Taxonogenomic description of four new Clostridium species isolated from human gut: ‘Clostridium amazonitimonense’, ‘Clostridium merdae’, ‘Clostridium massilidielmoense’ and ‘Clostridium nigeriense’

M. T. Alou1, S. Ndongo1, L. Frégère1, N. Labas1, C. Andrieu1, M. Richez1, C. Couderc1, J.-P. Baudoin1, J. Abrahão2, S. Brah3, A. Diallo1,4, C. Sokhna1,4, N. Cassir1, B. La Scola1, F. Cadoret1 and D. Raoult1,5
1) Aix-Marseille Université, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM63, CNRS 7278, IRD 198, INSERM 1095, Marseille, France, 2) Laboratório de Vírus, Departamento de Microbiologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, 3) Hopital National de Niamey, BP 247, Niamey, Niger, 4) Campus Commun UCAD-IRD of Hann, Route des pères Maristes, Hann Maristes, BP 1386, CP 18524, Dakar, Senegal and 5) Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract

Culturomics investigates microbial diversity of the human microbiome by combining diversified culture conditions, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and 16S rRNA gene identification. The present study allowed identification of four putative new Clostridium sensu stricto species: ‘Clostridium amazonitimonense’ strain LF2T, ‘Clostridium massilidielmoense’ strain MT26T, ‘Clostridium nigeriense’ strain Marseille-P2414T and ‘Clostridium merdae’ strain Marseille-P2953T, which we describe using the concept of taxonogenomics. We describe the main characteristics of each bacterium and present their complete genome sequence and annotation. © 2017 Published by Elsevier Ltd.

Keywords: ‘Clostridium amazonitimonense’, ‘Clostridium massilidielmoense’, ‘Clostridium merdae’, ‘Clostridium nigeriense’, culturomics, emerging bacteria, human microbiota, taxonogenomics

Original Submission: 18 August 2017; Revised Submission: 9 November 2017; Accepted: 16 November 2017

Article published online: 22 November 2017

Introduction

As showed by a remarkable number of studies, the human gut microbiota plays a tremendous role in human health and the homeostasis of the human body [1–3]. Conversely, an increasing number of links between the gut microbiota and several diseases has been uncovered; these diseases include nutritional disorders, metabolic diseases and pathologies of the intestine [1,4–6]. Techniques used in exploratory studies have evolved alongside technology from culture-based techniques in early studies to today’s culture-independent techniques [7]. A turning point in the study of the gut microbiota is in progress, with a return to culture-based techniques. In fact, cultured strains are needed for in vivo and in vitro experiments to apply the extensive knowledge acquired thanks to metagenomic studies. Culturomics, a method invented in our laboratory, is a part of this turning point [8]. It consists of the multiplication of culture conditions with a variation of media and physico-chemical parameters such as atmosphere, temperature, pH and salinity in order to exhaustively explore a microbial ecosystem [9].

As part of a global exploratory study designed to improve the known and cultivated bacterial diversity of the human gut microbiota, several previously unknown species of Clostridium were isolated. Exploratory studies have shown that the healthy gut microbiota mostly consists of anaerobic bacteria, which represent 95% of the bacterial diversity [2]. Among these
anaerobic bacteria, the genus *Clostridium* is one of the most represented in the human gut, although taxonomic chaos is associated with this genus and its higher ranks in taxonomy, such as the family *Clostridiaceae* and the *Clostridiales* order [7]. This genus was created in 1880 by Prazmowski with the type species *Clostridium butyricum* [10,11]. In early taxonomy, which was based on phenotypic characteristics, all Gram-positive, anaerobic and spore-forming bacilli were attributed to the genus *Clostridium*. Consequently, with the advent of DNA sequencing and phylogenetic analysis, taxonomic chaos was uncovered in this genus, and several clusters were subsequently built, among which cluster I, *Clostridium sensu stricto*, was built around *C. butyricum* [11,12]. Among the 211 *Clostridium* species with standing in nomenclature, only a third are *Clostridium sensu stricto* species [10,12]. These mesophilic species are mostly strict anaerobes with Gram-positive staining and the ability to sporulate. Their metabolism is highly variable because they are able to metabolize various compounds such as carbohydrates, amino acids, alcohols, amino acids and purines. Acid and alcohol can be produced from carbohydrates or via protein metabolism [10,11].

Here we propose the creation of four putative new *Clostridium sensu stricto* species: *Clostridium amazontimonense* strain LF27 (= CSUR P1445 = DSM 28600), *Clostridium massilibiolense* strain MT26 (= CSUR P2255 = CCUG 96993), *Clostridium nigeriense* strain Marseille-P2414T (= CSUR P2414 = DSM 102218) and one new putative *Clostridium* cluster IV, *Clostridium merdae* strain Marseille-P2953 (= CSUR P2953 = DSM 103669), which we describe using the concept of taxonomics [13]. This concept of new species description is based on the presentation of the phylogenetic, phenotypic and genomic features of the studied new species and their comparison to closely related species.

