Introduction

Campylobacter spp. can cause gastrointestinal and extra-intestinal infections [1]. Although the majority of cases (>90%) of intestinal campylobacteriosis are caused by Campylobacter jejuni and Campylobacter coli, a small number of these cases are also caused by Campylobacter fetus [2–5]. Among these, C. fetus is the most common cause of Campylobacter bacteremia. The frequency of detection in blood cultures varies between 19% and 53% [6–8] of all campylobacterioses. The reported case fatality rate of invasive C. fetus infections is at 14% [9]. Due to the high incidence rate of campylobacteriosis worldwide, this shows that C. fetus infections occur frequently and have the potential to become a significant public health issue. However, relatively little is known about the infection sources and the people at risk, so far. Most reported C. fetus infections were observed in AIDS patients and other immunocompromised individuals [1, 10].

C. fetus is a Gram-negative, microaerophilic bacterium, growing between 25 °C and 37 °C. Clinical symptoms of human C. fetus infection vary from acute diarrhea to systemic illness [11, 12], and the presentation of these symptoms depends on localization of the disseminated pathogen. Septicemia with fever, but without apparent localized infection, for example, is reported in 24% to 41% of cases [7, 9]. Other manifestations can be the result of neurological infections (i.e., meningitis, encephalitis, meningitis, or brain abscesses), arthritis, lung abscesses, osteomyelitis, and perinatal infections (i.e., abortion, infection in uterus, or placenta) [12]. Furthermore, C. fetus infections may also cause vascular pathology (i.e., endocarditis, pericarditis, vasculitis, and mycotic aneurysms) [13].

Currently, 3 subspecies of C. fetus are known. These are C. fetus subspecies fetus (Cff), C. fetus subspecies venerealis (Cfv), and C. fetus subspecies testudinum (Cft). For Cfv, also the biovar intermedius (Cfvi) has been identified in previous studies [14, 15]. Subspecies Cff and Cfv are primarily associated with mammals whereas Cft is associated with reptiles.

In combination, the 9 identified biomarkers allow the differentiation of Cff subspecies strains from Cfv and Cft subspecies strains. Biomarkers to distinguish between Cff and Cfv were not found. The results of the study show the potential of proteotyping to differentiate different subspecies, but also the limitations of the method.

Keywords: MALDI-TOF MS, Campylobacter fetus, below species differentiation, ICMS, proteotyping, MLST
similarities in mass spectra of unknown bacteria and biomarkers in existing databases, a procedure referred to as phylomicroproteomics [26]. Typing methods, which are based on mass spectrometric analysis, are generally known as

Figure 1. Illustration of the different proteotyping steps. 1) Recording of the ICMS mass spectra of the C. fetus test cohort and reference strain Cff LMG 6442 (NCTC 10842). 2) Establishment of a C. fetus-specific allelic isoform list by blasting the genome sequences obtained from the NCBI database against the genome of the C. fetus reference strain. Subsequently, allelic isoforms in the test cohort are identified by comparing with the newly established allelic isoform list. 3) For each strain in the test cohort, a specific set of biomarker isoforms is obtained. Subsequently the amino acid sequences of the biomarkers are fused into a single sequence that results in specific proteotyping-based sequence type for each of the strains and allows the calculation of a proteotyping-derived taxonomic dendrogram.
Proteotyping of *Campylobacter fetus* Subspecies

...proteotyping [27] and have previously been used for characterization of microbial communities, tissues, individual proteins, viruses, and bacteria for several years now [28–31]. Among clinically relevant bacteria *Salmonella* serotypes, *Clostridioides difficile* polymerase chain reaction (PCR) ribotypes and methicillin-resistant *Staphylococcus aureus* lineages have been shown to be detectable by proteotyping, to name just a few [32–34].

Previous studies of our working group demonstrated the potential of bacterial subtyping on *Campylobacter* species in the clinical context, as it was possible to differentiate clinically relevant from clinically less relevant subgroups (Figure 1) [35–39]. At the heart of our approach is a list of allele isoforms that resulted from non-synonymous mutations and post-translational modifications in biomarker gene sequences, which are detectable as mass shifts in MALDI-TOF spectra. In this way, a combination of amino-acid sequences specific for each of the isolates to be typed can be derived, in a similar manner as for multilocus sequence typing (MLST). By using proteotyping, only the changes in mass associated with a certain set of allelic isoforms of the same protein are taken into account for the derivation of phylogeny, whereas the visibility or absence of particular masses, as well as their intensity, is not considered. This improves the measurement accuracy, wherefore ICMS is a very promising subtyping approach and a realistic alternative to currently used sequence-based techniques [37].

