

Non-contiguous finished genome sequence and description of *Clostridium dakarensis* sp. nov.

Cheikh Ibrahima Lo¹, Ajay Kumar Mishra¹, Roshan Padhmanabhan¹, Bissoume Samb Ba², Amy Gassama Sow², Catherine Robert¹, Carine Couderc¹, Ngor Faye³, Didier Raoult¹, Pierre-Edouard Fournier¹, and Florence Fenollar^{1*}

¹Aix-Marseille Université, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), Marseille, France

²Unité de Bactériologie Expérimentale Institut Pasteur de Dakar 36, Dakar, Senegal

³Université Cheikh Anta Diop de Dakar, Laboratoire de Parasitologie générale, Fann (Senegal)

*Correspondence: Florence Fenollar (florence.fenollar@univ-amu.fr)

Keywords: *Clostridium dakarensis*, genome, culturomics, taxono-genomics

Clostridium dakarensis strain FF1^T, is the type strain of *Clostridium dakarensis* sp. nov., a new species within the genus *Clostridium*. This strain, whose genome is described here, was isolated from the fecal flora of a 4-month-old Senegalese child suffering from gastroenteritis. *C. dakarensis* sp. nov. strain FF1^T is an obligate anaerobic Gram-positive bacillus. Here we describe the features of this organism, together with the complete genome sequence and annotation. The 3,735,762 bp long genome (1 chromosome but no plasmid) exhibits a G+C content of 27.98% and contains 3,843 protein-coding and 73 RNA genes, including 8 rRNA genes.

Introduction

Clostridium dakarensis strain FF1^T (= CSUR P243 = DSM 27086), is the type strain of *Clostridium dakarensis* sp. nov. This bacterium is a Gram-positive, anaerobic, spore-forming, indole negative bacillus that was isolated from the stool of a 4-month-old Senegalese child suffering from gastroenteritis as part of a “culturomics” study aiming at cultivating individually all species within human feces.

The elevated cost and lack of intra- and inter-laboratory reproducibility of the “gold standard” of taxonomic tools, i.e. DNA-DNA hybridization and G+C content determination [1], put bacterial taxonomic classification in a precarious state. In addition, the internationally-validated cutoff values of 16S rRNA sequence comparison [2] do not apply to all validly published genera and species. Recently, high throughput genome sequencing and mass spectrometric analyses of bacteria have allowed unprecedented access to a wealth of genetic and proteomic information [3]. As a consequence, we proposed to use a polyphasic approach [4] to describe new bacterial taxa, including genome

sequence, MALDI-TOF spectrum and main phenotypic characteristics [5-11].

The genus *Clostridium* (Prazmowski, 1880), classified among the *Firmicutes*, was created in 1880 [12] and consists of obligate anaerobic rod-shaped bacilli capable of producing endospores [12]. More than 180 *Clostridium* species have been described to date [13]. Members of the genus *Clostridium* are mostly environmental bacteria or associated with the commensal digestive flora of mammals, but several are major human pathogens, including *C. botulinum*, *C. difficile*, *C. tetani* and *C. perfringens* [14,15]. A few species, such as *C. butyricum* and *C. pasteurianum*, fix nitrogen and have gained importance in agricultural and industrial applications [16,17].

Here we present a summary classification and a set of features for *C. dakarensis* sp. nov. strain FF1^T (= CSUR P243 = DSM 27086) together with the description of the complete genomic sequencing and annotation. These characteristics support the circumscription of the species *C. dakarensis* sp. nov.

Classification and features

A stool specimen was collected from a 4-month-old Senegalese child suffering from gastroenteritis. Informed consent was obtained from the child's parents and approval from the ethics committee from the Institut Federatif de Recherche 48 (Faculté de Médecine, Marseille, France). The fecal specimen was preserved at -20°C after collection and sent to Marseille. Strain FF1^T (Table 1) was isolated in July 2011 by anaerobic cultivation on 5% sheep blood-enriched Columbia agar (BioMerieux, Marcy

l'Etoile, France). This strain exhibited a 96.90% 16S rRNA nucleotide sequence similarity with *C. lituseburensis*, the phylogenetically closest validated *Clostridium* species (Figure 1). Although sequence similarity of the 16S rRNA is not uniform across taxa, this value was lower than the 98.7% threshold recommended by Stackebrandt and Ebers to delineate a new species without carrying out DNA-DNA hybridization [30]. In addition, it was consistent with 16S rRNA identity values observed among validated species within the *Clostridium* genus that range from 78.4 to 99.8%.

Table 1. Classification and general features of *Clostridium dakarensis* strain FF1^T according to the MIGS recommendations [18].

MIGS ID	Property	Term	Evidence code ^a
		Domain <i>Bacteria</i>	TAS [19]
		Phylum <i>Firmicutes</i>	TAS [20-22]
		Class <i>Clostridia</i>	TAS [23,24]
	Current classification	Order <i>Clostridiales</i>	TAS [25,26]
		Family <i>Clostridiaceae</i>	TAS [25,27]
		Genus <i>Clostridium</i>	TAS [12,25,28]
		Species <i>Clostridium dakarensis</i>	IDA
		Type strain FF1	IDA
	Gram stain	Positive	IDA
	Cell shape	Rod-shaped	IDA
	Motility	Motile	IDA
	Sporulation	Sporulating	IDA
	Temperature range	Mesophile	IDA
	Optimum temperature	37°C	IDA
MIGS-6.3	Salinity	Growth in BHI medium + 5% NaCl	IDA
MIGS-22	Oxygen requirement	Anaerobic	IDA
	Carbon source	Unknown	NAS
	Energy source	Unknown	NAS
MIGS-6	Habitat	Human gut	IDA
MIGS-15	Biotic relationship	Free living	IDA
	Pathogenicity	Unknown	
	Biosafety level	2	
MIGS-14	Isolation	Human feces	NAS
MIGS-4	Geographic location	Senegal	IDA
MIGS-5	Sample collection time	June 2011	IDA
MIGS-4.1	Latitude	13.7167	IDA
MIGS-4.1	Longitude	- 16.4167	IDA
MIGS-4.3	Depth	Surface	IDA
MIGS-4.4	Altitude	51 m above sea level	IDA

^aEvidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [29]. If the evidence is IDA, then the property was directly observed for a live isolate by one of the authors or an expert mentioned in the acknowledgements.