**Materials and methods**

**Sample information**

In order to explore the bacterial diversity of the human gut microbiota, stool samples were collected from several individuals: a healthy 8-month-old Senegalese boy, a healthy 44-month-old Senegalese girl, a healthy 37-year-old Amazonian woman from French Guyana and a healthy 29-year-old Senegalese woman. None of the donors was receiving antibiotic treatment at the time of sample collection. Oral consent was obtained from each adult. For the children, oral consent was provided by the parents. Samples were stored at −80°C upon arrival in the La Timone hospital in Marseille. This study received approval from the Institut Fédératif de Recherche 48 (Faculty of Medicine, Marseille, France) under agreement 09-022.

**Strain identification**

Each stool sample was cultured using the 18 culture conditions of standardized culturomics [8,9]. For each condition, pre-incubation in the medium of an anaerobic blood culture bottle from BD BACTEC (Becton Dickinson [BD], San Diego, CA, USA) was performed during 30 days with seeding on 5% sheep’s blood—enriched Colombia agar (bioMérieux, Marcy l’Étoile, France) every 3 days. Colonies were purified before identification using a Microflex spectrometer and a MALDI fingerprinting system (MALDI-TOF MS) target plate (Bruker Daltonics, Bremen, Germany) as described previously [14,15]. Each obtained spectrum was matched against the database from Bruker and from the La Timone hospital using MALDI Biotyper 3.0 software by standard pattern matching (with default parameter settings). Matching scores under 1.7 indicated the absence of a match in the database. For these unidentified colonies, the 16S rRNA gene was sequenced using fD1 and rP2 primers as previously described [16]. Kim et al. [17] recommended the creation of a new species at the 16S similarity level threshold of 98.65% with the closest species with standing in nomenclature. The corresponding spectra are then incremented in the database and compared to the spectra of closely related species.

**Growth conditions**

For each species, optimal growth conditions were determined by testing various temperatures (25, 28, 37, 45 and 56°C), atmospheres (aerobic with or without 5% CO₂, microaerophilic and anaerobic atmospheres using the GENbag anaer and GENbag microaer systems, respectively; bioMérieux), pH values (6, 6.5, 7 and 8.5) and NaCl concentrations (0.5%, 1%, 5%, 7.5% and 10%).

**Morphologic, biochemical and antibiotic susceptibility tests**

The morphologic characteristics of each species, namely Gram staining, motility and spore formation, were determined as previously described [8]. To determine the size, and to confirm the shape of the cells as well as spore formation, the cells were negatively stained and observed with electron microscopy as follows: cells were fixed with 2.5% glutaraldehyde in 0.1 M cacodylate buffer for at least 1 hour at 4°C. A drop of cell suspension was deposited for approximately 5 minutes on glow-discharged formvar carbon film on 400 mesh nickel grids (FCF400-Ni; Electron Microscopy Sciences (EMS), Hatfield, PA, USA). The grids were dried on blotting paper, and cells were
negatively stained for 10 seconds with 1% ammonium molybdate solution in filtered water at room temperature. Electron micrographs were acquired with a Tecnai G20 Cryo (FEI Company, Limeil-Brevannes, France) transmission electron microscope operated at 200 keV. Cells were then measured by ImageJ software (Image Processing and Analysis in Java; US National Institutes of Health, Bethesda, MD, USA; http://imagej.nih.gov/) [18,19].

Cellular fatty acid methyl ester (FAME) analysis was performed by gas chromatography/mass spectrometry (GC/MS). Two samples were prepared with approximately 23 mg of bacterial biomass per tube collected from several culture plates. Fatty acid methyl esters were separated as described by Sasser [20]. GC/MS analyses were carried out as described previously [21]. Briefly, fatty acid methyl esters were separated using an Elite 5-MS column and monitored by mass spectrometry (Clarus 500−SQ 8 S; PerkinElmer, Courtaboeuf, France). A spectral database search was performed using MS Search 2.0 operated with the Standard Reference Database 1A (National Institute of Standards and Technology, Gaithersburg, MD, USA) and the FAME mass spectral database (Wiley, Chichester, UK).

The metabolic capacities of the strains were determined using API ZYM, 20A and 50CCH strips according to the manufacturer’s instructions. After inoculation of the strips, they were incubated in anaerobic atmosphere for 4, 24 and 48 hours. Other enzymatic features were tested separately, namely oxidase and catalase, which were tested as previously described [8].

The disc diffusion method was used to test antibiotic susceptibility according to European Committee on Antimicrobial Susceptibility Testing recommendations [22].

Genomic DNA preparation and genome sequencing

Bacterial cell lysis was achieved using pretreatments such as lysozyme incubation at 37°C for 2 hours, followed by proteinase K action. DNA was extracted by the EZ1 biorobot (Qiagen, Germantown, MD, USA) with the EZ1 DNA tissues kit. The elution volume was 50 μL. Genomic DNA (gDNA) was quantified by a Qubit assay with the high sensitivity kit (Life Technologies, Carlsbad, CA, USA) to 123.7 ng/μL.

gDNA was sequenced with a MiSeq sequencer (Illumina, San Diego, CA, USA) using the mate-pair strategy. The gDNA was barcoded in order to be mixed with 11 other projects using the Nextera Mate Pair sample prep kit (Illumina). The mate-pair library was prepared with 1.5 μg of gDNA using the Nextera mate-pair Illumina guide. The gDNA sample was simultaneously fragmented and tagged with a mate-pair junction adapter. The fragmentation pattern was validated on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) with a DNA 7500 labchip. The DNA fragments ranged in size from 1.5 to 11 kb, with an optimal size at 4.031 kb. No size selection was performed, and 385.5 ng of tagged fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal size of 1070 bp on the Covaris device S2 in T6 tubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies), and the final concentration library was measured at 2.40 nmol/L. The libraries were normalized at 2 nM and pooled. After a denaturation step and dilution at 15 pM, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. The automated cluster generation and sequencing runs were performed in a single 2 × 301 bp run.

