ORIGINAL ARTICLE

Noncontiguous finished genome sequences and description of *Bacillus massiliglaciei, Bacillus mediterraneensis, Bacillus massilinigeriensis, Bacillus phocaeensis* and *Bacillus tuaregi,* five new species identified by culturomics

F. Cadoret[1], M. T. Alou[1], P. Afouda[1], I. S. Traore[1], L. Bréchard[1], C. Michelle[1], F. Di Pinto[1], C. Andrieu[1], J. Delerce[1], A. Levasseur[1], P.-E. Fournier[1] and D. Raoult[1,2]

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

Abstract

Microbial culturomics, which investigates microbial diversity by combining diversified culture conditions, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and 16S rDNA identification, allowed to identify five new species within the *Bacillus* genus. *Bacillus massiliglaciei* strain Marseille-P2600^T, *Bacillus mediterraneensis* strain Marseille-P2384^T, *Bacillus massilinigeriensis* strain Marseille-P2366^T, *Bacillus tuaregi* strain Marseille-P2489^T and *Bacillus phocaeensis* strain SIT16^T are each the type strain of the corresponding bacterial species. These strains, the genomes of which are described here, are facultative anaerobic Gram-positive bacilli. Here, we describe the main characteristics of each bacterium and present their complete genome sequence and annotation.

Keywords: *Bacillus massiliglaciei*, *Bacillus massilinigeriensis*, *Bacillus mediterraneensis*, *Bacillus phocaeensis*, *Bacillus tuaregi*, culturomics, emerging bacteria, human microbiota, taxonogenomics
Original Submission: 1 December 2016; **Revised Submission:** 6 April 2017; **Accepted:** 13 April 2017
Article published online: 20 April 2017

Corresponding author: D. Raoult, Aix-Marseille Université, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), CNRS 7278, IRD 198, INSERM 1095, UM63, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine, 27 Boulevard Jean Moulin, 13385, Marseille cedex 5, France.
E-mail: didier.raoult@gmail.com

Introduction

The genus *Bacillus* (Cohn, 1872), classified among the *Firmicutes,* was created in 1872 [1]. *Bacillus* species are strictly aerobic or anaerobic-tolerant rod-shaped bacteria that are able to form endospores [2–4]. Bacilli colonize a wide range of environments (soil, water) and human organisms. Several *Bacillus* species present a biotechnologic interest because of their metabolism, and some of them, as *Bacillus thuringiensis,* are known to be pathogenic for human beings [3]. Since the creation of the *Bacillus* genus, 221 new species with validly published names standing in nomenclature have been identified (Fig. 1). The development of PCR techniques at the end of the 1980s and now faster genome sequencing allow the number of *Bacillus* species identified to significantly increase; it also permits some strains to be reclassified as only one species because their description was based only on phenotypic observation. In this study, we used a new approach including genome sequencing, matrix-assisted desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) spectrum and main phenotypic characteristics [5–9] to describe five new *Bacillus* species.

Bacillus massiliglaciei strain Marseille-P2600^T, *Bacillus mediterraneensis* strain Marseille-P2384^T, *Bacillus massilinigeriensis* strain Marseille-P2366^T, *Bacillus tuaregi* strain Marseille-P2489^T and *Bacillus phocaeensis* strain SIT16^T are the type strains of the corresponding species. There are all Gram-positive bacilli and facultatively anaerobic. They were respectively isolated from a Siberian permafrost sample (*B. massiliglaciei*), the stool sample of a healthy Senegalese boy (*B. mediterraneensis*), the stool sample of a healthy Nigerien girl (*B. tuaregi* and *B. massilinigeriensis*) and

http://dx.doi.org/10.1016/j.nmni.2017.04.005

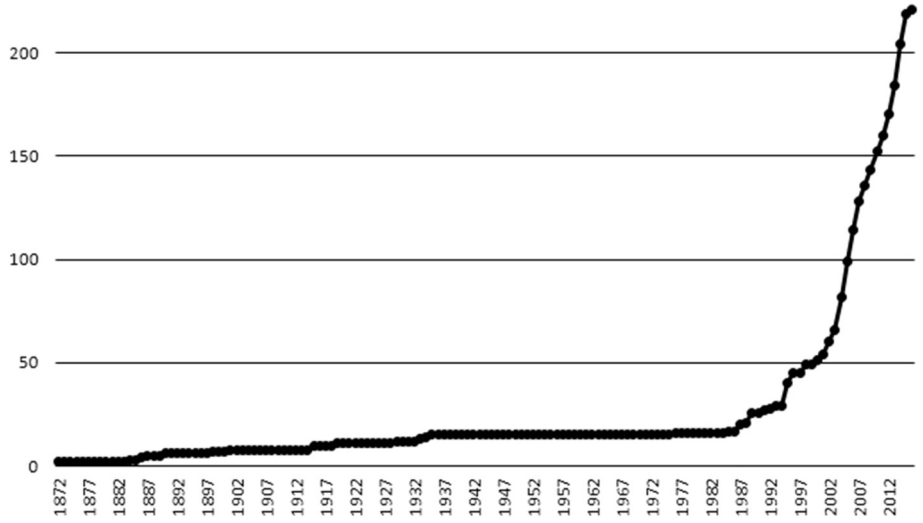


FIG. 1. Numbers of *Bacillus* species identifications per year, since the first one, *Bacillus cereus,* described by Cohn et al. in 1872 [1].

the stool sample of a Senegalese boy with kwashiorkor (*B. phocaeensis*). These isolations were part of the culturomics study aimed at exploring microbial diversity using multiple culture conditions [8].

Phylogenetic relationships based on the 16S ribosomal RNA gene have been used to classify these strains among the *Bacillus* genus. However, our study uses a new concept of bacterial description combining a proteomics analysis with the MALDI-TOF MS profile [8], associated with phenotypic and genomic descriptions of these five new species.

Here we present a summary of the classification, main features and complete genomic sequencing and annotation of *Bacillus massiliglaciei* strain Marseille-P2600^T (= Collection de souches de l'Unité des Rickettsies (CSUR) P2600 = Deutsche Sammlung von Mikroorganismen (DSM) 102861), *Bacillus mediterraneensis* strain Marseille-P2384^T (= CSUR P2384 = DSM 102091), *Bacillus massilinigeriensis* strain Marseille-P2366^T (= CSUR P2366 = DSM 102112), *Bacillus tuaregi* strain Marseille-P2489^T (= CSUR P2489 = DSM 103460) and *Bacillus phocaeensis* strain SIT16^T (= CSUR P2184 = CCUG 69739). These characteristics support the creation of the subsequent five new species.

Material and Methods

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

According to the culturomics approach, 18 conditions have been tested on samples to isolate these strains, as previously described by Lagier et al. [9]. The origin of the samples and conditions of isolation are summarized in Table 1. All the human samples were obtained after the child's parent's approval, and the study was approved by the Institut Fédératif de Recherche 48, Faculty of Medecine, Marseille, France, under agreement number 09-022.

Purified colonies were then identified by MALDI-TOF MS using a Microflex LT spectrometer and a MSP 96 MALDI-TOF target plate (Bruker Daltonics, Bremen, Germany), as previously described [6]. The obtained spectra were imported into

TABLE 1. Information for five *Bacillus* species

Characteristic	*B. massiliglaciei*	*B. mediterraneensis*	*B. massilinigeriensis*	*B. tuaregi*	*B. phocaeensis*
Strain	Marseille-P2600T	Marseille-P2384T	Marseille-P2366T	Marseille-P2489T	SIT16T
Sample origin	Permafrost	Human stool	Human stool	Human stool	Human stool
Patient information	—	Healthy Senegalese boy	Healthy Nigerien girl	Healthy Nigerien girl	Senegalese boy with kwashiorkor
Authorization/consent	No. 09-022 (IFR 48, Marseille)	No. 09-022 (IFR 48, Marseille), consent of boy's parents	No. 09-022 (IFR 48, Marseille), consent of girl's parents	No. 09-022 (IFR 48, Marseille), consent of girl's parents	No. 09-022 (IFR 48, Marseille), consent of boy's parents
Storage	−80°C	−80°C	−80°C	−80°C	−80°C
Isolation conditions	COS medium day 15 aerobic 19°C	Hemoculture + rumen + blood day 7 aerobic 37°C	Hemoculture + rumen day 7 aerobic 37°C	Marine medium day 7 aerobic 37°C	Marine medium day 15 aerobic 37°C

MALDI Biotyper 3.0 software (Bruker) and analyzed by standard pattern matching (with default parameter settings) against the main spectra of the 7537 bacteria included in the Bruker and Unité des Maladies Infectieuses et Tropicales Emergentes (URMITE) databases (the URMITE database is constantly updated). The resulting score enabled the identification (or not) of tested species: a score of $\geq$2 with a validly published species enabled identification at the species level, a score of $\geq$1.7 but <2 enabled identification at the genus level and a score of <1.7 did not enable any identification. No significant scores were obtained for our strains, suggesting that our isolates were not members of known species.

