Numidum massiliense gen. nov., sp. nov., a new member of the *Bacillaceae* family isolated from the human gut

M. Tidjani Alou[1], T.-T. Nguyen[1], N. Armstrong[1], J. Rathored[1], S. Khelaifia[1], D. Raoult[1,2], P.-E. Fournier[1] and J.-C. Lagier[1]

1) *Aix-Marseille Université, URMITE, UM63, CNRS7278, IRD198, Inserm 1095, Faculté de médecine, Marseille, France and* 2) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Numidum massiliense gen. nov., sp. nov., strain mt3^T is the type strain of *Numidum* gen. nov., a new genus within the family *Bacillaceae*. This strain was isolated from the faecal flora of a Tuareg boy from Algeria. We describe this Gram-positive facultative anaerobic rod and provide its complete annotated genome sequence according to the taxonogenomics concept. Its genome is 3 755 739 bp long and contains 3453 protein-coding genes and 64 RNA genes, including eight rRNA genes.

Keywords: Bacillaceae, Culturomics, gut microbiota, *Numidum massiliense* genome, taxonogenomics
Original Submission: 15 April 2016; **Revised Submission:** 10 May 2016; **Accepted:** 12 May 2016
Article published online: 17 May 2016

Corresponding author: J.-C. Lagier, Aix-Marseille Université, URMITE, UM63, CNRS7278, IRD198, Inserm 1095, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille Cedex 05, France
E-mail: jclagier@yahoo.fr

Introduction

Several microbial ecosystems are harboured by the human body, among which is the human gut microbiota. This particular ecosystem is so vast that its cell count (10^{14} cells) is evaluated at ten times the number of human cells in the human body, and its collective bacterial genome size is 150 times larger than the human genome [1–4]. Over the years, with the evolution of exploratory techniques of microbial ecosystems from culture to metagenomics, the gut microbiota has been shown to be involved in many conditions such as obesity, inflammatory bowel disease and irritable bowel disease [1]. It has also been shown to play key roles in digestion as well as metabolic and immunologic functions [1–3]. A better knowledge of the gut microbiota's composition is thus required for an improved understanding of its functions.

In order to extend the gut microbiota repertoire and bypass the noncultivable bacteria issue, the culturomics concept was developed in order to cultivate as exhaustively as possible the viable population of a bacterial ecosystem; it consists in the multiplication of culture conditions, as well as varying of media, temperature and atmosphere [5]. Using this technique, strain mt3^T was isolated and identified as a previously unknown member of the *Bacillaceae* family. Currently there are 53 validated genera in the *Bacillaceae* family. This family was created by Fisher in 1895 (http://www.bacterio.net/*Bacillaceae*.html). The genus *Bacillus* was described as its type genus. The genera that belong to this family are rod shaped, mostly aerobic and facultative anaerobic bacteria. They are found in various ecosystems like the human body, soil, water, air and other environmental ecosystems [6].

Bacterial classification is currently based on a polyphasic approach with phenotypic and genotypic characteristics such as DNA-DNA hybridization, G+C content and 16S rRNA sequence similarity [7–9]. Nevertheless, this classification system has its limits, among which is the high cost of the DNA-DNA hybridization technique and its low reproducibility [7,10]. With the recent development of genome sequencing technology [11], a new concept of bacterial description was developed in our laboratory [12–16]. This taxonogenomics

http://dx.doi.org/10.1016/j.nmni.2016.05.009

concept [17] combines a proteomic description with the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) profile [18] associated with a phenotypic description and the sequencing, annotation and comparison of the complete genome of the new bacterial species [19].

We describe strain mt3^T, a new genus *Numidum massiliense* gen. nov., sp. nov. (= CSUR P1305 = DSM 29571), a new member of the *Bacillaceae* family using the concept of taxonogenomics.

Materials and Methods

Organism information

A stool sample was collected from a healthy Tuareg boy living in Algeria. Verbal consent was obtained from the patient, and the study was approved by the Institut Fédératif de Recherche 48, Faculty of Medicine, Marseille, France, under agreement 09-022.

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

The sample was cultured using the 18 culture conditions of culturomics [20]. The colonies were obtained by seeding on solid medium, purified by subculture and identified using MALDI-TOF MS [18,21]. Colonies were deposited in duplicate on a MTP 96 MALDI-TOF MS target plate (Bruker Daltonics, Leipzig, Germany), which was analysed with a Microflex spectrometer (Bruker). The 12 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). An identification score over 1.9 with a validated species allows identification at the species level, and a score under 1.7 does not enable any identification. When identification by MALDI-TOF MS failed, the 16S rRNA was sequenced [22]. Stackebrandt and Ebers [23] suggest similarity levels of 98.7% and 95% of the 16s rRNA sequence as a threshold to define, respectively, a new species and a new genus without performing DNA-DNA hybridization.

Growth conditions

In order to determine our strain's ideal growth conditions, different temperatures (25, 28, 37, 45 and 56°C) and atmospheres (aerobic, microerophilic and anaerobic) were tested. GENbag anaer and GENbag miroaer systems (bioMérieux, Marcy l'Étoile, France) were used to respectively test anaerobic and microaerophilic growth. Aerobic growth was achieved with and without 5% CO_2.

Morphologic, biochemical and antibiotic susceptibility testing

Gram staining, motility, catalase, oxidase and sporulation were tested as previously described [20]. Biochemical description was performed using API 20 NE, ZYM and 50CH (bioMérieux) according to the manufacturer's instructions. Cellular fatty acid methyl ester (FAME) analysis was performed by gas chromatography/mass spectrometry (GC/MS). Two samples were prepared with approximately 70 mg of bacterial biomass per tube collected from several culture plates. FAMEs were prepared as previously described (http://www.midi-inc.com/pdf/MIS_Technote_101.pdf). GC/MS analyses were carried out as described before [24]. Briefly, fatty acid methyl esters were separated using an Elite 5-MS column and monitored by mass spectrometry (Clarus 500-SQ 8 S; Perkin Elmer, Courtaboeuf, France). Spectral database search was performed using MS Search 2.0 operated with the Standard Reference Database 1A (NIST, Gaithersburg, MD, USA) and the FAMEs mass spectral database (Wiley, Chichester, UK).

