Rubeoparvulum massiliense gen. nov., sp. nov., a new bacterial genus isolated from the human gut of a Senegalese infant with severe acute malnutrition

M. Tidjani Alou[1], J. Rathored[1], J.-C. Lagier[1,2,3], S. Khelaifia[1], C. Michelle[1], C. Sokhna[2], A. Diallo[2], A. B. Diallo[4], P.-E. Fournier[1], D. Raoult[1,5] and S. Edouard[1]

1) *Aix-Marseille Université, URMITE, UM63, CNRS7278, IRD198, INSERM 1095, Faculté de médecine,* 2) *Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes IRD 198, CNRS 7278, Aix-Marseille Université, Marseille, France,* 3) *Campus Commun UCAD-IRD of Hann, Dakar, Senegal,* 4) *Laboratoire de microbiologie, département de biologie, Université Abdou Moumouni de Niamey, Niamey, Niger and* 5) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Rubeoparvulum massiliense strain mt6^T was isolated from the gut microbiota of a severely malnourished boy from Senegal and consisted of facultative anaerobic, spore-forming, nonmotile and Gram-negative rods. *R. massiliense* showed a 92% similarity with the 16S rRNA of *Bacillus mannanilyticus.* The genome of strain mt6^T is 2 843 796 bp long with a 43.75% G+C content. It contains 2735 protein-coding genes and 76 RNA genes, among which are nine rRNA genes.

Keywords: Culturomics, genome, gut microbiota, *Rubeoparvulum massiliense*, taxonogenomics
Original Submission: 18 October 2016; **Accepted:** 9 November 2016
Article published online: 17 November 2016

Corresponding author: S. Edouard, Aix-Marseille Université, URMITE, UM63, CNRS7278, IRD198, INSERM 1095, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France
E-mail: sophie.edouard@univ-amu.fr

Introduction

The human microbiome is defined as the sum of all microbes colonizing the human body [1]. The gut microbiota is one of the largest microbial ecosystems of the human body, consisting of 10^{14} microbial cells with a microbiome 150 times larger than the human genome [2]. The gastrointestinal microbiota colonization starts before birth with the maternal microbiota, and its early composition is influenced by the mode of birth. Its composition matures rapidly for the first year and reaches adult form by 3 years [2,3]. A disruption of its equilibrium has been proven to be implicated in a growing number of pathologies such as inflammatory bowel disease, irritable bowel syndrome, obesity [3,4] and severe acute malnutrition [5–7].

A new cultural approach, microbial culturomics, based on the multiplication of culture conditions with a variation of temperature, media and atmosphere, was developed in our laboratory in order to explore as exhaustively as possible a microbial ecosystem [8,9]. Using this new approach, we isolated a new member of the *Bacillaceae* family. At this time, 52 validated genera are part of the *Bacillaceae* family, which was created in 1895 by Fisher; *Bacillus* is its type genus, described by Cohn in 1872 [10]. Most species of this family are found in the environment (soil, water and plants) and are opportunistic pathogens in humans, except *Bacillus anthracis,* which is well known as being highly pathogenic. The *Bacillaceae* family includes Gram-positive, rod-shaped, mostly aerobic and facultative anaerobic genera [11]. By adding the description of the assembled and annotated genome of the species and the proteomic description of the strain with the matrix-assisted desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) profile to the classical description principles (phylogenetic relationships based on the 16S rRNA sequence,

http://dx.doi.org/10.1016/j.nmni.2016.11.003

phenotypic and genotypic characteristics), a new concept of description called taxonogenomics was developed in our laboratory [12].

Here we describe the genus *Rubeoparvulum,* the type species of which is *Rubeoparvulum massiliense* strain mt6 (= CSUR P1473 = DSM 100479) from a stool sample collected in a 2-month-old infant living in Senegal and presenting with kwashiorkor, a type of severe acute malnutrition.

Materials and Methods

Ethics and sample collection

The strain mt6 was isolated from a stool taken from a severely malnourished 2-month-old boy with a height-for-age score of −5.87 who had nutritional edoema. Collection was performed in Senegal in April 2014. This sampling was undertaken as part of an exploratory study of the human gut microbiota in African children with malnutrition. The study was approved by the local IFR 48 ethics committee under agreement 09-022. The boy's parents provided informed oral consent. The sample was stored at −80°C after collection.

Strain identification by MALDI-TOF MS and 16S rRNA sequencing

In order to explore as exhaustively as possible the bacterial diversity of the faecal sample, the culturomics concept was used to culture this sample using 18 culture conditions [8]. The purified colonies obtained were identified using MALDI-TOF MS as described previously [13,14]. Colonies were deposited on a MTP 96 MALDI-TOF MS target plate (Bruker Daltonics, Leipzig, Germany), which was analysed with a Microflex spectrometer (Bruker Daltonics). The spectra obtained were matched against the references of the 7567 bacteria contained in the database by standard pattern matching (with default parameter settings) with MALDI BioTyper database software 2.0 (Bruker Daltonics). An identification score over 1.9 with a validated species allowed the identification at the species level, and a score under 1.7 did not enable any identification. The 16S rRNA gene was amplified and sequenced as previously described [15]. The obtained 16S rRNA sequence was compared to those in GenBank (http://blast.ncbi.nlm.nih.gov.gate1.inist.fr/Blast.cgi) to determine the percentage of sequence similarity with the closest bacteria. A new species or genus was defined by a similarity level of the 16S rRNA sequence under 98.65% or 95% respectively [16].

Growth conditions

The ideal growth conditions of strain mt6^T were determined by testing different culture conditions. Five growth temperatures (25, 30, 37, 45 and 56°C) were tested under anaerobic and microaerophilic atmospheres using GENbag anaer and GENbag microer systems respectively (bioMérieux, Marcy l'Étoile, France). Aerobic growth was tested with and without 5% CO_2. Growth was also tested at various pHs (6, 6.5, 7, 7.5, 8 and 8.5) using a pH-adjusted Colombia agar (bioMérieux). Salt tolerance was also tested with 0.5, 1, 5, 7.5 and 10% (w/v) NaCl.

Morphologic, biochemical and antibiotic susceptibility tests

Phenotypic characteristics (Gram staining, sporulation, motility) were determined as previously described [8]. The catalase (bioMérieux) and oxidase (Becton Dickinson, Le Pont de Claix, France) activities were also tested. Cell morphology was observed after negative staining of bacteria using a Tecnai G20 transmission electron microscope (FEI Company, Limeil-Brevannes, France). The biochemical features of strain mt6^T were investigated with API 50CH, API ZYM and API 20A strips (bioMérieux) according to the manufacturer's instructions. Cellular fatty acid methyl ester (FAME) analysis was performed by gas chromatography mass spectrometry (GC/MS). Strain mt6 was grown on 5% sheep's blood–enriched Colombia agar (bioMérieux) for the fatty acid analysis, which was carried out by GC/MS. Approximately 67 mg of bacterial biomass was each collected from several culture plates. Cellular FAMEs were prepared as described by Sasser (http://www.midi-inc.com/pdf/MIS_Technote_101.pdf).