The goal of this study was to complete the set of typing schemes for clinically relevant *Campylobacter* species by developing a *C. fetus*-specific proteotyping scheme. A set of 41 *C. fetus* isolates covering all currently known subspecies of *C. fetus* was used. All isolates were characterized by proteotyping and MLST, followed by the deduction of the phylogenetic relations.

**Materials and Methods**

*C. fetus* Isolates. The test cohort was compiled in way that all subspecies of the bacterial species were represented. In total, 41 *C. fetus* isolates were included in our study: 20 *Cfvi*, 11 *Cft*, and 3 *Cfv* isolates (Table 1). The isolates were of different biological origins, namely, preputial washing of cattle (4 *Cft*, 7 *Cfv*), vaginal mucus of cattle (2 *Cfv*), fetuses of cattle (2 *Cfv*), cattle (not further specified, 3 *Cfvi*), bovine sperm (1 *Cfvi*), bull genitals (1 *Cfvi*), calf fetus (2 *Cfv*), intestinal content of a calf (1 *Cfv*), intestinal content of a pig (1 *Cfv*), fetus brain of a sheep (1 *Cfvi*), reptile cloacal swab (3 *Cfvi*), human blood culture (7 *Cfi*, 4 *Cfv*), and 2 *Cfi* strains of unknown origin. Animal isolates were provided by the Friedrich-Loeffler-Institut Bundesforschungsanstalt für Tiergesundheit, Jena, Germany. The following strains were received from the Belgian coordinated collections of microorganisms (BCCM; http://bccm.belspo.be/about-us/bccm-lmg): LMG6443 (*Cfv*), LMG6442 (*Cff*), LMG6570 (*Cfv*), LMG27499 (*Cff*), LMG06569 (*Cfi*), LMG06571 (*Cff*), and LMG06727 (*Cff*). Human blood culture isolates were provided by the routine diagnostic laboratory of the University Medical Center, Göttingen, Germany (Table 1).

![Table 1. List of *C. fetus* isolates used in the study](https://example.com/table1.png)

| Isolate | Origin | Region | Date       | Other strain designations | MLST-ST |
|---------|--------|--------|------------|--------------------------|---------|
| Cft0018 | Preputial washing | Lower-Saxony | 28.04.2009 | 4                       |
| Cft14505 | Preputial washing | S-Bavaria | 02.08.2005 | 4                       |
| Cft0114 | Vaginal sample cattle | Lower-Saxony | 19.12.2006 | 4                       |
| Cft15105 | Preputial washing | S-Bavaria | 10.08.2005 | 4                       |
| Cft9305 | Preputial washing | Thuringia | 10.05.2005 | 6                       |
| Cft9405 | Preputial washing | Thuringia | 12.05.2005 | 6                       |
| Cft9605 | Preputial washing | Thuringia | 12.05.2005 | 4                       |
| Cft22504 | Fetus calf | Thuringia | 16.12.2004 | 3                       |
| Cft51299 | Calf intestinal content | Thuringia | 24.09.1999 | 5                       |
| Cft6305 | Preputial washing | N-Bavaria | 06.03.2005 | 4                       |
| Cft1105 | Preputial washing | N-Bavaria | 21.01.2005 | 4                       |
| Cft51222 | Fetus, cattle | Baden-W. | 14.06.2005 | 4                       |
| Cft07BS0007 | Preputial washing | Baden-W. | 26.09.2007 | 4                       |
| Cft13465 | Fetus cattle | S-Bavaria | 12.07.2005 | 4                       |
| Cft20105 | Preputial washing | Thuringia | 23.11.2005 | 2                       |
| Cft90105 | Preputial washing | Thuringia | 12.05.2005 | 6                       |
| Cft55560s | Preputial washing | Baden-W. | 07.09.2006 | 6                       |
| Cft22204 | Bovine sperm | Saxony | 16.12.2004 | 2                       |
| Cft8361 | Human blood culture | Germany | 1962 | 3                       |
| Cft69361 | Human blood culture | Germany | 1962 | 3                       |
| Cft645361 | Human blood culture | Germany | 1962 | 3                       |
| CftLMG6443 | Cow, vaginal mucus | United Kingdom | 1962 | 4                       |
| CftLMG6570 | Sheep fetus brain | Sweden, Göteborg | 1972 | ATCC 19438; CCGU 538; CIP 68.29; JCM 2528; NCDO 1876; NCTC 10354; Park X16115 |
| CftLMG6442 | Cattle | Belgium | 1985 | ATCC 27374; CCTM La3023; CCGU 6823A; CECT 564; CIP 53.96; JCM 2527; LMG 8849; NCTC 10842; NIAH 1049 |
| CftLMG6570 | Cattle | Belgium | 1985 | ATCC BAA-2539; Blaser 03 – 427 |
| CftLMG06571 | Bull genitals | Belgium | 1985 | CCGU 1769; CIP 68.8; Florum 7572; NIDO 7572 |
| CftLMG6277 | Calf fetus | Belgium | 1985 | CCGU 17695A; LMG 6628 t1 |