Antibiotic susceptibility testing was performed using the disk diffusion method according to European Committee on Antimicrobial Susceptibility Testing (EUCAST) 2015 recommendations (http://www.eucast.org/). To perform the negative staining of strain mt3^T, detection Formvar-coated grids were deposited on a 40 µL bacterial suspension drop, then incubated at 37°C for 30 minutes and on ammonium molybdate 1% for 10 seconds. The dried grids on blotted paper were observed with a Tecnai G20 transmission electron microscope (FEI Company, Limeil-Brevannes, France).

Growth conditions and genomic DNA preparation

N. massiliense strain mt3^T (= CSUR P1305 = DSM 29571) was grown on 5% sheep's blood–enriched Columbia agar (bioMérieux) at 37°C in aerobic atmosphere. Bacteria grown on three petri dishes were collected and 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 of *N. massiliense* 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 (Life Technologies, Carlsbad, CA, USA) to 66.2 ng/μL. The mate pair library was prepared with 1 μg of genomic DNA using the Nextera mate pair Illumina guide. The genomic DNA 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 11 kb, with an optimal size at 3.927 kb. No size selection, was performed and 505 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal at 597 bp on a Covaris device S2 in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies), and the final concentration library was measured at 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 were performed in a single 39-hour run in a 2 × 251 bp read length.

Genome annotation and comparison

Open reading frames (ORFs) were predicted using Prodigal [25] with default parameters, but the predicted ORFs were excluded if they were spanning a sequencing gap region (contains N). The predicted bacterial protein sequences were searched against the Clusters of Orthologous Groups (COGs) using BLASTP (*E* value 1e-03, coverage 70%, identity percent 30%). If no hit was found, it searched against the NR database using BLASTP with an *E* value of 1e-03 coverage 70% and identity percent of 30%. If sequence lengths were smaller than 80 amino acids, we used an *E* value of 1e-05. The tRNAScanSE tool [26] was used to find tRNA genes, whereas rRNAs were found by using RNAmmer [27]. Lipoprotein signal peptides and the number of transmembrane helices were predicted using Phobius [28]. ORFans were identified if all the performed BLASTP procedures did not give positive results (*E* value smaller than 1e-03 for ORFs with sequence size upper than 80 aa or *E* value smaller than 1e-05 for ORFs with sequence length smaller 80 aa). Such parameter thresholds have already been used in previous works to define ORFans.

Genomes were automatically retrieved from the 16s RNA tree using Xegen software (Phylopattern [29]). For each selected genome, complete genome sequence, proteome and ORFeome genome sequence were retrieved from the National Center for Biotechnology Information FTP site. All proteomes were analysed with proteinOrtho [30]. Then for each couple of genomes, a similarity score was computed. This score is the mean value of nucleotide similarity between all couples of orthologues between the two genomes studied (AGIOS) [19]. 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 the compared strains, we determined two parameters: digital DNA-DNA hybridization (dDDH), which exhibits a high correlation with DNA-DNA hybridization (DDH) [31,32], and AGIOS [19], which was designed to be independent from DDH.

Results

Strain identification and phylogenetic analyses

Strain mt3^T (Table 1) was first isolated in April 2014 by a preincubation of 21 days in brain–heart infusion supplemented with 5% sheep's blood and cultivated on 5% sheep's blood–enriched Colombia agar (bioMérieux) in an aerobic atmosphere at 37°C.

No significant score was obtained for strain mt3^T using MALDI-TOF MS, thus suggesting that our isolate's spectrum did not match any spectra in our database. The nucleotide sequence of the 16S r RNA of strain mt3^T (GenBank accession no. LK985385) showed a 90.5% similarity level with *Bacillus firmus,* the phylogenetically closest species with a validly published name (Fig. 1), therefore defining it as a new genus within the *Bacillaceae* family named *Numidum massiliense* (= CSUR P1305 = DSM29571). *N. massiliense* spectra (Fig. 2) were added as reference spectra to our database. The reference spectrum for *N. massiliense* was then compared to the spectra of phylogenetically close species, and the differences were exhibited in a gel view (Fig. 3).

Phenotypic description

Growth was observed from 25 to 56°C on blood-enriched Columbia agar (bioMérieux), with optimal growth being

TABLE 1. Classification and general features of *Numidum massiliense* strain mt3^T

Property	Term
Current classification	Domain: *Bacteria* Phylum: *Firmicutes* Class: *Bacilli* Order: *Bacillales* Family: *Bacillaceae* Genus: *Numidum* Species: *Numidum massiliense* Type strain: mt3
Gram stain	Positive
Cell shape	Rod
Motility	Nonmotile
Sporulation	Sporulating
Temperature range	Mesophilic
Optimum temperature	37°C