Genome annotation and analysis

Open reading frames (ORFs) were predicted using Prodigal [23] with default parameters, but the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched against the GenBank [24] and Clusters of Orthologous Groups (COGs) databases using BLASTP (E value 1e-03, coverage 0.7, identity percentage of 30%). If no hit was found, it was searched against the NR database using BLASTP with an E value of 1e-03, a coverage of 0.7 and an identity percentage of 30%, and if the sequence length was smaller than 80 aa, we used an E value of 1e-05. The tRNAScanSE tool [25] was used to find tRNA genes, while ribosomal RNAs were found using RNAmmer [26]. Lipoprotein signal peptides and the number of transmembrane helices were predicted using Phobius [27]. Mobile genetic elements were predicted using PHAST [28] and RAST [29]. ORFans were identified if all the BLASTP performed did not give positive results (E value smaller than 1e-03 for ORFs with sequence size larger than 80 aa or E value smaller than 1e-05 for ORFs with sequence length smaller than 80 aa). Such parameter thresholds have already been used in previous studies to define ORFans. Artemis [30] and DNA Plotter [31] were used for data management and the visualization of genomic features, respectively. The Mauve 2.3.1 alignment tool was used for multiple genomic sequence alignment [32].

Comparator species for genomic comparison were identified in the 16S RNA tree using PhyloPattern software [33]. For each selected genome, the complete genome sequence, proteome genome sequence and ORFeome genome sequence were retrieved from the National Center for Biotechnology Information FTP site. An annotation of the entire proteome was performed to define the distribution of functional classes of predicted genes according to the clusters of orthologous groups of proteins (using the same method as for the genome
Results

Strain identification and phylogenetic analyses
Isolation of strains Marseille-P2953 and LF2 occurred, respectively, after 7 days’ and 30 days’ preincubation in anaerobic blood culture bottles supplemented with rumen and sheep’s blood, respectively. Strain Marseille-P2414 was first isolated after 15 days’ preincubation in an aerobic blood culture bottle supplemented with rumen and sheep’s blood. Finally, strain MT26 was first isolated after 30 days’ preincubation in an anaerobic blood culture bottle supplemented with sheep’s blood. Because the routine identification using MALDI-TOF MS failed, the 16S rRNA gene of each strain was sequenced. The 16S rRNA gene sequences of strains Marseille-P2953 (LT732642), LF2 (LK021125), Marseille-P2414 (LT161894) and MT26 (LN998063) revealed a 98.41%, 96.95%, 98.23% and 98.44% similarity level with, respectively, Clostridium sporosphaeroides strain DSM 1294, Clostridium putrefaciens strain DSM 1291, Clostridium tertium strain JCM 6289 and Clostridium novyi strain JCM 1406 (Fig. 1, Table 1). All these similarity levels were lower than the 98.65 cutoff defined to create a new species without performing DDH. We thus propose the creation of four new species within the genus Clostridium with the following names: ‘Clostridium merdae’ (strain Marseille-P2953), ‘Clostridium amazontimonense’ (strain LF2), ‘Clostridium nigeriense’ (strain Marseille-P2414) and ‘Clostridium massilidielmoense’ (strain MT26). The MALDI-TOF MS spectra of the four strains in this study were incrementedin our database and compared to those of their closest relatives (Fig. 2).

Phenotypic description
Growth was observed between 28 and 45°C for all studied strains. While strain Marseille-P2414 grew in all tested atmospheres (namely aerobic, microaerophilic and anaerobic), growth was observed only in anaerobic atmosphere for strains Marseille-P2953, LF2 and MT26. All strains exhibited optimal growth at 37°C in anaerobic atmosphere with an incubation period of 24 hours for strain Marseille-P2414 and 48 hours for strains Marseille-P2953, LF2 and MT26. None of the strains was tolerant to NaCl, only growing at a concentration of 0.5%. Conversely, all strains grew at all tested pH concentrations (6, 6.5, 7, 8.5).

The phenotypic characteristics of the four strains in this study and their closest relatives are summarized in Table 2. All strains were Gram positive and were spore-forming, rod-shaped cells (Fig. 3). Strains LF2 and Marseille-P2414 were motile, whereas strains Marseille-P2953 and MT26 were nonmotile. Major fatty acids are presented in Table 3. We observed that strains Marseille-P2414, LF2 and MT26 present hexadecanoic acid (16:0) as the major fatty acid composing their cell envelope, confirming these three strains belong to the same group. For strain Marseille-P2953, its major fatty acid was 9-octadecenoic acid (18:1n9), which is minor or absent in the other strains. Although phylogeny clearly classified Marseille-P2953 as a Clostridium species, these data confirmed other phenotypical observations, suggesting that this strain could belong to another Clostridium cluster.

