



## OPEN ACCESS

## EDITED BY

Palmiro Poltronieri,  
National Research Council (CNR), Italy

## REVIEWED BY

Hoh Boon-Peng,  
International Medical University, Malaysia  
Pawan Kumar Kanaujia,  
Mahayogi Gorakhnath University, India  
Sidharth Prasad Mishra,  
University of South Florida, United States

## \*CORRESPONDENCE

Pascale Vonaesch  
✉ pascale.vonaesch@unil.ch

## †PRESENT ADDRESSES

Paul Tetteh Asare,  
Gnubiotics Sciences AG,  
SuperLab Suisse,  
Lausanne, Switzerland  
Chi-Hsien Lee,  
Louvain Drug Research Institute,  
Université Catholique de Louvain, Brussels,  
Belgium  
Vera Hürlimann,  
PharmaBiome AG,  
Schlieren, Switzerland  
Adrian Egli and Aline Cuénod  
Institute of Medical Microbiology,  
University of Zurich, Zurich,  
Switzerland

†These authors have contributed equally to this work

## SPECIALTY SECTION

This article was submitted to  
Microorganisms in Vertebrate Digestive Systems,  
a section of the journal  
Frontiers in Microbiology

RECEIVED 21 November 2022

ACCEPTED 31 January 2023

PUBLISHED 21 February 2023

## CITATION

Asare PT, Lee C-H, Hürlimann V, Teo Y, Cuénod A, Akduman N, Gekeler C, Afrizal A, Corthesy M, Kohout C, Thomas V, de Wouters T, Greub G, Clavel T, Pamer EG, Egli A, Maier L and Vonaesch P (2023) A MALDI-TOF MS library for rapid identification of human commensal gut bacteria from the class *Clostridia*. *Front. Microbiol.* 14:1104707. doi: 10.3389/fmicb.2023.1104707

## COPYRIGHT

© 2023 Asare, Lee, Hürlimann, Teo, Cuénod, Akduman, Gekeler, Afrizal, Corthesy, Kohout, Thomas, de Wouters, Greub, Clavel, Pamer, Egli, Maier and Vonaesch. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# A MALDI-TOF MS library for rapid identification of human commensal gut bacteria from the class *Clostridia*

Paul Tetteh Asare<sup>1,2,3,4†</sup>, Chi-Hsien Lee<sup>1,2,3†</sup>, Vera Hürlimann<sup>2,3†</sup>, Youzheng Teo<sup>1</sup>, Aline Cuénod<sup>4,5†</sup>, Nermin Akduman<sup>6,7</sup>, Cordula Gekeler<sup>6,7</sup>, Afrizal Afrizal<sup>8</sup>, Myriam Corthesy<sup>9</sup>, Claire Kohout<sup>10</sup>, Vincent Thomas<sup>11</sup>, Tomas de Wouters<sup>12</sup>, Gilbert Greub<sup>9</sup>, Thomas Clavel<sup>8</sup>, Eric G. Pamer<sup>10</sup>, Adrian Egli<sup>4,5†</sup>, Lisa Maier<sup>6,7</sup> and Pascale Vonaesch<sup>1,2,3\*</sup>

<sup>1</sup>Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland,

<sup>2</sup>Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland, <sup>3</sup>University of Basel, Basel, Switzerland, <sup>4</sup>Applied Microbiology Research, Department of Biomedicine, University of Basel, Basel, Switzerland, <sup>5</sup>Clinical Bacteriology and Mycology, University Hospital of Basel, Basel, Switzerland, <sup>6</sup>Interfaculty Institute of Microbiology and Infection Medicine Tübingen, University of Tübingen, Tübingen, Germany, <sup>7</sup>Cluster of Excellence 'Controlling Microbes to Fight Infections', University of Tübingen, Tübingen, Germany, <sup>8</sup>Functional Microbiome Research Group, Institute of Medical Microbiology, RWTH University Hospital, Aachen, Germany, <sup>9</sup>Institute of Microbiology of the University of Lausanne, University Hospital Centre (CHUV), Lausanne, Switzerland, <sup>10</sup>Duchossois Family Institute, Division of Infectious Diseases and Global Health, University of Chicago, Chicago, IL, United States, <sup>11</sup>BioAster, Paris, France, <sup>12</sup>PharmaBiome AG, Schlieren, Switzerland

**Introduction:** Microbial isolates from culture can be identified using 16S or whole-genome sequencing which generates substantial costs and requires time and expertise. Protein fingerprinting *via* Matrix-assisted Laser Desorption Ionization–time of flight mass spectrometry (MALDI-TOF MS) is widely used for rapid bacterial identification in routine diagnostics but shows a poor performance and resolution on commensal bacteria due to currently limited database entries. The aim of this study was to develop a MALDI-TOF MS plugin database (CLOSTRITOF) allowing for rapid identification of non-pathogenic human commensal gastrointestinal bacteria.

**Methods:** We constructed a database containing mass spectral profiles (MSP) from 142 bacterial strains representing 47 species and 21 genera within the class *Clostridia*. Each strain-specific MSP was constructed using >20 raw spectra measured on a microflex Biotyper system (Bruker-Daltonics) from two independent cultures.

**Results:** For validation, we used 58 sequence-confirmed strains and the CLOSTRITOF database successfully identified 98 and 93% of the strains, respectively, in two independent laboratories. Next, we applied the database to 326 isolates from stool of healthy Swiss volunteers and identified 264 (82%) of all isolates (compared to 170 (52.1%) with the Bruker-Daltonics library alone), thus classifying 60% of the formerly unknown isolates.

**Discussion:** We describe a new open-source MSP database for fast and accurate identification of the *Clostridia* class from the human gut microbiota. CLOSTRITOF expands the number of species which can be rapidly identified by MALDI-TOF MS.

## KEYWORDS

human gut microbiota, *Clostridia*, MALDI-TOF MS, commensal bacteria, anaerobic, bacterial identification, culturomics, next-generation probiotics

## Introduction

*Clostridia* is an important class of Gram-positive, often strictly anaerobic, rod-shaped, spore-forming bacteria within the phylum *Bacillota* (formerly *Firmicutes*). Some *Clostridia* spp. are pathogenic in humans and animals such as *C. botulinum*, *C. tetani*, and *C. difficile*. However, most members of this class play an important role in the human intestinal tract, where they contribute, among others, to the production of the short-chain fatty acid butyrate (Beyer-Sehlmeyer et al., 2003; Paparo et al., 2017; Cruz-Morales et al., 2019).

Depletion of butyrate-producing *Clostridia* from the intestinal microbiota has been associated with multiple intra- and extraintestinal diseases such as, i.e., childhood stunting (Vonaesch et al., 2018), inflammatory bowel disease (Eeckhaut et al., 2013; Gevers et al., 2014), colorectal cancer (Wu et al., 2013), and cardiometabolic disease (Karlsson et al., 2012; Wang et al., 2012; Le Chatelier et al., 2013). Further, depletion in butyrate producers was shown to decrease colonization resistance to enteric pathogens (Rivera-Chávez et al., 2016). Overall, there is thus a clear link between disease and the absence of butyrate producers. Owing to their important role in health and disease, efforts have multiplied to isolate and characterize these bacteria and eventually use them as so-called next-generation probiotics (NGPs); bacteria directly isolated from healthy humans, that are re-introduced in diseased individuals as health-promoting strains (O'Toole et al., 2017; Chang et al., 2019; Lin et al., 2019). The identification and characterization of novel NGPs heavily rely on culture collections and thus on tools that can rapidly identify commensal bacteria.