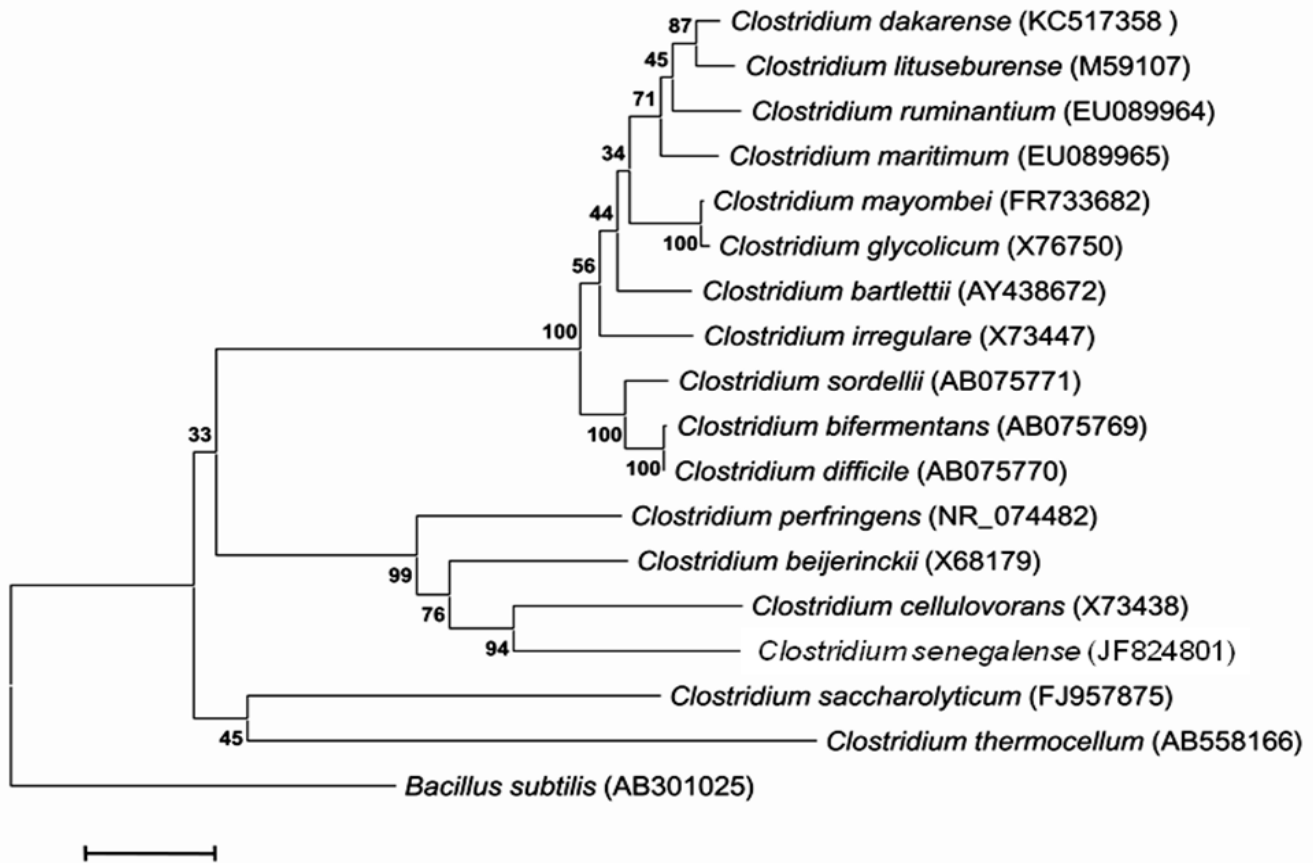


Figure 1. Phylogenetic tree highlighting the position of *C. dakarene* sp. nov. strain FF1^T relative to other type strains within the *Clostridium* genus. GenBank accession numbers are indicated in parentheses. Sequences were aligned using CLUSTALW, and phylogenetic inferences obtained using the maximum-likelihood method within the MEGA software. Numbers at the nodes are bootstrap values obtained by repeating 500 times the analysis to generate a majority consensus tree. *Bacillus subtilis* was used as an outgroup. The scale bar represents a 2% nucleotide sequence divergence.

Different growth temperatures (25, 30, 37, 45 and 56°C) were tested. Growth was observed between 25 and 37°C, with optimal growth at 37°C after 24 hours of inoculation in anaerobic conditions. Colonies were 1.5 mm in diameter and opaque and smooth appearance on blood-enriched Columbia agar. Growth of the strain was tested under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems, respectively (BioMerieux), and under aerobic conditions, with or without 5% CO₂. The strain growth was obtained only in anaerobic conditions. Gram staining showed rod-shaped Gram-positive bacilli able to form spores (Figure 2). The motility test was positive. Cells grown on agar have a mean diameter of 1.2 µm (Figure 3).

Strain FF1^T exhibited neither catalase nor oxidase activities. Using API Rapid ID 32A (BioMerieux, Marcy l’Etoile), a positive reaction were observed

for arginine dihydrolase, N-acetyl-β-glucosaminidase and pyroglutamic acid arylamidase. Negative reactions were observed for urease, indole and nitrate reduction. Using API 50 CH (BioMerieux, Marcy l’Etoile), positive reactions were observed for galactose, glucose, maltose and saccharose fermentation and negative reaction were observed for ribose, lactose and fructose. *C. dakarene* is susceptible to amoxicillin, metronidazole, vancomycin, imipenem and rifampicin and resistant to trimethoprim/ sulfamethoxazole. When compared with representative species from the genus *Clostridium*, *C. dakarene* strain FF1^T exhibited the phenotypic differences detailed in Table 2.

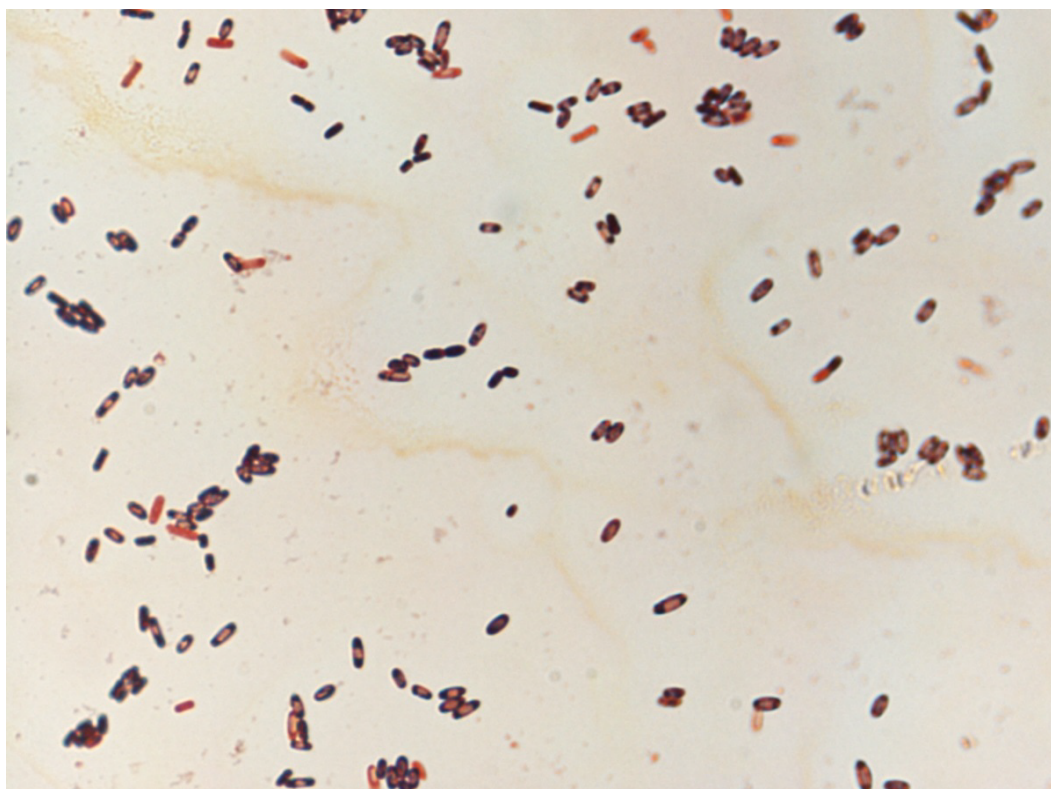


Figure 2. Gram staining of *C. dakarensis* sp. nov. strain FF1^T.