Genome description and comparison
The genomic properties of all strains are presented in Table 4 and Fig. 4. Genome sizes of the strains ranged from 2.75 Mb for strain Marseille-P2953 to 3.8 Mb for strain Marseille-P2414, while the G+C content ranged from 28.2% for strain Marseille-P2414 to 45.4% for strain Marseille-P2953. The proportion of protein-coding genes was similar in all four genomes, and the distribution into COGs categories was also similar (Table 5). However, we observed that the genome of ‘Clostridium merdae’ strain Marseille-P2953 presented a G+C% level significantly higher than the genomes of the other presented strains (Table 6). The G+C content of strain Marseille-P2953 sets it apart from the three other studied strains, as previously observed by its phenotypical features. Thus, the very large family of Clostridiaceae contains species with great phenotypical and genomic diversity, suggesting that this family, which was created before modern sequencing methods were in use, needs to be greatly reclassified. These observations demonstrate that phenotypical observations are no longer sufficient to classify all the newly discovered species. Whole genome sequencing must be used to characterize new bacterial species.

The DDH comparison presented in Table 7 demonstrates that the four studied new species are clearly distinct new species. Indeed, the DDH level between our four new species and the other Clostridium species comprises a range delimiting a distinct bacterial species.
FIG. 1. Phylogenetic tree highlighting position of ‘Clostridium amazonitimonense,’ ‘Clostridium mordae,’ ‘Clostridium massilielmoense’ and ‘Clostridium nigeriense’ relative to other type strains within Clostridium genus. Strains and their corresponding GenBank accession numbers for 16S rRNA genes are indicated in brackets. Sequences were aligned using Clustal W (http://www.clustal.org/clustal2/), and phylogenetic inferences were obtained using maximum-likelihood method within MEGA6 software (http://www.megasoftware.net/mega.php). Numbers at nodes are percentages of bootstrap values obtained by repeating analysis 1000 times to generate majority consensus tree. Butyricicoccus pullicaecorum (NR_044490) was used as outgroup. Scale bar = 0.02% nucleotide sequence divergence.

**TABLE 1. Species information**

| Characteristic | LF2                          | MT26                          | Marseille-P2953 | Marseille-P2414 |
|----------------|------------------------------|-------------------------------|----------------|-----------------|
| Patient information | Healthy 37-year-old Amazonian woman | Healthy 8-month-old Senegalese boy | Healthy 29-year-old Senegalese woman | Healthy 44-month-old girl from Niger |
| Isolation conditions | 30 days’ anaerobic preincubation in blood culture bottle with rumen and sheep’s blood | 30 days’ anaerobic preincubation in blood culture bottle with sheep’s blood | 7 days’ anaerobic preincubation in blood culture bottle with rumen and sheep’s blood | 15 days’ aerobic preincubation in blood culture bottle with rumen and sheep’s blood |
| Optimal growth temperature | 37°C | 37°C | 37°C | 37°C |
| Atmosphere | Strict anaerobe | Strict anaerobe | Strict anaerobe | Facultative anaerobe |
| Similarity percentage with closest species | 96.95% with Clostridium putrefaciens strain DSM 1291 (AJ127024) | 98.44% with Clostridium novyi strain DSM 1294 (Y66002) | 96.41% with Clostridium sporosphaeroides strain DSM 1294 (Y66002) | 98.23% with Clostridium tertium strain JCM 6289 (Y11874) |
Strain Marseille-P2414 shared 1342, 1434, 1553, 1564, 1628 and 1664 orthologous proteins with *Clostridium perfringens*, *Clostridium baratii*, *Clostridium dialis*, *Clostridium beijerinckii*, *Clostridium paraputri*cum and *Clostridium sartogoforme*, respectively (Supplementary Table S1a). Strain MT26 shared 1111, 1114, 1151, 1165, 1253, 1254, 1261, 1294 and 1372 orthologous proteins with *Clostridium acetireducens*, *Clostridium acetobutyli*cum, *Clostridium argentinense*, *Clostridium tetani*, *Clostridium carboxi*divorans, *Clostridium drakei*, *Clostridium homopropionicum*, *Clostridium novyi* and *Clostridium haemolyticum*, respectively.

**FIG. 2.** Reference mass spectra from *Clostridium amazonitimonense*, *Clostridium merdae*, *Clostridium massilidielmoense* and *Clostridium nigeriense.* Spectra from 12 individual colonies were compared and each reference spectrum generated (A). Gel view comparing *Clostridium amazonitimonense*, *Clostridium merdae*, *Clostridium massilidielmoense* and *Clostridium nigeriense* to other species within genus *Clostridium*. Gel view displays raw spectra of loaded spectrum files arranged in pseudo-gellike look. *x*-axis records *m/z* value. Left *y*-axis displays running spectrum number originating from subsequent spectra loading. Peak intensity expressed by greyscale scheme code. Colour bar and right *y*-axis indicate relation between colour peaks, with peak intensity in arbitrary units. Displayed species are indicated at left (B).
### TABLE 2. Differential characteristics of ‘Clostridium amazonitimonense,’ ‘Clostridium putrefaciens,’ ‘Clostridium merdae,’ ‘Clostridium sporosphaeroides,’ ‘Clostridium nigeriense,’ ‘Clostridium tertium, ‘Clostridium massilidielmoense’ and ‘Clostridium novyi’ [1,2]