Briefly, GC/MS analyses were realized by a Clarus 500 gas chromatograph equipped with a SQ8S MS detector (Perkin Elmer, Courtaboeuf, France). A total of 2 µL of FAME extracts were volatized at 250°C (split 20 mL/min) in a Focus liner with wool and separated on an Elite-5MS column (30 m, 0.25 mm i.d., 0.25 mm film thickness) using a linear temperature gradient (70 to 290°C at 6°C/min), allowing the detection of C4 to C24 fatty acid methyl esters. Helium flowing at 1.2 mL/min was used as carrier gas. The MS inlet line was set at 250°C and EI source at 200°C. Full scan monitoring was performed from 45 to 500 *m/z*. All data were collected and processed using Turbomass 6.1 (Perkin Elmer). FAMEs were identified by a spectral database search using MS Search 2.0 operated with the Standard Reference Database 1A (National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA) and the FAMEs mass spectral database (Wiley, Chichester, UK). Retention time correlations with estimated nonpolar retention indexes from the NIST database were obtained using a 37-component FAME mix (Supelco; Sigma-Aldrich, Saint-Quentin Fallavier, France); FAME identifications were confirmed using this index. Antibiotic susceptibility testing was performed using a disk

diffusion method according to European Committee on Antimicrobial Susceptibility Testing (EUCAST) 2015 recommendations [17]. Inhibition diameters were measured using the Scan1200 scanner (Interscience, Saint-Nom-La Bretêche, France).

Genomic DNA (gDNA) preparation

For gDNA preparation, *R. massiliense* strain mt6^T was cultured on 5% sheep's blood–enriched Columbia agar (bioMérieux) at 37°C aerobically. Bacteria grown on three petri dishes were resuspended in 4 × 100 µL of Tris-EDTA (TE) buffer. Then 200 µL of this suspension was diluted in 1 mL TE buffer for lysis treatment, which included a 30-minute incubation with 2.5 µg/µL lysozyme at 37°C, followed by an overnight incubation with 20 µg/µL proteinase K at 37°C. Extracted DNA was then purified using three successive phenol–chloroform extractions and ethanol precipitations at −20°C overnight. After centrifugation, the DNA was resuspended in 160 µL TE buffer.

Genome sequencing and assembly

Using the mate-pair strategy, the gDNA of *R. massiliense* strain mt6^T was sequenced on the MiSeq sequencer (Illumina, San Diego, CA, USA). The gDNA was barcoded in order to be mixed with 11 other projects with the Nextera Mate-Pair sample prep kit (Illumina). The mate-pair library was prepared with 1 µg of gDNA using the Nextera Mate-Pair Illumina guide, and the gDNA sample was simultaneously fragmented and tagged with a mate-pair junction adapter. The pattern of the fragmentation 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 to 10 kb, with an optimal size at 4.08 kb. No size selection was performed, and only 464 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal at 569 bp on the Covaris S2 device in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies), and the final library concentration was measured at 24.4 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. Automated cluster generation and sequencing run were performed in a single 39-hour run at 2 × 251 bp. Total information of 10.1 Gb was obtained from a 1189K/mm^2 cluster density with a cluster passing quality control filters of 99.1% (22 579 000 clusters). The reads obtained were trimmed; assembly was performed by CLC genomicsWB4 software.

Genome annotation and comparison

Open reading frames (ORFs) were predicted using Prodigal [18] 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 [19] and the Clusters of Orthologous Groups (COGs) databases using BLASTP (*E* value 1e-03, coverage 0.7 and identity percentage 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 [20] was used to find tRNA genes, while ribosomal RNAs were found using RNAmmer [21]. Lipoprotein signal peptides and the number of transmembrane helices were predicted using Phobius [22]. Mobile genetic elements were predicted using PHAST [23] and RAST [24]. 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 [25] and DNA Plotter [26] were used for data management and the visualization of genomic features respectively. The Mauve alignment tool (version 2.3.1) was used for multiple genomic sequence alignment [27].

Comparator species for genomic comparison were identified in the 16S RNA tree using Phylopattern software [28]. The genome of strain mt6^T was compared to those of *Alkaliphilus metalliredigens* strain QYMF, *Clostridium aceticum* strain DSM 1496, *Alkaliphilus transvaalensis* strain SAGM1 and *Alkaliphilus oremlandii* strain OhILAs.

For each selected genome, the complete genome sequence, proteome genome sequence and Orfeome genome sequence were retrieved from the FTP of NCBI. 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 annotation). Annotation and comparison processes were performed in the multiagent software system DAGOBAH [29], which includes Figenix [30] libraries that provide pipeline analysis. To evaluate the genomic similarity between studied genomes, we determined two parameters, digital DNA-DNA hybridization (DH), which exhibits a high correlation with DDH [31,32] and average genomic identity of orthologous gene sequences (AGIOS) [33], which was designed to be independent from DDH [33]. The AGIOS score is the mean value of nucleotide similarity between all couples of orthologous proteins between the two studied genomes [33].

Results

Strain identification and phylogenetic analyses

The mt6 strain was isolated after a 30-day preincubation at 37°C in an anaerobic blood culture bottle supplemented with 5 mL of rumen fluid filter-sterilized through a 0.2 µm pore filter (Thermo Fisher Scientific, Villebon sur Yvette, France). Strain mt6^T was subcultured on 5% blood-enriched Colombia agar (bioMérieux) in an anaerobic atmosphere at 37°C. The bacterium could not be identified by MALDI-TOF MS (score under 1.7), but sequencing of the 16S rRNA revealed that strain mt6^T's nucleotide sequence had a 92% similarity level with *Bacillus mannanilyticus,* the phylogenetically closest species with a validly published name (Fig. 1). According to Kim *et al.* [16], a new genus can be defined by a similarity level threshold lower than 95%. Consequently, strain mt6^T was classified as a new genus called *Rubeoparvulum,* its type species being *Rubeoparvulum massiliense* (Table 1). The 16S rRNA sequence of strain mt6 was deposited in GenBank under accession number LN828926.