The current gold standard to identify bacterial isolates are full-length 16S rRNA gene or whole-genome sequencing (WGS) (Browne et al., 2016; Lagkouvardos et al., 2019). WGS is time-consuming, costly and requires a specific expertise. For this reason, it is rarely used for species identification in clinical routine diagnostics (Rossen et al., 2018). Therefore, there is an ongoing interest in rapid, cheap, accurate, and easy-to-use methods to identify bacteria isolated from the human intestinal tract.

Over the last decade, matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) has been established as a technique for rapid and reliable bacterial identification and is mainly used in routine diagnostics (Bizzini and Greub, 2010; Li et al., 2019). The technique measures the mass spectral profiles (MSPs) within a range of 2000 to 20,000 Daltons, where also ribosomal proteins of bacteria occur. Usually, MSPs are compared within a commercial database to pre-recorded reference spectra and matching marker masses allow the identification of a species (Croxatto et al., 2012; Schumann and Maier, 2014; Seuylemezian et al., 2018). Quality of the spectra (Cuénod et al., 2021, 2022) and the database used are two critical factors for a reliable identification. The currently available commercial databases [we used the *MBT Compass reference library* (version 4.1.100)], which are widely used in routine diagnostics, only cover a minority of non-pathogenic *Clostridia*. Therefore, the current usage of MALDI-TOF MS for proper species identification is limited. This hurdle has been overcome in the past by researchers in other domains by the generation of in-house custom reference databases of bacterial strains relevant

to their field of study (Calderaro et al., 2014; Fergusson et al., 2020; Moussa et al., 2021).

Considering the robust classification, rapid acquisition times, and low cost per sample of MALDI-TOF MS, we aimed to close this important gap and extend this technology to the identification of commensal members of the gastrointestinal microbiota of the class *Clostridia*. We generated 142 mass spectral profiles (MSPs) for 21 genera and 47 species, using full-length 16S rRNA gene sequence-confirmed bacterial isolates from diverse geographic regions. The newly created MSP library was evaluated for accuracy by re-identifying 58 blind-coded, full-length 16S rRNA gene sequence validated isolates of *Clostridia* covering the same genera and species using the new database plugin. Finally, we applied the new MSP library on a set of 326 isolates from stool samples of healthy Swiss donors, showing its potential on identifying new isolates from human fecal samples.

## Materials and methods

### Bacterial strains in the CLOSTRI-TOF database

The number and composition of the mass spectrum profile (MSP) library used reference library have a significant impact on the accuracy of the microflex Biotyper taxonomy assignment. To extend the coverage to the identification of members of non-pathogenic *Clostridia*, we used 142 sequenced strains. These strains are the main representatives of the human gut commensals of the class *Clostridia* (Table 1; Supplementary Table 1) and have been chosen to include as many geographically diverse strains as possible, to cover each species as broadly as possible. The 142 strains belonging to 21 genera and 47 species were confirmed for their species identification using full-length 16S rRNA gene sequencing. Fifty-eight (58) 16S rRNA gene full-length identified strains were used to validate the newly generated MALDI-TOF MS database. Strain information is summarized in Supplementary Table 1 (strains used for library construction), and Supplementary Table 2 (validation strains).

### Bacterial culture

The culture medium used for each strain is given in Supplementary Table 1 and included the following: Brain Heart Infusion with Inulin (BHII Agar): BHI Agar (BD Difco, cat. number 279830, Franklin Lakes, NJ, USA) supplemented with inulin from chicory (Sigma, cat. number I2255, Darmstadt, Germany), 1g/l. Schaedler Agar (Oxoid, cat. number CM0497, Ireland), and modified Gifu Anaerobic Broth (mGAM) (HiMedia, cat. number M2079). All media were prepared according to manufacturer instructions and supplemented with 15 g agar/l whenever appropriate. Bacteria were cultured under strictly anaerobic conditions using an anaerobic workstation (Anaerobic chamber; Coy Laboratories, Ann Arbor, MI, USA) containing a gas mixture of 10% CO<sub>2</sub>, 5% H<sub>2</sub>, and 85% N<sub>2</sub>. Bacteria were incubated at 37°C for 3–4 days before MALDI-TOF MS analysis. Culture conditions are summarized in Supplementary Table 1 (strains used for library construction) and Supplementary Table 2 (validation strains).

TABLE 1 Number of bacterial strains used to construct the CLOSTRI-TOF database.

Genus	Species	Total no. of strains included (no. of reference strains)
<i>Agathobacter</i>	<i>rectalis</i>	5 (0)
<i>Anaerobutyricum</i>	<i>hallii</i>	2 (1)
<i>Anaerostipes</i>	<i>hadrus</i>	5 (1)
<i>Anaerotruncus</i>	<i>colihominis</i>	5 (1)
<i>Blautia</i>	<i>caecimuris</i>	2 (1)
	<i>coccoides</i>	1 (1)
	<i>faecis</i>	5 (1)
	<i>glucerasea</i>	3 (0)
	<i>hansenii</i>	1 (0)
	<i>luti</i>	2 (1)
	<i>massiliensis</i>	4 (0)
	<i>obeum</i>	5 (1)
	<i>producta</i>	6 (2)
	<i>pseudococcoides</i>	1 (0)
	<i>schinkii</i>	5 (1)
	<i>wexlerae</i>	4 (1)
<i>Clostridium</i>	<i>leptum</i>	1 (1)
	<i>nexilis</i>	1 (0)
	<i>scindens</i>	5 (2)
	<i>ymbiosum</i>	5 (1)
<i>Coprococcus</i>	<i>catus</i>	1 (1)
	<i>comes</i>	5 (1)
	<i>eutactus</i>	2 (0)
<i>Dorea</i>	<i>formicigenerans</i>	5 (1)
	<i>longicatena</i>	5 (1)
<i>Enterocloster</i>	<i>aldenensis</i>	3 (1)
	<i>clostridioformis</i>	5 (1)
<i>Eubacterium</i>	<i>ramulus</i>	2 (1)
<i>Faecalibacterium</i>	<i>prausnitzii</i>	4 (0)
	<i>longum</i>	1 (0)
<i>Faecalicatena</i>	<i>fissicatena</i>	5 (1)
<i>Fusicatenibacter</i>	<i>saccharivorans</i>	3 (1)
<i>Gemmiger</i>	<i>formicilis</i>	1 (0)
<i>Intestinibacter</i>	<i>bartlettii</i>	1 (1)
<i>Lachnospira</i>	<i>eligens</i>	3 (1)
<i>Lacrimispora</i>	<i>celerecrescens</i>	3 (1)
	<i>saccharolytica</i>	1 (1)
<i>Peptostreptococcus</i>	<i>stomatis</i>	1 (1)
<i>Roseburia</i>	<i>faecis</i>	2 (1)
	<i>hominis</i>	1 (0)
	<i>intestinalis</i>	1 (0)
	<i>inulinivorans</i>	2 (1)

(Continued)

TABLE 1 (Continued)

Genus	Species	Total no. of strains included (no. of reference strains)
<i>Ruminococcus</i>	<i>bromii</i>	3 (1)
	<i>gnavus</i>	4 (1)
	<i>lactaris</i>	3 (1)
	<i>torques</i>	2 (1)
<i>Sellimonas</i>	<i>intestinalis</i>	5 (1)

## DNA purification and 16S rRNA gene sequencing

Genomic DNA was extracted from bacterial cultures using a bead-beating device (SpeedMill BeadBeater for 30 s or with the vortex adapter on maximum speed for 10 min) followed by the Wizard genomic DNA purification kit (Promega, cat. number A2920, Dübendorf, Switzerland) according to manufacturer instructions. Total DNA was quantified by absorbance at 260 nm using the NanoDrop® ND-1000 Spectrophotometer (Witec AG, Littau, Switzerland). DNA samples were stored at -20°C until further analysis.