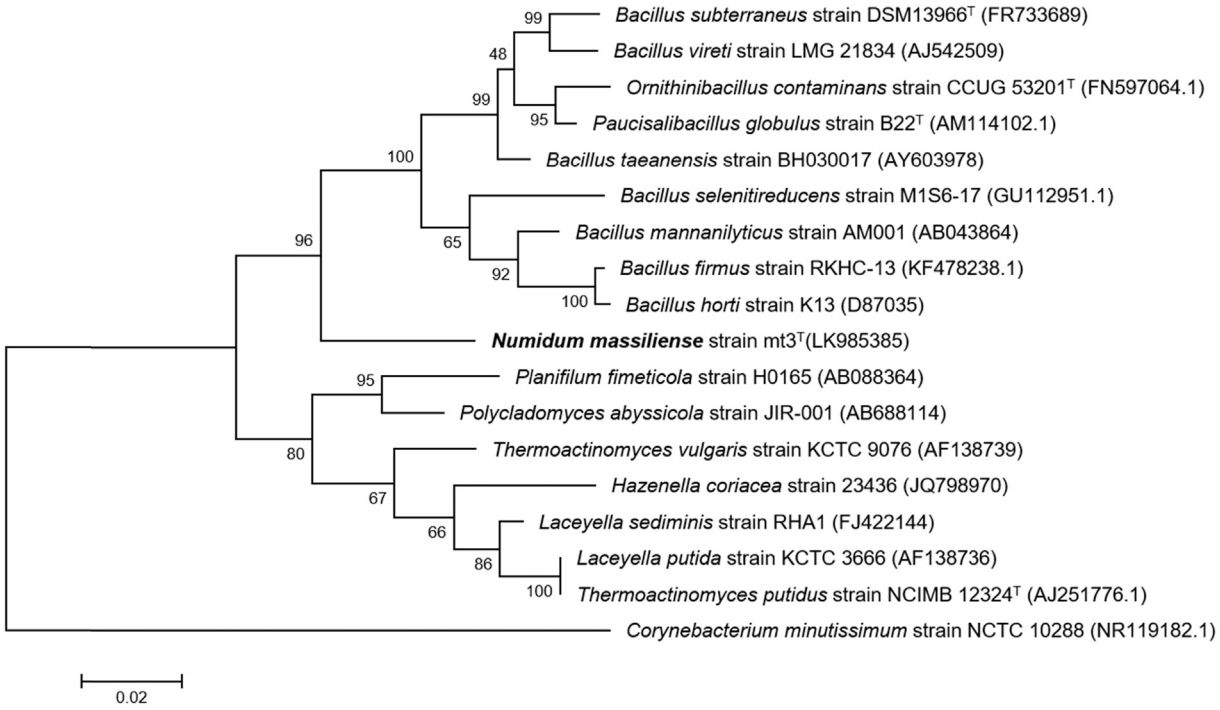

FIG. 1. Phylogenetic tree highlighting position of *Numidum massiliense* gen. nov., sp. nov. strain mt3^T (= CSUR P1305 = DSM 29571) relative to other strains within family *Bacillaceae*. Scale bar represents 1% nucleotide sequence divergence.

obtained aerobically at 37°C after 48 hours of incubation. Weak cell growth was observed under microaerophilic and anaerobic conditions. The cells were nonmotile and sporulating. Cells were Gram-positive rods (Fig. 4) and formed greyish colonies with a mean diameter of 10 mm on blood-enriched Columbia agar. Under electron microscopy, the bacteria had a mean diameter of 0.5 µm and length of 2.7 µm (Fig. 5).

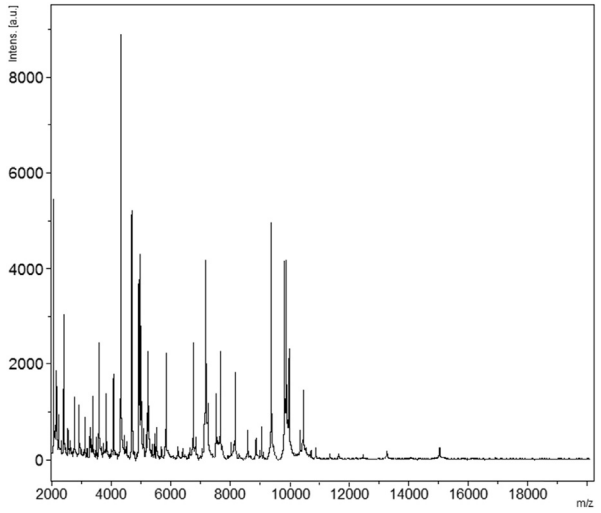

FIG. 2. Reference mass spectrum from *Numidum massiliense* strain mt3^T. Spectra from 12 individual colonies were compared and reference spectrum was generated.

The major fatty acid by far is the branched 13-methyl-tetradecanoic acid (88%). Other fatty acids are described with low abundances (below 6%). The majority of them were branched fatty acids (Table 2).

Strain mt3^T was positive for catalase and negative for oxidase. Alkaline phosphatase, esterase (C4), esterase lipase (C8), leucine arylamidase, trypsin, α-chymotrypsin, acid phosphatase, β-galactosidase, β-glucuronidase, α-glucosidase, β-glucosidase, protease and N-acetyl-β-glucosaminidase activities were exhibited. Nitrates were reduced into nitrites. D-Ribose, D-xylose, D-mannose, D-galactose, D-fructose, D-glucose, D-mannitol, N-acetylglucosamine, amygdalin, esculin ferric citrate, D-maltose, D-lactose, D-trehalose and D-tagatose and adipic acid were metabolized.

Cells were susceptible to doxycycline, ceftriaxone, gentamicin 500 µg, ticarcillin/clavulanic acid, rifampicin, teicoplanin, metronidazole and imipenem. Resistance was exhibited against erythromycin, colistin/polymyxin B, ciprofloxacin, penicillin, trimethoprim/sulfamethoxazole, nitrofurantoin and gentamicin 15 µg.

The biochemical and phenotypic features of strain mt3^T were compared to the corresponding features of other close representatives of the *Bacillaceae* family (Table 3).