Phenotypic description

The growth of the mt6 stains occurred between 25 and 56°C on 5% sheep's blood–enriched Colombia agar. Optimal growth was achieved at 37°C after 48 hours of incubation in both

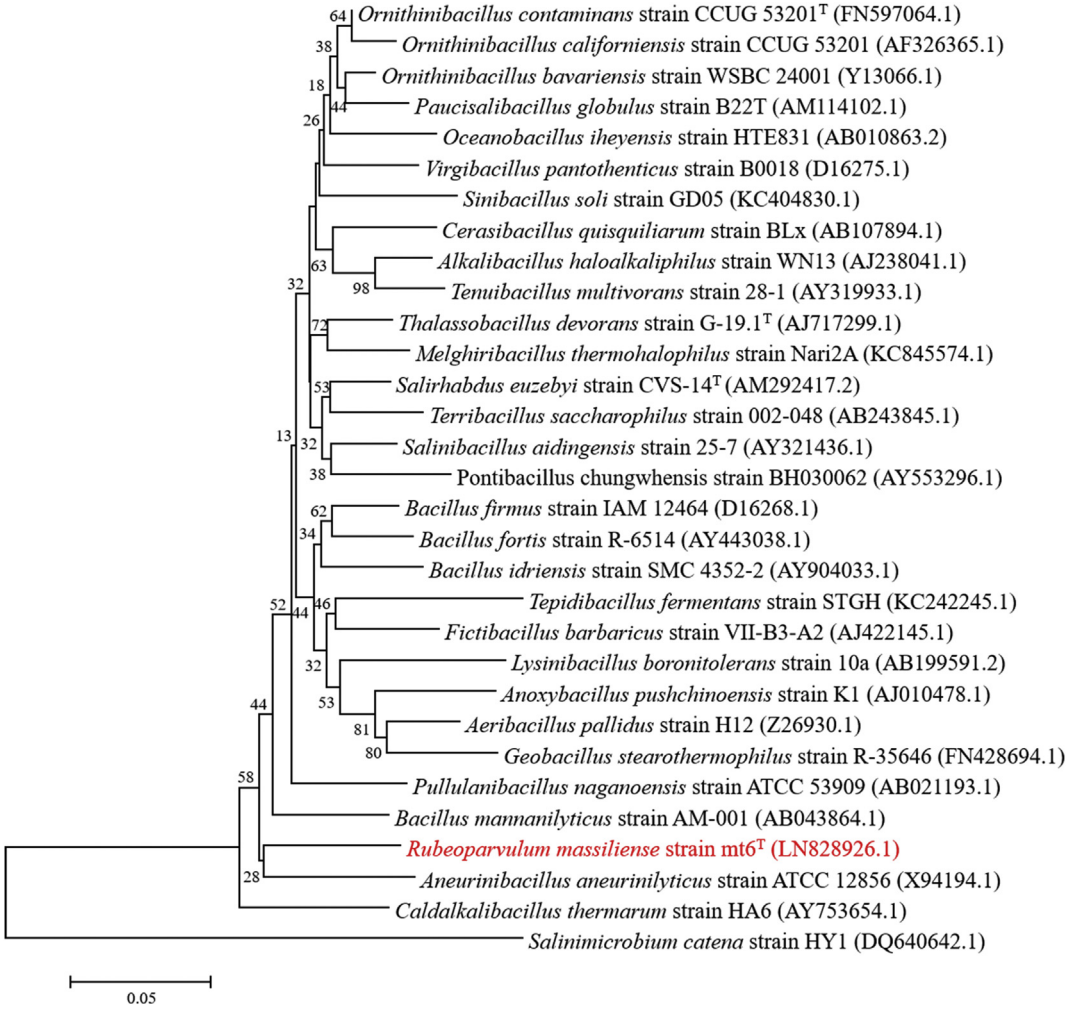

FIG. 1. Phylogenetic tree highlighting position of *Rubeoparvulum massiliense* strain mt6^T relative to other close strains. Respective GenBank accession numbers for 16S rRNA genes are indicated in parentheses. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using maximum-likelihood method within MEGA6 software. Numbers at nodes are percentages of bootstrap values obtained by repeating analysis 1000 times to generate majority consensus tree. *Salinimicrobium catena* strain HY1 was used as outgroup. Scale bar represents 5% nucleotide sequence divergence.

TABLE 1. Classification and general features of *Rubeoparvulum massiliense* strain mt6^T

Property	Term
Current classification	Domain: *Bacteria* Phylum: *Firmicutes* Class: *Bacilli* Order: *Bacillales* Family: *Bacillaceae* Genus: *Rubeoparvulum* Species: *Rubeoparvulum massiliense* Type strain: mt6
Gram stain	Negative
Cell shape	Rod
Motility	Motile
Sporulation	Sporulating
Temperature range	Mesophilic
Optimum temperature	37°C

anaerobic and microaerophilic conditions. The cell growth was weaker in aerobic conditions. Strain mt6 was able to grow at pH values ranging from 6 to 8.5 and 0.5 to 5% NaCl concentrations. Cells were spore forming and motile, and they formed translucent colonies with a mean diameter of 0.5 mm on blood-enriched Colombia agar. Microscopic observations showed Gram-stain-negative, rod-shaped cells (Fig. 2), and electron microscopy showed rods with a mean diameter of 1 µm and a mean length of 6.8 µm (Fig. 3). Our MALDI-TOF MS database was incremented with the reference spectrum obtained for strain mt6^T (Fig. 4). Comparisons of the aforementioned spectrum to that of other known species of the *Bacillaceae* family are represented in the gel view (Fig. 5). Reference spectra are available in our online database (http://www.mediterranee-infection.com/article.php?laref=256&titre=urms-database).

Strain mt6^T was negative for catalase activity and positive for oxidase activity. Using an API ZYM strip, positive reactions were recorded for alkaline phosphatase, esterase (C4), esterase lipase (C8), leucine arylamidase, α-chymotrypsin, acid phosphatase and naphtol-AS-BI-phosphohydrolase. Nitrate

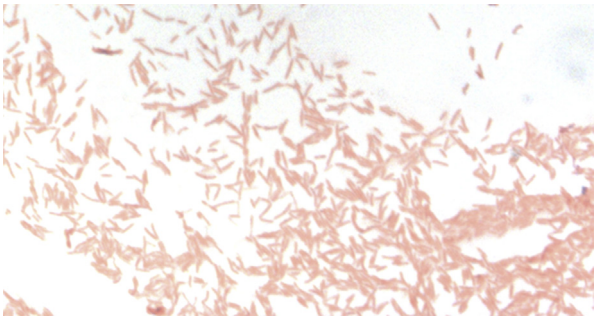

FIG. 2. Gram staining of *Rubeoparvulum massiliense* strain mt6^T (= CSUR P1473 = DSM 100479).