To confirm the identity of the isolates, the full region of the 16S rRNA gene was amplified by PCR using the FIREPol® DNA polymerase (Solis BioDyne, Tartu, Estonia) using primers 27F (5'-AGA GTT TGA TCC TGG CTC AG-3') and 1492R (5'-GGT TAC CTT GTT ACG ACT T-3') with the following conditions: 95°C for 3 min; 35 cycles of 95°C for 30 s, 50°C for 30 s, 72°C for 30 s; then 72°C for 7 min (Frank et al., 2008). PCR products were visualized using standard gel electrophoresis (1% agarose in 1× Tris-acetate-EDTA [TAE] buffer) at 100 V for 30 min (Bio-Rad, Cressier, Switzerland), purified using the Promega Wizard™ SV Gel, and PCR Cleanup System according to the manufacturer's protocol and sequenced bi-directionally to obtain near full-length 16S rRNA gene sequence using Sanger sequencing at Microsynth (Balgach, Switzerland). To identify the closest homologs, the obtained sequences were aligned using the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990). We defined sequence homology ≥99% as a species match, and sequence similarity ≥97% as a genus match (Becker et al., 2004). To build the phylogenetic tree, 16S rRNA gene sequences were aligned and trimmed using MUSCLE and a maximum-likelihood phylogenetic tree was constructed with 1,000 bootstrap replicates using MEGA 11 (Tamura et al., 2021).

## MALDI-TOF MS spectra preparation for MSP creation

As recommended by Bruker-Daltonics for new library creations, ethanol-formic acid extraction was used to prepare the sample for MSP creation. Formic acid is routinely used when using MALDI-TOF in bacteriology since acidic pH of the matrix improves the extraction of the ribosomal proteins (Croxatto et al., 2012). Briefly, two to three colonies were suspended in 300 µl of high-pressure liquid chromatography (HPLC)-grade water (Sigma-Aldrich, Buchs, Switzerland) and mixed with 900 µl of 100% ethanol (Sigma-Aldrich, Buchs, Switzerland). After centrifugation at 15,000× g for 2 min, the supernatant was removed, and the pellet was dried. For sample

extraction, 50 µl of formic acid (70% in water) was added to the bacterial pellet, the tube was thoroughly mixed by vortexing and 50 µl acetonitrile (Sigma) were added to the mixture. The mixture was centrifuged at 15,000 × *g* for 2 min. For each strain, 1 µl of the supernatant containing the bacterial extract was spotted on a 96-spot steel plate (Bruker-Daltonics) in 15 replicates and allowed to dry at room temperature. One microliter of bacterial test standard (BTS, Bruker-Daltonics) was pipetted on two MALDI target spots for each plate to allow for calibration during acquisition and processing. Subsequently, the samples were overlaid with 1 µl of α-cyano-4-hydroxycinnamic acid (HCCA) matrix and air dried again prior to measurement. A Bacterial Test Standard (BTS) was used to calibrate the instrument, before each acquisition session.

## MALDI-TOF MS data acquisition

MALDI-TOF MS was performed with a Bruker-Daltonics microflex LT/LH bench-top mass spectrometer. Protein mass spectra of samples were acquired in a mass range of 2000–20,000 Da using the flexControl 3.4 program (Bruker-Daltonics) and a laser frequency at 20 Hz with a linear positive mode. The default operating conditions were as follows: ion source 1, 20 kV; ion source 2, 18.25 kV; pulse ion extraction, 370 ns. Each spectrum was summed up to 240 laser shots (40 laser shots/position × 6 different positions). For each strain, 30 single spectra were generated from two independent cultures with 15 technical replicates. All raw spectra have been deposited on Zenodo with the access number 10.5281/zenodo.7573939.

## MALDI-TOF MS quality control and custom MSP library creation

The quality of raw spectra was evaluated using flexAnalysis 3.4 software (Bruker-Daltonics). To ease the quality check of the raw spectra, smoothing parameters and baseline subtraction parameters were applied. For baseline subtraction we used the multipolygon, minimum, penalized least squares, and penalized B-Spline baseline subtraction method, and for smoothing, we used the Savitzky-Golay smoothing method. For each analysis day, the eight known masses of the bacterial test standard (BTS) spectrum were controlled to be within the tolerance range of ± 300 ppm to prove mass accuracy. Spectra showing any flatline, outliers, dramatic mass shifts, and anomalies were deleted. Flatline spectra were defined as spectra with a completely flat line and outliers were defined as spectra with a unique peak that was detected in no other spectra of the same strain. Mass shifts were assessed on a random peak between 6,000 Da to 7,000 Da and we only accepted spectra with peak shifts below 500 ppm. After processing the raw spectra, a minimum of 27 high-quality spectra per strain were selected and merged using the MALDI Biotyper Compass Explorer software (Bruker-Daltonics) to create the mass spectral profiles (MSPs). The merged spectra per strain constituting the MSP were verified to have a log score greater than 2.7 and a peak frequency greater than 75%. The final library was compiled based on the MSP of 142 strains (Figure 1A). An MSP dendrogram was constructed using a Euclidean distance measure and an average linkage algorithm with the MALDI Biotyper Compass Explorer 4.1. The final library has been deposited on Zenodo with the access number 10.5281/zenodo.7573939.

## CLOSTRI-TOF database validation

To validate the CLOSTRI-TOF database, we performed a blind test of 58 full-length 16S rRNA gene sequenced isolates in two collaborating laboratories (Figure 1B; Supplementary Table 2). Validation was performed using the extended direct transfer (eDT) technique. Briefly, a few bacterial colonies were directly smeared onto a MALDI-TOF MS target and covered with a 1 µl of saturated HCCA matrix solution as is generally recommended for routine bacterial identification. In case no spectra were detected, we repeated the analysis using the ethanol-formic acid extraction method (as previously described). The raw spectra generated were compared to the two combined databases (Bruker-Daltonics MALDI Biotyper Compass MSP library v. 4.1.100 and CLOSTRI-TOF) (Figure 1B). The identification results were classified using the score values proposed by Bruker: the highest matching scores of the spectra were represented by ranges indicating high confidence identification (2,000–3,000), low confidence identification (1,700–1,999) and no organism identification possible (0,000–1,699).