Consequently, sequencing of 16S rRNA gene was realized in order to identify these strains. DNA was previously extracted by EZ1 DNA Tissue Kit using BioRobot EZ1 Advanced XL (Qiagen, Courtaboeuf, France). The amplification and purification of the 16S rRNA gene was done as previously described by means of the universal primer pair fD1 and rP2 (Eurogentec, Angers, France). Sequencing was then done using the Big Dye Terminator v1.1 Cycle Sequencing kit and ABI Prism 3230xl Genetic Analyzer capillary sequencer (Applied Biosystems; Thermo Fisher Scientific Life Sciences, Waltham, MA, USA), as previously described [10]. The 16S rRNA nucleotide sequences were assembled and corrected using CodonCode Aligner software (http://www.codoncode.com), and the BLASTn searches were performed against the GenBank National Center for Biotechnology Information (NCBI) database (http://blast.ncbi.nlm.nih.gov.gate1.inist.fr/Blast.cgi) to determine the percentage of similarity with the closest bacteria. A similarity threshold of <98.7% allows the definition of a new species, whereas a threshold of <95% allows the definition of a new genus without performing DNA-DNA hybridization (DDH) [11].

Phylogenetic analysis

A custom Python script was used to automatically retrieve all species from the same family of the new species and download 16S sequences from NCBI by parsing NCBI eUtils results and the NCBI taxonomy page. It only keeps sequences from type strains. In case of multiple sequences for one type strain, it selects the sequence obtaining the best identity rate from the BLASTn alignment with our sequence. The script then separates 16S sequences into two groups: one containing the sequences of strains from the same genus (group A) and one containing the others (group B). It finally considers all *Bacillus* strains with valid names from group A and the closest one from group B.

Phenotypic, biochemical and antibiotic susceptibility tests

The ideal growth conditions of our strains were determined by testing five growth temperatures (20, 25, 30, 37 and 45°C) in an aerobic atmosphere with or without 5% CO_2, and under anaerobic and microaerophilic conditions using the GENbag anaer and GENbag microaer systems respectively (bioMérieux, Marcy l'Étoile, France). Phenotypic characteristics such as Gram staining, motility, sporulation, and catalase and oxidase activities were tested as previously described [12]. Negative staining was done in order to observe cellular morphology. Cells were fixed with 2.5% glutaraldehyde in 0.1 M cacodylate buffer for at least 1 hour at 4° C. A drop of cell suspension was deposited for approximately 5 minutes on glow-discharged formvar carbon film on 400 mesh nickel grids (FCF400-Ni, EMS). The grids were dried on blotting paper, and the cells were negatively stained for 10 seconds with 1% ammonium molybdate solution in filtered water at room temperature. Electron micrographs were acquired with a Tecnai G20 Cryo (FEI Company, Limeil-Brevannes, France) transmission electron microscope operated at 200 keV.

Biochemical analysis of strains Marseille-P2600^T, Marseille-P2366^T, Marseille-P2384^T, Marseille-P2489^T and SIT16^T was carried out using API 50CH, API 20NE and API ZYM strips according to the manufacturer's instructions (bioMérieux).

Cellular fatty acid methyl ester (FAME) analysis was performed by gas chromatography/mass spectrometry (GC/MS). FAMEs were prepared, and GC/MS analyses were carried out as described previously [13,14]. Briefly, FAMEs were separated using an Elite 5-MS column and monitored by mass spectrometry (Clarus 500–SQ 8 S; Perkin Elmer, Courtaboeuf, France). A spectral database search was performed using MS Search 2.0 operated with the Standard Reference Database 1A (National Institute of Standards and Technology, Gaithersburg, MD, USA) and the FAME mass spectral database (Wiley, Chichester, UK). Antibiotic susceptibility was tested by Etest (bioMérieux).

Genome sequencing, annotation and comparison

Genomic DNA (gDNA) of strain SIT16^T and strain Marseille-P2489^T were first extracted through a mechanical treatment by acid-washed glass beads (G4649–500g; Sigma-Aldrich, St. Louis, MO, USA) using a FastPrep BIO 101 instrument (Qbiogene, Strasbourg, France) at maximum speed (6.5) for 3 × 30 seconds. Then for all the others strains 2 hours' lysozyme incubation at 37°C was done, and gDNA was extracted on the EZ1 biorobot (Qiagen) with EZ1 DNA tissues kit. The elution volume was 50 µL. gDNA was quantified by a Qubit assay with a high-sensitivity kit (Life Technologies, Carlsbad, CA, USA) (Supplementary Table S1).

gDNA 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).

The mate pair library was prepared with 1.5 µg of gDNA using the Nextera mate pair Illumina guide. The gDNA sample

was simultaneously fragmented and tagged with a mate pair junction adapter. The pattern of 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 1.5 to 11 kb (with an optimal size at 5.6, 8.032, 7.491, 6.747 and 8.680 kb for strains SIT16^T, Marseille-P2366^T, Marseille-P2384^T, Marseille-P2600^T and Marseille-P2489^T respectively). No size selection was performed, and 600 ng (450 ng for Marseille-P2366^T) of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments (with an optimal size at 1049, 883, 961 and 1079 bp for SIT16^T, Marseille-P2366^T, Marseille-P2384^T and Marseille-P2600^T, and with optima on a bimodal curve at 822 and 1823 bp for Marseille-P2489^T) on the Covaris device S2 in T6 tubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies), and the final concentration library was measured (Supplementary Table S1).

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 in a 2 × 151 bp (2 × 261 bp for Marseille-P2366^T and Marseille-P2384^T). The paired reads were finally trimmed and assembled. Complementary information is available in Supplementary Table S1.

Open reading frames (ORFs) were predicted using Prodigal [15] with default parameters, but the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched against the GenBank database [16] and the Clusters of Orthologous Groups database (COGs) database using BLASTP (*E* value 1e-03, coverage 0.7 and identity percentage of 30%). If no hit was found, it was searched against the NR database using BLASTP with an *E* value of 1e-03, coverage of 0.7 and an identity percentage of 30%, and if the sequence length was smaller than 80 amino acids (aa), we used an *E* value of 1e-05. The tRNAScanSE tool [17] was used to find tRNA genes, while ribosomal RNAs were found using RNAmmer [18]. Lipoprotein signal peptides and the number of transmembrane helices were predicted using Phobius [19]. Mobile genetic elements were predicted using PHAST [20] and RAST [21]. 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 [22] and DNAPlotter [23] 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 [24]. Closest species for genomic comparisons were identified in the 16S RNA tree using the PhyloPattern software [25]. For each selected genome, the complete genome sequence, proteome genome sequence and Orfeome genome sequence were retrieved from the NCBI FTP site. An annotation of the entire proteome was performed to define the distribution of functional classes of predicted genes according to the COGs of proteins (using the same method as for the genome annotation). Annotation and comparison processes were performed by the multiagent software system DAGOBAH [26], which includes Figenix [27] libraries which provide pipeline analysis. To evaluate the genomic similarity between studied genomes, we determined two parameters, digital DDH (dDDH), which exhibits a high correlation with DDH [28], and average genomic identity of orthologous gene sequences (AGIOS), which was designed to be independent from DDH. The AGIOS score is the mean value of nucleotide similarity between all couples of orthologous proteins and between the two studied genomes [29].

Results

Strain identification and phylogenetic analysis

The phylogenetic trees of our strains are provided in Fig. 2. All five studied strains present a similarity threshold of <98.7%, allowing us to define them as new species. Strain Marseille-P2600^T (accession no. LT223699) revealed 98% sequence similarity with the 16S rRNA of *Bacillus foraminis* strain CV53, the closest species with a validly published name. We therefore suggested that our strain is a representative strain of a new species within the genus *Bacillus* for which we suggest the name '*Bacillus massiliglaciei*' strain Marseille-P2600^T (= CSUR P2600 = DSM 102861). Strain Marseille-P2366^T (accession no. LT161887) revealed 97% sequence similarity with the 16S rRNA of *Bacillus acidicola* strain 105-2, the closest species with a validly published name. We therefore suggested that our strain is a representative strain of a new species within the genus *Bacillus* for which we suggest the name '*Bacillus massilinigeriensis*' strain Marseille-P2366^T (= CSUR P2366 = DSM 102112). Strain Marseille-P2384^T (accession no. LT16188) revealed 98% sequence similarity with the 16S rRNA of *Bacillus foraminis* strain CV53, the closest species with a validly published name. Strains Marseille-P2600^T and Marseille-P2384^T present 97% 16S rRNA sequence similarity. We therefore suggested that our strain is a representative strain of a new species within the genus *Bacillus* for which we suggest the name '*Bacillus mediterraneensis*' strain Marseille-P2384^T (= CSUR P2384 = DSM 102091). Strain Marseille-P2489^T (accession no. LT223701) revealed 97% sequence similarity with the 16S rRNA of *Bacillus benzoevorans* strain DSM

