Massilibacterium senegalense gen. nov., sp. nov., a new bacterial genus isolated from the human gut

M. Tidjani Alou, J. Rathored, J.-C. Lagier, S. Khelaifia, N. Labas, C. Sokhna, A. Diallo, D. Raoult and G. Dubourg

1) Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes IRD 198, CNRS 7278, Aix-Marseille Université, Marseille, France, 2) Campus Commun UCAD-IRD of Hann, Dakar, Senegal and 3) Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract

Massilibacterium senegalense gen. nov., sp. nov., strain mt8T, is the type strain of Massilibacterium gen. nov., a new genus within the Bacillaceae family. This Gram-negative facultative anaerobic rod was isolated from the gut microbiota of a severely malnourished boy. Its phenotypic description is hereby presented with a complete annotation of its genome sequence. This genome is 5 697 950 bp long and contains 5615 protein-coding genes and 178 RNA genes, among which are 40 rRNA genes.

New Microbes and New Infections © 2016 The Authors. Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases.

Keywords: Culturomics, genome, gut microbiota, Massilibacterium senegalense, taxonogenomics

Original Submission: 1 December 2015; Revised Submission: 11 January 2016; Accepted: 14 January 2016

Article published online: 22 January 2016

Introduction

The human gut microbiota is a complex and vast ecosystem harbouring eukaryotes, viruses, archaea and bacteria, these being by far the most abundant [1]. Its cell count is estimated to be approximately 10^{13}, representing ten times the human somatic cell count, and its collective bacterial genome size is 150 times the size of the human genome [1–4]. The development of metagenomics has allowed a better exploration of gut microbiota by bypassing the noncultivable bacteria problem and unveiling links between altered gut microbiota and several diseases such as obesity, inflammatory bowel disease and irritable bowel syndrome [2]. It has also been demonstrated that the microbiota plays key roles in digestion and in immunologic and metabolic functions [2–4]. Nevertheless, a cultivation approach would be a complementary way to explore the gut microbiome in order to have a better representation of the viable population. In addition, it would allow further knowledge about the gut bacterial repertoire.

A new approach was developed in our laboratory in order to explore as exhaustively as possible the human gut microbiota by multiplying culture conditions with different atmospheres, media and temperatures [5]. This approach, known as culturomics, allowed us to isolate a new member of the Bacillaceae family. This family was created by Cohn in 1872 and consists of 52 validated genera (http://www.bacterio.net/). Bacillus is the type genus of this family, containing genera that are mostly aerobic or facultative anaerobic, rod-shaped, spore-forming, Gram-positive bacteria. These ubiquitous species are found in many ecosystems—mainly soil but also other environmental and clinical samples. Most Bacillaceae species are harmless, but some can be opportunistic pathogens, and Bacillus anthracis, the agent of anthrax, is well known to be pathogenic for humans [6].

Bacterial classification is currently based on phylogenetic relationships built on the 16S ribosomal RNA gene, phenotypic and genotypic characteristics including G+C content and DNA-DNA hybridization [7–9]. However, a great breakthrough has been
made in the last years in the area of genome sequencing, partly due to its decreasing cost. In fact, to this day, almost 70,000 genomes have been sequenced (https://gold.jgi.doe.gov/). With the development of this innovation, we proposed a new concept of bacterial description, including a proteomic description with a matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) profile [10,11] alongside a biochemical and genomic description of the new species [12–17].

We describe here a new member of this family, the genus Massilibacterium, isolated in the faeces of a patient with kwashiorkor. Massilibacterium senegalense is the type species (= CSUR P1510 = DSM 100455) of this new genus.

Materials and Methods

Organism information
As part of a culturomics study of the gut microbiota of children with severe acute malnutrition, a stool sample was collected from a 2-month-old Senegalese boy with kwashiorkor (body mass index, 14 kg/m²) in April 2014. The patient was not treated with antibiotics at the time of sample collection; the sample was stored at −80°C. This study was authorized by the child’s parents and was approved by the Institut Fédératif de Recherche 48 (Faculty of Medicine, Marseille, France) under agreement 09-022.

Strain identification by MALDI-TOF and 16S rRNA sequencing
Using the 18 culture conditions of the culturomics concept, the fecal sample was cultivated, and the obtained colonies were identified by MALDI-TOF as described below [5]. Proteomic analysis of our strain was carried out with MALDI-TOF as previously described [10,11]. A Microflex spectrometer (Bruker Daltonics, Leipzig, Germany) was used with a MTP 96 MALDI-TOF target plate (Bruker) on which 12 individual colonies were deposited. Twelve spectra were thus obtained, imported into MALDI BioTyper 2.0 software (Bruker) and analysed by standard pattern matching (with default parameter settings) against the main spectra of 7567 bacteria. Comparison with the BioTyper database spectra enabled the identification and discrimination of the analysed species from those in the database in accordance with the obtained score: a score >2 with a validated species enabled the identification at the species level, and a score <1.7 did not enable any identification. After a failed identification of the colony with a clean spectrum, it was identified by sequencing the 16S ribosomal RNA as previously described [18]. A threshold of 98.7% similarity level was determined to define a new species without performing DNA-DNA hybridization [19].