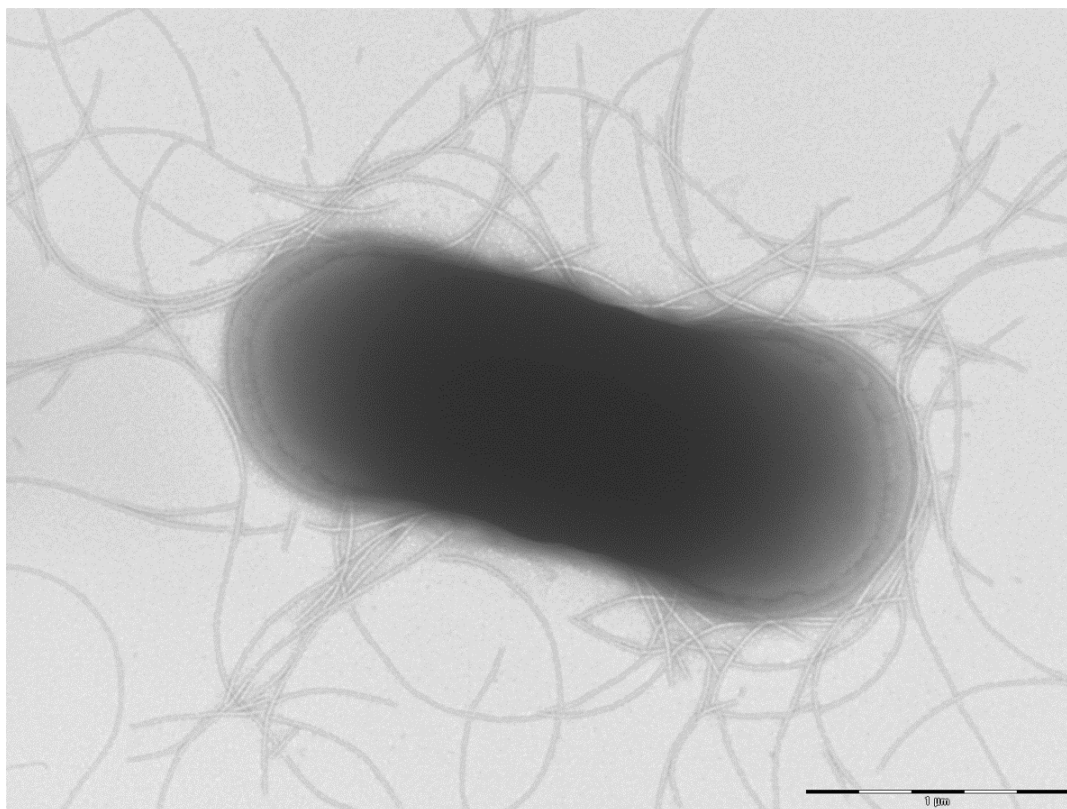


Figure 3. Transmission electron microscopy of *C. dakarensis* sp. nov. strain FF1^T, using a Morgani 268D (Philips) at an operating voltage of 60kV. The scale bar represents 1 μm.

Table 2. Differential characteristics of *C. dakareense* sp. nov. strain FF1^T (Cda)

Properties	CDa	CBa	CBe	CC	CDi	CG	CP	CSa	CSe	CT
Cell diameter (µm)	1.2	1.5	1.7	2.5	3.0	0.4-1.0	1.3	3.0	1.1	2.5
Oxygen requirement	-	-	-	-	-	-	-	-	-	na
Pigment production	-	-	-	-	+	+	+	na	-	+
Gram stain	+	+	V	-	+		+	-	+	-
Salt requirement	-	na	na	na	na	-	-	na	-	na
Motility	+	-	+	-	+	+	-	-	+	-
Endospore formation	+	+	+	+	+	+	w	+	+	+
Production of										
Acid phosphatase	+	+	na	na	na	na	+	na	na	na
Catalase	-	-	-	-	na	na	na	na	-	na
Oxidase	-	na	na	na	na	na	na	na	-	na
Nitrate reductase	-	-	-	na	-	-	+	+	-	-
Urease	-	-	-	na	na	na	na	na	-	na
β-galactosidase	-	+	na	na	na	-	+	na	-	na
Acid from										
L-Arabinose	-	na	+	-	-	-	-	+	na	na
Ribose	-	+	-		-	-	+	w	na	na
Mannose	-	-	+		+	-	+	na	na	na
Mannitol	-	+	+	+	+	-	-	w	na	na
Sucrose	-	+	+	+	+	-	+	w	na	na
D-glucose	+	+	+	+	na	+	+		na	na
D-fructose	-	+	+	+	+	+	+	+	na	na
D-maltose	+	+	+	+	-	+	+	w	na	na
D-lactose	-	na	+	+	-	-	+	w	na	na
Hydrolysis of										
Gelatin	na	-	+	-	na	-	na	na	na	+
Starch	na	na	+	-	-	-	+	-	na	
G+C content (mol%)										
	27.98	29.8	28	27	28	29	27	28	26.8	39
Habitat	Human gut	Human gut	Human gut	Poplar wood	Human gut	Mud, wastewater	Colonic flora	Sewage sludge	Human gut	Sewage sludge

C. bartlettii (CBa), *C. beijerinckii* (CBe), *C. cellulovorans* (CC), *C. difficile* (CDi), *C. glycolicum* (CG), *C. perfringens* (CP), *C. saccharolyticum* (CSa), *C. senegalense* (CSe) and *C. thermocellum* (CT).

na = data not available; w = weak, v = variable reaction

Matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) MS protein analysis was carried out as previously described [31]. Briefly, a pipette tip was used to pick one isolated bacterial colony from a culture agar plate, and to spread it as a thin film on a MTP 384 MALDI-TOF target plate (Bruker Daltonics, Leipzig, Germany). Eighteen distinct deposits were made for strain FF^T from eighteen isolated colonies. Each smear was overlaid with 2 μ L of matrix solution (saturated solution of alpha-cyano-4-hydroxycinnamic acid) in 50% acetonitrile, 2.5% tri-fluoroacetic-acid, and allowed to dry for five minutes. Measurements were performed with a Microflex spectrometer (Bruker). Spectra were recorded in the positive linear mode for the mass range of 2,000 to 20,000 Da (parameter settings: ion source 1 (IS1), 20 kV; IS2, 18.5 kV; lens, 7 kV). A spectrum was obtained after 675 shots at a variable laser power. The time of acquisition was between 30 seconds and 1 minute per spot. The eighteen spectra were imported into the MALDI BioTyper software (version 2.0,

Bruker) and analyzed by standard pattern matching (with default parameter settings) against the main spectra of 4,706 bacteria including 216 spectra from validly published species of *Clostridium*, that are part of the reference data contained in the BioTyper database. The method of identification included the m/z from 2,000 to 20,000 Da. For every spectrum, 100 peaks at most were taken into account and compared with spectra in the database. A score enabled the identification, or not, from the tested species: a score > 2 with a validly published species enabled the identification at the species level, and a score < 1.7 did not enable any identification at the genus level. For strain FF^T, the maximal obtained score was lower than 1.9, thus suggesting that our isolate was not a member of a known species. We added the spectrum from strain FF^T to our database for future reference (Figure 4). Finally, the gel view allows us to highlight the spectrum differences with other members of the genus *Clostridium* (Figure 5).