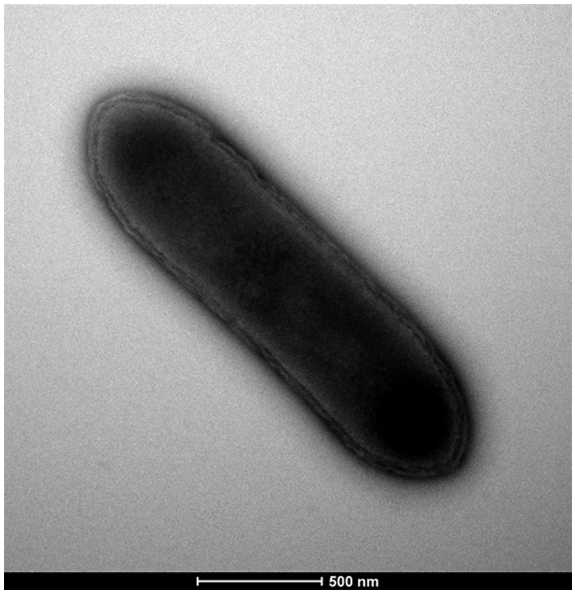

FIG. 3. Transmission electron microscopy of *Rubeoparvulum massiliense* strain mt6^T (= CSUR P1473 = DSM 100479) using Tecnai G20 transmission electron microscope (FEI Company) at operating voltage of 60 kV. Scale bar = 500 nm.

reduction was observed; urease, β-glucosidase and protease activities were positive using an API 20 NE strip. All other reactions were negative on both strips. An API 50CH was used to test the carbohydrates metabolization. The following carbohydrates were metabolized by strain mt6^T: glycerol, D-lactose, D-fucose, D-mannose, D-cellobiose, D-maltose, salicin, D-arabitol, *N*-acetyl-glucosamine and potassium-5-ketogluconate. Amygdalin, arbutin, D-fructose, inulin, D-sucrose, D-raffinose, erythritol, D-arabinose, L-arabinose, D-ribose, D-xylose, L-xylose, D-adonitol, methyl-βD-xylopyranoside, D-glucose, D-galactose, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl-αD-mannopyranoside, methyl-αD-glucopyranoside, esculin ferric citrate, D-melibiose, D-trehalose, D-melezitose, starch, glycogen, xylitol, gentiobiose, D-turanose, D-lyxose, D-tagatose, L-fucose, L-arabitol, potassium gluconate and potassium 2-ketogluconate showed negative reactions. Differences between the biochemical features of close members of the family *Bacillaceae* are listed in Table 2.

The major cellular fatty acids found for strain mt6 were 12-methyl-tetradecanoic acid (25%), 10-methyldodecanoic acid (15%) and hexadecanoic acid (12%). This strain was composed of many branched structures (iso/anteiso). A specific 3-OH fatty acid was also described (<1%, Table 3).

Among tested antibiotics, cells were susceptible to amoxicillin, gentamicin, ceftriaxone, ciprofloxacin, penicillin,

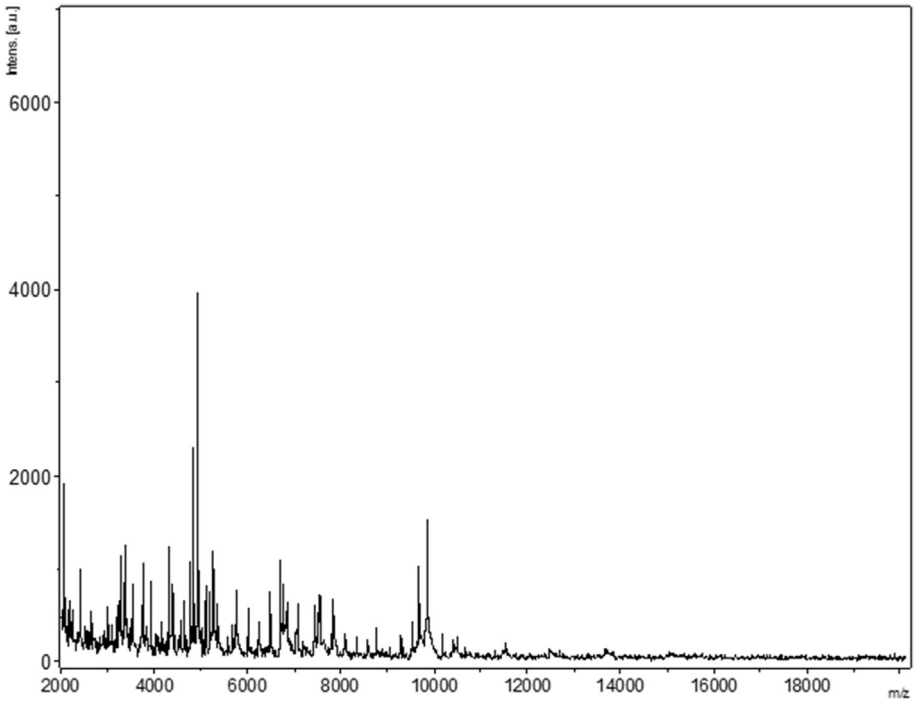

FIG. 4. Reference mass spectrum from *Rubeoparvulum massiliense* strain mt6^T (= CSUR P1473 = DSM 100479). Spectra from 12 individual colonies were compared and reference spectrum generated.