Growth conditions
In order to determine the ideal growth condition of M. senegalense, different growth temperatures (28, 30, 37, 45 and 56°C) were tested under anaerobic and microaerophilic atmospheres using GENbag anaer and GENbag miroaer systems respectively (bioMérieux, Marcy l’Étoile, France). The strain growth was also tested aerobically with or without 5% CO₂ supplementation.

Morphologic, biochemical and antibiotics susceptibility tests
The phenotypic characteristics (Gram staining, sporulation, motility, catalase, oxidase) were analysed as previously described [20]. Antibiotic susceptibility testing was performed using the disk diffusion method according to EUCAST 2015 recommendations (http://www.eucast.org/). Using API 20NE, API ZYM and API 50CH strips, we investigated the biochemical characteristics of the strain according to the manufacturer’s instructions (bioMérieux). Electronic microscopy was performed with detection Formvar-coated grids which were deposited on a 40 μL bacterial suspension drop and incubated at 37°C for 30 minutes. Then followed a 10-second incubation on ammonium molybdate 1%. The grids were dried on blotting paper and finally observed with a Tecnai G20 transmission electron microscope (FEI Company, Limeil-Brevannes, France).

Genomic DNA preparation
M. senegalense strain mt8 was cultured on 5% sheep’s blood–enriched Columbia agar 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 that 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
Genomic DNA (gDNA) of M. senegalense was sequenced on the MiSeq Technology (Illumina, San Diego, CA, USA) with the mate pair strategy. The gDNA was barcoded in order to be mixed with 11 other projects with the Nextera Mate Pair sample prep kit (Illumina). gDNA was quantified by a Qubit assay with a high sensitivity kit (Thermo Fisher Scientific, Waltham, MA, USA) to 66.2 ng/μL. The mate pair library was prepared with 1 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
pattern of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) with a DNA 7500 lab chip. The DNA fragments ranged in size from 1 to 11 kb, with an optimal size at 3.927 kb. No size selection was performed, and 505 ng of fragmented DNA was sheared. The circularized DNA was mechanically sheared to small fragments with an optimal size of 597 bp on the Covaris device S2 in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies, Santa Clara, CA, USA), and the final concentration library was measured at 59.2 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. An automated cluster generation and sequencing run was performed in a single 39-hour run in a 2 × 251 bp read length.

Genome annotation and analysis
Open reading frames (ORFs) were predicted using Prodigal [21] with default parameters, but the predicted ORFs were excluded if they were spanning a sequencing gap region (contain N). The predicted bacterial protein sequences were searched against the Clusters of Orthologous Groups (COGs) database using BLASTP (E value 1e-03, coverage 0.7 and identity percentage of 30%). If no hit was found, then it was searched against the NR database using BLASTP with an E value of 1e-03, coverage 0.7 and identity percentage of 30%, and if the sequence length was smaller than 80 amino acids, we used an E value of 1e-05. The tRNAscanSE tool [22] was used to find tRNA genes, whereas ribosomal RNAs were found by using RNAmmer [23]. Lipoprotein signal peptides and the number of transmembrane helices were predicted using Phobius [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 amino acids or E value smaller than 1e-05 for ORFs with sequence length smaller 80 amino acids). Such parameter thresholds have already been used in previous studies to define ORFans.

Genomes were automatically retrieved from the 16S rRNA tree using Xegen software (Phylopattern) [25]. For each selected genome, the complete genome sequence, proteome genome sequence and Orfeome genome sequence were retrieved from the FTP site of the National Center for Biotechnology Information (NCBI). All proteomes were analysed with proteinOrtho [26]. Then for each couple of genomes a similarity score was computed. This score is the mean value of nucleotide similarity between all couple of orthologues between the two genomes studied (average genomic identity of orthologous gene sequences, AGIOS) [27]. 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). To evaluate the genomic similarity among studied Bacillus strains, we determined two parameters: digital DNA-DNA hybridization (dDDH), which exhibits a high correlation with DNA-DNA hybridization (DDH) [28,29], and AGIOS [27], which was designed to be independent from DDH. Annotation and comparison processes were performed in the Multi-Agent software system DAGOBAH [30] that included Figenix [31] libraries that provide pipeline analysis.