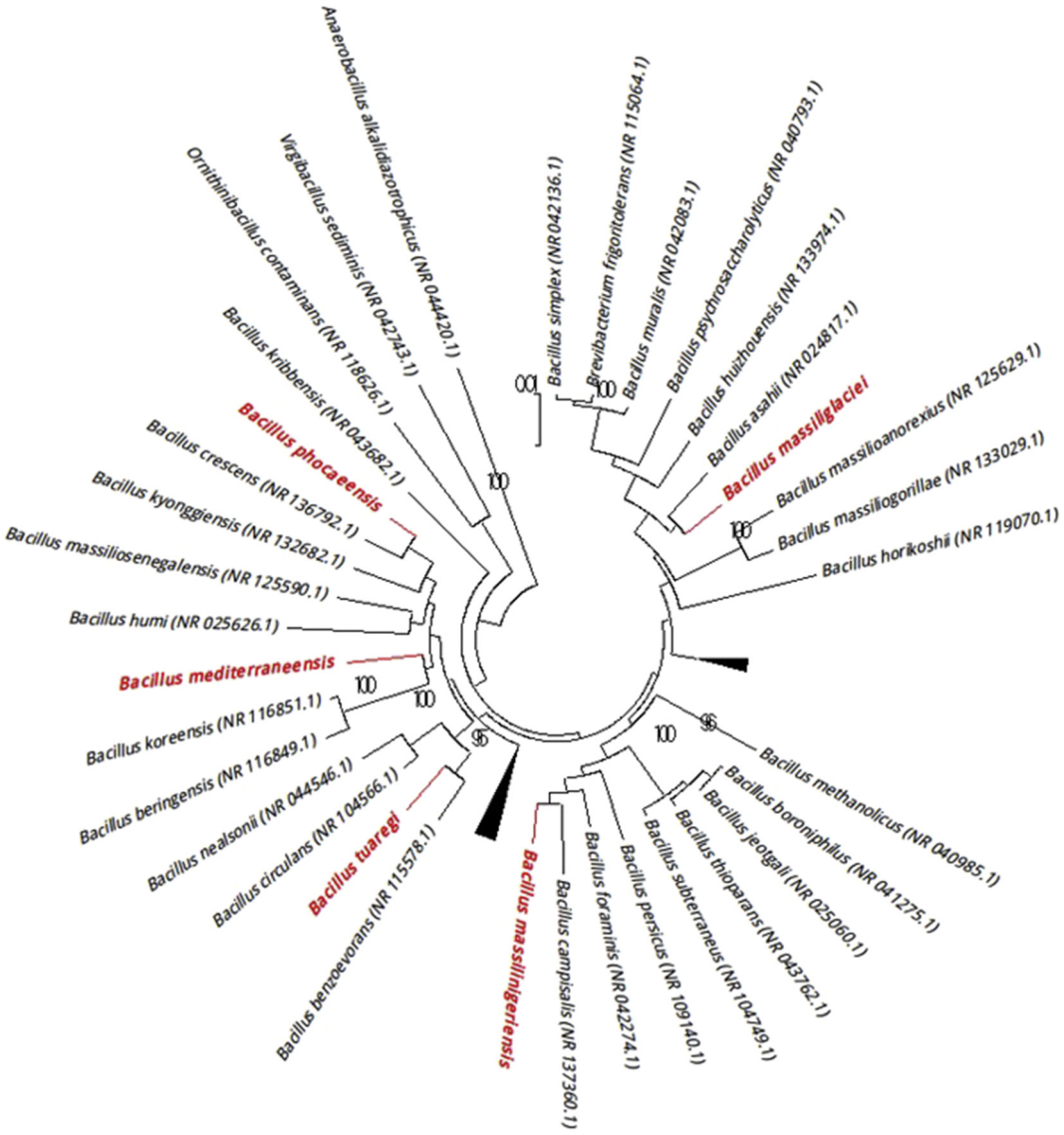
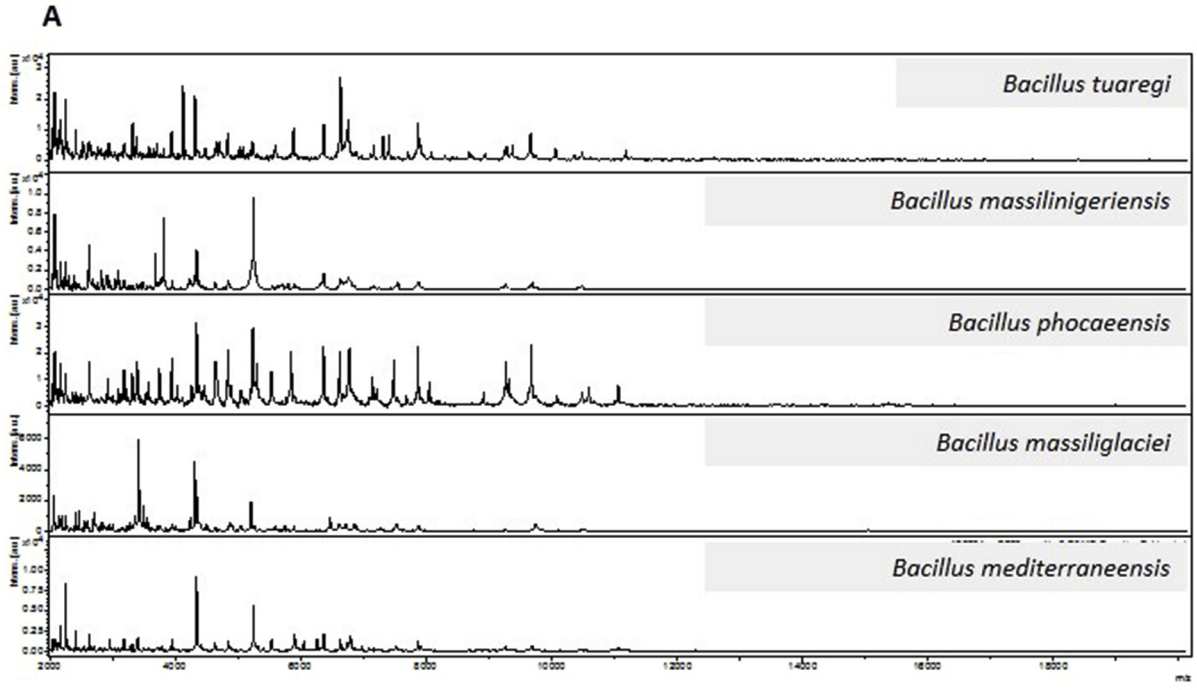


FIG. 2. Phylogenetic tree highlighting position of *Bacillus massiliglaciei* strain Marseille-P2600^T, *Bacillus mediterraneensis* strain Marseille-P2384^T, *Bacillus massilinigeriensis* strain Marseille-P2366^T, *Bacillus tuaregi* strain Marseille-P2489^T and *Bacillus phocaeensis* strain SIT16^T relative to other type strains within *Bacillus* genus. Strains and their corresponding GenBank accession numbers for 16S rRNA genes are indicated in brackets. Sequences were aligned by Clustal W (http://www.clustal.org/clustal2/), and phylogenetic inferences were obtained by maximum-likelihood method within MEGA6 (http://www.megasoftware.net/mega.php). Numbers at nodes are percentages of bootstrap values obtained by repeating analysis 1000 times to generate majority consensus tree. *Anoxybacillus vitaminiphilus* (NR_108379) was used as outgroup. Scale bar = 0.005% nucleotide sequence divergence. Due to high number of *Bacillus* species, only branches which included our studied strains are represented. Groups which are included in these branches but are not informative are indicated by black pictograms.

5391, the closest species with a validly published name. We therefore suggested that our strain is a representative strain of a new species within the genus *Bacillus* for which we suggest the name '*Bacillus tuaregi*' strain Marseille-P2489^T (= CSUR P2489 = DSM 103460). Strain SIT16^T (accession no. LN881595) revealed 96% sequence similarity with the 16S rRNA of *Bacillus acidicola* strain 105-2, the closest species with a validly published name. Strains Marseille-P2366^T and SIT16^T present 97% 16S rRNA sequence similarity. We therefore suggested that our strain is a representative strain of a new species within the genus *Bacillus* for which we suggest the name '*Bacillus phocaeensis*' strain SIT16^T (= CSUR P2184 = CCUG 69739).

All the spectra (Fig. 3A) of these new species have been integrated into the URMITE database (http://www.