## Identification of a set of 326 clinical isolates from healthy Swiss volunteers

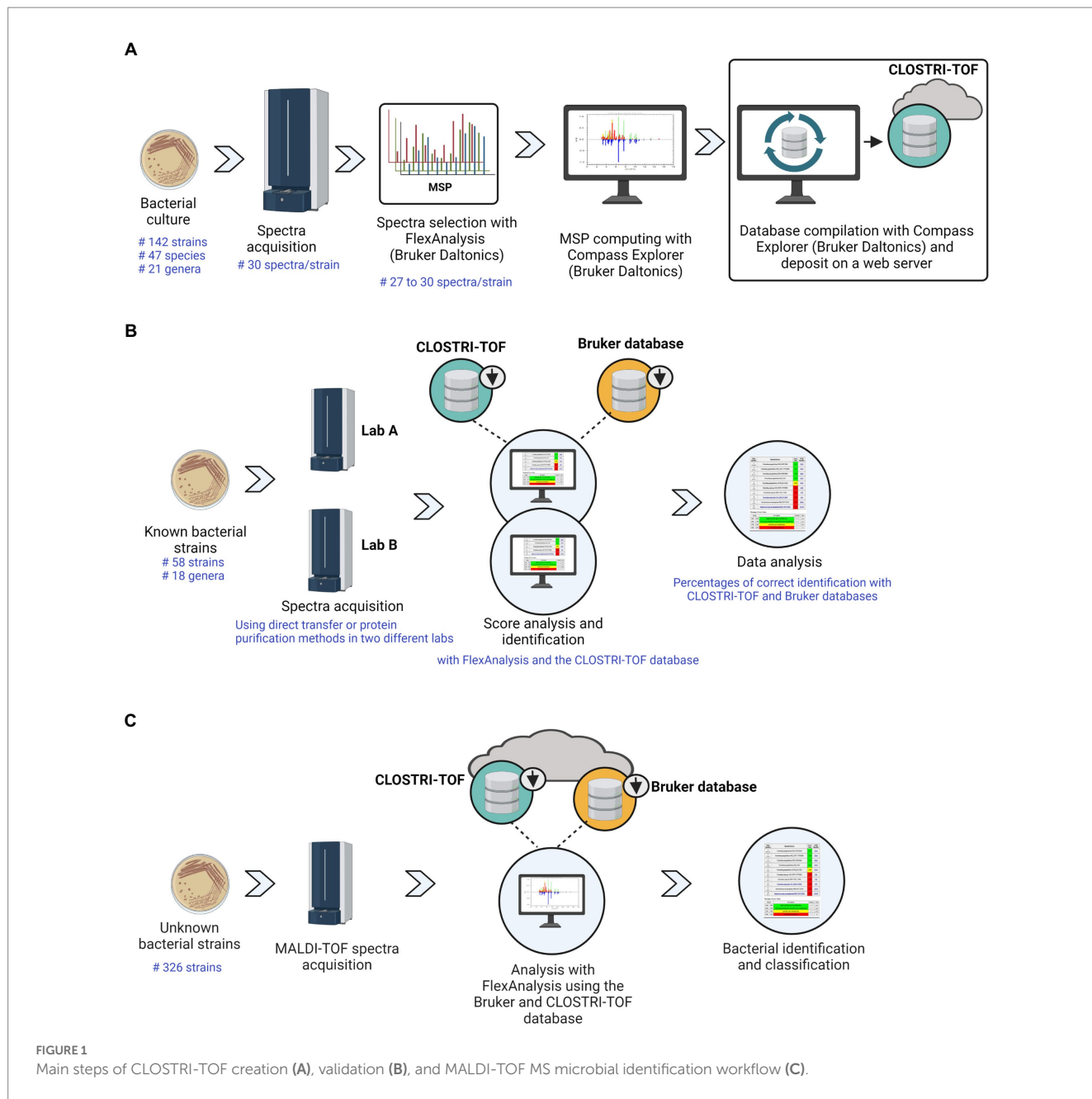
To test the library on a set of previously unknown isolates, we used the CLOSTRI-TOF database combined with the original Bruker-Daltonics MALDI Biotyper Compass MSP library v. 4.1.100 to identify a total of 326 anaerobic cultures isolated from stool samples of healthy Swiss individuals recruited in the EnrichBut project (Swiss Ethics Number: BASEC ID 2021-00199) (Figure 1C). In brief, stool samples were collected from four self-declared healthy Swiss adults between 30 and 60 years of age who did not consume antibiotics two weeks prior to sample collection. Each stool donor was provided with a custom-made stool sampling kit that allows for the preservation of the oxygen sensitive bacterial groups under anaerobic conditions. Stool samples were stored anaerobically at room temperature prior to processing. Upon receipt (within 48 h of collection), the stool samples were resuspended (1:10 w/v) in anaerobic phosphate-buffered saline (PBS). The suspensions were then serially diluted in PBS and 100 µL was plated onto Columbia Blood, BHII, Schaedler and mGAM Agar. The plates were incubated at 37°C for 5 days. Between 20 and 30 colonies showing different colony morphologies were picked from each medium and re-streaked three times on fresh agar plates to ensure purity of the isolates. The isolates were cultured in liquid broth and mixed with 20% glycerol and stored at – 80°C.

The isolates were prepared for MALDI-TOF MS identification using the extended direct transfer (eDT) technique and were identified based on three technical replicates. The obtained spectra were compared either against the commercial Bruker-Daltonics library or the combined library from Bruker-Daltonics and CLOSTRI-TOF.

## Results

### MSP library creation

For the 142 strains analyzed, we generated 2,130 target positions and a total of 4,260 individual mass spectra. The quality of the spectra was high, with an average of 1–3 of the raw spectra removed per strain prior to MSP generation. Thus, we could generate a robust library for the 47 species.