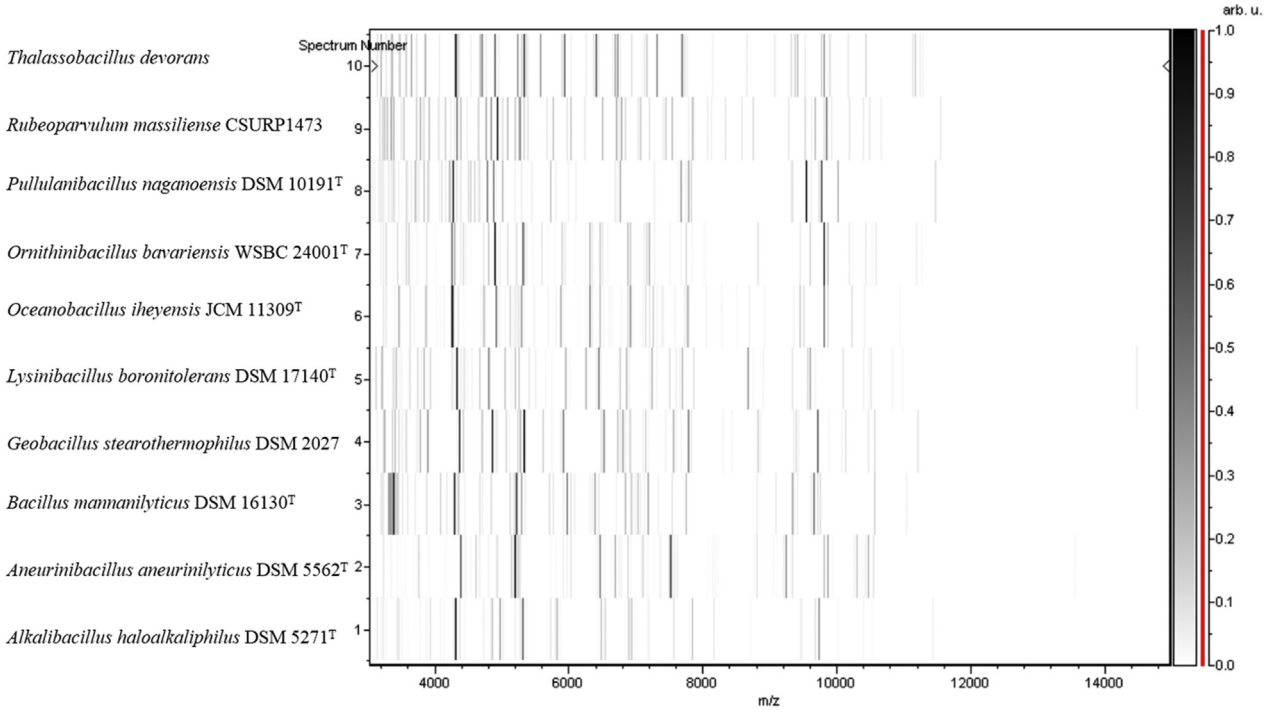

FIG. 5. Gel view comparing *Rubeoparvulum massiliense* strain mt6^T (= CSUR P1473 = DSM 100479) to other species within *Bacillaceae* family. Gel view displays raw spectra of loaded spectrum files arranged in pseudo-gel-like look. *x*-axis records *m*/*z* value. Left *y*-axis displays running spectrum number originating from subsequent spectra loading. Peak intensity is expressed by greyscale scheme code. Colour bar and right *y*-axis indicate relation between colour peak is displayed, with peak intensity in arbitrary units. Displayed species are indicated at left.

TABLE 2. Differential characteristics of *Rubeoparvulum massiliense* strain mt6^T CSUR P1473 = DSM 100479, *Bacillus mannanilyticus* strain AM-001^T DSM 16130^T, *Tepidibacillus fermentans* strain STGH^T DSM 23802^T, *Pullulanibacillus uraniitolerans* strain UG-2^T DSM 19429^T, *Alkalibacillus haloalkaliphilus* DSM 5271^T, *Tenuibacillus halotolerans* strain YIM 94025^T KCTC 33046^T, *Thalassobacillus devorans* strain G-19.1^T DSM 16966^T, *Salinibacillus aidingensis* strain 25-7^T JCM 12389^T, *Salinibacillus kushneri* strain 8-2^T JCM 12390^T, *Ornithinibacillus bavariensis* strain WSBC 24001^T DSM 15681^T [34–41]

Property	*Rubeoparvulum massiliense*	*Bacillus mannanilyticus*	*Tepidibacillus fermentans*	*Pullulanibacillus uraniitolerans*	*Alkalibacillus haloalkaliphilus*	*Tenuibacillus halotolerans*	*Thalassobacillus devorans*	*Salinibacillus kushneri*	*Salinibacillus aidingensis*	*Ornithinibacillus bavariensis*
Cell diameter (μm)	1.0	0.6–0.8	0.3	1.0	0.3–0.5	0.2–0.3	1.0–1.2	0.4–0.6	0.3–0.5	0.4
Oxygen requirement	+/–	+	–	+	+	+	+	+	+	+
Gram stain	–	+/–	+	+	–	+	+	+	+	+
Salt requirement	–	–	+	+	+	+	+	+	+	+
Indole	–	–	NA	NA	NA	–	–	NA	NA	–
Production of:										
Catalase	–	+	–	+	+	+	+	+	+	+
Oxidase	+	–	+	–	–	+	–	+/–	+/–	+
Nitrate reductase	+	–	+	–	–	–	+	–	–	–
Urease	+	NA	NA	NA	–	NA	–	–	–	–
Acid from:										
L-Arabinose	–	–	–	+	NA	–	–	+	+	–
Ribose	–	NA	NA	+	+	NA	–	NA	NA	+
Mannose	+	+	NA	+	+	NA	+	+	+	–
Mannitol	–	+	NA	NA	–	–	+	+	+	–
Sucrose	–	+	–	+	NA	+	+	+	+	–
D-Glucose	–	+	+	+	–	+	+	+	+	+
D-Fructose	–	+	+	+	+	+	+	+	+	–
D-Maltose	+	+	+	+	–	–	+	+	+	–
D-Lactose	+	–	–	+	NA	–	–	+	+	–
Habitat	Human stool	Industry	Gas storage	Mill tailing effluent	Salt lake	Salt lake	Hypersaline environments	Neutral saline lake	Neutral saline lake	Pasteurized milk

+, positive result; –, negative result; NA, data not available.

TABLE 3. Cellular fatty acid composition (%) of *Rubeoparvulum massiliense* strain mt6^T

Fatty acid	Name	Mean relative %[a]
15:0 anteiso	12-Methyl-tetradecanoic acid	25.2 ± 0.3
13:0 anteiso	10-Methyl-dodecanoic acid	15.2 ± 0.2
16:0	Hexadecanoic acid	12.0 ± 0.6
18:1n9	9-Octadecenoic acid	9.3 ± 0.4
13:0 iso	11-Methyl-dodecanoic acid	7.1 ± 0.1
18:0	Octadecanoic acid	6.3 ± 0.1
15:0 iso	13-Methyl-tetradecanoic acid	5.8 ± 0.1
14:0 iso	12-Methyl-tridecanoic acid	5.6 ± 0.3
18:2n6	9,12-Octadecadienoic acid	5.3 ± 0.2
5:0 iso	3-Methyl-butanoic acid	1.7 ± 0.1
14:0	Tetradecanoic acid	1.6 ± 0.1
11:0 anteiso	8-Methyl-decanoic acid	1.1 ± 0.1
18:1n6	12-Octadecenoic acid	TR
15:0	Pentadecanoic acid	TR
12:0 iso	10-Methyl-undecanoic acid	TR
17:0	Heptadecanoic acid	TR
18:1n7	11-Octadecenoic acid	TR
13:0	Tridecanoic acid	TR
15:0 3-OH anteiso	3-Hydroxy-12-methyl-tetradecanoic acid	TR
11:0 iso	9-Methyl-decanoic acid	TR
16:0 iso	14-Methyl-pentadecanoic acid	TR
10:0	Decanoic acid	TR
16:1n7	9-Hexadecenoic acid	TR
20:4n6	5,8,11,14-Eicosatetraenoic acid	TR
17:0 anteiso	14-Methyl-hexadecanoic acid	TR
12:0	Dodecanoic acid	TR
17:0 iso	15-Methyl-hexadecanoic acid	TR

TR, trace amounts <1%.
[a]Mean peak area percentage.

imipenem, tobramycin and oxacillin but were resistant to metronidazole, trimethoprim/sulfamethoxazole, rifampicin, doxycycline, vancomycin, nitrofurantoin and erythromycin.

Genome properties

With an estimated size of 2 843 796 bp, the *R. massiliense* genome had a G+C content of 43.75% (Table 4, Fig. 6). It was composed of six scaffolds composed of six contigs. Out of 2811 predicted genes, 2735 were protein-coding genes, and 76 were RNAs (seven 5S rRNA, one 16S rRNA gene, one 23S rRNA gene, 67 tRNA genes). A putative function was assigned to 1873 genes (66.63%) by COGs or NR blast. A total of 233 genes (8.28%) were identified as ORFans. The remaining 402 genes (14.70%) were annotated as hypothetical proteins. Table 4 shows the statistics of the genome, while Table 5 presents the distribution of genes into COGs functional categories.