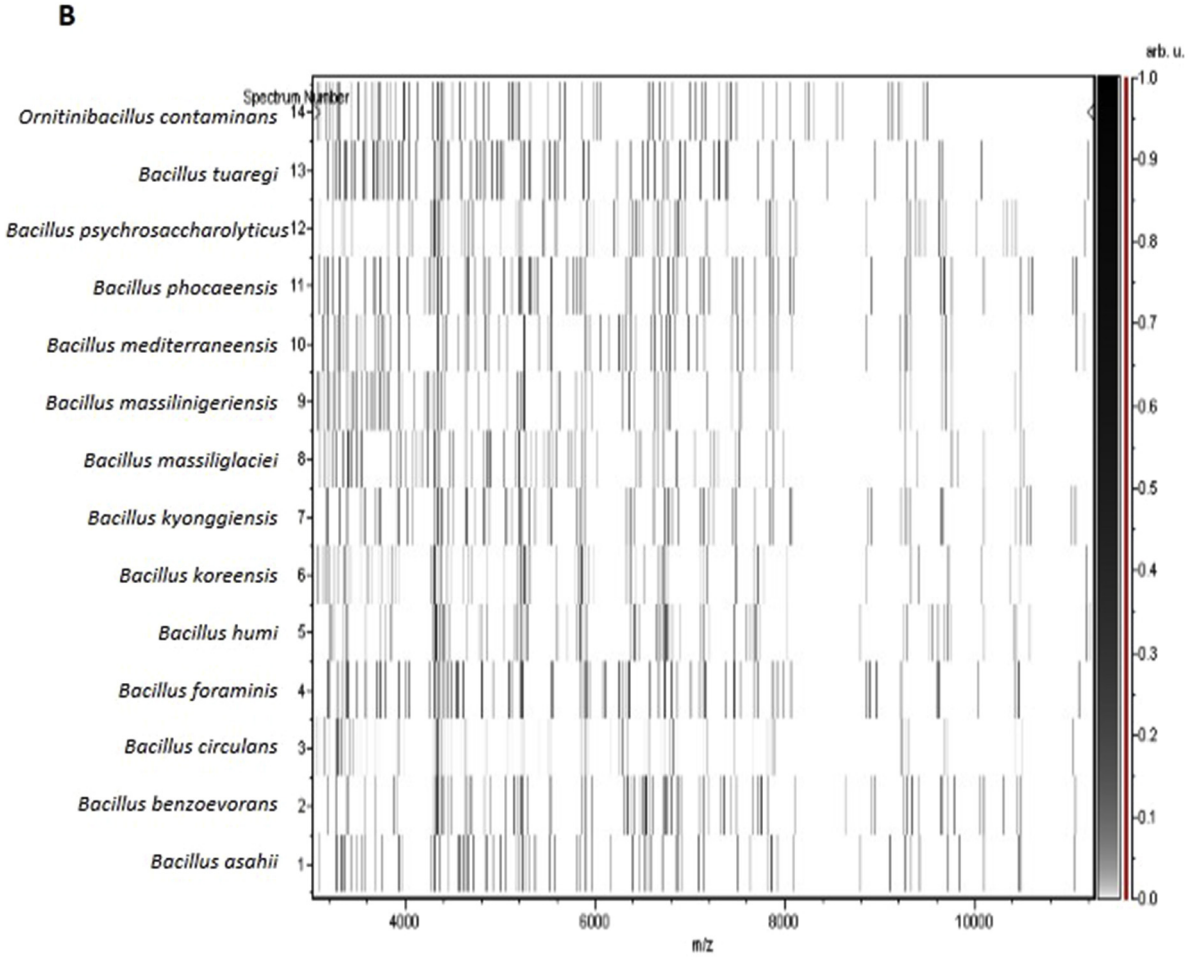


TABLE 2. Comparison of phenotypic characteristics of five *Bacillus* species as well as *Bacillus foraminis* [30], *Bacillus acidicola* [31] and *Bacillus nealsonii* [32]

Property	*B. massiliglaciei*	*B. massilinigeriensis*	*B. foraminis*	*B. mediterraneensis*	*B. phocaeensis*	*B. acidicola*	*B. tuaregi*	*B. nealsonii*
Optimal temperature	19°C	37°C	40°C	37°C	37°C	37°C	37°C	32°C
Atmosphere	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic
Colony aspect	Small, smooth, convex and grey	Smooth and grey	Small, smooth, convex and grey	Smooth and grey	Smooth and grey	Smooth, shiny, circular and slight yellow tinted	Small, smooth and grey	Beige, irregular
Cell shape	Rod shaped	Rod shaped	Rod shaped	Rod shaped	Rod shaped	Rod shaped	Rod shaped	Rod shaped
Cell size (µm)	2.0–5.0	3.5–5.0	2.4–3.9	2.0–5.0	2.5–4.0	3–5.9	3.0–5.0	4.0–5.0
Cell diameter (µm)	0.5–0.6	0.5	1.0	0.4–0.6	0.5–0.7	1.0–1.6	0.5–1.0	1.0
Gram stain	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive
Motility	+	+	ND	+	+	+	+	+
Endospore formation	+	–	–	+	+	+	–	+
Major cellular fatty acid	15:0 iso	15:0 iso	15:0 iso	15:0 iso	15:0 iso	15:0 iso	14:0 iso	15:0 anteiso
Production of:								
Alkaline phosphatase	–	–	ND	–	–	ND	–	+
Catalase	–	+	–	–	+	+	–	+
Oxidase	+	–	–	–	+	–	–	–
Nitrate reductase	–	–	+	–	–	ND	–	+
Urease	–	–	+	–	–	ND	–	–
β-Galactosidase	–	–	+	–	–	ND	–	+
N-acetyl-glucosamine	–	–	+	+	–	ND	–	+
Acid from:								
L-Arabinose	–	–	+	+	–	ND	+	–
Ribose	–	+	+	+	–	ND	+	–
Mannose	–	–	+	+	–	ND	–	+
Mannitol	+	–	+	+	–	+	+	–
D-Saccharose	–	–	+	+	–	ND	–	ND
D-Glucose	+	+	+	+	+	+	+	+
D-Fructose	–	–	+	+	+	+	+	ND
D-Maltose	–	–	+	+	+	+	+	ND
D-Lactose	–	–	+	–	+	v	–	ND
Habitat	Permafrost	Human gut	Alkaline groundwater	Human gut	Human gut	Sphagnum peat bogs	Human gut	Soil

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

mediterranee-infection.com/article.php?laref=266&titre=urms-database). A comparison of the proteomic profiles was made between our strains and their closest species (Fig. 3B). An analysis of the gel view shows that all the profiles of our studied strains share similar general characteristics with the other *Bacillus* species used for comparison. Furthermore, the outsider species *Ornithinibacillus contaminans* profile shows several unique differences.

Phenotypic features

In Table 2, data from our five new species are compared to published data of close species: *Bacillus foraminis* type strain CV53 [30], *Bacillus acidicola* type strain 105-2 [31] and *Bacillus nealsonii* type strain DSM 15077 [32]. Results show that morphologic description and notably cell shape and colonies' aspect support that these five new species are members of the *Bacillus* genus. These morphologic observations have been allowed by electronic microscopy (Fig. 4) which reveals a similar morphology. These results are supported by API profiles (Supplementary Tables S2–S4) which are coherent for *Bacillus* species.

The cellular fatty acid compositions of our strains are provided in Table 3; antibiotic analyses are presented in Table 4. We can observe in Table 3 that the major cellular fatty acid in all the presented *Bacillus* strains is 13-methyl-tetradecanoic acid, except for strain Marseille-P2489^T, for which 12-methyl-tridecanoic acid is the most abundant cellular fatty acid. However, 13-methyl-tetradecanoic acid is the second most abundant. In Table 4, the minimum inhibitory concentrations (in µg/mL) of five antibiotics classically used with Gram-positive bacteria are presented. We notice that the five *Bacillus* species show very weak minimum inhibitory concentrations, as described for the

FIG. 3. Reference mass spectra from *Bacillus tuaregi* strain Marseille-P2489^T, *Bacillus massilinigeriensis* strain Marseille-P2366^T, *Bacillus phocaeensis* strain SIT16^T, *Bacillus massiliglaciei* strain Marseille-P2600^T and *Bacillus mediterraneensis* strain Marseille-P2384^T. Spectra from 12 individual colonies were compared and each reference spectrum generated (A). Gel view comparing *Bacillus tuaregi* strain Marseille-P2489^T, *Bacillus massilinigeriensis* strain Marseille-P2366^T, *Bacillus phocaeensis* strain SIT16^T, *Bacillus massiliglaciei* strain Marseille-P2600^T and *Bacillus mediterraneensis* strain Marseille-P2384^T to other species within genus *Bacillus*. Gel view displays raw spectra of loaded spectrum files arranged in pseudo-gel-like look. *x*-axis records *m*/*z* value, left *y*-axis running spectrum number originating from subsequent spectra loading. Peak intensity expressed by greyscale scheme code. Color bar and right *y*-axis indicate relation between color peaks, with peak intensity in arbitrary units. Displayed species are indicated at left (B).