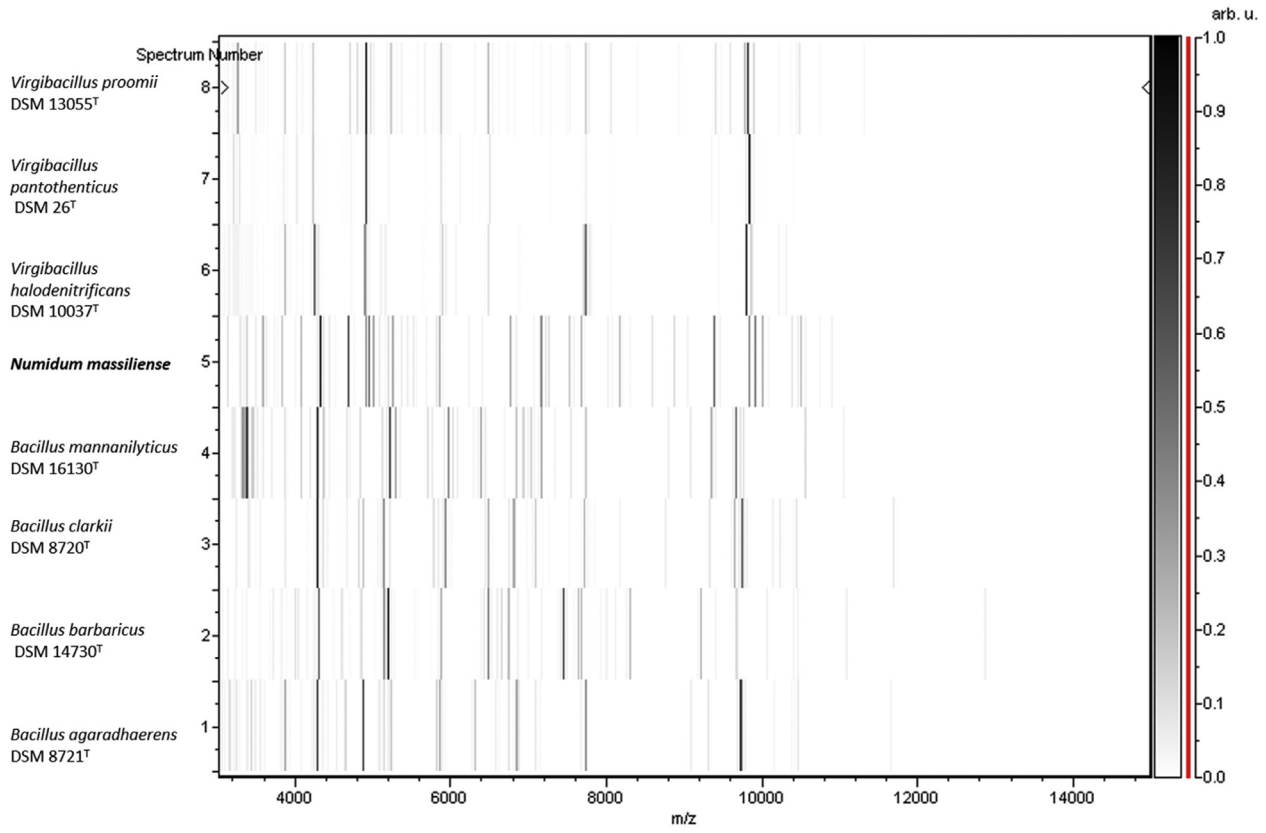

FIG. 3. Gel view comparing *Numidum massiliense* strain mt3^T (= CSUR P1305 = DSM 29571) to other species within *Bacillaceae* family. Gel view displays raw spectra of loaded spectrum files arranged in pseudo-gel-like look. X-axis records m/z value. Left y-axis displays running spectrum number originating 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 and peak intensity in arbitrary units. Displayed species are indicated on left.

Genome properties

The genome of *N. massiliense* strain mt3^T is 3 757 266 bp long with a 52.05% G+C content (Table 4, Fig. 6). Of the 3513 predicted genes, 3448 were protein-coding genes and 65 were RNAs (three genes are 5S rRNA, four genes are 16S rRNA,

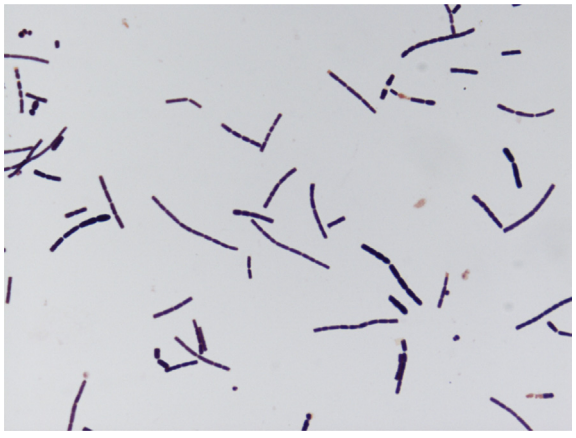

FIG. 4. Gram staining of *Numidum massiliense* strain mt3^T.

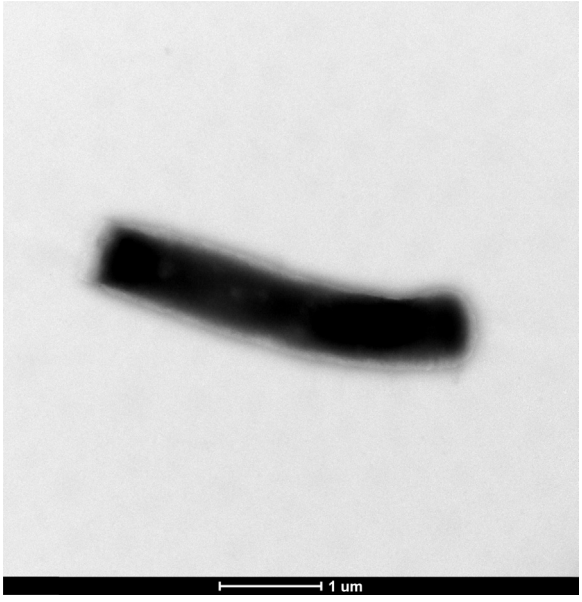

FIG. 5. Transmission electron microscopy of *Numidum massiliense* strain mt3^T, using Morgani 268D (Philips, Amsterdam, The Netherlands) at operating voltage of 60 kV. Scale bar represents 1 µm.

TABLE 2. Cellular fatty acid composition (%)

Fatty acid	IUPAC name	Mean relative %[a]
15:0 iso	13-methyl-tetradecanoic acid	87.6 ± 1.6
15:0 anteiso	12-methyl-tetradecanoic acid	5.5 ± 0.4
17:0 iso	15-methyl-Hexadecanoic acid	3.0 ± 0.9
16:0	Hexadecanoic acid	0.9 ± 0.1
16:0 iso	14-methyl-Pentadecanoic acid	0.6 ± 0.1
18:1n9	9-Octadecenoic acid	0.6 ± 0.2
13:0 anteiso	10-methyl-Dodecanoic acid	TR
5:0 iso	3-methyl-butanoic acid	TR
17:0 anteiso	14-methyl-Hexadecanoic acid	TR
18:0	Octadecanoic acid	TR
18:2n6	9,12-Octadecadienoic acid	TR
14:0 iso	12-methyl-Tridecanoic acid	TR
14:0	Tetradecanoic acid	TR

IUPAC, International Union of Pure and Applied Chemistry; TR, trace amounts (<1%).
[a]Mean peak area percentage ± standard deviation.