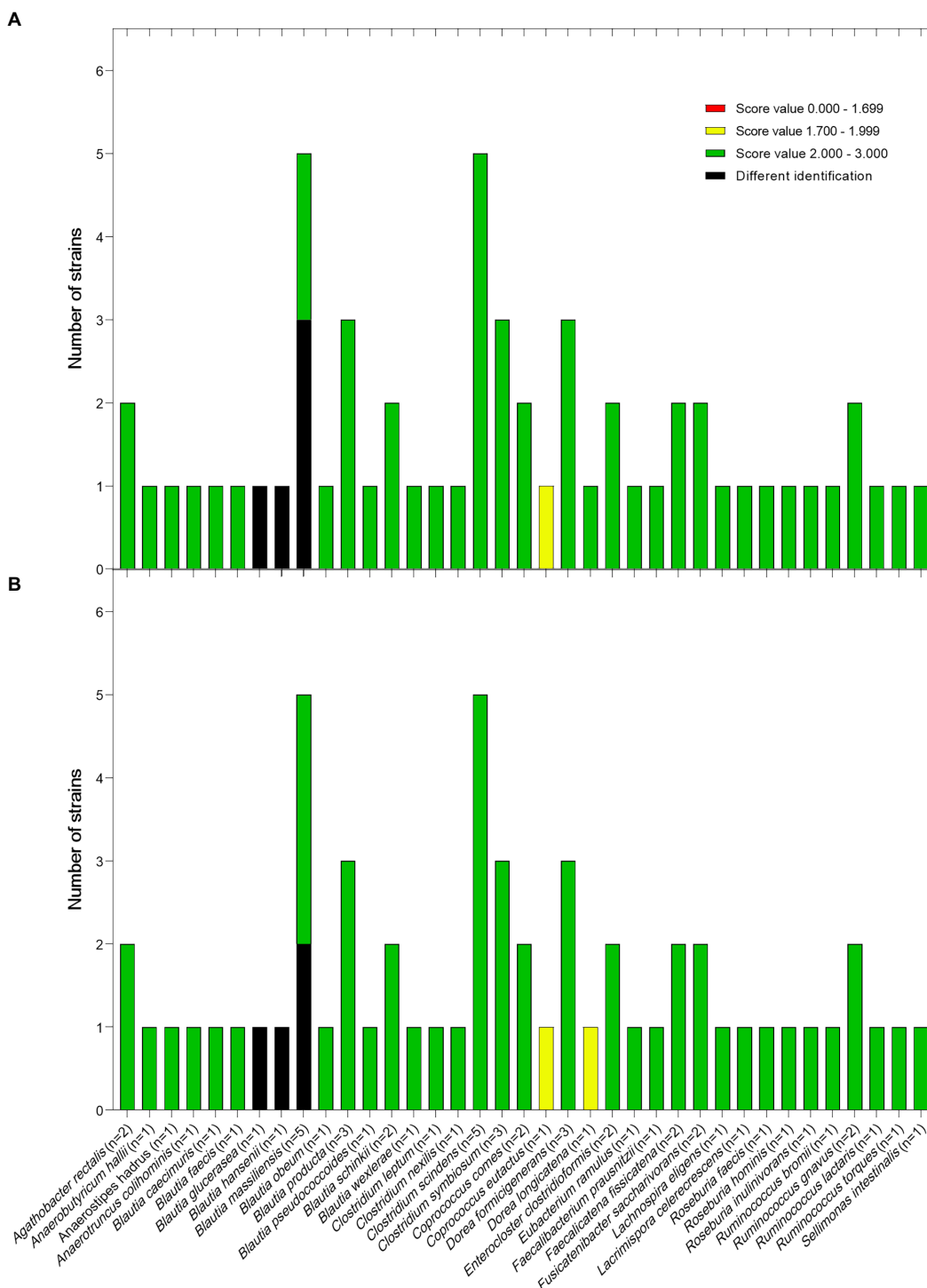
## Evaluation of MSP library performance

To test the newly established library, we subjected 58 additional strains (validation strains, [Supplementary Table S2](#)), not included in the set of strains used for library construction but previously identified based on their 16S rRNA gene sequences to MALDI-TOF MS. Combining the original Bruker database with the CLOSTRI-TOF database plugin allowed reliable identification of at the species level (score value  $\geq 2.0$ ) of 57/58 strains (98.3%) in laboratory A and 54/58 (93.1%) in laboratory B ([Figures 2](#)). Of note, several strains belonging to the genus *Blautia* were identified with a high score ( $\geq 2.0$ ) but were incorrectly assigned at the species level. Further, *Coprococcus eutactus* was correctly identified in both laboratories, but with a score between 1.7 and 2.0. All the validation strains (58/58) (100%) were reliably identified at the genus level

(score  $\geq 1.7$  but  $< 2.0$ ) in both laboratory A and B. Thus, using the combination of the original Bruker database and the CLOSTRI-TOF database, we were able to reidentify most of the validation strains to species level and all strains to genus level.

## 16S rRNA gene-based phylogeny and MALDI-TOF MS profiling

To investigate the taxonomic assignment relationship between the microflex Biotyper and 16S rRNA gene-based sequencing, we compared dendrograms generated by either of these techniques. In most cases, the two methods provided identical results where strains belonging to the same species clustered together ([Figures 3, 4](#)). However, different strains of *Agathobacter rectalis*, *Blautia* spp.,



**FIGURE 2**  
Blind test of *Clostridia* strains using the newly created CLOSTRI-TOF MS performed in lab A (A) and lab B (B). The highest matching scores of the spectra were represented by ranges indicating high confidence identification (2.000–3.000), low confidence identification (1.700–1.999) and no organism identification possible (0.000–1.699).

*Faecalicatena fissicatena*, and *Lacrimispora celerecrescens* formed two or three distinct MALDI-TOF-based clusters (Figure 3). Likewise, the maximum-likelihood phylogenetic tree based on 16S rRNA gene sequences was unable to cluster together strains of the same species of *Agathobacter rectalis*, *Blautia* spp., *Faecalicatena fissicatena*, and *Lacrimispora celerecrescens* (Figure 4).

### Identification of anaerobic gut bacterial isolates using MALDI-TOF MS enhanced by CLOSTRI-TOF

We next analyzed a bank of 326 strains isolated from the feces of healthy Swiss volunteers (Supplementary Table 5)