Results

Strain identification and phylogenetic analyses
Strain mt8T (Table 1) was first isolated in January 2015 after 30-day preincubation in a blood culture bottle with sheep’s blood and cultivation on 5% sheep’s blood–enriched Colombia agar in an aerobic atmosphere at 37°C. MALDI-TOF displayed an identification score under 1.7 for strain mt8T, suggesting that the obtained spectra was not matched to any spectra in our database. The 16S ribosomal RNA sequence (accession no. LN828943) of strain mt8T showed a 93% nucleotide sequence similarity with Bacillus halodurans, which is the phylogenetically closest species with a validly published name (Fig. 1). Consequently, as this 16S rRNA nucleotide sequence similarity was lower than the threshold of 95% recommended by Stackebrandt and Ebers [19] to delineate a new genus, it was classified as a new genus called Massilibacterium, type species Massilibacterium senegalense strain mt8T. The reference spectrum for strain mt8T (Fig. 2) was thus incremented in our database and then compared to other known species of the genus Bacillus. The differences exhibited are shown in the obtained gel view (Fig. 3).

| Property | Term |
|----------|------|
| Domain   | Bacteria |
| Phylum   | Firmicutes |
| Class    | Bacilli |
| Order    | Bacillales |
| Family   | Bacillaceae |
| Genus    | Massilibacterium |
| Species  | Massilibacterium senegalense |
| Type strain | mt8 |
| Gram stain | Negative |
| Cell shape | Rod |
| Motility | Motile |
| Sporulation | Sporulating |
| Temperature range | Mesophilic |
| Optimum temperature | 37°C |

TABLE 1. Classification and general features of Massilibacterium senegalense strain mt8T
**FIG. 1.** Phylogenetic tree highlighting the position of *Massilibacterium senegalense* strain mt8T relative to other close strains. The respective GenBank accession numbers for 16S rRNA genes are indicated in parenthesis. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using the maximum-likelihood method within the MEGA6. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 1,000 times to generate a majority consensus tree. *Paenibacillus polymyxa* strain KCTC3717 was used as an outgroup. The scale bar represents a 1% nucleotide sequence divergence.

**FIG. 2.** Reference mass spectrum from *Massilibacterium senegalense* strain mt8T. Spectra from 12 individual colonies were compared and a reference spectrum was generated.
**Phenotypic description**

Growth of strain mt8T was observed between 28 and 45°C on 5% sheep’s blood–enriched Columbia agar, and optimal growth was achieved at 37°C after 24 hours’ incubation in aerobic conditions. Poor growth occurred under microaerophilic and anaerobic conditions. Cells were motile and sporulating. Colonies were irregular white colonies with a mean diameter of 5 mm on blood-enriched Colombia agar. The Gram staining (Fig. 4) showed Gram-negative rods. Using electron microscopy, the rods had a mean diameter of 1.8 μm and a length of 5.9 μm (Fig. 5).

Catalase and oxidase activities were negative for strain mt8T. Using API ZYM, positive reactions were observed for esterase (C4) and acid phosphatase. Reactions for alkaline phosphatase, esterase lipase (C8), lipase (C14), leucine arylamidase, valine arylamidase, trypsin, α-chymotrypsin, naphthol-AS-BI-phosphorylaser, β-galactosidase, N-acetyl-β-glucosaminidase, α-galactosidase, β-glucuronidase, α-glucosidase, β-glucosidase, α-fucosidase and α-mannosidase were negative. Using API 20NE, nitrate reduction and esculin hydrolysis were observed. All other reactions were negative, including indole formation and urease. An API 50CH strip showed positive reactions for N-acetyl-glucosamine, amygdalin, arbutin, salicin, d-maltose, d-fructose, inulin, d-mannose, d-sucrose and d-raffinose. Negative reactions were recorded for glycerol, erythritol, d-arabinose, l-arabinose, d-ribose, d-xylose, l-xylose, d-adenitol, methyl-β-d-xylopyranoside, d-glucose, d-galactose, d-lactose, l-sorbose, l-rhamnose, dulcitol, inositol, d-mannitol, d-sorbitol, methyl-α-d-mannopyranoside, methyl-α-d-glucopyranoside, esculin ferric citrate, d-cellobiose, d-melibiose, d-trehalose, d-melezitose, starch, glycogen, xylitol, gentiobiose, d-turanose, d-lyxose, d-tagatose, d-fucose, d-

**FIG. 3.** Gel view comparing *Massilibacterium senegalense* strain mt8T to other species within the Bacillaceae family. The gel view displays the raw spectra of 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 link between the color in which a peak is displayed and the peak intensity in arbitrary units. Displayed species are indicated on the left.

