset is indicated. Please click here to download this file.

Discussion

The most critical step in the establishment of an MSPP scheme is the unequivocal genetic determination of biomarker ion identities. If it is not possible to identify a biomarker undoubtedly, then it should be excluded from the scheme[13].

The *C. jejuni* ssp. *doylei* scheme includes 14 different biomarker ions. These are 5 less compared to the *C. jejuni* ssp. *jejuni* MSPP scheme[13].The most significant difference between the detectable *C. jejuni* ssp. *jejuni* and *C. jejuni* ssp. *doylei* biomarkers was the posttranslational removal of the amino-terminal methionine in case of L31(-M) and L35(-M)[13].

As shown in the tested *C. jejuni* ssp. *doylei* isolate collection, MSPP was able to discriminate all seven different MLST sequence types. This means that every tested isolate belonged to a specific MSPP sequence type. Additionally, MSPP allows subspecies differentiation between *C. jejuni* ssp. *jejuni* and *C. jejuni* ssp. *doylei*.

In general, the potential to discriminate isolates at the below-strain level by MSPP is lower compared to DNA sequence-based methods like MLST. Assuming a well-established isoform database, the subtyping of bacterial isolates is much faster and much cheaper compared to DNA sequencing methods[4,13].

The biggest advantage of MSPP in comparison to hierarchical clustering methods of ICMS-spectra such as using the principal component analysis (PCA)[4,10,28] or UPGMA analysis of the mathematical matrix that results of the comparison of binary peak matching profiles[29,30] is its high intrinsic reproducibility. While different cultural conditions such as different culture media, different agar charges, different incubation temperature, and different incubation periods significantly affect the phyloproteomic relations calculated by PCA, they do not affect the MSPP deduced relations [6,13]. The main reason for this high reproducibility is the independence of the peak intensity and mass spectrum quality in general[13]. If a MSPP relevant peak in the mass spectrum of an isolate cannot be identified indubitably, it is necessary to record the mass spectrum of the individual isolate again.

The MSPP method can be adapted, in principle, to every microbial species. Especially, the subtyping of clinical relevant microbial species like *Clostridium difficile*, *Escherichia coli*, *Staphylococcus aureus*, and *Salmonella enterica* ssp. *enterica* are potential future applications. If the expression of virulence factors or resistance genes correlates with the phyloproteomic relationship, MSPP can become an innovative tool in routine clinical diagnostics. The number of detectable peaks and thereby the number of biomarker ions included in the MSPP scheme could potentially be increased by using specific extraction and/or lysis protocols, especially in the case of fungi[31].

The recently used MALDI-TOF techniques can mainly detect only the highly abundant ribosomal proteins. These highly abundant proteins interfere with nearly all other smaller proteins such as virulence factors in the mass spectrum. Smaller proteins could be detected by linking ESI-TOF-MS (Electrospray ionization time of flight mass spectrometry) and HPLC (high performance liquid chromatography)[32]. However, currently this method is too labor and cost intensive to characterize larger isolate cohorts.

Disclosures

The authors have nothing to disclose.

Acknowledgements

We are grateful to Hannah Kleinschmidt for excellent technical support. This paper was funded by the Open Access support program of the Deutsche Forschungsgemeinschaft and the publication fund of the Georg August Universität Göttingen.