TABLE 4. Nucleotide content and gene count levels of genome

	Genome (total)	
Attribute	**Value**	**% of total[a]**
Size (bp)	2 843 796	100
G+C content (%)	1 244 365	43.75
Coding region (bp)	2 498 460	87.85
Total genes	2811	100
RNA genes	76	2.70
Protein-coding genes	2735	97.29
Genes with function prediction	1873	66.63
Genes assigned to COGs	862	30.66
Genes with peptide signals	319	11.34
No. of pseudogenes	13	0.46
CRISPR repeats	0	0
No. genes with Pfam-A domains	2557	90.96
ORFans genes	233	8.28
Genes associated with PKS or NRPS	12	0.42
No. of antibiotic resistance genes	0	0

COGs, Clusters of Orthologous Groups database.
[a]Total is based on either size of genome in base pairs or total number of protein-coding genes in annotated genome.

Genome comparison

The genome of strain mt6^T was compared to those of closely related species (Table 6) by comparing their main genomic characteristics (size, G+C content, protein-coding genes, total number of genes). The genome size of strain mt6^T (2.84 Mb) is smaller than *B. agri* (5.51 Mb), *B. borstelensis* (5.16 Mb), *B. mannanilyticus* (4.53 Mb), *B. thermoruber* (4.43 Mb) and *C. thermarum* (2.9Mb). Strain mt6^T had a higher G+C content (43.75%) than *B. mannanilyticus* (39.6%) but lower than *B. thermoruber* (58.4%), *B. agri* (54.2%), *B. borstelensis* (52%) and *C. thermarum* (47.6%). Strain mt6^T has the smallest number of protein-coding genes as well as the smallest number of total genes than all of the other compared genomes, as summarized in Table 6.

Among species with standing in nomenclature, AGIOS values ranged from 75.55 between *B. borstelensis* and *B. thermoruber* to 59.20 between *B. thermoruber* and *B. mannanilyticus*. The comparison of the AGIOS value of strain mt6^T with the other species gave AGIOS values ranging from 60.30 with *B. thermoruber* to 63.12 with *B. mannanilyticus* (Tables 7 and 8). In addition, strain mt6^T shared 1296, 1316, 1039, 1079 and 1605 orthologous genes with *B. thermoruber, B. borstelensis, C. thermarum, B. mannanilyticus* and *B. agri* respectively. Finally, we observed that in each COGs categories, all compared genomes have nearly the same number of genes (Fig. 7).

Conclusion

The proteomic analysis of strain mt6^T with its MALDI-TOF MS spectrum, the 92% similarity level of the 16S rRNA nucleotide sequence to *Bacillus mannanilyticus* and the analysis of its complete assembled and annotated genome allowed us to propose the creation of a new genus called *Rubeoparvulum* gen. nov. *Rubeoparvulum massiliense* sp. nov. and strain mt6^T are the type species and type strain respectively of *Rubeoparvulum* gen. nov.

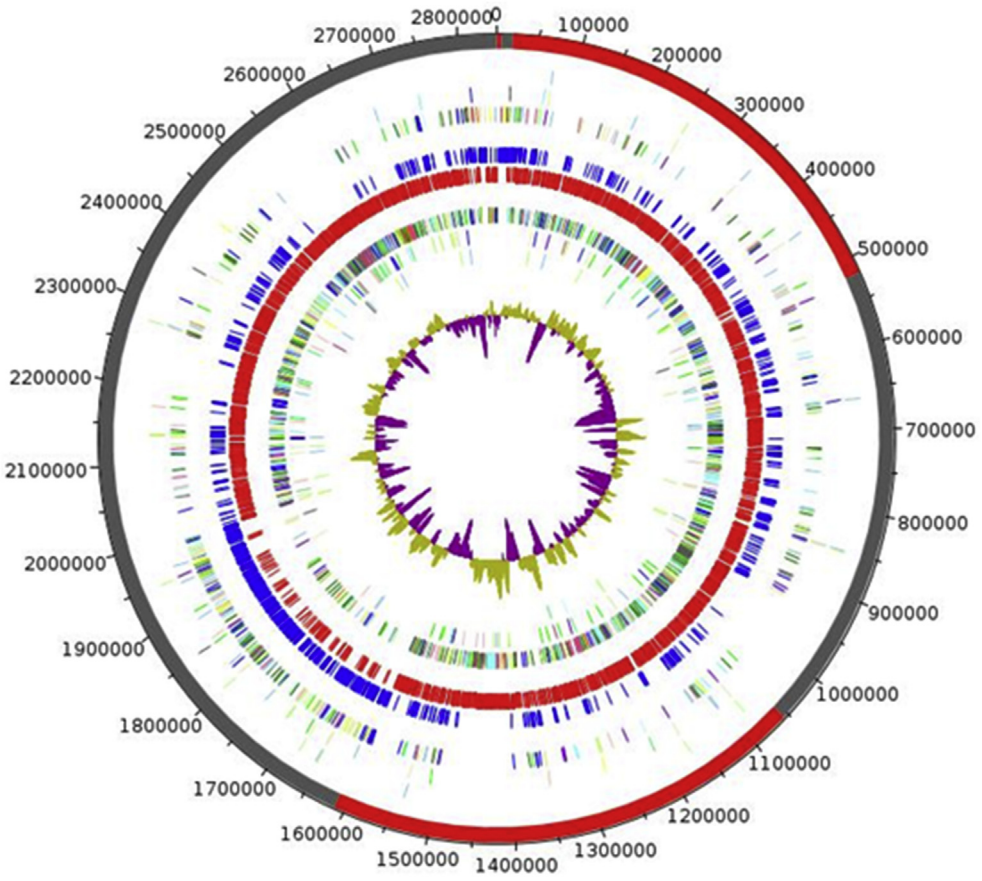

FIG. 6. Graphical circular map of chromosome. From outside to centre, genes on forward strain are coloured by COGs categories (only gene assigned to COGs), RNA genes (tRNAs green, rRNAs red), G+C content and G+C skew. COGs, Clusters of Orthologous Groups database.