TABLE 4. Nucleotide content and gene count levels of genome

	Genome (total)	
Attribute	Value	% of total[a]
Size (bp)	3 757 266	100
G+C content (bp)	1 955 657	52.05
Coding region (bp)	3 181569	84.67
Total genes	3513	100
RNA genes	65	1.85
Protein-coding genes	3448	98.14
Genes with function prediction	2570	73.15
Genes assigned to COGs	2314	65.86
Genes with peptide signals	229	6.51
Genes with transmembrane helices	977	27.81

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.

two genes are 23S rRNA and 56 genes are tRNA genes). A total of 2570 genes (73.15%) were assigned as putative function (by COGs or by NR blast). Four hundred twelve genes were identified as ORFans (11.93%). The remaining 503 genes were annotated as hypothetical proteins (14.57%). The National Center for Biotechnology Information ID project is PRJEB8811, and the genome is deposited under accession number CTDZ01000000. The distribution of genes into COGs functional categories is presented in Table 5.

Genome comparison

N. massiliense genomic characteristics were compared to other close species (Table 6).

The draft genome sequence of *N. massiliense* strain mt3^T (3.76 MB) is smaller than the draft genome sequences of *Bacillus vireti* LMG 21834, *Bacillus mannanilyticus* JCM 10596, *Paucisalibacillus globulus* DSM 18846 *and Bacillus subterraneus* DSM 13966^T (5.29, 4.53, 4.24 and 3.9 MB respectively) and larger than those of *Bacillus selenitireducens* MLS10 and *Laceyella sacchari* 1-1 (3.59 and 3.32 MB respectively). The G+C content of *N. massiliense* (52.05%) is larger than the G+C contents of *L. sacchari* 1-1, *B. selenitireducens* MLS10, *B. subterraneus* DSM 13966^T, *B. vireti* LMG 21834, *B. mannanilyticus* JCM 10596 and *P. globulus* DSM 18846 (48.9, 48.7, 42.1, 39.7, 39.6 and 35.8% respectively).

TABLE 3. Differential characteristics of *Numidum massiliense* strain mt3^T, *Bacillus mannanilyticus* strain DSM 16130, *Virgibacillus pantothenticus* strain ATCC 14576, *Virgibacillus dokdonensis* DSM 16826, *Ornithinibacillus contaminans* DSM 22953, *Bacillus polygoni* strain NCIMB 14282^T, *Bacillus agaradhaerens* strain DSM 8721, *Paucisalibacillus globulus* strain LMG 23148^T, *Bacillus barbaricus* strain DSM 14730^T and *Virgibacillus koreensis* strain JCM 12387^T [33–41]

Property	*N. massiliense*	*B. mannanilyticus*	*V. pantothenticus*	*V. dokdonensis*	*O. contaminans*	*B. polygoni*	*B. agaradhaerens*	*P. globulus*	*B. barbaricus*	*V. koreensis*
Cell diameter (μm)	0.5–0.6	0.6–0.8	0.5–0.7	0.6–0.8	0.8–1	0.4–0.5	0.5–0.6	0.5	0.5	0.5–0.7
Oxygen requirement	+	+	+	–	+	+	+	+	+	–
Gram stain	+	+/–	+	+/–	+	+	NA	+	+	+
Salt requirement	–	–	+	–	–	+	+	–	–	+
Motility	–	+	+	+	–	–	NA	–	–	+
Endospore formation	+	+	+	+	+	+	+	+	+	+
Indole	–	–	–	–	NA	–	NA	NA	–	–
Production of:										
Alkaline phosphatase	+	NA	NA	–	NA	NA	NA	NA	NA	NA
Catalase	+	+	+	+	+	+	–	+	+	+
Oxidase	–	–	NA	+	+	–	NA	–	–	+
Nitrate reductase	+	–	+/–	–	NA	+	+	–	–	–
Urease	–	NA	NA	–	NA	NA	–	–	–	–
β-Galactosidase	+	NA	NA	–	NA	NA	NA	NA	NA	+
N-acetyl-glucosamine	+	NA	+	–	NA	NA	+	+	+	+
Acid from:										
L-Arabinose	–	–	–	–	NA	NA	+	–	–	+
Ribose	+	NA	+	+	–	–	NA	–	+	NA
Mannose	+	+	+	+	–	+	+	+	+	+
Mannitol	+	+	–	–	w	+	+	+	–	+
Sucrose	–	+	+/–	+	NA	+	+	+	–	NA
D-Glucose	+	+	+	+	+	+	+	+	+	–
D-Fructose	+	+	+	+	NA	+	+	+	+	+
D-Maltose	+	+	+	+	NA	+	+	+	+	+
D-Lactose	+	+	+/–	+	NA	–	NA	+	+/–	NA
Habitat	Human gut	Industry	Soil	Seawater	Blood	Indigo balls	Industry	Soil	Paint	Salt