---

64
Bacterial Culture Conditions. *C. fetus* isolates used in the experiments were kept as cryobank stocks (Mast Diagnostica, Reinfeld, Germany) at −80 °C. For the subsequent MALDI-TOF MS analysis, the isolates were incubated under microaerophilic conditions (5% O2, 10% CO2, and 85% N2) in Mueller–Hinton agar supplemented with horse blood at 37 °C for 2–3 days.

Preparation of Matrix Solution. As part of the measurement preparation α-cyano-4-hydroxy-cinnamic acid (HCCA) purified matrix substance (Bruker Daltonics, Bremen, Germany) was dissolved in standard solvent (acetonitrile 50%, trifluoroacetic acid 2.5% in ddH2O) to 10 mg HCCA/mL. Purified recombinant human insulin (Sigma-Aldrich, Taufkirchen, Germany) was added to the HCCA solution as an internal calibrant to a final concentration of 10 pg/μL. The exact mass of the internal calibrant was experimentally determined (m/z = 5806.1) with reference to the Bruker Test Standard (BTS). The calibrant did not overlap with any of the biomarker masses of interest and allowed a very precise internal mass calibration of the spectra.

MALDI-TOF Mass Spectrometry. To prepare samples for the measurements, 2 different variants were used: smear preparation and formic acid/acetonitrile extraction. Smear preparation by experience yields clearer peaks in the m/z range <10,000 Da, whereas the extraction variant allows more precise analysis in the field >10,000 Da [39].

The samples for the measurements were prepared as described before [37, 39]. In the measurement process, 600 spectra (mass range 2 to 20 kDa) were obtained in 100-shot steps on an Autoflex III system and summed up. If the MALDI Biotyper (Database release 2016) identification score values were ≥2.00, they were considered correct.

Identification of Biomarkers in ICMS Spectra. The obtained mass spectra were analyzed by standard algorithms of FlexAnalysis (Bruker Daltonics, Bremen, Germany). Initially, spectra were internally calibrated to the spiked human insulin peak. Subsequently, the baseline was subtracted, and the spectra were smoothened (standard MBT method).

For determination of the theoretical average weight of the amino acid sequences corresponding to the respective open reading frames of ribosomal proteins, the amino acid sequences were uploaded one by one to the ExPASy Bioinformatics Resource Portal (https://web.expasy.org/compute_pi/), where a molecular weight calculator tool is provided.

Proteins used for previous proteotyping schemes sometimes underwent posttranslational modifications [40, 41]; therefore, further molecular weights were calculated for each biomarker, taking into account potential proteolytic removal of the N-terminal methionine (−131.04 Da), acetylation, phosphorylation, formylation, and methylation (Table 2). Biomarker masses observed in the reference genome of reference strain LMG 6442 (NCTC 10842) (Figure 2) were matched to the calculated masses. In contrast, biomarker masses are observed in the spectrum of clinical isolates, which could not be assigned to the calculated masses from the