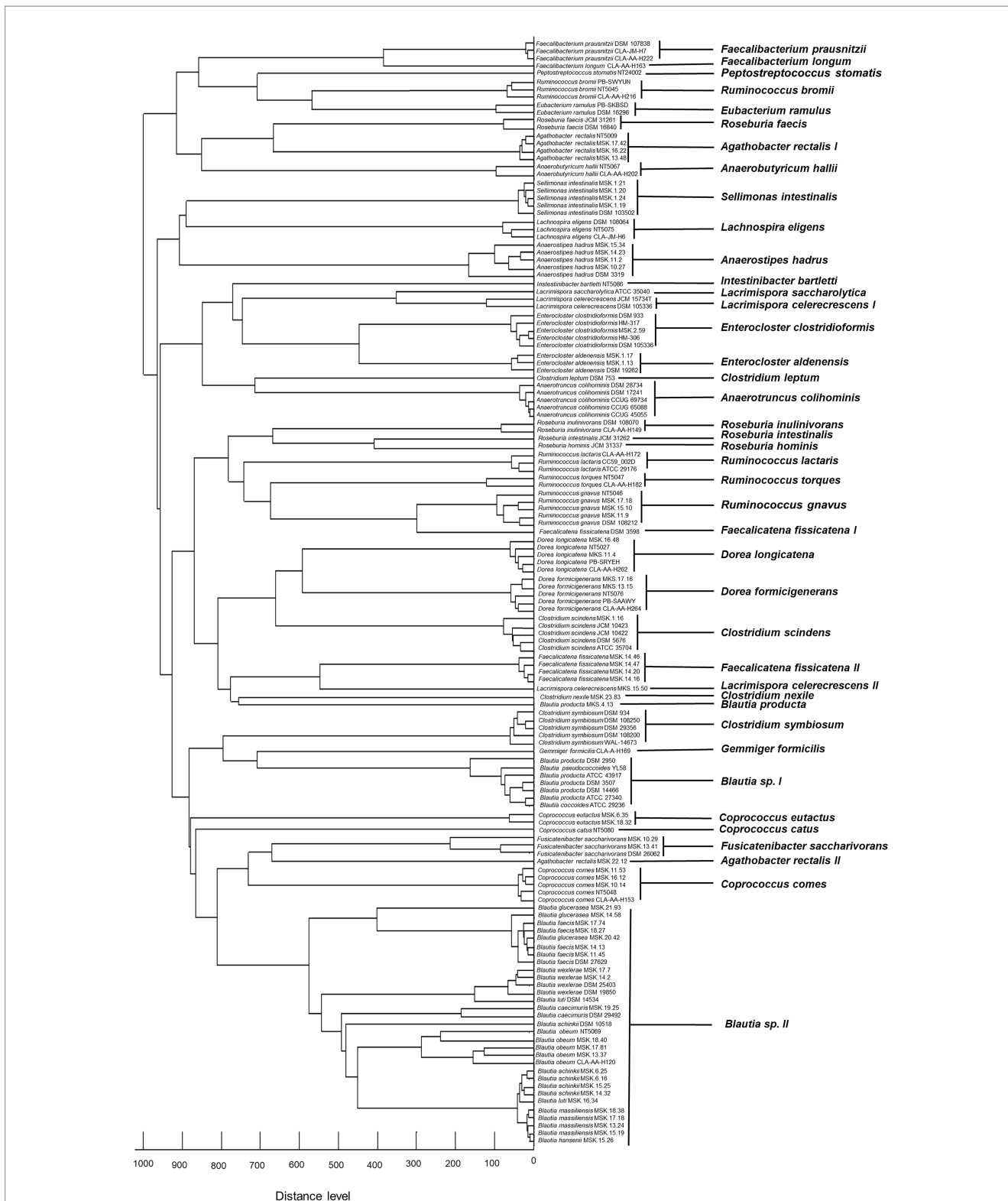


FIGURE 3 Distribution of different species with similar spectral profiles according to MSP dendrogram cluster analysis.

using the microflex Biotyper. Of the 326 analyzed strains, 170 (52.1%) were identified with the original Bruker database, while 156 (47.9%) could not be identified/had no match in the library at a score >1.7. Combining the original library with the

CLOSTRI-TOF library plugin, we increased the number of strains identified at a score >2.0 to 264 (80.9%) (Figure 5), which corresponds to an identification of formerly unknown strains of 60%.

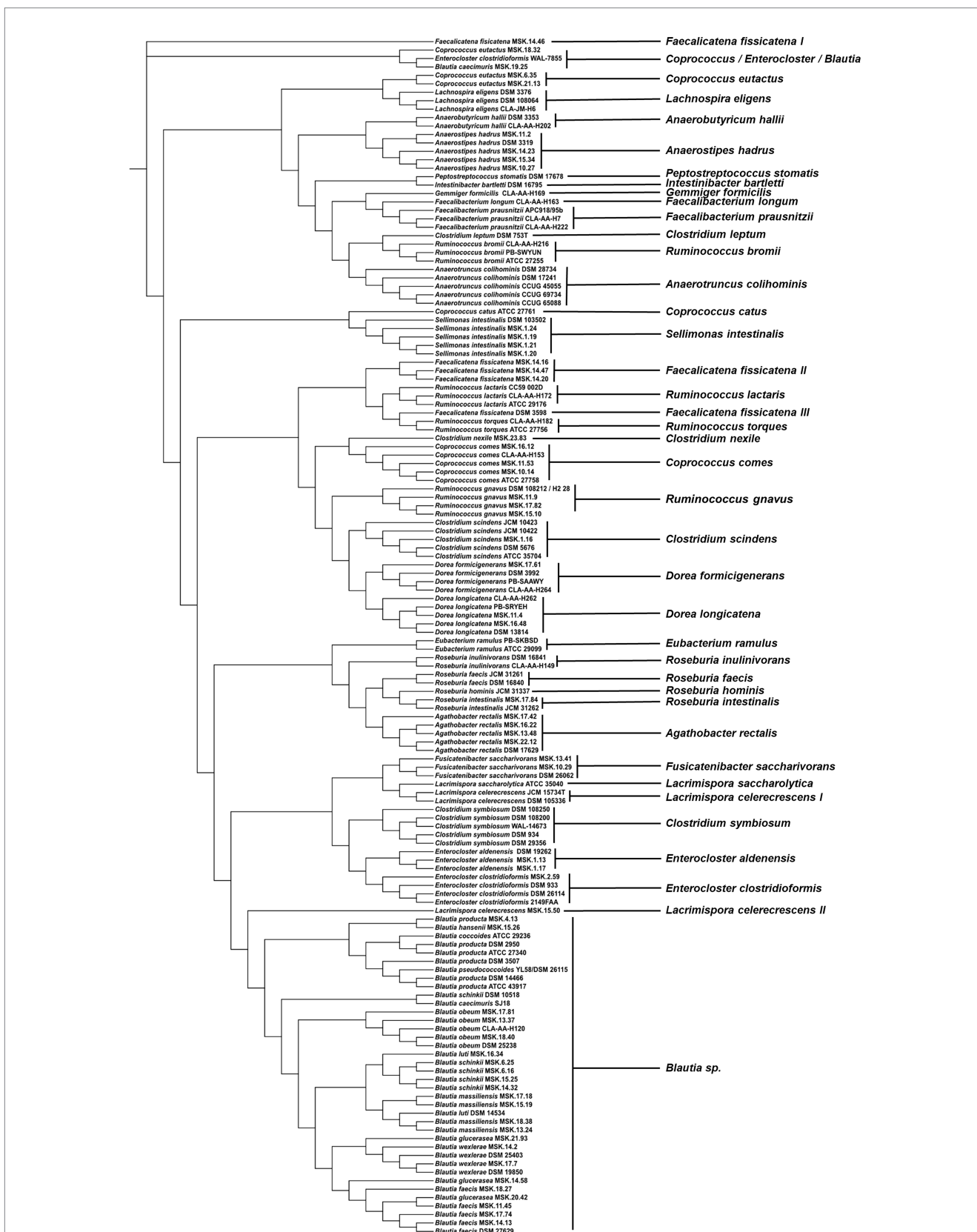
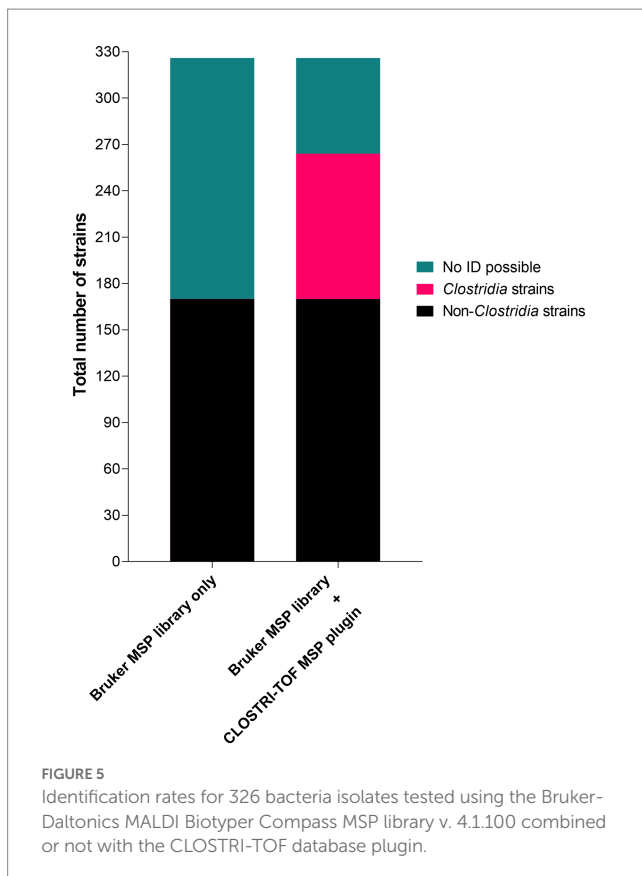


FIGURE 4  
Maximum-likelihood phylogenetic tree based on complete 16S rRNA gene sequences of the strains used in the construction of CLOSTRI-TOF database.