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

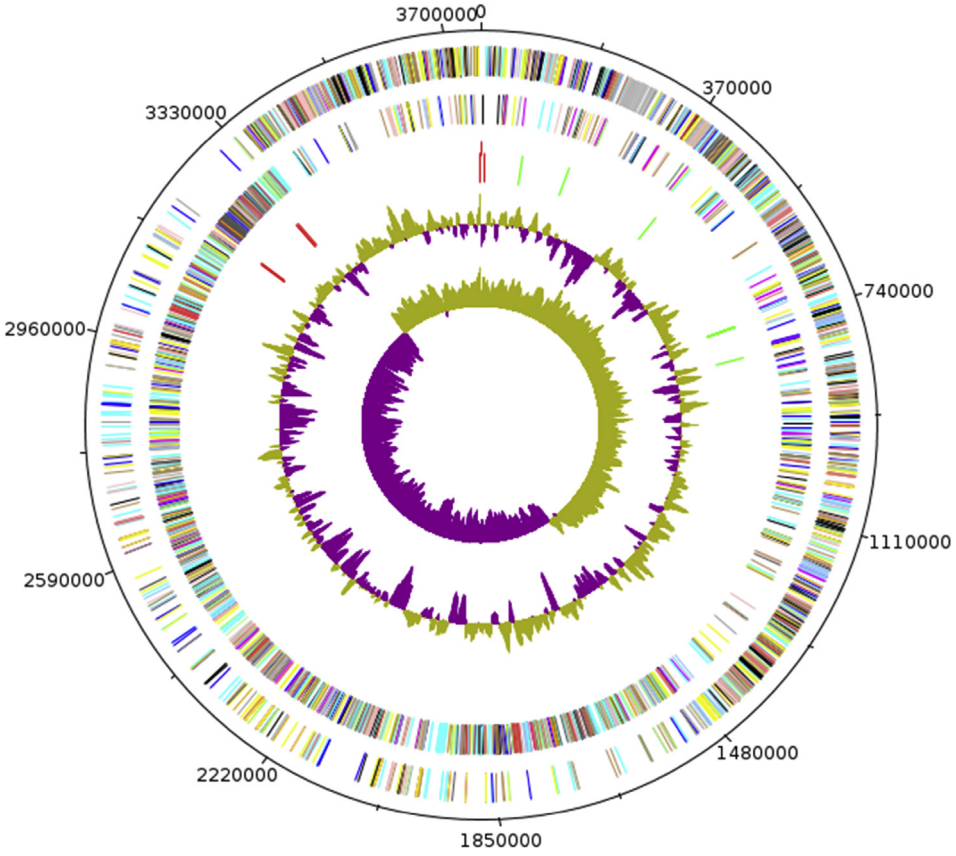

FIG. 6. Graphical circular map of chromosome. From outside to center: Genes on forward strain coloured by COGs categories (only gene assigned to COGs), RNA genes (tRNAs green, rRNAs red), GC content and GC skew.

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

Code	Value	% of total[a]	Description
J	150	4.35	Translation
A	0	0	RNA processing and modification
K	247	7.16	Transcription
L	169	4.90	Replication, recombination and repair
B	1	0.03	Chromatin structure and dynamics
D	30	0.87	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	85	2.47	Defense mechanisms
T	123	3.57	Signal transduction mechanisms
M	143	4.15	Cell wall/membrane biogenesis
N	8	0.23	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	34	0.99	Intracellular trafficking and secretion
O	93	2.70	Posttranslational modification, protein turnover, chaperones
C	156	4.52	Energy production and conversion
G	234	6.79	Carbohydrate transport and metabolism
E	278	8.06	Amino acid transport and metabolism
F	63	1.83	Nucleotide transport and metabolism
H	95	2.76	Coenzyme transport and metabolism
I	117	3.39	Lipid transport and metabolism
P	163	4.73	Inorganic ion transport and metabolism
Q	75	2.18	Secondary metabolites biosynthesis, transport and catabolism
R	370	10.73	General function prediction only
S	269	4.80	Function unknown
—	2903	84.19	Not in COGs

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

The gene content of *N. massiliense* (3513) is smaller than the gene contents of *B. vireti* LMG 21834, *B. mannanilyticus* JCM 10596, *P. globulus* DSM 18846 and *B. subterraneus* DSM 13966^T (5050, 4369, 4127 and 3772 respectively) but larger than those of *B. selenitireducens* MLS10 and *L. sacchari* 1-1 (3368 and 3256 respectively).

TABLE 6. Genome comparison of closely related species to *Numidum massiliense* strain mt3^T

Organism	INSDC	Size (Mb)	G+C (%)	Total genes
Numidum massiliense strain mt3^T	CTDZ00000000.1	3.76	52.05	3513
Bacillus vireti strain LMG 21834	ALAN00000000.1	5.29	39.7	5050
Bacillus mannanilyticus JCM 10596	BAMO00000000.1	4.53	39.6	4369
Paucisalibacillus globulus DSM 18846	AXVK00000000.1	4.24	35.8	4127
Bacillus subterraneus DSM 13966^T	JXIQ00000000.1	3.9	42.1	3772
Bacillus selenitireducens strain MLS10	CP001791.1	3.59	48.7	3368
Laceyella sacchari strain 1-1	ASZU00000000.1	3.32	48.9	3256

INSDC, International Nucleotide Sequence Database Collaboration.

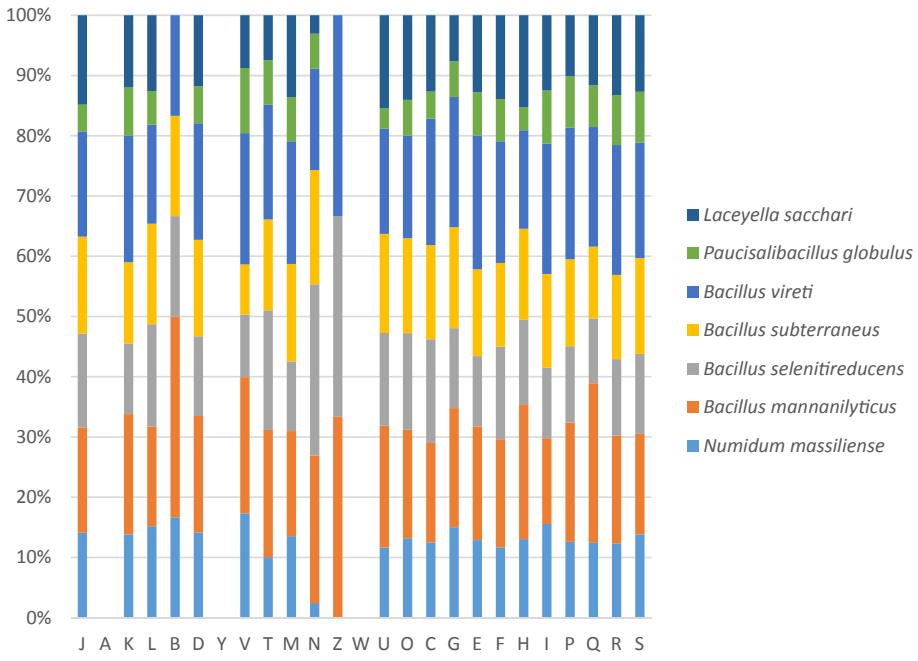

FIG. 7. Distribution of functional classes of predicted genes according to clusters of orthologous groups of proteins.