References

1. Seng, P., *et al.* Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin Infect Dis* **49**(4), 543-51 (2009).

2. Bader, O. MALDI-TOF-MS-based species identification and typing approaches in medical mycology. *Proteomics.* **13**(5), 788-99 (2013).
3. Sandrin, T. R., Goldstein, J. E., & Schumaker, S. MALDI TOF MS profiling of bacteria at the strain level: a review. *Mass Spectrom Rev.* **32**(3), 188-217 (2013).
4. Zautner, A. E., *et al.* Discrimination of multilocus sequence typing-based Campylobacter jejuni subgroups by MALDI-TOF mass spectrometry. *BMC Microbiol* **13**, 247 (2013).
5. Reil, M., *et al.* Recognition of Clostridium difficile PCR-ribotypes 001, 027 and 126/078 using an extended MALDI-TOF MS system. *Eur J Clin Microbiol Infect Dis* **30**(11), 1431-6 (2011).
6. Kuhns, M., & Zautner, A. E., *et al.* Rapid discrimination of Salmonella enterica serovar Typhi from other serovars by MALDI-TOF mass spectrometry. *PLoS One* **7**(6), e40004 (2012).
7. Wolters, M., *et al.* MALDI-TOF MS fingerprinting allows for discrimination of major methicillin-resistant Staphylococcus aureus lineages. *Int J Med Microbiol* **301**(1), 64-8 (2011).
8. Josten, M., *et al.* Analysis of the matrix-assisted laser desorption ionization-time of flight mass spectrum of Staphylococcus aureus identifies mutations that allow differentiation of the main clonal lineages. *J Clin Microbiol* **51**(6), 1809-17 (2013).
9. Lu, J. J., Tsai, F. J., Ho, C. M., Liu, Y. C., & Chen, C. J. Peptide biomarker discovery for identification of methicillin-resistant and vancomycin-intermediate Staphylococcus aureus strains by MALDI-TOF. *Anal Chem.* **84**(13), 5685-92 (2012).
10. Novais, A., *et al.* MALDI-TOF mass spectrometry as a tool for the discrimination of high-risk Escherichia coli clones from phylogenetic groups B2 (ST131) and D (ST69, ST405, ST393). *Eur J Clin Microbiol Infect Dis* (2014).
11. Matsumura, Y., *et al.* Detection of extended-spectrum-beta-lactamase-producing Escherichia coli ST131 and ST405 clonal groups by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol* **52**(4), 1034-40 (2014).
12. Christner, M., *et al.* Rapid MALDI-TOF Mass Spectrometry Strain Typing during a Large Outbreak of Shiga-Toxigenic Escherichia coli. *PLoS One* **9**(7), e101924 (2014).
13. Zautner, A. E., Masanta, W. O., Weig, M., Groß, U., & Bader, O. Mass Spectrometry-based PhyloProteomics (MSPP): A novel microbial typing Method. *Scientific Reports.* **5** (2015).
14. World Health Organisation (WHO). The global view of campylobacteriosis. *WHO.* at http://www.who.int/foodsafety/publications/campylobacteriosis/en/ (2013).
15. Dasti, J. I., Tareen, A. M., Lugert, R., Zautner, A. E., & Gross, U. Campylobacter jejuni: a brief overview on pathogenicity-associated factors and disease-mediating mechanisms. *Int J Med Microbiol.* **300**(4), 205-11 (2010).
16. Zautner, A. E., *et al.* Seroprevalence of campylobacteriosis and relevant post-infectious sequelae. *Eur J Clin Microbiol Infect Dis* **33**(6), 1019-27 (2014).
17. Zautner, A. E., Herrmann, S., & Groß, U. Campylobacter jejuni - The Search for virulence-associated factors. *Archiv Fur Lebensmittelhygiene.* **61**(3), 91-101 (2010).
18. Dingle, K. E., *et al.* Multilocus sequence typing system for Campylobacter jejuni. *J Clin Microbiol* **39**(1), 14-23 (2001).
19. Dingle, K. E., *et al.* Molecular characterization of Campylobacter jejuni clones: a basis for epidemiologic investigation. *Emerg Infect Dis* **8**(9), 949-55 (2002).
20. Cody, A. J., *et al.* Real-time genomic epidemiological evaluation of human Campylobacter isolates by use of whole-genome multilocus sequence typing. *J Clin Microbiol* **51**(8), 2526-34 (2013).
21. Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* **30**(12), 2725-9 (2013).
22. Jolley, K. A., Chan, M. S., & Maiden, M. C. mlstdbNet - distributed multi-locus sequence typing (MLST) databases. *BMC Bioinformatics.* **5**, 86 (2004).
23. Verroken, A., *et al.* Evaluation of Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry for Identification of Nocardia Species. *J Clinl Microbiol* **48**(11), 4015-4021 (2010).
24. Khéchine, A. El, Couderc, C., Flaudrops, C., Raoult, D., & Drancourt, M. Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry Identification of Mycobacteria in Routine Clinical Practice. *PLoS ONE.* **6**(9), e24720 (2011).
25. Goujon, M., *et al.* A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research* **38**(Web Server issue), W695-699 (2010).
26. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series.* **41**, 95-98 (1999).
27. Jolley, K. A., *et al.* Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* **158**(Pt 4), 1005-15 (2012).
28. Suarez, S., *et al.* Ribosomal proteins as biomarkers for bacterial identification by mass spectrometry in the clinical microbiology laboratory. *J Microbiol Methods* **94**(3), 390-6 (2013).
29. Teramoto, K., *et al.* Phylogenetic classification of Pseudomonas putida strains by MALDI-MS using ribosomal subunit proteins as biomarkers. *Anal Chem* **79**(22), 8712-9 (2007).
30. Teramoto, K., Kitagawa, W., Sato, H., Torimura, M., Tamura, T., & Tao, H. Phylogenetic analysis of Rhodococcus erythropolis based on the variation of ribosomal proteins as observed by matrix-assisted laser desorption ionization-mass spectrometry without using genome information. *J Biosci Bioeng.* **108**(4), 348-53 (2009).
31. Bernhard, M., Weig, M., Zautner, A. E., Gross, U., & Bader, O. Yeast on-target lysis (YOTL), a procedure for making auxiliary mass spectrum data sets for clinical routine identification of yeasts. *J Clin Microbiol.* **52**(12), 4163-7 (2014).
32. Stark, T., *et al.* Mass spectrometric profiling of Bacillus cereus strains and quantitation of the emetic toxin cereulide by means of stable isotope dilution analysis and HEp-2 bioassay. *Anal Bioanal Chem* **405**(1), 191-201 (2012).