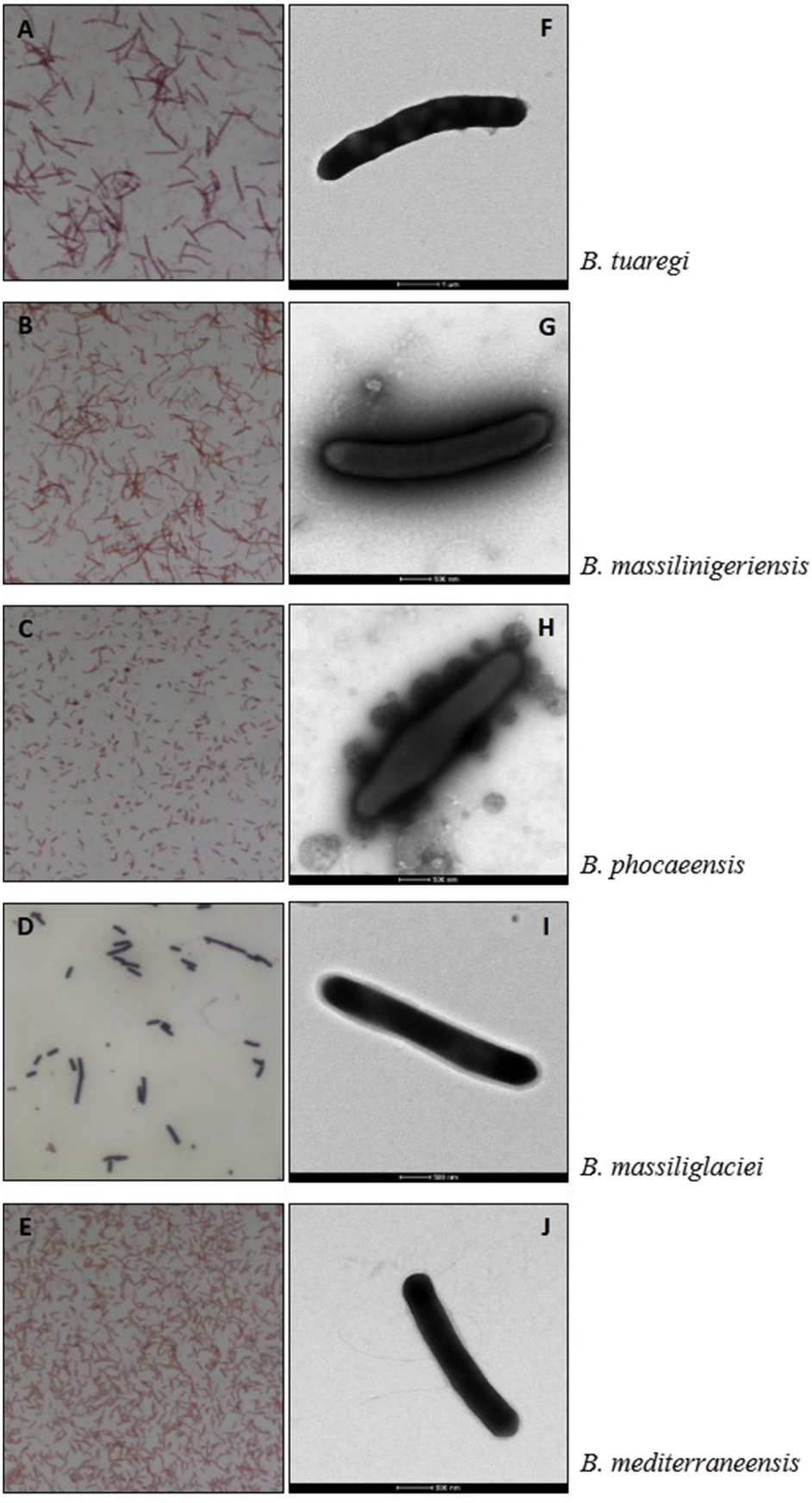


TABLE 3. Cellular fatty acid composition (in %[a]) for five *Bacillus* species

Fatty acids	IUPAC name	*B. massiliglaciei*	*B. massilinigeriensis*	*B. mediterraneensis*	*B. phocaeensis*	*B. tuaregi*
5:0 iso	3-methyl-butanoic acid	TR	1.3 ± 0	9 ± 0.9	No	No
14:0.	Tetradecanoic acid	TR	TR	1.6 ± 0	No	7 ± 0.9
14:0 iso	12-methyl-Tridecanoic acid	4.2 ± 0	4.8 ± 0.2	1.6 ± 0.2	4.4 ± 0.4	41 ± 2.0
15:0.	Pentadecanoic acid	No	TR	No	No	No
15:0 iso	13-methyl-tetradecanoic acid	56.3 ± 0.2	45.0 ± 0	47.7 ± 1.6	43.4 ± 1.4	2.4 ± 0.5
15:0 anteiso	12-methyl-tetradecanoic acid	32.3 ± 0.4	25.2 ± 0.8	19.0 ± 0.5	31.0 ± 2.0	7.5 ± 0.7
15:1n5 iso	13-Methyltetradec-9-enoic acid	1.0 ± 0	No	No	No	No
16:0.	Hexadecanoic acid	TR	8.3 ± 0	5 ± 0.6	3.8 ± 1.0	39.2 ± 1.2
16:0 iso	14-methyl-Pentadecanoic acid	TR	3.7 ± 0	3.0 ± 0.2	9.6 ± 0.8	1.7 ± 0.3
16:1n6 iso	14-Methylpentadec-9-enoic acid	1.3 ± 0	No	No	No	No
16:1n7	9-Hexadecenoic acid	TR	No	No	No	No
16:1n9	7-Hexadecenoic acid	No	No	3.4 ± 0.3	TR	1.0 ± 0.7
16:1n9 iso	14-methyl-7-Hexadecenoic acid	No	No	No	TR	No
16:0 9,10-methylene	2-octyl-Cyclopropaneoctanoic acid	No	No	2.5 ± 0	No	No
17:0	Heptadecanoic acid	No	TR	No	No	No
17:0 iso	15-methyl-Hexadecanoic acid	TR	TR	2.5 ± 0	1.8 ± 0.6	No
17:0 anteiso	14-methyl-Hexadecanoic acid	TR	1.6 ± 0.2	4.7 ± 0.6	4.7 ± 1.3	No
17:1n7 iso	15-Methylhexadec-9-enoic acid	TR	No	No	No	No
17:1n7 anteiso	14-Methylhexadec-9-enoic acid	1.0 ± 0.2	No	No	No	No
18:0.	Octadecanoic acid	No	1.2 ± 0.2	No	No	No
18:1n5	13-Octadecenoic acid	5.0 ± 0.2	No	No	No	No
18:1n7	11-Octadecenoic acid	No	TR	No	No	No
18:1n9	9-Octadecenoic acid	1.0 ± 0.2	No	No	No	No
18:2n6	9,12-Octadecadienoic acid	No	1.2 ± 0	No	No	No

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

TABLE 4. Results of antibiotic resistance tests for five *Bacillus* species

Antibiotic	*B. massiliglaciei*	*B. massilinigeriensis*	*B. mediterraneensis*	*B. phocaeensis*	*B. tuaregi*
Amoxicillin	0.160 (NA)	0.5 ± 0.35 (NA)	<0.16 (NA)	<0.16 (NA)	<0.16 (NA)
Minocycline	0.640 (S)	0.5 ± 0.35 (S)	0.213 ± 0.04 (S)	<0.16 (S)	<0.16 (S)
Imipenem	0.320 (S)	0.001 (S)	0.055 ± 0.04 (S)	0.083 (S)	0.004 (S)
Vancomycin	0.125 (S)	0.293 ± 0.06 (S)	0.158 ± 0.02 (S)	0.19 ± 0.13 (S)	0.293 ± 0.03 (S)
Penicillin G	0.527 ± 0.04 (R)	0.001 (S)	0.003 (S)	0.005 (S)	0.005 (S)

NA, no data; R, resistant; S, sensitive. Data are presented as minimum inhibitory concentrations (in µg/mL).

TABLE 5. Nucleotide content and gene count levels of genome for five *Bacillus* species

Attribute	*B. massiliglaciei*		*B. massilinigeriensis*		*B. mediterraneensis*		*B. phocaeensis*		*B. tuaregi*	
	19;28		13;21		28;37		12;13		16;19	
Scaffolds/contigs	Value	% of total[a]	Value	% of total[a]	Value	% of total[a]	Value	% of total[a]	Value	% of total[a]
Size (bp)	4 145 712	100	4 216 915	100	3 340 955	100	4 561 140	100	4 863 528	100
G + C content (%)	1 742 019	42.02	1 544 049	36.65	1 410 912	42.26	1 744 925	38.26	1 918 250	39.45
Coding region (bp)	3 493 570	84.27	3 554 893	84.30	2 810 993	84.14	3 869 074	84.83	4 088 300	84.06
Total genes	4291	100	4197	100	3488	100	4577	100	4808	100
Protein-coding genes	4210	98.11	4042	96.31	3387	97.10	4473	97.73	4,63	96.30
Total RNA genes	81	1.89	155	3.69	101	2.90	104	2.27	178	3.70
Proteins with function prediction	3023	71.81	2882	71.30	2354	69.50	3005	67.18	3253	70.26
Proteins assigned to COGs	2722	64.66	2528	62.54	2090	61.71	2827	63.20	2836	61.25
Proteins with peptide signals	410	9.74	416	10.29	381	11.25	503	11.25	491	10.60
No. of protein associated to ORFan	207	4.92	164	4.06	175	5.17	160	3.58	201	4.34
Genes with transmembrane helices	912	21.66	952	23.55	790	23.32	1139	25.46	1151	24.86
Genes associated with PKS or NRPS	12	0.29	30	0.74	9	0.27	30	0.67	14	0.30
No. of antibiotic resistance genes	0	0	2	0.05	0	0	0	0	6	0.13
No. of genes associated with Pfam-A domains	3832	89	373	88	3088	88	4126	90	4248	88

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

FIG. 4. Gram staining and electron micrographs respectively of *Bacillus* species *B. tuaregi* strain Marseille-P2489^T (A–F), *B. massilinigeriensis* strain Marseille-P2366^T (B–G), *B. phocaeensis* strain SIT16^T (C–H), *B. massiliglaciei* strain Marseille-P2600^T (D–I) and *B. mediterraneensis* strain Marseille-P2384^T (E–J).