**FIG. 4.** Gram staining of *Massilibacterium senegalense* strain mt8T.
arabitol, L-arabitol, potassium gluconate, potassium 2-ketogluconate and potassium 5-ketogluconate.

Cells were susceptible to imipenem, doxycycline, rifampicin, vancomycin, amoxicillin, ceftriaxone, gentamicin (500 μg), trimethoprim/sulfamethoxazole, erythromycin, ciprofloxacin and gentamicin (15 μg) but were resistant to metronidazole.

Table 2 shows the differences between the biochemical properties of close relatives of Massilibacterium senegalense strain mt8T, members of the Bacillaceae family.

### Genome properties

The genome of M. senegalense strain mt8T genome (accession no. HTW00000000) was 5 697 950 bp long with a 35.67% G+C content (Fig. 6, Table 3). It was composed of ten scaffolds and 12 contigs. There were 5793 predicted genes, among which 5615 were protein-coding genes and 178 RNAs (14 5S rRNA genes, 16 16S rRNA genes, ten 23S rRNA genes and 138 tRNA genes). A total of 4262 genes were assigned a putative function, and 208 genes were identified as ORFans. The remaining 386 genes were annotated as hypothetical proteins. Using ARG-ANNOT [32], no resistance genes were found. The properties and statistics of the genome are summarized in Table 3, and the gene distribution into COGs functional categories is presented in Table 4.

### Genome comparison

The genomic characteristics (size, percentage of G+C content, protein-coding genes and total number of genes) were used to compare strain mt8T with the genome of closely related species (Table 5). The size of Massilibacterium senegalense strain mt8T (5.69 Mb) is larger than Bacillus wakoensis strain N_1, Lysinibacillus fusiformis strain DSM 2898T, Bacillus halodurans strain C-125, Bacillus pseudofirmus strain OF4, Anoxybacillus tepidamans strain PS2 and Bacillus smithii strain 7_3_47FAA (5.53, 4.84, 4.2, 3.86, 3.36 and 3.24 Mb respectively). The G+C content of M. senegalense (35.6%) is smaller than those of B. halodurans, B. tepidamans, B. smithii, B. pseudofirmus, B. wakoensis and L. fusiformis (43.7, 43.0, 40.7, 40.3, 38.3 and 37.6%). The gene content of M. senegalense (5793) is bigger than the gene content of L. fusiformis, B. wakoensis, B. halodurans, B. pseudofirmus, A. tepidamans and B. smithii (4764, 4460, 4076, 3841, 3400 and 3235 respectively) (Table 5). There are more protein-coding genes (5615) in the genome of M. senegalense than in the genomes of L. fusiformis, B. wakoensis, B. halodurans, B. pseudofirmus, A. tepidamans, and B. smithii (4548, 3912, 3903, 3704, 3245 and 2832 respectively). The distribution of genes into COGs categories was similar in all compared genomes (Fig. 7, Table 4). M. senegalense also shared 1368, 1244, 1263, 1318, 1321 and 1231 orthologous genes with B. pseudofirmus, L. fusiformis, B. wakoensis, A. tepidamans, B. halodurans, and B. smithii respectively (Table 6). Among species with standing in nomenclature, AGIOS values ranged from 63.43 to 70.67% among compared species except M. senegalense. When compared to other species, the AGIOS values ranged from 65.40% with B. halodurans to 66.37% with A. tepidamans (Table 6). To evaluate the genomic similarity among studied Bacillaceae strains, we determined two parameters, dDDH, which exhibits a high correlation with DDH [28,29], and AGIOS [27], which was designed to be independent from DDH (Table 7).

### Conclusion

Given the 93% similarity level to Bacillus halodurans for the 16S rRNA sequence of strain mt8T, its MALDI-TOF spectrum and the analysis of its annotated genome, we created a new genus, Massilibacterium. Massilibacterium senegalense is the type strain.

### Description of Massilibacterium gen. nov.

Massilibacterium (from Massilia, Marseille’s old Roman and Greek name; Marseille is the city in which the strain was isolated).

Facultative anaerobic rod-shaped bacteria. Gram negative. Optimal growth in aerobic conditions at 37°C. Catalase and oxidase negative. Nitrites were reduced into nitrates. Negative for indole formation. β-Glucosidase positive. Urease negative. The type species is Massilibacterium senegalense strain mt8T.