TABLE 5. Number of genes associated with 25 general COGs functional categories

Code	Value	% of total[a]	Description
J	150	5.48	Translation
A	0	0	RNA processing and modification
K	141	5.15	Transcription
L	139	5.08	Replication, recombination and repair
B	1	0.03	Chromatin structure and dynamics
D	28	1.02	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	50	1.82	Defense mechanisms
T	90	3.29	Signal transduction mechanisms
M	82	2.99	Cell wall/membrane biogenesis
N	49	1.79	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	39	1.42	Intracellular trafficking and secretion
O	82	2.99	Posttranslational modification, protein turnover, chaperones
C	129	4.71	Energy production and conversion
G	81	2.96	Carbohydrate transport and metabolism
E	227	8.29	Amino acid transport and metabolism
F	62	2.26	Nucleotide transport and metabolism
H	76	2.77	Coenzyme transport and metabolism
I	66	2.41	Lipid transport and metabolism
P	155	5.66	Inorganic ion transport and metabolism
Q	30	1.09	Secondary metabolites biosynthesis, transport and catabolism
R	273	9.98	General function prediction only
S	189	6.91	Function unknown
—	862	31.51	Not in COGs

COGs, Clusters of Orthologous Groups database.
[a]Total is based on total number of protein-coding genes in annotated genome.

Description of *Rubeoparvulum* gen. nov.

Rubeoparvulum (ru.be.o, adj. 'red'; par.vu.lum, n. 'infant') strain mt6^T was isolated from the stool of a patient with kwashiorkor. The term 'red infant' refers to the hair discolouration observed in kwashiorkor patients).

Cells are rod-shaped, Gram-stain-negative bacteria. Optimal growth in anaerobic and microaerophilic conditions is at 37°C. The organism is catalase negative and oxidase positive; nitrate reduction, urease, β-glucosidase and alkaline phosphatase were positive. The type species is *Rubeoparvulum massiliense* strain mt6^T.

Description of *Rubeoparvulum massiliense* strain mt6^T gen. nov., sp. nov.

Rubeoparvulum massiliense (mas.si.li.en'se, L. adj. *massiliense,* of Massilia, the old Greek and Roman name of Marseille, France, where the strain was isolated).

Cells are spore-forming, motile and facultative anaerobe, Gram-stain-negative, rod-shaped bacilli with a mean diameter of 1 µm and a mean length of 6.8 µm. Colonies were small (mean diameter of 0.5 mm) and translucent on 5% sheep's blood–

TABLE 6. Genome comparison of closely related species to *Rubeoparvulum massiliense* strain mt6^T

Organism	INSDC	Size (Mb)	G+C (%)	Protein-coding genes	Total genes
Rubeoparvulum massiliense strain mt6^T	CVPE00000000	2.84	43.75	2735	2811
Bacillus mannanilyticus strain AM-001	BAMO00000000.1	4.53	39.6	3846	4454
Brevibacillus agri strain DSM 6348^T	JATL00000000.1	5.51	54.2	5047	5297
Brevibacillus borstelensis strain DSM 6347^T	APBN00000000.1	5.16	52.0	4817	5039
Brevibacillus thermoruber strain DSM 7064	ATNE00000000.1	4.43	58.4	4072	4269
Caldalkalibacillus thermarum strain HA6	AFCE00000000.1	2.9	47.6	2741	2969

INSDC, International Nucleotide Sequence Database Collaboration.

TABLE 7. Numbers of orthologous protein shared between genomes (upper right)[a]

	Bacillus mannanilyticus strain AM-001	*Brevibacillus agri* strain DSM 6348^T	*Brevibacillus borstelensis* strain DSM 6347^T	*Brevibacillus thermoruber* strain DSM 7064	*Caldalkalibacillus thermarum* strain HA6	*Rubeoparvulum massiliense* strain mt6
B. mannanilyticus strain AM-001	**4842**	1605	1606	1533	1178	1079
B. agri strain DSM6348^T	60.62	**5273**	2713	2625	1368	1286
B. borstelensis strain DSM 6347T	61.19	73.07	**5019**	2645	1376	1316
B. thermoruber strain DSM 7064	59.20	75.04	75.55	**4253**	135	1269
C. thermarum strain HA6	66.10	62.97	63.15	63.13	**2986**	1039
R. massiliense strain mt6	63.12	61.17	61.57	60.30	62.87	**2733**

[a]Average percentage similarity of nucleotides corresponding to orthologous protein shared between genomes (lower left) and numbers of proteins per genome (bold).

TABLE 8. Pairwise comparison of *Rubeoparvulum massiliense* strain mt6^T with other species using GGDC, formula 2 (DDH estimates based on identities/HSP length),[a] upper right

	Bacillus mannanilyticus	*Brevibacillus agri*	*Brevibacillus borstelensis*	*Brevibacillus thermoruber*	*Caldalkalibacillus thermarum*	*Rubeoparvulum massiliense*
B. mannanilyticus	100% ± 00	32.9% ± 2.52	32.2% ± 2.52	31.6% ± 2.52	26.2% ± 2.52	26.3% ± 2.52
B. agri		100% ± 00	18.8% ± 2.70	19.4% ± 2.73	30% ± 2.52	35.4% ± 2.52
B. borstelensis			100% ± 00	20.2% ± 2.80	29.9% ± 2.52	30.7% ± 2.52
B. thermoruber				100% ± 00	35.1% ± 2.52	23.2% ± 2.52
C. thermarum					100% ± 00	27.2% ± 2.52
R. massiliense						100% ± 00

[a]Confidence intervals indicate inherent uncertainty in estimating DDH values from intergenomic distances based on models derived from empirical test data sets (which are always limited in size). These results are in accordance with 16S rRNA (Fig. 1) and phylogenomic analyses as well as GGDC results.

enriched Colombia agar. The organism is catalase negative and oxidase positive.

Positive reactions were recorded for alkaline phosphatase, esterase (C4), esterase lipase (C8), leucine arylamidase, α-chymotrypsin, acid phosphatase and naphtol-AS-BI-phosphohydrolase. Urease, β-glucosidase, protease activities and nitrate reduction were also positive. Glycerol, D-lactose, D-fucose, D-mannose, D-cellobiose, D-maltose, salicin, *N*-acetyl-glucosamine, potassium-5-ketogluconate and D-arabitol were metabolized. Cells were susceptible to amoxicillin, gentamicin, ceftriaxone, ciprofloxacin, penicillin, imipenem, tobramycin and oxacillin but were resistant to metronidazole, trimethoprim/sulfamethoxazole, rifampicin, doxycycline, vancomycin, nitrofurantoin and erythromycin.

The G+C content of the genome is 43.75%. The 16S rRNA gene sequence and whole-genome shotgun sequence of *R. massiliense* strain mt6^T are deposited in European Molecular Biology Laboratory/European Bioinformatics Institute under accession numbers LN828926 and CVPE00000000 respectively. The type strain mt6^T (= CSUR P1473 = DSM 100479) was isolated from the faecal matter of a 2-month-old boy from Senegal with kwashiorkor.