## Discussion

To date, 16S rRNA gene sequencing is one of the most widely used techniques to identify unknown microbial isolates from the human gastrointestinal tract. Here, we propose a MALDI-TOF MS library plugin, CLOSTRI-TOF, enhancing the power of MALDI biotyping as a rapid, accurate, and economical alternative.

MALDI biotyping is strongly dependent on spectral databases, thus allowing to identify only taxa that already have spectral information for a given species. This hurdle has been overcome by the generation of in-house custom reference databases for specific groups of microorganisms. As an example, previous research has generated database plugins for different species of the genera *Borrelia* (Calderaro et al., 2014), *Burkholderia* (Fergusson et al., 2020) or *Vibrio* (Moussa et al., 2021), different strains of *Clostridium tybutyricum* (Burtscher et al., 2020) as well as for helminths (Sy et al., 2022).

The CLOSTRI-TOF library plugin presented here allowed to identify all the 47 bacterial species included in the library plugin at the genus level and most even down to species level. Further, we could show that the combination of the original Bruker library combined with the CLOSTRI-TOF library plugin significantly increase the number of identified strains from a collection of fecal bacterial isolates of healthy Swiss volunteers. One of the species included in our library (*Coprococcus eutactus*) did not yield reliable identification using the validation strains. Of note, *C. eutactus* is represented by only two strains in the CLOSTRI-TOF library. It has been shown previously that the number of included strains in the database is critical to account for intraspecific

diversity, thus allowing reliable identification (Erler et al., 2015). Future updates to the CLOSTRI-TOF database should therefore expand the number of strains included for species with a low strain coverage.

While our database plugin allowed identification of *Blautia* at the genus level, it did not allow reliable assignment at species-level. As the strains of *Blautia* were also polyphyletic in the 16S rRNA gene phylogenetic tree, it would be interesting to compare the phylogenetic relationship within this genus in more detail. The library plugin presented here is focused on the main members of the class *Clostridia* colonizing the human gastrointestinal tract. We have chosen this group of microorganisms, as they are frequently involved in the production of the key metabolite butyrate (Louis and Flint, 2009) and reduced in dysbiotic diseases, including undernutrition (Vonaesch et al., 2018), ulcerative colitis (Machiels et al., 2014), and type 2 diabetes (Wang et al., 2012). Expanding the database beyond the class of *Clostridia* to other frequent members of the human microbiome will be key for rapid identification of all bacterial strains isolated from human fecal samples.

Of note, MALDI biotyping allows for rapid identification of bacteria, yet is not able to give the same amount of information on each isolate as would be gained by whole-genome sequencing. Thus, MALDI identification should be seen as a low cost, rapid pre-identification tool to select the strains that will be characterized in greater detail using more advanced methods such as whole-genome sequencing.

## Conclusion

Our data show that MALDI-TOF MS is a promising tool to rapidly identify isolates of commensal bacteria. The new open-source library plugin developed here allows to discriminate all taxa included to the genus and most taxa even down to species level.

## Data availability statement

The datasets presented in this study can be found on Zenodo under the following link <https://zenodo.org/deposit/7573939#>.

## Author contributions

PA, LM, and PV: conceptualization. VT, TW, GG, TC, EP, and AE: resources. PA, CL, VH, YT, AC, NA, CG, AA, MC, CK, and PV: methodology. PA, CL, and PV: data curation, formal analysis, visualization, and writing of original draft. PA, CL, GG, TC, EP, AE, LM, and PV: review and editing. All authors contributed to the article and approved the submitted version.

## Funding

This work was funded by the Bill and Melinda Gates Foundation (grant number INV-004352 to PV and LM), the Forschungsfonds of the University of Basel (PV) and an

Eccellenza Fellowship from the Swiss National Science Foundation to PV (grant number PCEFP3\_194545). The Vonaesch and Greub labs are part of NCCR Microbiomes, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (grant number 180575). TC received funding from the German Research Foundation (DFG), project ID 403224013—SFB 1382, project ID 395357507—SFB 1371, and project ID 460129525—NFDI4Microbiota. NA, CG, and LM are supported by DFG (CMFI Cluster of Excellence EXC 2124 and Emmy Noether Program).

## Acknowledgments

The authors would like to thank members of the Vonaesch and the Maier lab as well as Sören Becker, Johanna Burtscher, and Jennifer Keiser for helpful discussions and feedback. They further wish to thank laboratory technicians at the Clinical Bacteriology and Mycology of the University Hospital Basel and the Centre Hospitalo-Universitaire Vaudois (CHUV) for help with the operation of the Bruker MALDI-TOF MS machine. Figure 1 was created using BioRender.