Habitat is human gut.
| Property                          | Massilibacterium senegalense | Bacillus halodurans | Bacillus acidicola | Bacillus wakoensis | Bacillus hemicellulosilyticus | Bacillus cellulosilyticus | Bacillus akibai | Bacillus mannanilyticus | Bacillus okuhidensis | Bacillus sonorensis |
|---------------------------------|-----------------------------|---------------------|-------------------|-------------------|---------------------------|--------------------------|----------------|--------------------------|-------------------|----------------------|
| Cell diameter (μm)              | 1.7–1.9                     | 0.5–0.6             | 1.0–1.6           | 0.5–0.8           | 0.3–0.5                   | 0.6–0.8                  | 0.6–0.8        | 0.6–0.8                  | 0.5–1.0           | 1.0                  |
| Gram stain                      | −                           | +                   | +                 | +                 | +                         | +                        | +              | +                        | +                 | +                    |
| Salt requirement                | −                           | +                   | −                 | −                 | −                         | −                        | −              | −                        | −                 | −                    |
| Motility                        | +                           | NA                  | +                 | +                 | +                         | +                        | +              | +                        | +                 | +                    |
| Endospore formation             | +                           | +                   | +                 | +                 | +                         | +                        | +              | +                        | +                 | +                    |
| Indole                          | −                           | NA                  | −                 | −                 | −                         | −                        | −              | −                        | −                 | −                    |
| Production of:                  |                             |                     |                   |                   |                           |                          |                 |                           |                   |                      |
| Alkaline phosphatase            | −                           | NA                  | −                 | −                 | −                         | −                        | −              | −                        | −                 | −                    |
| Catalase                        | −                           | NA                  | +                 | +                 | +                         | +                        | +              | +                        | +                 | +                    |
| Oxidase                         | −                           | NA                  | −                 | −                 | −                         | +                        | N              | −                        | −                 | −                    |
| Nitrate reductase               | +                           | −                   | +                 | +                 | +                         | +                        | +              | +                        | +                 | +                    |
| Urease                          | −                           | NA                  | −                 | −                 | −                         | −                        | −              | −                        | −                 | −                    |
| β-Galactosidase                 | −                           | NA                  | −                 | −                 | −                         | −                        | −              | −                        | −                 | −                    |
| N-acetyl-glucosamine           | +                           | NA                  | −                 | −                 | −                         | −                        | −              | +                        | −                 | −                    |
| Acid from:                      |                             |                     |                   |                   |                           |                          |                 |                           |                   |                      |
| L-Arabinose                     | −                           | +                   | −                 | −                 | −                         | −                        | −              | −                        | −                 | −                    |
| Ribose                          | −                           | +                   | −                 | −                 | −                         | −                        | −              | −                        | −                 | −                    |
| Mannose                         | +                           | +                   | +                 | +                 | +                         | +                        | +              | +                        | +                 | +                    |
| Mannitol                        | −                           | +                   | −                 | +                 | +                         | +                        | +              | +                        | +                 | +                    |
| Sucrose                         | +                           | +                   | +                 | +                 | +                         | +                        | +              | +                        | +                 | +                    |
| D-Glucose                       | −                           | +                   | +                 | +                 | +                         | +                        | +              | +                        | +                 | +                    |
| D-Fructose                      | +                           | +                   | +                 | +                 | +                         | +                        | +              | +                        | +                 | +                    |
| D-Maltose                       | +                           | +                   | +                 | +                 | +                         | +                        | +              | +                        | +                 | +                    |
| D-Lactose                       | −                           | +                   | −                 | +                 | +                         | +                        | +              | +                        | +                 | +                    |
| Habitat                         | Human gut                   | Soil                | Acidic peat bogs  | Industry          | Industry                  | Industry                 | Industry       | Industry                  | Industry          | Industry |
| **Table 2. Differential characteristics of Massilibacterium senegalense strain mt8T, Bacillus halodurans DSM 497, Bacillus acidicola DSM 14745T, Bacillus wakoensis DSM 2521T, Bacillus hemicellulosilyticus DSM 16731T, Bacillus cellulosilyticus DSM 2522T, Bacillus akibai ATCC 43226T, Bacillus mannanilyticus DSM 16130T, Bacillus okuhidensis DSM 13666T, Bacillus sonorensis DSM 13779T**
**Description of *Massilibacterium senegalense* strain mt8T gen. nov., sp. nov.**

*Massilibacterium senegalense* (se.ne.gal.e.nis. L. gen. masc., meaning originating from Senegal, the country from which the stool sample was collected).

Cells are sporulating, motile and facultative anaerobic, Gram-negative, rod-shaped bacilli with a mean diameter of 1.8 μm and a length of 5.9 μm. Colonies were 5 mm diameter white irregular colonies on 5% sheep’s blood-enriched Colombia agar. Catalase and oxidase negative.