TABLE 6. Number of genes associated with 25 general COGs functional categories for five *Bacillus* species

		B. massiliglaciei		*B. massilinigeriensis*		*B. mediterraneensis*		*B. phocaeensis*		*B. tuaregi*	
Code	**Description**	**Value**	**% of total[a]**	**Value**	**% of total[a]**	**Value**	**% of total[a]**	**Value**	**% of total[a]**	**Value**	**% of total[a]**
J	Translation	232	5.51	224	5.54	203	5.99	235	5.25	203	4.38
A	RNA processing and modification	0	0	0	0	0	0	0	0	0	0
K	Transcription	198	4.7	208	55	137	4.04	224	5.01	208	4.49
L	Replication, recombination and repair	135	3.21	112	2.78	114	3.37	103	2.30	112	2.42
B	Chromatin structure and dynamics	1	0.02	1	0.02	1	0.03	1	0.02	1	0.02
D	Cell cycle control, mitosis and meiosis	53	1.26	53	1.31	49	1.45	53	18	53	14
Y	Nuclear structure	0	0	0	0	0	0	0	0	0	0
V	Defense mechanisms	51	1.21	67	1.66	55	1.62	76	1.70	71	1.53
T	Signal transduction mechanisms	130	3.09	159	3.93	110	3.25	191	4.27	182	3.93
M	Cell wall/membrane biogenesis	125	2.97	138	3.41	103	3.04	129	2.88	142	3.07
N	Cell motility	51	1.21	63	1.56	51	1.51	63	1.41	66	1.43
Z	Cytoskeleton	0	0	0	0	0	0	0	0	0	0
W	Extracellular structures	2	0.05	14	0.35	2	0.06	10	0.22	16	0.35
U	Intracellular trafficking and secretion	31	0.74	33	0.82	33	0.97	41	0.92	41	0.89
O	Post-translational modification, protein turnover, chaperones	101	2.4	121	2.99	103	3.04	125	2.79	129	2.79
X	Mobilome: prophages, transposons	175	4.16	54	1.34	49	1.45	33	0.74	51	1.1
C	Energy production and conversion	159	3.78	152	3.76	136	4.02	200	4.47	236	5.1
G	Carbohydrate transport and metabolism	202	4.8	171	4.23	107	3.16	192	4.29	209	4.51
E	Amino acid transport and metabolism	329	7.81	301	7.45	203	5.99	300	6.71	299	6.46
F	Nucleotide transport and metabolism	94	2.23	79	1.95	82	2.42	104	2.33	83	1.79
H	Coenzyme transport and metabolism	161	3.82	131	3.24	147	4.34	161	3.6	185	4
I	Lipid transport and metabolism	147	3.49	174	4.3	101	2.98	199	4.45	145	3.13
P	Inorganic ion transport and metabolism	180	4.28	150	3.71	132	3.9	191	4.27	199	4.3
Q	Secondary metabolites biosynthesis, transport and catabolism	84	2.00	81	2	40	1.18	98	2.19	75	1.62
R	General function prediction only	258	6.13	236	5.84	179	5.28	285	6.37	272	5.87
S	Function unknown	195	4.63	184	4.55	175	5.17	223	4.99	210	4.54
—	Not in COGs	1488	35.34	1514	37.46	1297	38.29	1646	36.8	1794	38.75

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

majority of known *Bacillus* species [33,34]. These results support the notion that these strains are all members of the *Bacillus* genus.

Genome description and comparison

Maps of genomes of our different strains are presented in Supplementary Fig. 1. The properties and statistics of the genomes are summarized in Table 5, and the distribution of predicted genes of our strains according to COGs categories is shown in Table 6. We can observe that for all the 26 general COGs functional categories, the values of our five new *Bacillus* species are in the same range. The genomic characteristics of our strains are compared to those of closely related species with an available genome in Table 7. Although the genome of

TABLE 7. Genome comparison of closely related species

Organism	**Strain**	**INSDC**	**Size (Mb)**	**G + C%**	**Total genes**
Bacillus massiliglaciei	Marseille-P2600^T	FMSO00000000	4.0	42.02	4210
Bacillus psychrosaccharolyticus	ATCC 23296	AJTN00000000	4.6	38.83	4828
Bacillus kribbensis	BT080	AUMQ00000000	5.1	43.00	4926
Bacillus simplex	DSM 1321T	CP011008	5.2	42.79	4711
Bacillus butanolivorans	K9	LGYA00000000	5.7	37.91	4893
Bacillus muralis	LMG 20238	CP017080	5.9	40.94	4985
Bacillus flexus	IFO15715	BCVD00000000	5.7	37.73	5474
Bacillus massilinigeriensis	Marseille-P2366^T	FPDX00000000	4.2	36.75	4197
Bacillus mediterraneensis	Marseille-P2384^T	FOJL00000000	3.3	42.26	3488
Bacillus subterraneus	DSM13966	JXIQ00000000	3.9	42.30	3465
Bacillus bataviensis	LMG 21833	AJLS00000000	5.4	39.61	5207
Bacillus selenatarsenatis	SF-1	BASE00000000	4.8	42.20	4894
Bacillus vireti	LMG 21834	ALAN00000000	5.3	39.74	5084
Bhargavaea cecembensis	DSE10	CDGP00000000	3.2	54.83	3169
Bacillus phocaeensis	SIT16^T	FBXX00000000	4.6	38.26	4577
Bacillus niacini	IFO15566	BAWM00000000	2.2	38.3	1922
Bacillus methanolicus	PB1	CP007739	3.4	39.06	3410
Bacillus firmus	IAM 12464	GCA_000565285	4.9	41.46	4922
Bacillus humi	LMG 22167	LMBW00000000	5.4	38.21	4842
Bacillus ginsengihumi	Gsoil	JAGM00000000	3.9	35.85	3832
Bacillus licheniformis	DSM 13	CP000002.3	4.2	46.90	4172
Bacillus shackletonii	LMG 18435	LJJC00000000	5.3	36.71	4727
Bacillus tuaregi	Marseille-P2489^T	FLKE00000000	4.9	39.45	4808
Oceanobacillus picturae	LMG 19492	CCAX000000000	3.7	39.50	3666
Bacillus nealsonii	DSM 15077	ASRU00000000	5.0	35.00	4789
Bacillus korlensis	ZLC-26	BCVH00000000	5.0	38.90	4888
Bacillus gottheilii	WCC 4585	LUUP00000000	5.0	38.98	4450

INSDC, International Nucleotide Sequence Database Collaboration.

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

Species	*B. fumarioli*	*B. massili-nigeriensis*	*B. kribbensis*	*B. mediter-raneensis*	*Oceanobacillus picturae*	*B. tuaregi*	*B. nealsonii*	*B. humi*	*B. massiliglaciei*	*B. muralis*	*B. methanolicus*	*B. phocaeensis*
B. boroniphilus	18.80% ± 2.25	21.10% ± 2.3	19.50% ± 2.3	20.90% ± 2.35	18.00% ± 2.25	38.60% ± 2.5	18.00% ± 2.25	18.20% ± 2.25	20.90% ± 2.3	19.40% ± 2.3	18.20% ± 2.25	18.00% ± 2.25
B. fumarioli	100% ± 00	21.20% ± 2.3	18.90% ± 2.25	21.60% ± 2.35	19.20% ± 2.25	31.20% ± 2.45	18.20% ± 2.25	18.00% ± 2.25	19.50% ± 2.3	18.40% ± 2.2	18.40% ± 2.3	18.90% ± 2.25
B. massilinigeriensis		100% ± 00	21.50% ± 2.35	21.70% ± 2.35	48.50% ± 2.6	32.70% ± 2.45	64.00% ± 2.85	21.10% ± 2.3	25.50% ± 2.4	32.20% ± 2.45	16.80% ± 2.25	10.60% ± 1.8
B. kribbensis			100% ± 00	19.80% ± 2.3	18.40% ± 2.25	16.80% ± 2.2	19.80% ± 2.3	20.80% ± 2.35	22.90% ± 2.35	19.90% ± 2.3	20.70% ± 2.3	21.00% ± 2.35
B. mediterraneensis				100% ± 00	28.20% ± 2.4	35.60% ± 2.5	27.90% ± 2.4	23.80% ± 2.35	23.20% ± 2.35	24.30% ± 2.4	21.80% ± 2.35	22.60% ± 2.35
O. picturae					100% ± 00	32.20% ± 2.45	49.80% ± 2.65	21.30% ± 2.35	27.60% ± 2.45	26.50% ± 2.4	28.80% ± 2.4	32.40% ± 2.45
B. tuaregi						100% ± 00	59.50% ± 2.8	20.90% ± 2.35	29.30% ± 2.45	34.80% ± 2.45	28.50% ± 2.4	34.10% ± 2.45
B. nealsonii							100% ± 00	18.60% ± 2.25	22.30% ± 2.4	22.90% ± 2.4	18.70% ± 2.25	23.70% ± 2.35
B. humi								100% ± 00	26.60% ± 2.45	25.60% ± 2.45	22.50% ± 2.35	21.90% ± 2.35
B. massiliglaciei									100% ± 00	19.70% ± 2.3	22.60% ± 2.35	26.40% ± 2.45
B. muralis										100% ± 00	29.50% ± 2.45	27.30% ± 2.4
B. methanolicus											100% ± 00	21.40% ± 2.35
B. phocaeensis												100% ± 00

DDH, DNA-DNA hybridization; GGDC, Genome-to-Genome Distance Calculator; HSP, high-scoring pair.
[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 the 16S rRNA and phylogenetic analyses as well as GGDC results.