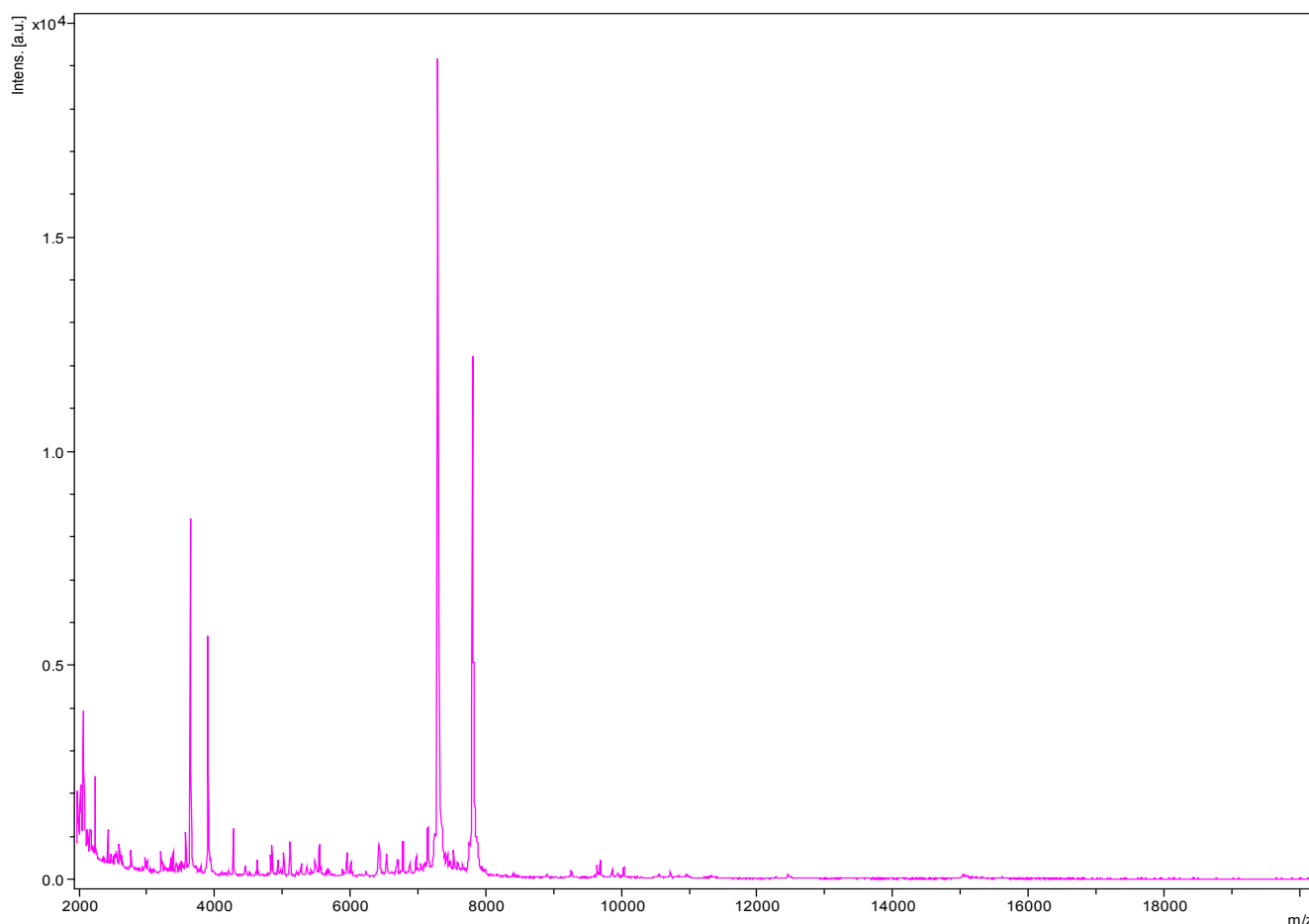


Figure 4. Reference mass spectrum from *C. dakarensis* strain FF^T. Spectra from 18 individual colonies were compared and a reference spectrum was generated.