Positive reactions were observed for esterase (C4) and acid phosphatase. Nitrate reduction and aesculin hydrolysis were positive. N-acetyl-glucosamine, amygdalin, arbutin, salicin, D-maltose, D-fructose, inulin, D-mannose, D-sucrose and D-raffinose were metabolized. Cells were susceptible to imipenem, doxycycline, rifampicin, vancomycin, amoxicillin, ceftriaxone, gentamicin (500 μg), trimethoprim/sulfamethoxazole, erythromycin, ciprofloxacin and gentamicin (15 μg) but were resistant to metronidazole.

The G+C content of the genome is 35.67%. The 16S rRNA gene sequence and whole genome shotgun sequence of *M. senegalense* strain mt8T are deposited in GenBank under accession numbers LN828943 and CTRN01000000 respectively.

**TABLE 3. Nucleotide content and gene count levels of the genome**

| Attribute                      | Genome (total) | % of total\(^a\) |
|--------------------------------|----------------|------------------|
| Size (bp)                      | 5,697,950      | 100              |
| G+C content (%)                | 2,034,168      | 35.7             |
| Coding region (bp)             | 4,442,019      | 77.95            |
| Total genes                    | 5,793          | 100              |
| RNA genes                      | 178            | 3.07             |
| Protein-coding genes           | 5,615          | 96.92            |
| Genes with function prediction | 4,362          | 73.57            |
| Genes assigned to COGs         | 3,833          | 66.25            |
| Genes with peptide signals     | 210            | 3.62             |
| Genes with transmembrane helices | 504        | 8.70             |
| CRISPR repeats                 | 0              | 0                |
| ORFans genes                   | 208            | 3.59             |
| Genes associated with PKS or NRPS | 13            | 0.22             |
| No. of antibiotic resistance genes | 0            | 0                |

COGs, Clusters of Orthologous Groups database; CRISPR, clustered regularly interspaced short palindromic repeat.

\(^a\)Total is based on either size of genome (bp) or total number of protein-coding genes in annotated genome.

**TABLE 4. Number of genes associated with 25 general COGs functional categories**

| Code | Value | % of total\(^a\) | Description |
|------|-------|------------------|-------------|
| J    | 298   | 10.63            | Translation |
| A    | 0     | 0.0              | RNA processing and modification |
| K    | 334   | 11.91            | Transcription |
| L    | 406   | 14.48            | Replication, recombination and repair |
| B    | 2     | 0.07             | Chromatin structure and dynamics |
| D    | 68    | 2.43             | Cell cycle control, mitosis and meiosis |
| Y    | 0     | 0.0              | Nuclear structure |
| V    | 80    | 2.85             | Defense mechanisms |
| T    | 252   | 8.99             | Signal transduction mechanisms |
| M    | 224   | 7.99             | Cell wall/membrane biogenesis |
| N    | 132   | 4.71             | Cell motility |
| Z    | 0     | 0.0              | Cytoskeleton |
| W    | 0     | 0.0              | Extracellular structures |
| U    | 98    | 3.50             | Intracellular trafficking and secretion |
| O    | 190   | 6.78             | Posttranslational modification, protein turnover, chaperones |
| C    | 248   | 8.84             | Energy production and conversion |
| G    | 140   | 4.99             | Carbohydrate transport and metabolism |
| E    | 370   | 13.20            | Amino acid transport and metabolism |
| F    | 138   | 4.92             | Nucleotide transport and metabolism |
| H    | 226   | 8.06             | Coenzyme transport and metabolism |
| I    | 190   | 6.78             | Lipid transport and metabolism |
| P    | 286   | 10.20            | Inorganic ion transport and metabolism |
| Q    | 92    | 3.28             | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 602   | 21.47            | General function prediction only |
| S    | 432   | 15.41            | Function unknown |
| —    | 424   | 7.31             | Not in COGs |

COGs, Clusters of Orthologous Groups database. \(^a\)Total is based on total number of protein-coding genes in annotated genome.
### TABLE 5. Genome comparison of species closely related to *Massilibacterium senegalense* strain mt8<sup>T</sup>

| Organism                                      | INSDC          | Size (Mb) | G+C%  | Protein-coding genes | Total genes |
|------------------------------------------------|----------------|-----------|-------|----------------------|-------------|
| *Massilibacterium senegalense* strain mt8<sup>T</sup> | CTRN00000000.1 | 5.69      | 35.6  | 5615                 | 5793        |
| *Bacillus pseudofirmus* strain OF4            | CP001878.2     | 3.86      | 40.3  | 3704                 | 3041        |
| *Lysinibacillus fusiformis* strain DSM 2898<sup>T</sup> | CP010820.1     | 4.84      | 37.6  | 4548                 | 4764        |
| *Bacillus wakoensis* strain N_1               | BAUT00000000.1 | 5.53      | 38.3  | 3912                 | 4460        |
| *Anoxybacillus tepidamans* strain PS2         | JHV00000000.1  | 3.36      | 43.0  | 3245                 | 3400        |
| *Bacillus halodurans* strain C-125            | BA000004.3     | 4.2       | 43.7  | 3903                 | 4076        |
| *Bacillus smithii* strain 7_3_47FAA           | ACWF00000000.1 | 3.24      | 40.7  | 2832                 | 3235        |