**Table 2. Theoretical biomarker masses predicted by the genome sequence of *C. fetus* reference strain LMG 6442 (NCTC 10842) under consideration of possible posttranslational modifications**

| Biomarker | [-Met M + H+] | [-Met mM + H+] | [-Met + PO4 M + H+] | [M + H+] | [fM + H+] |
|-----------|---------------|----------------|---------------------|---------|---------|
| L36       | 4197          | 4211           | 4277                | 4332    | 4360    |
| L34       | 5083          | 5097           | 5163                | 5218    | 5246    |
| L32-M     | 5527          | 5541           | 5607                | 5662    | 5690    |
| L33-M     | 6202          | 6216           | 6282                | 6337    | 6365    |
| S14-M     | 6725          | 6739           | 6805                | 6860    | 6888    |
| L29       | 6759          | 6773           | 6839                | 6894    | 6922    |
| L24-M     | 8023          | 8037           | 8103                | 8158    | 8186    |
| S20-M     | 9738          | 9752           | 9818                | 9873    | 9901    |
| S19-M     | 10,274        | 10,288         | 10,354              | 10,409  | 10,437  |

[-Met mM + H+] = methylated mass - demethioninated form. [-Met +PO4 M + H+] = phosphorylated mass - demethioninated form. [M + H+] = unmodified mass. [fM + H+] = formylated mass.

**Figure 2. ICMS spectrum of *C. fetus* subsp. fetus reference strain LMG 6442 (NCTC 10842).** Singularly charged biomarkers that were part of the *C. fetus* proteotyping scheme labeled with a black arrow or, in the case of an N-terminal methionine cleavage (posttranslational modification) with a red arrow. Multiple charged ions are not marked in this illustration.
C. fetus reference genome, and the spectra were considered as novel isoforms of the particular biomarker. For each isolate of the C. fetus test cohort, all biomarker genes were amplified by PCR using primers listed in Table 3, and the amplicon was sequenced (Microsynth Seqlab, Göttingen, Germany). To confirm the respective allelic isoforms, the gene sequences obtained from the amplicons were translated in silico, and the amino acid sequences were subsequently aligned.

### Multilocule Sequence Typing (MLST). For MLST, a procedure modified from the original typing schemes was used [18, 23]. In brief, the annealing temperature of the PCR was decreased from 48 °C to 47 °C, and the glyS4 oligonucleotide primer for the amplification of the glyS4 locus was replaced with the primer glyS4 [18]. After concatenating the MLST gene sequences for each strain, the software MEGA X was also used to construct an MLST-based UPGMA dendrogram [42].

### Phylogenetic and Phylproteomic Analyses. An amino acid sequence list of all allelic isoforms of the 9 identified biomarkers was compiled (Table 4). GenBank accession numbers for the biomarker sequences observed in this study are listed in Table 5. To analyze the biomarkers' protein sequences translated from the National Center for Biotechnology Information (NCBI) nucleotide database (Geneious V10.1.3) they were concatenated for each strain and an unweighted pair group method with arithmetic mean (UPGMA) dendrogram (MEGA X) was constructed [42].

### Ethical Approval. Ethical approval for the study was obtained from Ethics Commission of the University Medical Center Göttingen, Germany. No humans, animals, or personalized data were used for this study.

### Results and Discussion

In 2015, our working group set up a new proteotyping workflow for the proteotyping of microorganisms (Figure 1) [37]. Now, the established procedure was used to develop a C. fetus-specific proteotyping scheme. According to the standard workflow, masses emerging in the mass spectrum of the genome sequenced Cff reference strain LMG 6442 (NCTC 10842) were analyzed, and MS biomarker ions were related with gene products consistent with the observed mass. By evaluating the 67 C. fetus nucleotide sequences available in the NCBI database, a collection of allelic isoforms for all biomarkers observed in the reference spectrum was set up (Table 4). In accordance with the established proteotyping procedure, mass spectra of all strains included in the test cohort were recorded. Subsequently, spectra were edited (baseline subtraction and smoothing) and overlaid with the spectrum of Cff reference strain LMG 6442 (NCTC 10842). Recorded biomarker masses were matched with the calculated average protein masses, and mass shifts in relation to the masses of the references strain were analyzed. After concatenation of amino acid sequences of the biomarkers included in the C. fetus typing scheme, a UPGMA tree based on these strain-specific proteotyping-based types was calculated.