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

	Numidum massiliense	*Bacillus mannanilyticus*	*Bacillus selenitireducens*	*Bacillus subterraneus*	*Bacillus vireti*	*Laceyella sacchari*	*Paucisalibacillus globulus*
N. massiliense	**3453**	1162	1028	1191	1294	1121	456
B. mannanilyticus	53.11	**3710**	1194	1369	1471	1174	511
B. selenitireducens	55.11	54.58	**3212**	1301	1318	972	461
B. subterraneus	54.81	56.4	58.15	**3648**	1632	1141	558
B. vireti	54.39	56.91	57.56	66.1	**4963**	1244	656
L. sacchari	57.96	54.52	55.66	55.94	55.59	**3152**	412
P. globulus	52.26	55.47	54.19	58.18	58.89	52.72	**4000**

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

However, the distribution of genes into COGs categories was similar in all compared genomes except for those corresponding to the cytoskeleton category, which were only present in *B. vireti, B. selenitireducens* and *B. mannanilyticus* (Fig. 7). *N. massiliense* strain mt3^T shared 1162, 1028, 1191, 1294, 1121 and 456 orthologous genes with *B. mannanilyticus, B. selenitireducens, B. subterraneus, B. vireti, L. sacchari* 1-1 and *P. globulus* respectively (Table 7). Among species with standing in nomenclature, AGIOS values ranged from 52.26% between *N. massiliense* and *P. globulus* to 66.1% between *B. vireti* and *B. subterraneus*. When *N. massiliense* was compared to the other species, AGIOS values ranged from 52.26% with *P. globulus* to

TABLE 8. Pairwise comparison of *Bacillus niameyensis* with eight other species using GGDC, formula 2 (DDH estimates based on identities/HSP length)[a]

	Numidum massiliense	*Bacillus mannanilyticus*	*Bacillus selenitireducens*	*Bacillus subterraneus*	*Bacillus vireti*	*Laceyella sacchari*	*Paucisalibacillus globulus*	*Corynebacterium minutissimum*
N. massiliense	100% ± 00	2.52% ± 0.13	2.53% ± 0.15	2.52% ± 0.14	2.52% ± 0.13	2.52% ± 0.13	2.52% ± 0.16	2.52% ± 0.21
B. mannanilyticus		100% ± 00	2.53% ± 0.10	2.52% ± 0.18	2.52% ± 0.13	2.52% ± 0.12	2.52% ± 0.19	2.52% ± 0.20
B. selenitireducens			100% ± 00	2.53% ± 0.13	2.52% ± 0.12	2.53% ± 0.16	2.52% ± 0.11	2.52% ± 0.21
B. subterraneus				100% ± 00	2.55% ± 0.23	2.52% ± 0.18	2.52% ± 0.21	2.52% ± 0.21
B. vireti					100% ± 00	2.52% ± 0.17	2.52% ± 0.23	2.52% ± 0.20
L. sacchari						100% ± 00	2.52% ± 0.06	2.52% ± 0.21
P. globulus							100% ± 00	2.52% ± 00
C. minutissimum								100% ± 00

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

57.96% with *L. sacchari*. To evaluate the genomic similarity among the compared strains, dDDH was also determined (Table 8).

Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Numidum massiliense* which contains the type strain mt3^T. This bacterial strain has been isolated from the faecal flora of a Tuareg boy living in Algeria.

Description of *Numidum* gen. nov.

Numidum (nu.mi'dum, from Numidum, which relates to a nomad people from Africa), is a Gram-positive, sporulating, facultative anaerobic bacilli. Optimal growth in aerobic condition at 37°C. Catalase positive and oxidase negative. Nitrates were reduced into nitrites. It is urease negative. The type strain is *Numidum massiliense* strain mt3^T.

Description of *Numidum massiliense* strain mt3^T gen. nov., sp. nov.

Numidum massiliense (mas.il'ien'se. L. gen. masc., massiliense, of Massilia, the Latin name of Marseille, where strain mt3^T was isolated) cells have a mean diameter of 0.5 µm. Colonies are greyish and 10 mm in diameter on 5% sheep's blood–enriched Columbia agar (bioMérieux). Positive reactions are observed for alkaline phosphatase, esterase (C4), esterase lipase (C8), leucine arylamidase, trypsin, α-chymotrypsin, acid phosphatase, β-galactosidase, β-glucuronidase, α-glucosidase and N-acetyl-β-glucosaminidase. D-Ribose, D-xylose, D-mannose, D-galactose, D-fructose, D-glucose, D-mannitol, N-acetylglucosamin, amygdalin, esculin ferric citrate, D-maltose, D-lactose, D-trehalose and D-tagatose and adipic acid were metabolized.

Cells were susceptible to doxycycline, ceftriaxone, gentamicin 500 µg, ticarcillin/clavulanic acid, rifampicin, teicoplanin, metronidazole and imipenem.

The G+C content of the genome is 52.05%. The 16S rRNA gene sequence and whole-genome shotgun sequence of *N. massiliense* strain mt3^T are deposited in GenBank under accession numbers LK985385 and CTDZ01000000, respectively. The type strain mt3^T (= CSUR P1305 = DSM 29571) was isolated from the stool of a Tuareg boy living in Algeria.

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.

Conflict of Interest

None declared.