*DDH, DNA-DNA hybridization; GGDC, Genome-to-Genome Distance Calculator; HSP, high-scoring segment pairs.*

### TABLE 6. Number of orthologous proteins shared between genomes (upper right)<sup>a</sup>

| Bacillus pseudofirmus | Lysinibacillus fusiformis | Massilibacterium senegalense | Bacillus wakoensis | Anoxybacillus tepidamans | Bacillus halodurans | Bacillus smithii |
|-----------------------|---------------------------|------------------------------|-------------------|--------------------------|---------------------|-----------------|
| B. pseudofirmus       | 4335                      | 1496                         | 1368              | 1336                     | 1615                | 1959            |
| L. fusiformis         | 64.28                     | 4767                         | 1244              | 1356                     | 1456                | 1464            |
| M.                    | 65.99                     | 65.59                        | 5615              | 1263                     | 1318                | 1321            |
| B. wakoensis          | 70.67                     | 64.46                        | 66.35             | 4576                     | 1495                | 1819            |
| A. tepidamans         | 65.38                     | 64.71                        | 65.37             | 65.28                    | 3463                | 1611            |
| B. halodurans         | 68.48                     | 63.43                        | 65.40             | 68.03                    | 65.66               | 4066            |
| B. smithii            | 65.06                     | 65.14                        | 65.95             | 64.98                    | 68.03               | 65.00           |

<sup>a</sup>Average percentage similarity of nucleotides corresponding to orthologous proteins shared between genomes (lower left) and numbers of proteins per genome.

### TABLE 7. Pairwise comparison of *Massilibacterium senegalense* strain mt8<sup>T</sup> with other species using GGDC, formula 2 (DDH estimates based on identities/HSP length)<sup>b</sup> (upper right)

| Bacillus pseudofirmus | Lysinibacillus fusiformis | Massilibacterium senegalense | Bacillus wakoensis | Anoxybacillus tepidamans | Bacillus halodurans | Bacillus smithii |
|-----------------------|---------------------------|------------------------------|-------------------|--------------------------|---------------------|-----------------|
| B. pseudofirmus       | 100%                      | 29.1% ± 2.54                | 26.3% ± 2.55      | 21.2% ± 2.59             | 21.9% ± 2.53        | 27.5% ± 2.56    |
| L. fusiformis         | 100%                      | 29.5% ± 2.54                | 26.1% ± 2.54      | 23.2% ± 2.53             | 23.3% ± 2.56        | 28.3% ± 2.53    |
| M.                    | 100%                      | 25% ± 2.54                  | 21% ± 2.54        | 19.5% ± 2.53             | 23.3% ± 2.56        | 21.6% ± 2.55    |
| B. wakoensis          | 100%                      | 100%                         | 100%              | 100%                     | 100%                | 100%            |
| A. tepidamans         | 100%                      | 100%                         | 100%              | 100%                     | 100%                | 100%            |
| B. halodurans         | 100%                      | 22.2% ± 2.53                | 21.6% ± 2.55      | 22.2% ± 2.53             | 22.3% ± 2.53        | 25.9% ± 2.54    |
| B. smithii            | 100%                      | 100%                         | 100%              | 100%                     | 100%                | 100%            |

<sup>b</sup>DDH, DNA-DNA hybridization; GGDC, Genome-to-Genome Distance Calculator; HSP, high-scoring segment pairs.

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
The type strain mt8T (= CSUR P1510 = DSM 100455) was isolated from the stool of a young Senegalese boy with kwashiorkor.

Conflict of Interest

None declared.