| Characteristic          | ‘C. amazonitimonense’ | ‘C. putrefaciens’ | ‘C. merdae’ | ‘C. sporosphaeroides’ | ‘C. nigeriense’ | ‘C. tertium’ | ‘C. massilidielmoense’ | ‘C. novyi’ |
|-------------------------|-----------------------|-------------------|------------|-----------------------|----------------|-------------|-----------------------|-----------|
| Property                | ‘Bacilli’             | ‘Bacilli’         | ‘Bacilli’  | ‘Bacilli’             | ‘Bacilli’      | ‘Bacilli’   | ‘Bacilli’             | ‘Bacilli’ |
| Cell size (μm)          | 2–2.5                 | 1.2–1.5           | 0.8–1.5    | 0.5–0.6               | 0.5–1.4        | 0.5–1.4     | 0.6–1.4               |           |
| Gram stain              | +                     | +                 | +          | +                     | +              | +           | +                     | +         |
| Oxygen tolerance        | –                     | –                 | –          | –                     | –              | –           | –                     | –         |
| Salt tolerance          | –                     | –                 | –          | –                     | –              | –           | –                     | –         |
| Motility                | +                     | +                 | +          | +                     | +              | –           | +                     | +         |
| Endospore formation     | +                     | +                 | +          | +                     | +              | +           | +                     | +         |
| Major cellular fatty acid | 15:0 iso | 16.00 | NA | 16.00 | NA | 16.00 | NA | 16.00 | NA |
| Production of:          |                       |                   |            |                       |                |             |                       |           |
| Catalase                | –                     | NA                | –          | NA                    | –              | NA          | –                     | NA        |
| Oxidase                 | –                     | NA                | –          | NA                    | +              | NA          | –                     | NA        |
| Alkaline phosphatase    | +                     | NA                | –          | NA                    | +              | NA          | +                     | NA        |
| Urase                   | –                     | NA                | –          | NA                    | –              | NA          | –                     | NA        |
| β-Galactosidase         | –                     | NA                | –          | NA                    | –              | NA          | –                     | NA        |
| N-acetyl-glucosamine    | +                     | NA                | –          | NA                    | +              | NA          | +                     | NA        |
| Acid from:              |                       |                   |            |                       |                |             |                       |           |
| L-Arabinose             | –                     | NA                | –          | NA                    | –              | NA          | –                     | NA        |
| Ribose                  | +                     | NA                | –          | NA                    | +              | NA          | +                     | NA        |
| Mannose                 | +                     | NA                | –          | NA                    | +              | NA          | +                     | NA        |
| Mannitol                | –                     | NA                | –          | NA                    | –              | NA          | –                     | NA        |
| Sucrose                 | –                     | NA                | –          | NA                    | +              | NA          | –                     | NA        |
| d-Glucose               | +                     | +                 | +          | NA                    | +              | NA          | +                     | NA        |
| d-Fructose              | +                     | NA                | –          | NA                    | +              | NA          | +                     | NA        |
| d-Maltose               | –                     | NA                | –          | NA                    | +              | NA          | +                     | NA        |
| d-Lactose               | –                     | –                 | –          | NA                    | +              | NA          | +                     | NA        |
| Habitat                 | Human gut             | Hog muscle tissue| Human gut  | Human gut             | Human gut      | Human gut   | Human gut             | Soil      |

*+, positive result; −, negative result; NA, data not available.

**FIG. 3.** Electron micrographs of ‘Clostridium nigeriense’ (A), ‘Clostridium amazonitimonense’ (B), ‘Clostridium massilidielmoense’ (C) and ‘Clostridium merdae’ (D).
(Supplementary Table S1b). Strain LF2 shared 774, 1005, 1017 and 1138 orthologous proteins with Clostridium clari\textit{f}lavum, Clostridium perfringens, Clostridium cellulovorans and Clostridium beijerinckii, respectively (Supplementary Table S1c). Finally, strain Marseille-P2953 shared 735 orthologous proteins with Ruminococcus albus and Clostridium thermocellum as well as 740, 810, 833, 867, 889 and 907 orthologous proteins with Eubacterium siraeum, Clostridium leptum, Pseudoflavonifractor capillosus, Intestinimonas butyriciproducens, Flavoni\textit{f}ractor plautii, Oscillibacter ruminantium and Oscillibacter\textit{valericigenes}, respectively (Supplementary Table S1d). Strain Marseille-P2953 shared fewer orthologous proteins than the other described genomes, as this strain belongs to Clostridium cluster IV and has been compared to more closely related species with available genomes, which mostly belong to other genera, unlike the other described strains. AGIOS value ranges were similar among species with standing in nomenclature with compared species and with the described genomes, thus confirming that the described strains are new species.