### Identification of Biomarker Ions. In total, the analysis based on the genome of Cff reference strain LMG 6442 (NCTC 10842) yielded nine, single charged biomarker masses between m/z = 4300 and 10,300, which were presumptively correlated with a specific gene product. To provide reliable statements on reproducibility of our measurements, the standard deviation was calculated on the basis of 6 measurements. The highest standard deviation (0.959) was observed for isoform 1 of biomarker S20-M, whereas the lowest standard deviation (0.271) was observed for isoform 5 of biomarker L33-M (Table 6). The following biomarkers were identified: L36 (4331.35 Da), L34 (4217.26 Da), L32-M (5530.47 Da), L33-M (6205.31 Da), S20-M (6728.11 Da), S29-M (6893.22 Da), L24-M (8026.59 Da), S20-M (9741.33 Da), and S19-M (10,277.10 Da). De-methionation was observed for biomarkers L36 (4331.35 Da), L34 (4217.26 Da), L32-M (5530.47 Da), L33-M (6205.31 Da), S20-M (6728.11 Da), S29-M (6893.22 Da), L24-M (8026.59 Da), S20-M (9741.33 Da), and S19-M (10,277.10 Da). De-methionation was observed for biomarkers L32-M, L33-M, S14-M, L24-M, S19-M, and S20-M (Table 2, Figures 2 & 3). In the case of MLST, the established markers are distributed over the whole genome of the reference strain. As the biomarkers identified in this study show a comparable distribution, they were suitable for the deduction of phylogenetic relations.

Comparing C. fetus proteotyping biomarkers to biomarkers identified within the context of C. jejuni subsp. jejuni, C. jejuni subsp. doylei, and C. coli proteotyping [37–39], several differences can be noted: In the case of C. jejuni subsp. jejuni, 19 biomarkers were identified and associated with the respective peak in the ICMS spectrum, whereas less than half (9) were found for C. fetus. Furthermore, biomarker L33 lacked N-terminal methionine in the case of C. fetus (L33-M) but it was present in C. jejuni subsp. jejuni, C. jejuni subsp. doylei, and C. coli. These observations confirm the results published by Fagerquist et al., in which posttranslational modification patterns are microbial species-specific. Within the isolate collection, biomarker mass shifts were observed in 7 out of 9 biomarkers [43].

### Establishment of an Allelic Isoform Database. Following the identification of biomarker ions, an amino acid sequence isoform list for each of the biomarkers identified in the previous step was compiled. In this context, we analyzed the 67 C. fetus genome sequences that can be found on NCBI. The number of identified isoforms for the respective biomarker varied. The highest number was 6, whereas 1 biomarker showed just a single isoform. Differences were also observed regarding frequency of occurrence; whereas some isoforms occurred in >99% of the cases, other isoforms were only found once. Regarding single occurrence of isoforms, a sequencing error is possible. Except for biomarker L36, all identified biomarkers showed at least 3 different isoforms, demonstrating their suitability in the C. fetus subtyping context.
| Locus         | Full name/product                  | Calc. average mass [Da] | Frequency in database |
|--------------|-----------------------------------|------------------------|-----------------------|
| RpmJ/L36     | **C. fetus**-specific allelic isoform list |
| **Locus**    | **Calc. average mass**            | **Frequency in database** |
|              | **Full name/product**             | **database**           |
|              | **(ORF Locus tag in LMG 6442)**  |                        |
|              | **Calc. average mass**            | **Frequency in database** |
|              | **[Da]**                          | **database**           |
|              | **RpmJ/L34**                      |                        |
| **Locus**    | **Calc. average mass**            | **Frequency in database** |
|              | **Full name/product**             | **database**           |
|              | **(ORF Locus tag in LMG 6442)**  |                        |
|              | **Calc. average mass**            | **Frequency in database** |
|              | **[Da]**                          | **database**           |
|              | **RpmJ/L32-M**                    |                        |
| **Locus**    | **Calc. average mass**            | **Frequency in database** |
|              | **Full name/product**             | **database**           |
|              | **(ORF Locus tag in LMG 6442)**  |                        |
|              | **Calc. average mass**            | **Frequency in database** |
|              | **[Da]**                          | **database**           |
|              | **RpsN/S14-M**                    |                        |
| **Locus**    | **Calc. average mass**            | **Frequency in database** |
|              | **Full name/product**             | **database**           |
|              | **(ORF Locus tag in LMG 6442)**  |                        |
|              | **Calc. average mass**            | **Frequency in database** |
|              | **[Da]**                          | **database**           |
|              | **RpsS/S19-M**                    |                        |

*Observed in test population AA numbering including start-methionine, if mass spectrometry indicates its absence it is written in brackets (M).
The amino acid sequences of all biomarker isoform are listed in Table 4. Variations of the amino acid sequences obtained by alignment of the sequences are indicated in red; additionally, the computed average protein mass for each isoform is listed. It should be noted that due to some draft genomes in GenBank, the number of available sequences may vary, as there were no contigs with the sequences coding for each biomarker in all genomes.