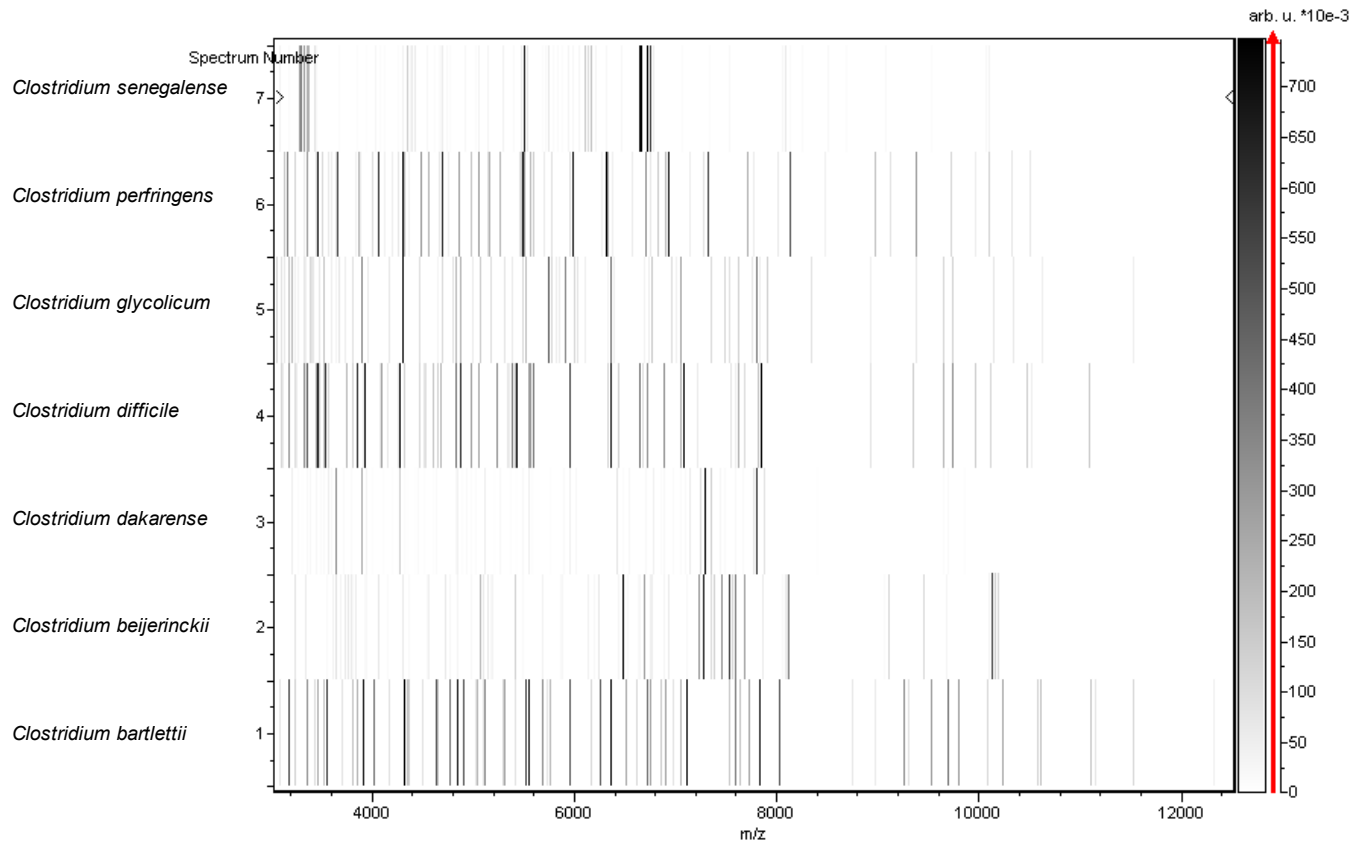


Figure 5. Gel view comparing *C. dakareense* sp. nov. strain FF1^T spectra with other members of the *Clostridium* genus (*C. bartlettii*, *C. beijerinckii*, *C. difficile*, *C. glycolicum*, *C. perfringens*, *C. senegalense*). The Gel View displays the raw spectra of all loaded spectrum files arranged in a pseudo-gel like look. The x-axis records the m/z value. The left y-axis displays the running spectrum number originating from subsequent spectra loading. The peak intensity is expressed by a Gray scale scheme code. The color bar and the right y-axis indicate the relation between the color a peak is displayed with and the peak intensity in arbitrary units.

Genome sequencing information

Genome project history

The organism was selected for sequencing on the basis of its phylogenetic position and 16S rRNA similarity to other members of the genus *Clostridium*, and is part of a “culturomics” study of the human digestive flora aiming at isolating all bacterial species within human feces. It was the 94th genome of a *Clostridium* species and the first genome of *Clostridium dakareense* sp. nov. The Genbank accession number is CBTZ00000000 and consists of 257 contigs. Table 3 shows the project information and its association with MIGS version 2.0 compliance [32].

Growth conditions and DNA isolation

C. dakareense sp. nov. strain FF1^T (= CSUR P243 = DSM 27086), was grown anaerobically on sheep blood-enriched Columbia agar medium at 37°C. Eight petri dishes were spread and resuspended in 4x100µl of G2 buffer (EZ1 DNA Tissue kit, Qiagen). A first mechanical lysis was performed by glass powder on the Fastprep-24 device (Sample Preparation system) from MP Biomedicals, USA) using 2x20 seconds cycles. DNA was then treated with 2.5 µg/µL lysozyme (30 minutes at 37°C) and extracted through the BioRobot EZ 1 Advanced XL (Qiagen). The DNA was then concentrated and purified on a Qiap kit (Qiagen). DNA concentration was 70.7ng/µl as determined by the Genios Tecan fluorometer, using the Quant-it Picogreen kit (Invitrogen).

Table 3. Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality draft
MIGS-28	Libraries used	One 454 paired end 3-kb library
MIGS-29	Sequencing platforms	454 GS FLX Titanium
MIGS-31.2	Fold coverage	35
MIGS-30	Assemblers	Newbler version 2.5.3
MIGS-32	Gene calling method	Prodigal
	Genbank ID	CBTZ00000000
	Genbank Date of Release	
MIGS-13	Project relevance	Study of the human gut microbiome

Genome sequencing and assembly

This project was loaded twice on a 1/4 region for the paired-end application and once on a 1/8 region for the shotgun on PTP Picotiterplates. The shotgun library was constructed with 500 ng of DNA as described by the manufacturer (Roche) with the GS Rapid library Prep kit. For the paired-end sequencing, 5 µg of DNA was mechanically fragmented on the Hydroshear device (Digilab, Holliston, MA, USA) with an enrichment size of 3-4kb. The DNA fragmentation was visualized using an Agilent 2100 BioAnalyzer on a DNA labchip 7500, which yield an optimal size of 3.6 kb. The library was constructed according to the 454_Titanium paired-end protocol and manufacturer. Circularization and nebulization were performed and generated a pattern with an optimum at 561 bp. After PCR amplification through 15 cycles followed by double size selection, the single stranded paired end library was then quantified with Quant-it Ribogreen kit (Invitrogen) on the Genios_Tecan fluorometer at 52 pg/µL. The library concentration equivalence was calculated as 1.7E+08 molecules/µL. The library was stored at -20°C until use.

The shotgun library was clonally amplified with 3cpb in 3 emPCR reactions and the paired end library was amplified with lower cpb (1cpb) in 4 emPCR reactions with the GS Titanium SV emPCR Kit (Lib-L) v2. The yield of the emPCR was 5.37% for the shotgun reads and 19.27% for the paired-end reads, according to the quality expected by the range of 5 to 20% from the Roche procedure. A total of 340,000 beads from the 1/8 region of

the shotgun reads and 790,000 beads from the 1/4 region of the paired-end reads were loaded on the GS Titanium PicoTiterPlates (PTP Kit 70×75) and sequenced with the GS Titanium Sequencing Kit XLR70.

The runs were performed overnight and then analyzed on the cluster through the gsRunBrowser and gsAssembler_Roche. The global 383,079 passed filter sequences generated 96.50 Mb with a length average of 277 bp. These sequences were assembled using the Newbler software from Roche with 90% identity and 40 bp as overlap. Fourteen scaffolds and 257 large contigs (>1500bp) were obtained, for a genome size of 3,735,762 bp.

Genome annotation

Open Reading Frames (ORFs) were predicted using Prodigal [33] 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 database [34] and the Clusters of Orthologous Groups (COG) databases using BLASTP. The tRNAScanSE tool [35] was used to find tRNA genes, whereas ribosomal RNAs were found by using RNAmmer [36] and BLASTn against the GenBank database. Lipoprotein signal peptides and numbers of transmembrane helices were predicted using SignalP [37] and TMHMM [38] respectively. ORFans were identified if their BLASTP *E*-value was lower than $1e^{-03}$ for alignment length greater than 80 amino acids. If alignment lengths

were smaller than 80 amino acids, we used an *E*-value of 1e-05. Such parameter thresholds have already been used in previous works to define ORFans. Artemis [39] was used for data management and DNA Plotter [40] was used for visualization of genomic features. Mauve alignment tool was used for multiple genomic sequence alignment and visualization [41].

To estimate the mean level of nucleotide sequence similarity at the genome level between *C. dakarensis* and nine other members of the genus *Clostridium* (Table 6), we use the Average Genomic Identity of gene Sequences (AGIOS) homemade software. Briefly, this software combines the Proteinortho software [42] for detecting orthologous proteins between genomes compared two by two, then retrieves the corresponding genes and determines the mean percentage of nucleotide sequence identity among orthologous ORFs using the Needleman-Wunsch global alignment algorithm. *Clostridium dakarensis* strain FF1^T, was compared to *C. bartlettii* strain DSM 16795 (GenBank accession number NZ_DS499569), *C. beijerinckii* strain NCIMB 8052 (NC_009617), *C. cellulovorans* strain 743B (NC_014393), *C. difficile*

strain 630 (NC8009089), *C. glycolicum* strain ATCC 14880 (ARES01000000), *C. perfringens* strain ATCC 13124 (BA000016), *C. saccharolyticum* strain WM1 (NC_014376), *C. senegalense* strain JC122^T (CAEV00000000), and *C. thermocellum* strain ATCC 27405 (CP000568).