**Conclusion**

Considering the 16S rRNA gene sequence similarity level with the closest species, all under the 98.65% cutoff to define a new species without performing DDH; considering the phenotypic characteristics distinguishing these strains from other known Clostridium species; and confirming that they belong to the

| Table 3. Cellular fatty acid composition (%) |
|---------------------------------------------|
| Fatty acid   | Name                      | 'Clostridium nigeriense' | 'Clostridium merdae' | 'Clostridium massiliolmoense' | 'Clostridium amazonitimonense' |
|--------------|---------------------------|--------------------------|----------------------|-------------------------------|-------------------------------|
| 16:00        | Hexadecanoic acid         | 38.9 ± 0.8               | 38.5 ± 2.1           | 55.0 ± 2.5                    | 15.3 ± 2.3                    |
| 14:00        | Tetradecanoic acid        | ND                       | ND                   | 15.7 ± 1.7                    | 12.0 ± 0.6                    |
| 18:00        | Octadecanoic acid         | 6.7 ± 1.1                | 3.2 ± 0.6            | 3.6 ± 0.9                     | TR                            |
| 13:00        | Tridecanoic acid          | ND                       | ND                   | 3.0 ± 0.6                     | TR                            |
| 12:00        | Dodecanoic acid           | 1.8 ± 0.1                | TR                   | 1.5 ± 1.0                     | TR                            |
| 04:00        | Butanoic acid             | ND                       | ND                   | 1.7 ± 1.0                     | TR                            |
| 15:00        | Pentadecanoic acid        | TR                       | ND                   | ND                            | TR                            |
| 18:1n9       | 9-Octadecanoic acid       | 20.1 ± 0.4               | 35.4 ± 0.4           | 14.4 ± 1.0                    | 2.3 ± 0.6                     |
| 13:0 iso     | 11-methyl-Dodecanoic acid | ND                       | ND                   | 22.7 ± 1.8                    | 11.2 ± 2.1                    |
| 5:0 iso      | 3-methyl-Butanoic acid    | ND                       | ND                   | 23.1 ± 0.3                    | ND                            |
| 18:1n7       | 11-Octadecanoic acid      | 7.9 ± 0.2                | ND                   | TR                            | ND                            |
| 14:01        | Tetradecanoic acid        | 6.6 ± 0.3                | 5.3 ± 1.4            | ND                            | ND                            |
| 18:2n6       | 9,12-Octadecanoic acid    | 5.5 ± 0.4                | 15.2 ± 0.6           | 3.9 ± 1.3                     | TR                            |
| 16:1n7       | 9-Hexadecanoic acid       | 5.3 ± 0.2                | TR                   | ND                            | ND                            |
| 16:1n9       | 7-Hexadecenoic acid       | 3.9 ± 0.2                | ND                   | ND                            | ND                            |
| 13:0 anteiso | 10-methyl-Dodecanoic acid | ND                       | ND                   | 2.3 ± 0.3                     | ND                            |
| 15:0 anteiso | 12-methyl-Tetradecanoic   | TR                       | TR                   | 4.8 ± 0.5                     | ND                            |
| 15:0 iso     | 13-methyl-Tetradecanoic   | TR                       | TR                   | 22.0 ± 1.5                    | ND                            |

ND, not detected. TR, traces.

| Table 4. Nucleotide content and gene count levels of genomes |
|-------------------------------------------------------------|
| Attribute               | 'Clostridium amazonitimonense' | 'Clostridium merdae' | 'Clostridium nigeriense' | 'Clostridium massiliolmoense' | 'Clostridium amazonitimonense' |
|-------------------------|--------------------------------|---------------------|-------------------------|-------------------------------|-------------------------------|
| No. of reads            | 5 235 372                      | 1 841 718           | 1 832 750               | 3 097 182                     |
| Scaffolds/contigs        | 4/15                           | 4/4                 | 4/19                    | 9/9                           |
| Size (bp)               | 3 761 417                      | 2 748 860           | 3 799 474               | 3 759 325                     |
| G+C content (%)         | 1 145 890 (30.55)              | 1 246 285           | 1 064 556               | 1 263 307                     |
| Coding region (bp)      | 3 129 293 (83.2)               | 2 373 508           | 3 322 705               | 2 453 243                     |
| Total genes             | 3553 (100)                     | 2699 (100)          | 2659 (100)              | 2772 (100)                    |
| RNA genes               | 147 (4.1)                      | 59 (2.2)            | 187 (5.1)               | 91 (3.3)                      |
| Protein-coding genes    | 3406 (95.9)                    | 264 (97.8)          | 3472 (94.9)             | 2681 (96.7)                   |
| Genes with function prediction | 2477 (72.7)                   | 1966 (74.4)         | 245 (70.5)              | 1989 (74.2)                   |
| Genes assigned to COGs  | 2241 (65.8)                    | 167 (63.2)          | 2149 (61.9)             | 1639 (61.1)                   |
| Genes with peptide signals | 270 (7.9)                     | 271 (10.2)          | 307 (8.8)               | 228 (2.8)                     |
| ORFans genes            | 189 (5.5)                      | 117 (4.4)           | 119 (3.4)               | 40 (1.5)                      |
| Genes with transmembrane helices | 849 (24.9)                     | 569 (21.5)          | 935 (26.9)              | 639 (23.8)                    |
| Genes associated with PKS or NRPS | 3 (0.08)                      | 0 (0)               | 0 (0)                   | 0 (0)                         |
| No. of antibiotic resistance genes | 1 (0.03)                       | 0 (0)               | 0 (0)                   | 0 (0)                         |
| No. of genes associated with Pfam-A domains | 2386 (88)                     | 3182 (86)          | 3182 (86)               | 246 (88)                      |

COGs, Clusters of Orthologous Groups database; NRPS, nonribosomal peptide synthase; ORF, open reading frame; PKS, polyketide synthase.