TABLE 9. Number of orthologous proteins shared between *Bacillus* genomes (upper right)[a]

	B. boroniphilus	*B. fumarioli*	*B. massili-nigeriensis*	*B. kribbensis*	*B. mediter-raneensis*	*Oceanobacillus picturae*	*B. tuaregi*	*B. nealsonii*	*B. simplex*	*B. massili-glaciei*	*B. muralis*	*B. butano-livorans*	*B. phocaeensis*	*B. ginsengy-ihumi*
B. boroniphilus	**5391**	1286	1449	1463	1441	1170	1469	1333	1495	1356	1515	1482	1641	1180
B. fumarioli	68.44	**3374**	1547	1578	1453	1236	1543	1360	1572	1447	1583	1578	1573	1399
B. massilinigeriensis	68.65	70.23	**4042**	1670	1603	1356	1800	1560	1718	1557	1720	1719	1927	1475
B. kribbensis	66.99	67.57	67.18	**4926**	1570	1428	1614	1618	1894	1706	1912	1861	1769	1579
B. mediterraneensis	69.78	68.72	68.79	67.29	**3387**	1294	1611	1412	1622	1508	1624	1644	1655	1337
O. picturae	55.71	56.48	57.00	55.59	56.16	**3666**	1390	1375	1516	1392	1486	1468	1502	1296
B. tuaregi	68.24	69.77	70.99	67.33	68.51	56.61	**4630**	1624	1717	1605	1709	1780	1942	1515
B. nealsonii	59.63	61.03	62.02	59.16	60.24	55.98	61.35	**4789**	1690	1543	1683	1633	1685	1496
B. simplex	56.68	57.05	56.99	57.14	57.02	55.12	56.92	69.34	**4711**	1917	2619	2505	1810	1574
B. massiliglaciei	67.05	67.88	67.58	68.43	57.45	55.98	67.69	59.76	58.63	**4210**	1901	1874	1660	1526
B. muralis	66.73	67.68	67.77	67.60	67.18	55.95	67.63	59.63	65.65	71.58	**4985**	2420	1796	1557
B. butanolivorans	57.92	58.73	69.35	58.40	58.03	53.64	58.72	60.20	66.67	60.04	65.71	**4893**	1841	1569
B. phocaeensis	68.44	69.80	71.70	66.87	68.62	56.75	70.26	61.38	56.76	67.42	67.53	59.03	**4473**	1506
B. ginsengihumi	66.21	68.52	69.07	65.90	66.90	56.30	68.07	61.19	56.58	66.79	66.72	58.51	68.39	**3832**

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

strain Marseille-P2384^T is smaller than that of the other species, they all have the same G + C percentage of approximately 40% (± 2%) as the other characterized genomes of known *Bacillus* species [30–32].

Furthermore, dDDH values are <70% and confirm that all the studied species are distinct species (Table 8). These results are supported by AGIOS values (Table 9). Indeed, the ten closest species of the five studied strains have been compared among each other. Their AGIOS values are between 50.13 and 74.38%. When the five studied strains are compared with the ten same species, the AGIOS values of strains SIT16^T, Marseille-P2366^T, Marseille-P2384^T, Marseille-P2600^T, and Marseille-P2489^T, are respectively 52.49–70.51%, 54.75–69.78%, 55.34–69.83%, 54.73–71.58% and 53.01–71.70%. These results being all included between 50.13 and 74.38% that confirm their new species status.

Conclusion

In this study we used a new polyphasic approach, developed in our laboratory, to describe five new species of the *Bacillus* genus. This concept is based on the genome sequence, MALDI-TOF MS identification and main phenotypic characteristics of the studied new species. As previously observed, the presented strains, which have been isolated from diverse origins, possess close phenotypic results, including notable morphologic and biochemical properties. Their cellular fatty acid composition and their profile of resistance to antibiotics support that these five strains belong to the same genus. Their 16S rRNA gene sequencing, supported by genomic analysis, compared to other characterized strains of *Bacillus* genus indicate that *Bacillus massiliglaciei* strain Marseille-P2600^T, *Bacillus massilinigeriensis* strain Marseille-P2366^T, *Bacillus mediterraneensis* strain Marseille-P2384^T, *Bacillus tuaregi* strain Marseille-P2489^T and *Bacillus phocaeensis* strain SIT16^T are all members of the *Bacillus* genus.

Description of *Bacillus massiliglaciei* sp. nov.

Bacillus massiliglaciei (*massiliglaciei* is composed of mas.si.li, L. masc. adj. *massili,* from Massilia, the old Roman name for Marseille, where the strain was isolated, and gla.ci′ei, L. gen. n. *glaciei,* 'of ice,' referring to the isolation source of the strain).

Cells are Gram-positive bacilli and are fusiform shaped, with a length ranging from 2 to 5 µm and a width ranging from 0.5 to 0.6 µm. This strain exhibited catalase activity but no oxidase activity. *Bacillus massiliglaciei* is motile and spore forming. Colonies are circular and grey, with a diameter of 0.4 to 0.5 mm. Optimum growth occurs at 37°C in an aerobic atmosphere on Colombia agar enriched with 5% sheep's blood after 24 hours' growth. API ZYM analysis shows that *B. massiliglaciei* has positive activity for only esterase, esterase lipase, naphthol-AS-BI-phosphohydrolase and *N*-acetyl-β-glucosaminidase. All the results of API ZYM galleries are indicated in Supplementary Table S2. API 20NE analysis revealed that *B. massiliglaciei* has a positive hydrolysis activity for protease and is able to assimilate only glucose, arabinose, mannitol, potassium gluconate, malate and trisodium citrate. All the results of API 20NE galleries are indicated in Supplementary Table S3. API 50CH analysis shows that *B. massiliglaciei* is able to ferment only D-ribose, D-glucose and D-mannitol. All the results of the API 50CH galleries are provided in Supplementary Table S4.

Strain Marseille-P2600^T was susceptible to amoxicillin (0.16 µg/mL), minocycline (0.64 µg/mL), penicillin G (0.47 µg/mL), imipenem (0.32 µg/mL) and vancomycin (0.125 µg/mL). The major fatty acid is 13-methyl-tetradecanoic acid.

The genome of strain Marseille-P2600^T is 4 145 712 bp long with 42.02% G + C content. The 16S rRNA and genome sequences are available in the European Molecular Biology Laboratory–European Bioinformatics Institute (EMBL-EBI) database under accession numbers LT223699 and FMSO00000000 respectively. The type strain Marseille-P2600^T (= CSUR P2600 = DSM 102861) was isolated from the Siberian permafrost. Its habitat is ice.

Description of *Bacillus massilinigeriensis* sp. nov.

Bacillus massilinigeriensis (*massilinigeriensis* is composed of mas.si.li, L. masc. adj. *massili,* 'of Massilia,' the old Roman name for Marseille, where the strain was isolated, and ni.ge.rien′sis, L. gen. adj. *nigeriensis,* 'of Niger,' referring to the nationality of the people who provided the stool samples).

Cells are Gram-positive bacilli and are fusiform shaped with a length ranging from 3.5 to 5.0 µm and a width of 0.5 µm. This strain exhibited catalase activity but no oxidase activity. *Bacillus massilinigeriensis* is motile and non–spore forming. Colonies are circular, smooth and grey, with a diameter of 0.6 to 0.8 mm. Optimum growth occurs at 37°C + 5% CO_2 in an aerobic atmosphere on Colombia agar enriched with 5% sheep's blood after 24 hours' growth. API ZYM analysis shows that *B. massilinigeriensis* has positive activity for only esterase, esterase lipase, naphthol-AS-BI-phosphohydrolase and α-glucosidase (Supplementary Table S2). API 20NE analysis shows that all results for *B. massilinigeriensis* are negative (Supplementary Table S3). API 50CH analysis shows that *B. massilinigeriensis* is able to ferment only D-glucose (Supplementary Table S4).

Strain Marseille-P2366^T was susceptible to amoxicillin (0.5 µg/mL), minocycline (0.5 µg/mL), penicillin G (0.47 µg/mL), imipenem (0.001 µg/mL) and vancomycin (0.293 µg/mL). The major fatty acid is 13-methyl-tetradecanoic acid.

The genome of strain Marseille-P2366^T is 4 216 915 bp long with 36.7% G + C content. The 16S rRNA and genome sequences are available in the EMBL-EBI database under accession numbers LT161887 and FOJL00000000 respectively. The type strain Marseille-P2366^T (= CSUR P2366 = DSM 102112) was isolated from the stool of a healthy Nigerien girl. Its habitat is the human digestive tract.

Description of *Bacillus mediterraneensis* sp. nov.

Bacillus mediterraneensis (me.di.ter.ra.neen′sis, L. masc. adj. *mediterraneensis,* from the Mediterranean Sea, which borders Marseille, where the strain was found).