MLST and Proteotyping of the Isolate Collection. To prove functionality of the C. fetus Δ Biomarker Isoform Measured mass (Da) Standard deviation Δ Measured mass/average mass Monoisotopic mass (Da) Average mass (Da)

Table 5. Accession numbers of C. fetus-specific proteotyping biomarker isoforms

| Biomarker | Isoform | Gene Bank Accession | Locus Tag | Protein ID |
|-----------|---------|---------------------|-----------|------------|
| L36       | 1       | MK463617            |           |            |
| L34       | 1       | CP000487.1:557520-557654 | CFF8240_0551 | CFF8240_0551 | ABK9601.7 |
| L34       | 2       | CP027287.1:608973-609077 | C6B32_03095 | C6B32_03095 | AVK96018.1 |
| L32-M     | 1       | CP000487.1:706648-706848 | CFF8240_0235 | CFF8240_0235 | ABK9818.94 |
| L32-M     | 6       | MK463615            |           |            |
| L33-M     | 1       | CP000487.1:1313847-1313947 | CFF8240_1324 | CFF8240_1324 | ABK9624.14 |
| L3-M      | 3       | CP027287.1:1398913-1399013 | C6B32_06940 | C6B32_06940 | AVK96015.0 |
| L3-M      | 5       | MK463616            |           |            |
| S14-M     | 1       | CP000487.1:735926-735936 | CFF8240_0047 | CFF8240_0047 | ABK9824.98 |
| L29       | 1       | CP000487.1:737925-737935 | CFF8240_0042 | CFF8240_0042 | ABK96084.1 |
| L29       | 3       | CP027287.1:376898-377098 | C6B32_02000 | C6B32_02000 | AVK96031.9 |
| L24-M     | 1       | CP000487.1:38746-38756 | CFF8240_0445 | CFF8240_0445 | ABK96333.3 |
| L24-M     | 3       | CP027287.1:37719-37729 | C6B32_02015 | C6B32_02015 | AVK96032.1 |
| S20-M     | 1       | CP000487.1:1678191-1678291 | CFF8240_1718 | CFF8240_1718 | ABK9624.35 |
| S19-M     | 3       | CP000487.1:136187-136197 | CFF8240_0038 | CFF8240_0038 | ABK9618.69 |
| L33-M     | 3       | CP027287.1:35160-35170 | C6B32_00180 | C6B32_00180 | AVK96031.5 |

The most interesting findings were that proteotyping-derived type D consisted only of Cff isolates. Regarding MLST sequence types, it comprised particularly 1 isolate of ST16, 2 isolates of ST15, 1 isolate of ST27, 1 isolate of ST30, 1 isolate of sequence type 31, and 1 isolate of ST61.

Identification of Allelic Isoforms. The test cohort was measured in exactly the same manner as it was done for the reference strain LMG 6442 (NCTC 10842). The evaluation of the measurements of mass spectra of the strains was done based on the comparison with the spectrum of this reference strain. Observed mass shifts were compared to the sequence list of amino acid isoforms, whereby a particular allelic isoform could be identified.

If two different isoforms with the same mutation at different positions were observed, which though did not differ regarding mass difference to the reference isoform, the variants were further examined by DNA sequencing. In the test cohort, 3 allelic isoforms for biomarker L33-M (RpmG) and 2 for biomarkers L34 (RpmM) and L32-M (RpmN), L24-M (RplX), S20-M (RpsT), and S19-M (RpsS) were detected. Only for biomarkers L36 (RpmJ) and S14-M (RpsN) only one allelic isoform was identified (Table 2, Figure 3).