The total is based on either the size of the genome in base pairs or the total number of protein-coding genes in the annotated genome.
genus *Clostridium* and have the genomic characteristics of these strains, we propose the creation of four new *Clostridium* species: ‘*Clostridium nigeriense*’, ‘*Clostridium massilidielmoense*’, ‘*Clostridium amazonitimonense*’ and ‘*Clostridium merdae*.’

**Description of the new species ‘*Clostridium nigeriense*’**

‘*Clostridium nigeriense*’ (ni.ge.rien’se, L. gen. adj., from nigeriense, ‘of Niger,’ referring to where the stool sample was collected) was isolated from the stool sample of a healthy girl from Niger and was deposited in the DSMZ collection under accession number DSM 102218 and in the CSUR collection under accession number CSUR P2414.

‘*Clostridium nigeriense*’ is a Gram-positive bacillus that is motile, spore forming and facultative anaerobic. The test of catalase activity was negative but was positive for oxidase. ‘*Clostridium nigeriense*’ forms translucent colonies with a mean...
The 16S rRNA gene and genome sequences are available in the European Molecular Biology Laboratory–European Bioinformatics Institute (EMBL-EBI) database under accession numbers LT161894 and FLKB00000000, respectively.

Strain Marseille-P2414T is the type strain of the new species ‘Clostridium amazonitimonense’.

Description of the new species ‘Clostridium amazonitimonense’

‘Clostridium amazonitimonense’ (am.a.zo.ni.ti.mo.nen’se, N.L. neut. adj., from amazoni (a.ma.zo.ni, N.L. neut. adj.), referring to the Amazonian forest, the location where the stool sample was collected, and timonense (t.i.mo.nen’se, N.L. neut. adj., referring to La Timone hospital, where the strain was isolated) was isolated from the stool sample of a healthy middle-aged woman from Amazonia and is deposited in the DSMZ collection under accession number DSM 28600 and in the CSUR collection under accession number CSUR P1445.

Strain LF2 is a Gram-positive bacillus that is motile, spore forming and a strict anaerobe. It is negative for catalase and oxidase activity. ‘Clostridium amazonitimonense’ forms white colonies with a mean diameter of 5 mm.

The genome is 3 761 417 bp long with 30.56% G+C content. The 16S rRNA gene and genome sequences are available in the EMBL-EBI database under accession numbers LK021125 and CCNN00000000, respectively.

The type strain LF2T is the type strain of the new species ‘Clostridium amazonitimonense.’
Description of the new species ‘Clostridium massilidielmoense’

‘Clostridium massilidielmoense’ (mass.si.li.di.el.mo.en’se, L. neut. adj., a combination of massili (ma.ssi.li, N.L. neut. adj.), from Massilia, the Roman name of Marseille, where strain MT26 was isolated; and dielmoense (di.el.mo.en’se, N.L. neut. adj.), from Dielmo, the Senegalese village where the boy from whom the strain was cultivated lived) was isolated from the stool sample of a healthy 8-month-old Senegalese boy and is deposited in the CCUG collection under accession number CCUG 69893 and in the CSUR collection under accession number CSUR P2255.

Strain MT26 is a Gram-positive bacillus that is nonmotile, spore forming and a strict anaerobe. It shows positive catalase activity but no oxidase activity. ‘Clostridium massilidielmoense’ forms a translucent biofilm with cells of an average diameter of 0.54 μm.

The genome is 2 759 335 bp long with 28.35% G+C content. The 16S rRNA gene and genome sequences are available in the EMBL-EBI database under accession numbers LN998063 and FXLO00000000, respectively.

The type strain MT26T is the type strain of the new species ‘Clostridium massilidielmoense.’

Description of the new species ‘Clostridium merdae’

‘Clostridium merdae’ (merdae, L. gen. n., from merdae, ‘of faeces,’ referring to the fact that the strain was isolated from a stool sample) was isolated from the stool sample of a healthy 29-year-old Senegalese woman and is deposited in the DSMZ collection under accession number DSM 103669 and in the CSUR collection under accession number CSUR P2953.

Strain Marseille-P2953 is a Gram-positive bacillus that is motile, spore forming and a strict anaerobe. It does not show catalase or oxidase activity. ‘Clostridium merdae’ forms small white colonies with a mean diameter of 1 mm.

The genome is 2 748 860 bp long with 45.34% G+C content. The 16S rRNA gene and genome sequences are available in the EMBL-EBI database under accession numbers LT732642 and FXLN00000000, respectively.

The type strain Marseille-P2953 is the type strain of the new species ‘Clostridium merdae.’

Conflict of interest

None declared.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.nmni.2017.11.003.
References

[1] Salazar N, Arboleya S, Valdés L, Stanton C, Ross P, Ruiz L, et al. The human intestinal microbiome at extreme ages of life. Dietary intervention as a way to counteract alterations. Front Genet 2014;5:406.