Cells are Gram-positive bacilli and are fusiform shaped with a length ranging from 2.0 to 5.0 µm and a width ranging from 0.4 to 0.6 µm. This strain exhibited no catalase and no oxidase activities. *Bacillus mediterraneensis* is motile and spore forming. Colonies are circular, smooth and grey, with a diameter of 0.7 to 1.0 mm. Optimum growth occurs at 37°C in an aerobic atmosphere on Colombia agar enriched with 5% sheep's blood after 24 hours' growth. API ZYM analysis shows that *B. mediterraneensis* has positive activity for only naphthol-AS-BI-phosphohydrolase (Supplementary Table S2). API 20NE analysis shows that *B. mediterraneensis* has a positive hydrolysis activity for protease (Supplementary Table S3). API 50CH analysis shows that *B. mediterraneensis* is able to ferment only L-arabinose, D-ribose, D-xylose, D-glucose, D-fructose, D-mannose, D-mannitol, D-sorbitol, *N*-acetylglucosamine, arbutin, esculin ferric citrate, salicin, D-cellobiose, D-maltose, D-saccharose, D-trehalose, D-melezitose, D-arabitol, potassium gluconate and potassium 5-ketogluconate (Supplementary Table S4).

Strain Marseille-P2384^T was susceptible to amoxicillin (<0.16 µg/mL), minocycline (0.213 µg/mL), penicillin G (0.003 µg/mL), imipenem (0.006 µg/mL) and vancomycin (0.158 µg/mL). The major fatty acid is 13-methyl-tetradecanoic acid.

The genome of strain Marseille-P2384^T is 3 340 955 bp long with 42.3% G + C content. The 16S rRNA and genome sequences are available in the EMBL-EBI database under accession numbers LT161888 and FPDX00000000 respectively. The type strain Marseille-P2384^T (= CSUR P2384 = DSM 102091) was isolated from the stool of a healthy Senegalese boy. Its habitat is the human digestive tract.

Description of *Bacillus tuaregi* sp. nov.

Bacillus tuaregi (tua.reg′i, L. masc. adj. *tuaregi,* from the Tuareg people, the ethnicity of the donor whose stool was used to isolate the strain).

Cells are Gram-positive bacilli and are fusiform shaped with a length ranging from 3.5 to 5.0 µm and a width ranging from 0.5 to 1.0 µm. This strain exhibited no catalase and no oxidase activities. *Bacillus tuaregi* is motile and non–spore forming. Colonies are small, circular and grey, with a diameter of 0.2 to 0.3 mm. Optimum growth occurs at 37°C in an aerobic atmosphere on Colombia agar enriched with 5% sheep's blood after 24 hours' growth. API ZYM analysis shows that *B. tuaregi* has positive activity for only esterase lipase and naphthol-AS-BI-phosphohydrolase (Supplementary Table S2). API 20NE analysis shows that *B. tuaregi* has any positive results (Supplementary Table S3). API 50CH analysis shows that *B. tuaregi* is able to ferment only L-arabinose, D-ribose, D-glucose, D-fructose, D-mannitol, Methyl-αD-glucopyranoside, D-maltose and D-trehalose (Supplementary Table S4).

Strain Marseille-P2489^T was susceptible to amoxicillin (<0.16 µg/mL), minocycline (<0.16 µg/mL), penicillin G (0.005 µg/mL), imipenem (0.004 µg/mL) and vancomycin (0.293 µg/mL). The major fatty acid is 12-methyl-tridecanoic acid.

The genome of strain Marseille-P2489^T is 4 863 528 bp long with 39.45% G + C content. The 16S rRNA and genome sequences are available in the EMBL-EBI database under accession numbers LT223701 and FNLH00000000 respectively. The type strain Marseille-P2489^T (= CSUR P2489 = DSM 103460) was isolated from the stool of a healthy Nigerien girl. Its habitat is the human digestive tract.

Description of *Bacillus phocaeensis* sp. nov.

Bacillus phocaeensis (pho.cae.en′sis, L. masc. adj. *phocaeensis,* from Phocea, to refer to the city of Marseille, where the strain was isolated).

Cells are Gram-positive bacilli and are fusiform shaped with a length ranging from 2.5 to 4.0 µm and a width ranging from 0.5 to 0.7 µm. This strain exhibited catalase but no oxidase activities. *Bacillus phocaeensis* is motile and spore forming. Colonies are circular, smooth and grey, with a diameter of 0.5 to 0.8 mm. Optimum growth occurs at 37°C in an aerobic atmosphere on Colombia agar enriched with 5% sheep's blood after 24 hours' growth. API ZYM analysis shows that *B. phocaeensis* has positive activity for only esterase, esterase lipase and α-glucosidase (Supplementary Table S2). API 20NE analysis shows that *B. phocaeensis* has a positive hydrolysis activity for protease

(Supplementary Table S3). API 50CH analysis shows that *B. phocaeensis* is able to ferment only D-glucose, D-fructose, salicin, D-cellobiose, D-maltose, D-lactose and inulin (Supplementary Table S4).

Strain SIT16^T was susceptible to amoxicillin (<0.16 µg/mL), minocycline (<0.16 µg/mL), penicillin G (0.005 µg/mL), imipenem (0.083 µg/mL) and vancomycin (0.19 µg/mL). The major fatty acid is 13-methyl-tetradecanoic acid.

The genome of strain SIT16^T is 4 561 140 bp long with 38.3% G + C content. The 16S rRNA and genome sequences are available in the EMBL-EBI database under accession numbers LN881595 and FBXX00000000 respectively. The type strain SIT16^T (= CSUR P2184 = CCUG 69739) was isolated from the stool of a Senegalese boy with kwashiorkor. Its habitat is the human digestive tract.

Acknowledgements

The authors thank the Xegen Company (http://www.xegen.fr/) for automating the genomic annotation process. The study was funded by the Fondation Méditerranée Infection. We thank M. Lardière for English-language revision and A. Oren for the etymology review.

Conflict of Interest

None declared.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.nmni.2017.04.005.

References

[1] Claus D, Berkeley RCW. Genus *Bacillus* Cohn, 1872. In: Sneath PHA, Mair NS, Sharpe ME, Holt JG, editors. Bergey's manual of sytematic bacteriology vol. 2. Baltimore, MD: Williams & Wilkins; 1986. p. 1105–39.

[2] Fritze D. Taxonomy of the genus *Bacillus* and related genera: the aerobic endospore-forming bacteria. Physiopathology 2004;94: 1245–8.

[3] Gordon RE, Haynes WC, Pang CHN. The genus *Bacillus*. Washington, DC: US Department of Agriculture; 1973.

[4] Priest FG, Goodfellow M, Todd C. A numerical classification of the genus *Bacillus*. J Gen Microbiol 1988;134:1847–82.

[5] Lagier JC, El Karkouri K, Mishra AK, Robert C, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Enterobacter massiliensis* sp. nov. Stand Genomic Sci 2013;7:399–412.

[6] 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.

[7] 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.

[8] Dubourg G, Lagier JC, Armougom F, Robert C, Hamad I, Brouqui P, et al. The gut microbiota of a patient with resistant tuberculosis is more comprehensively studied by culturomics than by metagenomics. Eur J Clin Microbiol Infect Dis 2013;32:637–45.

[9] Lagier JC, Edouard S, Pagnier I, Mediannikov O, Drancourt M, Raoult D. Current and past strategies for bacterial culture in clinical microbiology. Clin Microbiol Rev 2015;28:208–36.

[10] Morel AS, Dubourg G, Prudent E, Edouard S, Gouriet F, Casalta JP, et al. Complementarity between targeted real-time specific PCR and conventional broad-range 16S rDNA PCR in the syndrome-driven diagnosis of infectious diseases. Eur J Clin Microbiol Infect Dis 2015;34:561–70.

[11] 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.

[12] 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.

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

[14] 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.

[15] 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.

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

[17] 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.

[18] 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.

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

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

[21] 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.

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

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

[24] Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004;14:1394–403.

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

[26] Gouret P, Paganini J, Dainat J, Louati D, Darbo E, Pontarotti P, et al. Integration of evolutionary biology concepts for functional annotation

and automation of complex research in evolution: the multi-agent software system DAGOBAH. In: Pontarotti P, editor. Evolutionary biology—concepts, biodiversity, macroevolution and genome evolution. Berlin: Springer; 2011. p. 71–87.

[27] 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.

[28] 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.

[29] 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.

[30] Tiago I, Pires C, Mendes V, Morais PV, da Costa MS, Veríssimo A. *Bacillus foraminis* sp. nov., isolated from a non-saline alkaline groundwater. Int J Syst Evol Microbiol 2006;56:2571–4.

[31] Albert RA, Archambault J, Rosselló-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.

[32] Venkateswaran K, Kempf M, Chen F, Satomi M, Nicholson W, Kern R. *Bacillus nealsonii* sp. nov., isolated from a spacecraft-assembly facility, whose spores are gamma-radiation resistant. Int J Syst Evol Microbiol 2003;53:165–72.

[33] Ikeda M, Yagihara Y, Tatsuno K, Okazaki M, Okugawa S, Moriya K. Clinical characteristics and antimicrobial susceptibility of *Bacillus cereus* blood stream infections. Ann Clin Microbiol Antimicrob 2015;14:43.

[34] Reva ON, Vyunitskaya VA, Reznik SR, Kozachko IA, Smirnov VV. Antibiotic susceptibility as a taxonomic characteristic of the genus *Bacillus*. Int J Syst Bacteriol 1995;45:409–11.