Genome properties

The genome of *C. dakarensis* sp. nov. strain FF1^T is 3,735,762 bp long (1 chromosome, but no plasmid) with a 27,98% G + C content of (Figure 6 and Table 4). Of the 3,916 predicted genes, 3,843 protein-coding genes, and 73 were RNAs. Eight rRNA genes (one 16S rRNA, one 23S rRNA and six 5S rRNA) and 65 predicted tRNA genes were identified in the genome. A total of 2,769 genes (72.05%) were assigned a putative function (by COG or NR blast). Two hundred ninety-eight genes were identified as ORFans (7.75%). The remaining 515 genes were annotated as hypothetical proteins (13, 40%). The distribution of genes into COGs functional categories is presented in Table 4. The properties and the statistics of the genome are summarized in Tables 4 and 5.

Table 4. Nucleotide content and gene count levels of the genome.

Attribute	Value	% of total ^a
Genome size (bp)	3,735,762	100
DNA coding region (bp)	3,239,020	86.70
DNA G+C content (bp)	1,045,424	27.98
Total genes	3,916	100
RNA genes	73	1.86
Protein-coding genes	3,843	98.14
Genes with function prediction	2,769	72.05
Genes assigned to COGs	2,849	74.13
Genes with peptide signals	410	10.67
Genes with transmembrane helices	1,016	26.44

^aThe 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

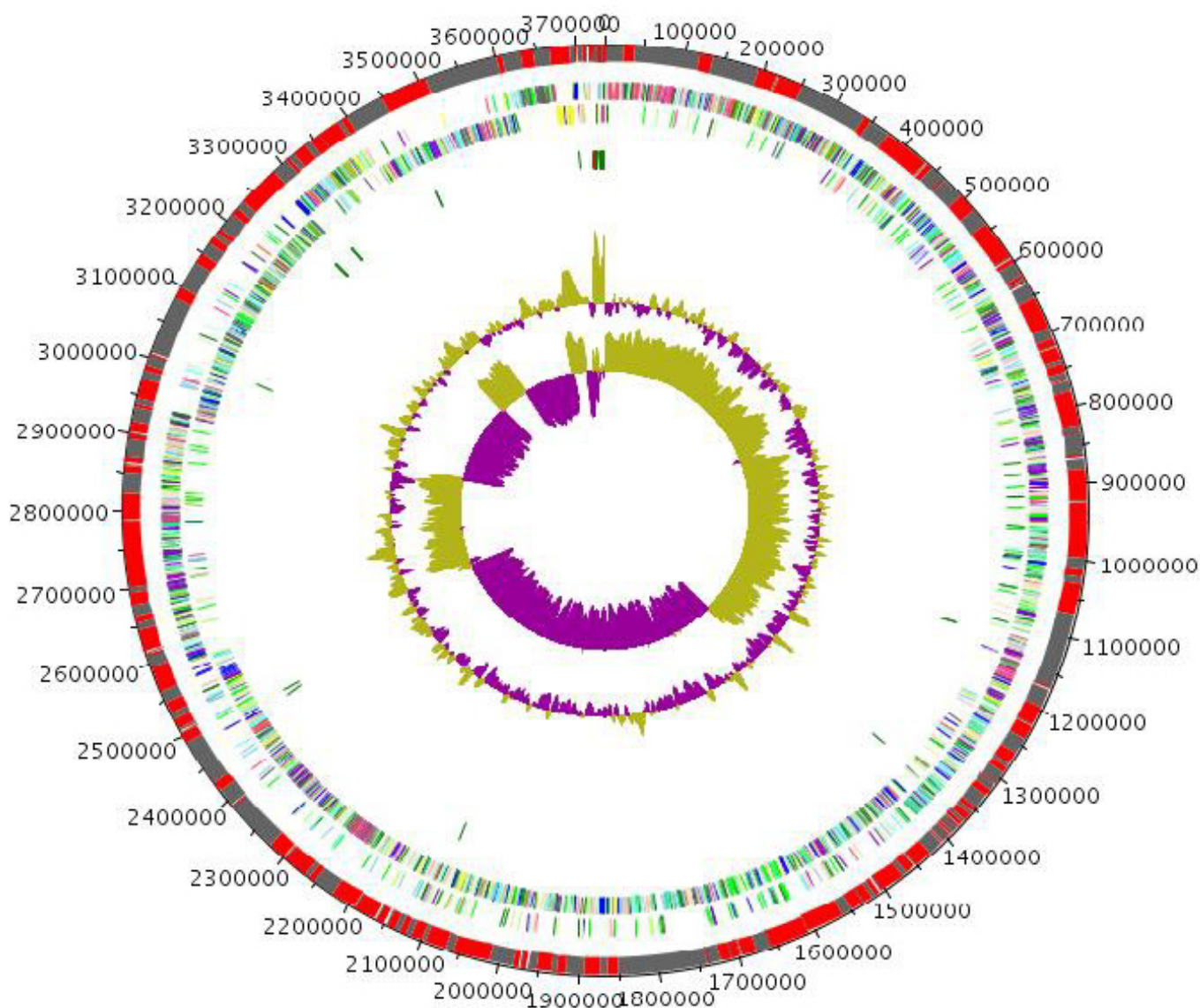


Figure 6. Graphical circular map of the chromosome. From the outside in, the outer two circles show open reading frames oriented in the forward and reverse directions (colored by COG categories), respectively. The third circle marks the rRNA gene operon (red) and tRNA genes (green). The fourth circle shows the G+C% content plot. The inner-most circle shows the GC skew, purple and olive indicating negative and positive values, respectively.

Comparison with the genomes from other *Clostridium* species

The genome sequence of *Clostridium* sp. is currently available for more than seventy-five *Clostridium* species. Here we compared the genome sequence of *C. dakarensis* strain FF1^T with those of *C. bartlettii*, *C. beijerinckii*, *C. cellulovorans*, *C. difficile*, *C. glycolicum*, *C. perfringens*, *C. saccharolyticum*, *C. senegalense*, and *C. thermocellum*.