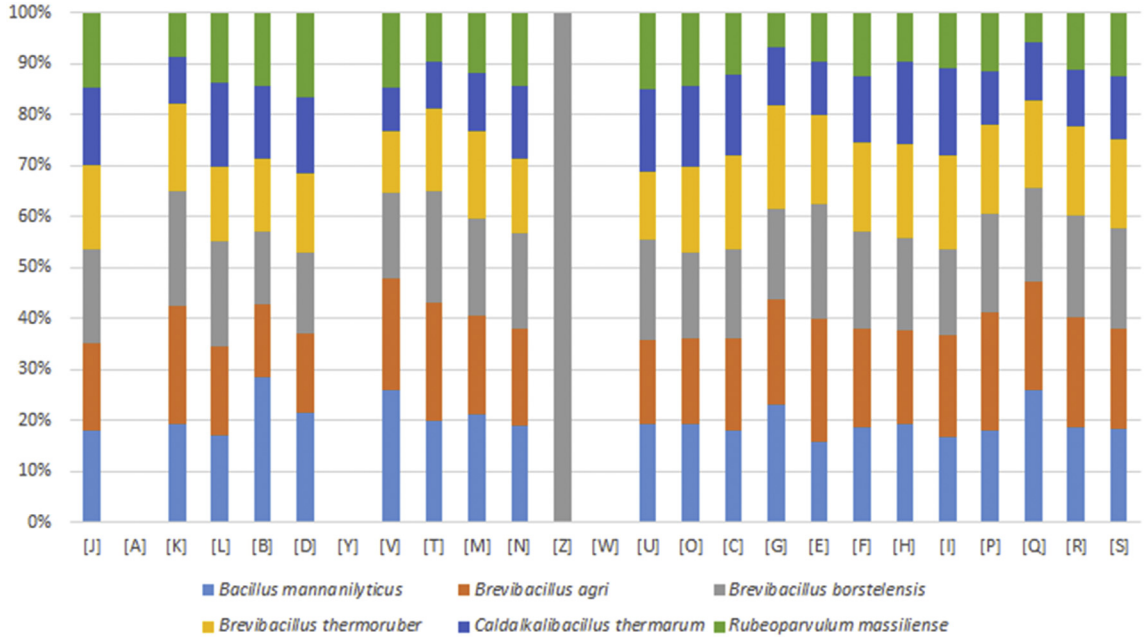

FIG. 7. Distribution of functional classes of predicted genes according COGs of protein. COGs, Clusters of Orthologous Groups database.

Acknowledgements

The authors thank the Xegen Company (www.xegen.fr) for automating the genomic annotation process. This study was funded by the Fondation Méditerranée Infection. We thank K. Griffiths for English-language review and C. Andrieu for administrative assistance.

Conflict of Interest

None declared.

References

[1] Fricke WF. The more the merrier? Reduced fecal microbiota diversity in preterm infants treated with antibiotics. J Pediatr 2014;165:8–10.
[2] Simpson HL, Campbell BJ. Review article: dietary fibre–microbiota interactions. Aliment Pharmacol Ther 2015;42:158–79.
[3] Gritz EC, Bhandari V. The human neonatal gut microbiome: a brief review. Front Pediatr 2015;3:60.
[4] 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.
[5] Million M, Diallo A, Raoult D. Gut microbiota and malnutrition. Microb Pathog 2016.
[6] Million M, Tidjani Alou M, Khelaifia S, Bachar D, Lagier JC, Dione N, et al. Increased gut redox and depletion of anaerobic and methanogenic prokaryotes in severe acute malnutrition. Sci Rep 2016;6:26051.
[7] 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.
[8] Lagier JC, Hugon P, Khelaifia 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] Fournier PE, Lagier JC, Dubourg G, Raoult D. From culturomics to taxonomogenomics: a need to change the taxonomy of prokaryotes in clinical microbiology. Anaerobe 2015;36:73–8.
[13] 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.
[14] 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.
[15] Drancourt M, Bollet C, Carlioz A, Martelin R, 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.
[16] 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.
[17] Matuschek E, 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.

[18] Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:119.
[19] Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2012;40:D48–53.
[20] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955–64.
[21] Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 2007;35:3100–8.
[22] Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol 2004;338: 1027–36.
[23] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res 2011;39:W347–52.
[24] 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.
[25] Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. Bioinforma Oxf Engl 2000;16:944–5.
[26] Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinforma Oxf Engl 2009;25:119–20.
[27] Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004;14:1394–403.
[28] Gouret P, Thompson JD, Pontarotti P. PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinformatics 2009;10:298.
[29] 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. New York: Springer; 2011. p. 71–87.
[30] 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 Bioinformatics 2005;6:198.
[31] Auch AF, von Jan M, Klenk HP, 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.
[32] Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence–based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 2013;14:60.
[33] Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan 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.
[34] Gao XH, Gao S, Zhou Y, Guan HL, Zhang YJ, Jia M, et al. *Tenuibacillus halotolerans* sp. nov., a novel bacterium isolated from a soil sample from a salt lake in Xinjiang, China and emended description of the genus *Tenuibacillus*. Antonie Van Leeuwenhoek 2013;103:207–15.
[35] Garcia MT. *Thalassobacillus devorans* gen. nov., sp. nov., a moderately halophilic, phenol-degrading, Gram-positive bacterium. Int J Syst Evol Microbiol 2005;55:1789–95.
[36] Jeon CO. Reclassification of *Bacillus haloalkaliphilus* Fritze 1996 as *Alkalibacillus haloalkaliphilus* gen. nov., comb. nov. and the description of *Alkalibacillus salilacus* sp. nov., a novel halophilic bacterium isolated from a salt lake in China. Int J Syst Evol Microbiol 2005;55:1891–6.
[37] Mayr R. *Ornithinibacillus* gen. nov., with the species *Ornithinibacillus bavariensis* sp. nov. and *Ornithinibacillus californiensis* sp. nov. Int J Syst Evol Microbiol 2006;56:1383–9.
[38] Nogi Y. Characterization of alkaliphilic *Bacillus* strains used in industry: proposal of five novel species. Int J Syst Evol Microbiol 2005;55: 2309–15.
[39] Pereira SG, Albuquerque L, Nobre MF, Tiago I, Verissimo A, Pereira A, et al. *Pullulanibacillus uraniitolerans* sp. nov., an acidophilic, U(VI)-resistant species isolated from an acid uranium mill tailing effluent and emended description of the genus *Pullulanibacillus*. Int J Syst Evol Microbiol 2013;63:158–62.
[40] Ren PG. *Salinibacillus aidingensis* gen. nov., sp. nov. and *Salinibacillus kushneri* sp. nov., moderately halophilic bacteria isolated from a neutral saline lake in Xin-Jiang, China. Int J Syst Evol Microbiol 2005;55: 949–53.
[41] Slobodkina GB, Panteleeva AN, Kostrikina NA, Kopitsyn DS, Bonch-Osmolovskaya EA, Slobodkin AI. *Tepidibacillus fermentans* gen. nov., sp. nov.: a moderately thermophilic anaerobic and microaerophilic bacterium from an underground gas storage. Extremophiles 2013;17: 833–9.