Acknowledgements

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

References

[1] Xu Z, Knight R. Dietary effects on human gut microbiome diversity. Br J Nutr 2015;113:51–5.
[2] Salazar N, Arboleya S, Valdés L, et al. The human intestinal microbiome at extreme ages of life. Dietary intervention as a way to counteract alterations. Front Genet 2014;4:406.
[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] Simpson HL, Campbell PJ. Review article: dietary fibre–microbiota interactions. Aliment Pharmacol Ther 2015;42:518–78.
[5] Lagier JC, Armougom F, Million M, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 2012;18:1185–93.
[6] De Vos P, Garrity GM, Jones D, et al. Bergey’s manual of systematic bacteriology. The firmicutes. 2nd ed., Vol. 3. New York: Springer; 2009.
[7] Viale AM, Arakaki AK, Sancini FC, Ferreyra RG. Evolutionary relationships among eubacterial groups as inferred from GroEL (Chaperonin) sequences comparison. Int J Syst Bacteriol 1944;44:527–33.
[8] Woese CR, Kandler O, Wheelis ML. Towards a natural system of organic classification: proposals for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 1990;87:4576–9.
[9] Wolf M, Muller T, Dandekar T, Pollack JD. Phylogeny of Firmicutes with special reference to Mycoplasma (Mollicutes) as inferred from phosphoglycerate kinase amino acid sequence data. Int J Syst Evol Microbiol 2004;54:871–5.
[10] Seng P, Drancourt M, Gouret F, et al. Ongoing evolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 2009;49:543–51.
[11] Seng P, Abat C, Rolain JM, 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.
[12] Hugon P, Mishra AK, Lagier JC, et al. Non-contiguous finished genome sequence and description of Brevibacillus massiliensis sp. nov. Stand Genomic Sci 2013;8:1–14.
[13] Kolcha S, Ramasamy D, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of Brevibacterium senegalense sp. nov. Stand Genomic Sci 2012;7:233–45.
[14] 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;6:315–24.
[15] Lagier JC, Elkarkouri 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;6:116–25.
[16] Lagier JC, Elkarkouri K, Rivet R, Couderc C, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of Sene glemassilia anaerobia gen. nov., sp. nov. Stand Genomic Sci 2013;7:399–412.
[17] Drancourt M, Bollet C, Carliaz O, Martelini 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(10):3623–30.
[18] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152–5.
[19] Lagier JC, Edouard S, Paglieri I, Mediannikov O, Drancourt M, Raoult D. Current and past strategies for bacterial culture in clinical microbiology. Clin Microbiol Rev 2015;28:208–36.
[20] Hyatt D, Chen GL, Laccasio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:1:19–30.
[21] 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.
[22] 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;35:3100–8.
[23] Kall L, Krogh A, Sonnhammer EL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol 2004;338:1027–36.
[24] Gouret P, Thompson JD, Pontarotti P. PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinformatics 2009;10:298.
[25] Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. A combined transmembrane topology and signal peptide prediction method. J Mol Biol 2004;338:1027–36.
[26] Gouret P, Vitiello V, Balandraud N, Gilles A, Pontarotti P. PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinformatics 2009;10:298.
[27] Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics 2011;12:124.
[28] Ramasy D, Mishra AK, Lagier JC, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of new bacterial species. Int J Syst Evol Microbiol 2014;64:384–91.
[29] Meier-Kolthoff JP, Auch AF, von Jan M, Klenk HP, Goker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome–genome sequence comparison. Stand Genomic Sci 2010;2:117–34.
[30] Auch AF, von Jan M, Klenk HP, Goker M. Genome sequence-based species delineation with confidence intervals and improved distance functions. BMC Bioinformatics 2013;14:60.
[31] Gouret P, Paganini J, Dainat J, 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 Verlag; 2011. p. 71–87.
[32] Gouret P, Vitelli V, Balandraud N, Gilles A, Pontarotti P, Danchin E. FIGENIX: intelligent automation of genomic annotation: expertise integration in a new software platform. BMC Bioinformatics 2005;6:198.
[33] Gouret P, Vitelli V, Balandraud N, Gilles A, Pontarotti P, Danchin E. FIGENIX: intelligent automation of genomic annotation: expertise integration in a new software platform. BMC Bioinformatics 2005;6:198.
[33] Li Z, Kawamura Y, Shida O, Yamagata S, Deguchi T, Ezaki T. Bacillus okuhidensis sp. nov. isolated from the Okuhida spa area of Japan. Int J Syst Evol Microbiol 2002;52:1205–9.

[34] Nielsen P, Fritze D, Priest FG. Phenetic diversity of alkaliphilic Bacillus strains: proposal of nine new species. Microbiology 1995;141:1745–61.

[35] Noqi Y, Takami H, Horikoshi K. Characterization of alkaliphilic Bacillus strains used in industry: proposal of five novel species. Int J Syst Evol Microbiol 2005;55:2309–15.

[36] Palmisano MM, Nakamura LK, Duncan KE, Istock CA, Cohan FM. Bacillus sonorense sp. nov., a close relative of Bacillus licheniformis, isolated from soil in the Sonoran desert, Arizona. Int J Syst Evol Microbiol 2001;51:1671–9.

[37] Albert RA, Archambault J, Rosselö-Mora R, Tindall BJ, Matheny M. Bacillus acidicola sp. nov., a novel mesophilic, acidophilic species isolated from acidic sphagnum peat bogs in Wisconsin. Int J Syst Evol Microbiol 2005;55:2125–30.