[2] Lagier JC, Million M, Hugon P, Armougom F, Raoult D. Human gut microbiota: repertoire and variations. Front Cell Infect Microbiol 2012;2:136.

[3] Sankar SA, Lagier JC, Pontarotti P, Raoult D. Fournier PE. The human gut microbiome: a taxonomic conundrum. Syst Appl Microbiol 2015;38:276–86.

[4] Tidjani Alou M, Lagier JC, Raoult D. Diet influence on the gut microbiota and dysbiosis related to nutritional disorders. Hum Microbiome J 2016;1:3–11.

[5] Million M, Diao A, Raoult D. Gut microbiota and malnutrition. Microb Pathog 2017;106:127–38.

[6] Angelakis E, Lagier JC. Samples and techniques highlighting the links between obesity and microbiota. Microb Pathog 2017;106:119–26.

[7] Rajilić-Stojić M, de Vos WM. The first 1000 cultured species of the human gastrointestinal microbiota. FEMS Microbiol Rev 2014;38:996–1047.

[8] Lagier JC, Hugon P, Khelifa S, Fournier PE, La Scola B, Raoult D. The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. Clin Microbiol Rev 2015;28:237–64.

[9] Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 2012;18:1185–93.

[10] Parte AC. LPSN—list of prokaryotic names with standing in nomenclature. Nucleic Acids Res 2014;42:D613–6.

[11] De Vos P, Garrity GM, Jones D, Krieg NR, Ludwig W, Rainey FA, et al., editors. Systematic bacteriology. New York: Springer; 2009.

[12] Lawson PA, Citron DM, Tyrrell KL, Finegold SM. Reclassification of Clostridium difficile as Clostridoides difficile (Hall and O’Toole 1935) Prévet 1938. Anaerobe 2016;40:95–9.

[13] Fournier PE, Lagier JC, Dubourg G, Raoult D. From culturomics to microbiota and dysbiosis related to nutritional disorders. Hum Microbiota J 2015;3:518–30.

[14] Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 2009;49:543–51.

[15] Seng P, Abat C, Rolain JM, Colson P, Lagier JC, Gouriet F, et al. Identification of rare pathogenic bacteria in a clinical microbiology laboratory: impact of matrix-assisted laser desorption ionization—time of flight mass spectrometry. J Clin Microbiol 2013;51:2182–94.

[16] Drancourt M, Bollet C, Carlia J, Armougom F, Gayral JP, Raoult D. 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. J Clin Microbiol 2000;38:3623–30.

[17] Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol 2014;64:346–51.

[18] Schindelin J, Rueden CT, Hiner MC, Eliceiri KW. The ImageJ ecosystem: an open platform for biomedical image analysis. Mol Reprod Dev 2015;82:518–29.

[19] Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 2012;9:671–5.

[20] Sasser M. Bacterial identification by gas chromatographic analysis of fatty acids methyl esters (GC-FAME). Newark, NJ: Microbial ID; 2006.

[21] Dione N, Sankar SA, Lagier JC, Khelifa S, Michele C, Armstrong N, et al. Genome sequence and description of Anaerobiospirillum massiliensis sp. nov. New Microbe New Infect 2016;10:66–76.

[22] Matsushita K, Brown DFJ, Kahlmeter G. Development of the EUCAST disk diffusion antimicrobial susceptibility testing method and its implementation in routine microbiology laboratories. Clin Microbiol Infect 2014;20:O255–66.

[23] Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform 2010;11:119.

[24] Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2012;40:D48–53.

[25] Lowe TM, Eddy SR. rRNA Scan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955–64.

[26] Lagesen K, Hallin P, Rødland EA, Sætherfeldt HH, Rognes T, Lysseth DW. RNAHammer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 2007;35:3100–8.

[27] Kall L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol 2004;338:1027–36.

[28] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phase search tool. Nucleic Acids Res 2011;39:W347–52.

[29] Aziz RK, Bartels D, Best AA, Dejongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 2008;9:75.

[30] Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. Bioinforma 2000;16:944–5.

[31] Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinforma 2009;25:119–20.

[32] Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004;14:4394–403.

[33] Gouret P, Thompson JD, Pontarotti P. PhylotoPlot: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinform 2009;10:298.

[34] Gouret P, Paganini J, Dainat J, Louati D, Darbo E, Pontarotti P, et al. Integration of evolutionary biology concepts for functional annotation and automation of complex research in evolution: the multi-agent software system DAGOBAH. In: Pontarotti P, editor. Evolutionary biology—concepts, biodiversity, macroevolution and genome evolution. Berlin: Springer; 2011. p. 71–87.

[35] Gouret P, Vitiello V, Balandraud N, Gilles A, Pontarotti P, Danchin EG. FIGENIX: intelligent automation of genomic annotation: expertise integration in a new software platform. BMC Bioinform 2005;6:198.

[36] Auch AF, von Jan M, Klenk H-P, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci 2010;2:117–34.

[37] Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence—based species delineation with confidence intervals and improved distance functions. BMC Bioinform 2013;14:60.

[38] Ramasamy D, Mishra AK, Lagier JC, Padmanabhan R, Rossi M, Sentausa E, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol 2014;64:384–91.

[39] Sturges WS, Drake ET. A complete description of Clostridium putrefaciens (McBryde). J Bacteriol 1927;10:175–9.