## References

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Becker, K., Harmsen, D., Mellmann, A., Meier, C., Schumann, P., Peters, G., et al. (2004). Development and evaluation of a quality-controlled ribosomal sequence database for 16S ribosomal DNA-based identification of *Staphylococcus* species. *J. Clin. Microbiol.* 42, 4988–4995. doi: 10.1128/JCM.42.11.4988-4995.2004
- Beyer-Sehlmeyer, G., Glei, M., Hartmann, E., Hughes, R., Persin, C., Böhm, V., et al. (2003). Butyrate is only one of several growth inhibitors produced during gut flora-mediated fermentation of dietary fibre sources. *Br. J. Nutr.* 90, 1057–1070. doi: 10.1079/bjn20031003
- Bizzini, A., and Greub, G. (2010). Matrix-assisted laser desorption ionization time-of-flight mass spectrometry, a revolution in clinical microbial identification. *Clin. Microbiol. Infect.* 16, 1614–1619. doi: 10.1111/J.1469-0691.2010.03311.X
- Browne, H. P., Forster, S. C., Anonye, B. O., Kumar, N., Neville, B. A., Stares, M. D., et al. (2016). Culturing of “unculturable” human microbiota reveals novel taxa and extensive sporulation. *Nature* 533, 543–546. doi: 10.1038/nature17645
- Burtscher, J., Küller, F., Dreier, M., Arias-Roth, E., Drissner, D., and Domig, K. J. (2020). Characterization of *Clostridium tyrobutyricum* strains using three different typing techniques. *Microorganisms* 8, 1–17. doi: 10.3390/microorganisms8071057
- Calderaro, A., Gorrini, C., Piccolo, G., Montecchini, S., Buttrini, M., Rossi, S., et al. (2014). Identification of *Borrelia* species after creation of an in-house MALDI-TOF MS database. *PLoS One* 9:e88895. doi: 10.1371/journal.pone.0088895
- Chang, C. J., Lin, T. L., Tsai, Y. L., Wu, T. R., Lai, W. F., Lu, C. C., et al. (2019). Next generation probiotics in disease amelioration. *J. Food Drug Anal.* 27, 615–622. doi: 10.1016/j.jfda.2018.12.011
- Croxatto, A., Prod'homme, G., and Greub, G. (2012). Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. *FEMS Microbiol. Rev.* 36, 380–407. doi: 10.1111/J.1574-6976.2011.00298.X
- Cruz-Morales, P., Orellana, C. A., Moutafis, G., Moonen, G., Rincon, G., Nielsen, L. K., et al. (2019). Revisiting the evolution and taxonomy of *Clostridia*, a phylogenomic update. *Genome Biol. Evol.* 11, 2035–2044. doi: 10.1093/gbe/evz096
- Cuénod, A., Aerni, M., Bagutti, C., Bayraktar, B., Boz, E. S., Carneiro, C. B., et al. (2022). Quality of MALDI-TOF mass spectra in routine diagnostics: results from an international external quality assessment including 36 laboratories from 12 countries using 47 challenging bacterial strains. *Clin. Microbiol. Infect.*, 190–199. doi: 10.1016/j.cmi.2022.05.017
- Cuénod, A., Foucault, F., Pflüger, V., and Egli, A. (2021). Factors associated with MALDI-TOF mass spectral quality of species identification in clinical routine diagnostics. *Front. Cell. Infect. Microbiol.* 11:646648. doi: 10.3389/fcimb.2021.646648
- Eeckhaut, V., Machiels, K., Perrier, C., Romero, C., Maes, S., Flahou, B., et al. (2013). *Butyricoccus pulliaecorum* in inflammatory bowel disease. *Gut* 62, 1745–1752. doi: 10.1136/gutjnl-2012-303611
- Erler, R., Wichels, A., Heinemeyer, E. A., Hauk, G., Hippelein, M., Reyes, N. T., et al. (2015). VibrioBase: a MALDI-TOF MS database for fast identification of *Vibrio* spp. that are potentially pathogenic in humans. *Syst. Appl. Microbiol.* 38, 16–25. doi: 10.1016/j.syapm.2014.10.009
- Fergusson, C. H., Coloma, J. M. F., Valentine, M. C., Haeckl, F. P. J., and Linington, R. G. (2020). Custom matrix-assisted laser desorption ionization-time of flight mass spectrometric database for identification of environmental isolates of the genus *Burkholderia* and related genera. *Appl. Environ. Microbiol.* 86. doi: 10.1128/AEM.00354-20
- Frank, J. A., Reich, C. I., Sharma, S., Weisbaum, J. S., Wilson, B. A., and Olsen, G. J. (2008). Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl. Environ. Microbiol.* 74, 2461–2470. doi: 10.1128/AEM.02272-07
- Gevers, D., Kugathasan, S., Denson, L. A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., et al. (2014). The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe* 15, 382–392. doi: 10.1016/j.chom.2014.02.005
- Karlsson, F. H., Fåk, F., Nookaew, I., Tremaroli, V., Fagerberg, B., Petranovic, D., et al. (2012). Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nat. Commun.* 3:1245. doi: 10.1038/ncomms2266
- Lagkouvardos, I., Lesker, T. R., Hitch, T. C. A., Gálvez, E. J. C., Smit, N., Neuhaus, K., et al. (2019). Sequence and cultivation study of *Muribaculaceae* reveals novel species, host preference, and functional potential of this yet undescribed family. *Microbiome* 7, 28–15. doi: 10.1186/s40168-019-0637-2
- Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., et al. (2013). Richness of human gut microbiome correlates with metabolic markers. *Nature* 500, 541–546. doi: 10.1038/nature12506
- Li, Y., Shan, M., Zhu, Z., Mao, X., Yan, M., Chen, Y., et al. (2019). Application of MALDI-TOF MS to rapid identification of anaerobic bacteria. *BMC Infect. Dis.* 19, 941–911. doi: 10.1186/s12879-019-4584-0
- Lin, T.-L., Shu, C.-C., Lai, W.-F., Tzeng, C.-M., Lai, H.-C., and Lu, C.-C. (2019). Investiture of next generation probiotics on amelioration of diseases – strains do matter. *Med. Microcol.* 1-2:100002. doi: 10.1016/j.medmic.2019.100002
- Louis, P., and Flint, H. J. (2009). Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiol. Lett.* 294, 1–8. doi: 10.1111/j.1574-6968.2009.01514.x
- Machiels, K., Joossens, M., Sabino, J., de Preter, V., Arijis, I., Eeckhaut, V., et al. (2014). A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis. *Gut* 63, 1275–1283. doi: 10.1136/gutjnl-2013-304833
- Moussa, M., Cauvin, E., Le Piouffle, A., Lucas, O., Bidault, A., Paillard, C., et al. (2021). A MALDI-TOF MS database for fast identification of *vibrio* spp. potentially pathogenic to marine mollusks. *Appl. Microbiol. Biotechnol.* 105, 2527–2539. doi: 10.1007/s00253-021-11141-0

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1104707/full#supplementary-material>

- O'Toole, P. W., Marchesi, J. R., and Hill, C. (2017). Next-generation probiotics: the spectrum from probiotics to live biotherapeutics. *Nat. Microbiol.* 2:17057. doi: 10.1038/nmicrobiol.2017.57
- Paparo, L., Calignano, A., Tocchetti, C. G., Di Scala, C., Russo, R., Bonaduce, D., et al. (2017). The influence of fiber on gut microbiota: butyrate as molecular player involved in the beneficial interplay between dietary fiber and cardiovascular health. Amsterdam: Elsevier Inc.
- Rivera-Chávez, F., Zhang, L. F., Faber, F., Lopez, C. A., Byndloss, M. X., Olsan, E. E., et al. (2016). Depletion of butyrate-producing *Clostridia* from the gut microbiota drives an aerobic luminal expansion of *Salmonella*. *Cell Host Microbe* 19, 443–454. doi: 10.1016/j.chom.2016.03.004
- Rossen, J. W. A., Friedrich, A. W., and Moran-Gilad, J. (2018). Practical issues in implementing whole-genome-sequencing in routine diagnostic microbiology. *Clin. Microbiol. Infect.* 24, 355–360. doi: 10.1016/j.cmi.2017.11.001
- Schumann, P., and Maier, T. (2014). *MALDI-TOF Mass Spectrometry Applied to Classification and Identification of Bacteria*. 1st Amsterdam Elsevier Ltd
- Seuylemezian, A., Aronson, H. S., Tan, J., Lin, M., Schubert, W., and Vaishampayan, P. (2018). Development of a custom MALDI-TOF MS database for species-level identification of bacterial isolates collected from spacecraft and associated surfaces. *Front. Microbiol.* 9:780. doi: 10.3389/fmicb.2018.00780
- Sy, I., Conrad, L., and Becker, S. L. (2022). Recent advances and potential future applications of MALDI-TOF mass spectrometry for identification of helminths. *Diagnostics* 12, 3035. doi: 10.3390/diagnostics12123035
- Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120
- Vonaesch, P., Morien, E., Andrianonimiadana, L., Sanke, H., Mbecko, J. R., Huus, K. E., et al. (2018). Stunted childhood growth is associated with decompartmentalization of the gastrointestinal tract and overgrowth of oropharyngeal taxa. *Proc. Natl. Acad. Sci. U. S. A.* 115, E8489–E8498. doi: 10.1073/pnas.1806573115
- Wang, J., Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490, 55–60. doi: 10.1038/nature11450
- Wu, N., Yang, X., Zhang, R., Li, J., Xiao, X., Hu, Y., et al. (2013). Dysbiosis signature of fecal microbiota in colorectal cancer patients. *Microb. Ecol.* 66, 462–470. doi: 10.1007/s00248-013-0245-9