The draft genome sequence of *C. dakarensis* strain FF1^T is smaller than those of *C. cellulovorans*, *C. beijerinckii*, *C. senegalense*, *C. saccharolyticum*, *C. thermocellum*, *C. difficile*, *C. glycolicum* (3.73, 5.26,

6.0, 3.89, 4.66, 3.84, 4.3 and 3.99 Mb, respectively) but larger than those of *C. perfringens* and *C. bartlettii* (3.26 and 2.97 Mb, respectively). The G+C content of *C. dakarensis* is lower than those of *C. cellulovorans*, *C. beijerinckii*, *C. perfringens*, *C. saccharolyticum*, *C. thermocellum*, *C. difficile* (31.2, 29.9, 28.4, 45, 39 and 29.1%, respectively) but higher than those of *C. bartlettii*, *C. glycolicum* and *C. senegalense* (28.8, 28 and 26.8%, respectively). The gene content of *C. dakarensis* is larger than those of *C. thermocellum*, *C. senegalense*, *C.*

perfringens, *C. glycolicum*, *C. bartlettii* (3,916, 3,173, 3,761, 2,876, 3,840 and 2,787, respectively) and smaller than those of *C. cellulovorans*, *C. beijerinckii*, *C. saccharolyticum* and *C. difficile*, (4,501, 5,243, 4,154 and 4,019, respectively). The ratio of genes per Mb of *C. dakarensis* is larger to those of *C. cellulovorans*, *C. beijerinckii*, *C. senegalense*, *C. saccharolyticum*, *C. thermocellum*, *C. difficile*, *C. bartlettii*, *C. glycolicum* and *C.*

perfringens (1,049, 856, 874, 966, 891, 826, 934, 938, 962 and 882, respectively).

The number of orthologous genes shared between *C. dakarensis* and other compared *Clostridium* species has been summarized in Table 6. The average percentage of nucleotide sequence identity ranged from 62.05 to 74.5% among previously published *Clostridium* species, and from 61.94 to 75.7% between *C. dakarensis* and other studied *Clostridium* species, thus confirming its new species status.

Table 5. Number of genes associated with the 25 general COG functional categories.

Code	Value	%age ^a	Description
J	171	4.45	Translation
A	0	0	RNA processing and modification
K	325	8.46	Transcription
L	158	4.11	Replication, recombination and repair
B	1	0.03	Chromatin structure and dynamics
D	34	0.88	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	111	2.89	Defense mechanisms
T	225	5.85	Signal transduction mechanisms
M	165	4.29	Cell wall/membrane biogenesis
N	58	1.51	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	45	1.17	Intracellular trafficking and secretion
O	95	2.47	Posttranslational modification, protein turnover, chaperones
C	194	5.04	Energy production and conversion
G	248	6.45	Carbohydrate transport and metabolism
E	248	6.45	Amino acid transport and metabolism
F	88	2.29	Nucleotide transport and metabolism
H	117	3.04	Coenzyme transport and metabolism
I	72	1.87	Lipid transport and metabolism
P	181	4.71	Inorganic ion transport and metabolism
Q	52	1.35	Secondary metabolites biosynthesis, transport and catabolism
R	386	10.04	General function prediction only
S	261	6.79	Function unknown
-	994	25.87	Not in COGs

^aThe total is based on the total number of protein coding genes in the annotated genome.

Table 6. Numbers of orthologous proteins shared between genomes (upper right)

	CDa	CC	CBe	CP	CSe	CSa	CT	CBa	CG	CDi
CDa	3,808	1,045	1,230	1,089	1,131	1,013	806	1,324	1,690	1,203
CC	68.22	4,254	1,490	1,163	1,181	1,057	967	871	1,038	1,021
CBe	68.84	70.36	5,020	1,300	1,289	1,207	968	989	1,204	1,129
CP	70.02	70.43	72.15	2,660	1,168	920	777	845	1,005	1,147
CSe	69.91	70.37	70.82	70.13	3,704	930	821	856	1,134	1,008
CSa	61.94	62.50	62.44	62.22	62.05	4,154	854	833	1,004	998
CT	64.49	64.84	64.56	64.78	64.53	63.83	3,173	713	840	952
CBa	74.98	68.22	68.84	69.46	69.52	62.15	64.73	2,787	1,517	1,303
CG	75.70	68.28	68.83	69.49	69.57	62.26	64.59	76.04	3,840	1,568
CDi	71.34	69.57	68.52	71.52	65.49	66.37	64.32	74.45	74.50	3,798

average percentage similarity of nucleotides corresponding to orthologous protein shared between genomes (lower left) and numbers of proteins per genome (bold). CDa: *C. dakarensis*; CC: *C. cellulovorans*; CBe: *C. beijerinckii*; CP: *C. perfringens*; CSe: *C. senegalensis*; CSa: *C. saccharolyticum*; CT: *C. thermocellum*; CBa: *C. bartlettii*; CG: *C. glycolicum*; CDi: *C. difficile*.

Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Clostridium dakarensis* sp. nov. which contains strain FF1^T. This bacterium strain has been isolated from the fecal flora of a 4-months-old Senegalese child suffering from gastroenteritis.

Description of *Clostridium senegalense* sp. nov.

Clostridium dakarensis (da.kar.e'n.se. L. gen. neutr. n. *dakarensis*, pertaining to, or originating from Dakar, the capital of Senegal, where the type strain was isolated).

Colonies were 1.5 mm in diameter on blood-enriched Columbia agar and Chocolate agar + PolyViteX. Cells are rod-shaped with a mean diameter of 1.2 µm. Optimal growth is achieved anaerobically. No growth is observed in aerobic conditions. Growth occurs between 25-37°C, with

optimal growth observed at 37°C, in medium 5% sheep blood-enriched Columbia agar. Cells stain Gram-positive, are endospore-forming, and motile. Catalase, oxidase, urease, indole and nitrate reduction activity are absent. Arginine dihydrolase, N-acetyl-β-glucosanimidase and pyroglutamic acid arylamidase activity are present. Cells are susceptible to amoxicillin, metronidazole, vancomycin, imipenem and rifampicin but resistant to trimethoprim/sulfamethoxazole.

The G+C content of the genome is 27.98%. The 16S rRNA gene sequence and whole-genome shotgun sequence of *C. dakarensis* strain FF1^T (= CSUR P243 = DSM 27086) are deposited in GenBank under accession numbers KC517358 and CBTZ00000000, respectively. The type strain FF1^T (= CSUR P243 = DSM 27086) was isolated from the fecal flora of a 4-months-old child in Dakar, Senegal.

Acknowledgements

The authors thank the Xegen Company (www.xegen.fr) for automating the genomic annotation process. This

study was funded by the Mediterranean-Infection Foundation.