Construction of an UPGMA-Dendrogram. To deduce the phylogenetic relationships of the species, amino acid sequences of the 9 identified proteotyping biomarkers were fused into a single sequence. The concatenated sequence was then further processed with the MEGA X software to calculate a phyloproteomic tree (UPGMA). The 9 identified
biomarkers allowed a clear differentiation of a group of \textit{Cff} and \textit{Cfv} strains from a group of utterly \textit{Cft} strains. In order to assess the quality of the proteotyping results, another UPGMA tree was calculated based on MLST data (Figure 4). Comparative analysis of the trees revealed some differences between the two resulting phylogenies. While the test cohort was differentiated into 14 MLST sequence types, the proteotyping-based analysis led to a division into only 4 different groups. The most interesting finding was that proteotyping-based type D comprised all of the \textit{Cft} isolates, showing that our approach here is comparable to the quality of the current gold standard MLST.

Unfortunately, the MLST-ST4 corresponding to the subspecies \textit{Cfv} could not be differentiated by means of proteotyping. Here, proteotyping proves to be inferior to MLST in its discriminatory resolution.

A previous study by Fitzgerald et al. showed that it is possible to distinguish \textit{Cft} from other \textit{C. fetus} subspecies. Based on multiple unidentified biomarker peaks, a dendrogram was calculated using Pearson correlation \cite{15}. A factor, which reduces the informative value of these results, was the lack of knowledge about the proteins responsible for each of the discriminating peaks.

In contrast to this study, we were able to identify at least 9 defined ribosomal proteins as biomarkers. As \textit{Cft} strains exhibited different biomarker isoforms compared to the other two \textit{C. fetus} subspecies, they could be clearly differentiated. PCR and subsequent Sanger sequencing of the respective biomarkers further confirmed these differences.

Regarding the limitations of proteotyping, the number of sequence data available is decisive for the quality of the typing scheme. In the case of \textit{C. fetus}, much less sequences (67) were available as compared to \textit{C. jejuni} subsp. \textit{jejuni} (more than 3000) \cite{37}. Another factor affecting the quality of the typing scheme is the number of biomarkers it comprises. Further studies should therefore focus on the identification of additional reliable biomarkers that can be included in the existing scheme.

The prerequisite for the application of the technique is the visibility of all biomarkers of the typing scheme. If this is not the case, it is advisable to use sequence-based techniques.
Conclusion

As the results obtained so far demonstrate, proteotyping is a promising tool for microbial typing at the species, subspecies, and even below subspecies levels. A smart bioinformatics solution and the development of an easy-to-handle user interface would allow the application of the technique in daily diagnostic routine, as the corresponding equipment for proteotyping is available in modern clinical laboratories anyway. The rapidly growing sequence databases due to next generation sequencing (NGS) are opening up a wide range of opportunities for the development of further proteotyping schemes that possibly allow a rapid detection in the case of a disease outbreak.

Authors’ Contributions

M.F.E. and M.K. contributed equally to this work. M.F.E., O.B., and A.E.Z. wrote the manuscript and established the biomarker isoform database in silico. M.K. performed MALDI measurements and confirmatory PCRs. H.H., L.vd.G.B. and A.E.Z. collected bacterial isolates and performed data interpretation, bioinformatics, and correction of the manuscript. O.B., U.G. and A.E.Z. designed the experiments and evaluated the data.

Conflicts of Interest

There are no conflicts of interest.

References

1. Forbes BA, Sahm DF, Weissfeld AS. Diagnostic microbiology. In: Bailey & Scott’s Diagnostic Microbiology. 2002. pp. 11–14.
2. Gillespie IA, O’Brien SJ, Frost JA, Adak GK, Horby P, Swan AV, et al. A Case-Case Comparison of Campylobacter coli and Campylobacter jejuni Infection: A Tool for Generating Hypotheses. Emerg Infect Dis. 2002;8:937–42.

Figure 4. Comparison of MLST- and proteotyping-derived phylogenies. On the left: Evolutionary tree calculated based on MLST by means of the maximum composite likelihood method (UPGMA). In total, 14 different MLST sequence types were identified which are illustrated in different colors. On the right: Evolutionary tree based on proteotyping and calculated using UPGMA. Four different proteotyping-derived types were identified. Type A contains most of the C. fetus subsp. fetus and C. fetus subsp. venerealis strains. Type B and C contain 2 MLST ST 5 and one MLST ST 20 strain. The most interesting proteotyping-derived type is type D, which contains all C. fetus subsp. testudinum strains and thereby allows the differentiation of the subspecies from other C. fetus subspecies. The different proteotyping-based sequence types are marked at the branches of the evolutionary tree (A, B, C, and D).