References

[1] Salazar N, Arboleya S, Valdès L, Stanton C, Ross P, Ruiz L, et al. The human intestinal microbiome at extreme ages of life. Dietary intervention as a way to counteract alterations. Front Genet 2014;5:406.
[2] Sankar SA, Lagier JC, Pontarotti P, Raoult D, Fournier PE. The human gut microbiome, a taxonomic conundrum. Syst Appl Microbiol 2015;38:276–86.
[3] Simpson HL, Campbell BJ. Review article: dietary fibre-microbiota interactions. Aliment Pharmacol Ther 2015;42:158–76.
[4] Xu Z, Knight R. Dietary effects on human gut microbiome diversity. Br J Nutr 2015;113(Suppl.):S1–5.
[5] 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.
[6] Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, et al. Bergey's manual of systematic bacteriology. The firmicutes. 2nd ed., vol. 3. New York: Springer; 2009.
[7] Rosselló-Móra R. DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation. In: Stackebrandt E, editor. Molecular identification, systematics, and population structure of prokaryotes. Berlin: Springer; 2006. p. 23–50.
[8] Wayne LG, Brenner DJ, Colwell PR, Grimont PAD, Kandler O, Krichevsky MI, et al. Report of the ad hoc committee on reconciliation of approaches to bacterial systematic. Int J Syst Bacteriol 1987;37:463–4.
[9] Woese R, Kandler O, Wheelis ML. Towards a natural system of organisms: proposals for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 1990;87:4576–9.
[10] Viale AM, Arakaki AK, Soncini FC, Ferreyra RG. Evolutionnary relationships among eubacterial groups as inferred from GroEL (Chaperonin) sequences comparison. Int J Syst Bacteriol 1944;44:527–33.
[11] Reddy TB, Thomas AD, Stamatis D, Bertsch J, Isbandi M, Jansson J, et al. The Genome OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. Nucleic Acids Res 2015;43(Database issue):D1099–106.
[12] Caputo A, Lagier JC, Azza S, Robert C, Mouelhi D, Fournier PE, et al. *Microvirga massiliensis* sp. nov., the human commensal with the largest genome. Microbiologyopen 2016;5:307–22.
[13] Tidjani Alou M, Rathored J, Khelaifia S, Michelle C, Brah S, Diallo BA, et al. *Bacillus rubiinfantis* sp. nov. strain mt2^T, a new bacterial species isolated from human gut. New Microbes New Infect 2015;8:51–60.
[14] Seck E, Rathored J, Khelaifia S, Croce O, Robert C, Couderc C, et al. *Virgibacillus senegalensis* sp. nov., a new moderately halophilic bacterium isolated from human gut. New Microbes New Infect 2015;8:116–26.
[15] Hugon P, Mishra AK, Lagier JC, Nguyen TT, Couderc C, Raoult D, et al. Non-contiguous finished genome sequence and description of *Brevibacillus massiliensis* sp. nov. Stand Genomic Sci 2013;8:1–14.
[16] 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.

[17] Fournier PE, Lagier JC, Dubourg G, Raoult D. From culturomics to taxonogenomics: a need to change the taxonomy of prokaryotes in clinical microbiology. Anaerobe 2015;36:73–8.
[18] 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.
[19] 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(Pt 2):384–91.
[20] 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.
[21] 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.
[22] 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.
[23] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152–5.
[24] Dione N, Sankar SA, Lagier JC, Khelaifia S, Michele C, Armstrong N, et al. Genome sequence and description of *Anaerosalibacter massiliensis* sp. nov. New Microbes New infect 2016;10:66–76.
[25] 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.
[26] 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.
[27] 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.
[28] Käll L, Krogh A, Sonnhammer EL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol 2004;338:1027–36.
[29] Gouret P, Thompson JD, Pontarotti P. PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinformatics 2009;10:298.
[30] 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.
[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;2:117–37.
[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] Aino K, Hirato K, Matsuno T, Morita N, Nodosaka Y, Fujiwara T, et al. *Bacillus polygoni* sp. nov., a moderately halophilic, non-motile obligate alkaliphile isolated from indigo balls. Int J Syst Evol Microbiol 2008;58:120–4.
[34] Heyndrickx M, Lebbe L, Kersters K, De Vos P, Forsyth G, Logan NA. Virgibacillus: a new genus to accommodate *Bacillus pantothenticus* (Prom and Knight 1950). Emended description of *Virgibacillus pantothenticus*. Int J Syst Evol Microbiol 1998;48:99–106.
[35] Lee JS, Lim JM, Lee KC, Lee JC, Park YH, Kim CJ. *Virgibacillus koreensis* sp. nov., a novel bacterium from a salt field, and transfer of *Virgibacillus picturae* to the genus *Oceanobacillus picturae* comb. nov. with emended descriptions. Int J Syst Evol Microbiol 2006;56:251–7.
[36] Nielsen P, Fritze D, Priest FG. Phenetic diversity of alkaliphilic *Bacillus* strains: proposal for nine species. Microbiology 1995;141:1745–61.
[37] 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.
[38] Nunes I, Tiago I, Pires AL, Da Costa MS, Verissimo A. *Paucisalibacillus globulus* gen. nov., sp. nov., a Gram-positive bacterium isolated from potting soil. Int J Syst Evol Microbiol 2006;56:1841–5.
[39] Täubel M, Kämpfer P, Buczolits S, Lubitz W, Busse HJ. *Bacillus barbaricus* sp. nov. isolated from an experimental wall painting. Int J Syst Evol Microbiol 2003;53:725–30.
[40] Yoon JH, Kang SJ, Lee SY, Lee MH, Oh TK. *Virgibacillus dokdonensis* sp. nov., isolated from a Korean island, Dokdo, located at the edge of the East Sea in Korea. Int J Syst Evol Microbiol 2005;55:1833–7.
[41] Kämpfer P, Falsen E, Lodders N, Langer S, Busse HJ, Schumann P. *Ornithinibacillus contaminans* sp. nov., an endospore-forming species. Int J Syst Evol Microbiol 2010;60:2930–4.