References

1. Rossello-Mora R. DNA-DNA Reassociation Methods Applied to Microbial Taxonomy and Their Critical Evaluation. *In: Stackebrandt E (ed), Molecular Identification, Systematics, and Population Structure of Prokaryotes.* Springer, Berlin, 2006, p. 23-50.
2. Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* 2006; **33**:152-155.
3. Welker M, Moore ER. Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. *Syst Appl Microbiol* 2011; **34**:2-11. [PubMed](http://dx.doi.org/10.1016/j.syapm.2010.11.013) <http://dx.doi.org/10.1016/j.syapm.2010.11.013>
4. Tindall BJ, Rosselló-Móra R, Busse HJ, Ludwig W, Kämpfer P. Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol* 2010; **60**:249-266. [PubMed](http://dx.doi.org/10.1099/ijs.0.016949-0) <http://dx.doi.org/10.1099/ijs.0.016949-0>
5. Kokcha S, Mishra AK, Lagier JC, Million M, Leroy Q, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Bacillus timonensis* sp. nov. *Stand Genomic Sci* 2012; **6**:346-355. [PubMed](http://dx.doi.org/10.4056/sigs.2776064) <http://dx.doi.org/10.4056/sigs.2776064>
6. Lagier JC, El Karkouri K, Nguyen TT, Armougom F, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Anaerococcus senegalensis* sp. nov. *Stand Genomic Sci* 2012; **6**:116-125. [PubMed](http://dx.doi.org/10.4056/sigs.2415480) <http://dx.doi.org/10.4056/sigs.2415480>
7. Mishra AK, Gimenez G, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Alistipes senegalensis* sp. nov. *Stand Genomic Sci* 2012; **6**:304-314. <http://dx.doi.org/10.4056/sigs.2625821>
8. Lagier JC, Armougom F, Mishra AK, Nguyen TT, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Alistipes timonensis* sp. nov. *Stand Genomic Sci* 2012; **6**:315-324. [PubMed](http://dx.doi.org/10.4056/sigs.2415480)
9. Mishra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Clostridium senegalense* sp. nov. *Stand Genomic Sci* 2012; **6**:386-395. [PubMed](http://dx.doi.org/10.4056/sigs.2415480)
10. Mishra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Peptoniphilus timonensis* sp. nov. *Stand Genomic Sci* 2012; **7**:1-11. [PubMed](http://dx.doi.org/10.4056/sigs.2956294) <http://dx.doi.org/10.4056/sigs.2956294>
11. Mishra AK, Lagier JC, Rivet R, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Paenibacillus senegalensis* sp. nov. *Stand Genomic Sci* 2012; **7**:70-81. [PubMed](http://dx.doi.org/10.4056/sigs.2956294)
12. Prazmowski A. "Untersuchung über die Entwicklungsgeschichte und Fermentwirkung einiger Bakterien-Arten." Ph.D. Dissertation, University of Leipzig, Germany, 1880, p. 366-371.
13. List of prokaryotic names with standing with nomenclature. <http://www.bacterio.net>
14. Evaluations and Standards Laboratory. "Identification of *Clostridium* species", 2008 pp. 14.
15. Wells CL, Wilkins TD. (1996). *Clostridia: Spore forming Anaerobic Bacilli* In: *Baron's Medical Microbiology* (Baron S et al., eds.) (4th ed.). University of Texas Medical Branch.
16. Keis S, Shaheen R, David T. Jones Emended descriptions of *Clostridium acetobutylicum* and *Clostridium beijerinckii*, and descriptions of *Clostridium saccharoperbutylacetonicum* sp. nov. and *Clostridium saccharobutylicum* sp. nov. *Int J Syst Evol Microbiol* 2001; **51**:2095-2103. [PubMed](http://dx.doi.org/10.1099/00207713-51-6-2095) <http://dx.doi.org/10.1099/00207713-51-6-2095>
17. Carnahan JE, Castle JE. Nitrogen Fixation. *Annu Rev Plant Physiol* 1963; **14**:125-136. <http://dx.doi.org/10.1146/annurev.pp.14.060163.001013>
18. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. [PubMed](http://dx.doi.org/10.1038/nbt1360) <http://dx.doi.org/10.1038/nbt1360>
19. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eukarya. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. [PubMed](http://dx.doi.org/10.1073/pnas.87.12.4576) <http://dx.doi.org/10.1073/pnas.87.12.4576>
20. Gibbons NE, Murray RGE. Proposals Concerning the Higher Taxa of Bacteria. *Int J Syst Bacteriol*

- 1978; **28**:1-6. <http://dx.doi.org/10.1099/00207713-28-1-1>
21. Garrity GM, Holt JG. The Road Map to the Manual. In: Garrity GM, Boone DR, Castenholz RW (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 1, Springer, New York, 2001, p. 119-169.
 22. Murray RGE. The Higher Taxa, or, a Place for Everything...? In: Holt JG (ed), *Bergey's Manual of Systematic Bacteriology*, First Edition, Volume 1, The Williams and Wilkins Co., Baltimore, 1984, p. 31-34.
 23. List of new names and new combinations previously effectively, but not validly, published. List no. 132. *Int J Syst Evol Microbiol* 2010; **60**:469-472. <http://dx.doi.org/10.1099/ijs.0.022855-0>
 24. Rainey FA. Class II. *Clostridia* class nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 3, Springer-Verlag, New York, 2009, p. 736.
 25. Skerman VBD, Sneath PHA. Approved list of bacterial names. *Int J Syst Bacteriol* 1980; **30**:225-420. <http://dx.doi.org/10.1099/00207713-30-1-225>
 26. Prévot AR. In: Hauderoy P, Ehringer G, Guillot G, Magrou J., Prévot AR, Rosset D, Urbain A (eds), *Dictionnaire des Bactéries Pathogènes*, Second Edition, Masson et Cie, Paris, 1953, p. 1-692.
 27. Pribram E. Klassifikation der Schizomyceten. Klassifikation der Schizomyceten (Bakterien), Franz Deuticke, Leipzig, 1933, p. 1-143.
 28. Smith LDS, Hobbs G. Genus III. *Clostridium* Prazmowski 1880, 23. In: Buchanan RE, Gibbons NE (eds), *Bergey's Manual of Determinative Bacteriology*, Eighth Edition, The Williams and Wilkins Co., Baltimore, 1974, p. 551-572.
 29. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29. [PubMed](http://dx.doi.org/10.1038/75556) <http://dx.doi.org/10.1038/75556>
 30. Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* 2006; **33**:152-155.
 31. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, Raoult D. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin Infect Dis* 2009; **49**:543-551. [PubMed](http://dx.doi.org/10.1086/600885) <http://dx.doi.org/10.1086/600885>
 32. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. [PubMed](http://dx.doi.org/10.1038/nbt1360) <http://dx.doi.org/10.1038/nbt1360>
 33. Prodigal. <http://prodigal.ornl.gov>
 34. GenBank database. <http://www.ncbi.nlm.nih.gov/genbank>
 35. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997; **25**:955-964. [PubMed](http://dx.doi.org/10.1093/nar/gkm160)
 36. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007; **35**:3100-3108. [PubMed](http://dx.doi.org/10.1093/nar/gkm160) <http://dx.doi.org/10.1093/nar/gkm160>
 37. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004; **340**:783-795. [PubMed](http://dx.doi.org/10.1016/j.jmb.2004.05.028) <http://dx.doi.org/10.1016/j.jmb.2004.05.028>
 38. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 2001; **305**:567-580. [PubMed](http://dx.doi.org/10.1006/jmbi.2000.4315) <http://dx.doi.org/10.1006/jmbi.2000.4315>
 39. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. Artemis: sequence visualization and annotation. *Bioinformatics* 2000; **16**:944-945. [PubMed](http://dx.doi.org/10.1093/bioinformatics/16.10.944) <http://dx.doi.org/10.1093/bioinformatics/16.10.944>
 40. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 2009; **25**:119-120. [PubMed](http://dx.doi.org/10.1093/bioinformatics/btn578) <http://dx.doi.org/10.1093/bioinformatics/btn578>
 41. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004; **14**:1394-1403. [PubMed](http://dx.doi.org/10.1101/gr.2289704) <http://dx.doi.org/10.1101/gr.2289704>
 42. Lechner M, Findeib S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics* 2011; **12**:124. [PubMed](http://dx.doi.org/10.1186/1471-2105-12-124) <http://dx.doi.org/10.1186/1471-2105-12-124>