Noncontiguous finished genome sequence and description of *Virgibacillus massiliensis* sp. nov., a moderately halophilic bacterium isolated from human gut

S. Khelaifia[1], O. Croce[1], J.-C. Lagier[1], C. Robert[1], C. Couderc[1], F. Di Pinto[1], B. Davoust[1], F. Djossou[2], D. Raoult[1,3] and P.-E. Fournier[1]

1) *Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM 63, CNRS 7278, IRD 198, Inserm 1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine, Aix-Marseille Université, Marseille, France,* 2) *Centre Hospitalier André Rosemon, Cayenne, French Guiana and* 3) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Strain Vm-5^T was isolated from the stool specimen of a 10-year-old Amazonian boy. This bacterium is a Gram-positive, strictly aerobic rod, motile by a polar flagellum. Here we describe its phenotypic characteristics and complete genome sequence. The 4 353 177 bp long genome exhibits a G + C content of 36.87% and contains 4394 protein-coding and 125 predicted RNA genes. Phylogenetically and genetically, strain Vm-c is a member of the genus *Virgibacillus* but is distinct enough to be classified as a new species. We propose the creation of *V. massiliensis* sp. nov., whose type strain is strain Vm-5^T (CSUR P971 = DSM 28587).

Keywords: Culturomics, genome, human gut, moderately halophilic bacteria, taxonogenomics, *Virgibacillus massiliensis*
Original Submission: 14 March 2015; **Revised Submission:** 21 September 2015; **Accepted:** 23 September 2015
Article published online:

Corresponding author: P.-E. Fournier, URMITE, UMR CNRS 7278, L'Institut de Recherche pour le Développement 198, INSERM U1095, Faculté de Médecine, Aix-Marseille Université, 27 Boulevard Jean Moulin, 13385 Marseille Cedex 5, France
E-mail: pierre-edouard.fournier@univ-amu.fr

Introduction

Virgibacillus massiliensis strain Vm-5^T (= CSUR P971 = DSM 28587) is the type strain of *V. massiliensis* sp. nov. This bacterium is a Gram-positive, strictly aerobic rod, motile by a polar flagellum, isolated from the stool specimen of a healthy Amazonian boy as part of the culturomics study aiming at cultivating halophilic bacteria from the human feces using a high-salt-concentration medium [1].

The usual parameters used to delineate a bacterial species include 16S rRNA sequence identity and phylogeny [2,3], genomic G + C content diversity and DNA-DNA hybridization [4,5]. Nevertheless, these methods have limitations, notably because these similarity values vary greatly between species and genera [6]. In addition, chemotaxonomic analyses such as fatty acid profile, cell wall diagnostic diamino acid and sporangium morphology are only performed by a few laboratories, are only partially reproducible and thus are of no practical value to identify clinical isolates. Therefore, we deliberately decided not to use these methods but rather include parameters that could be compared among laboratories, including widely used phenotypic criteria, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) spectrum and genome sequence.

The introduction of high-throughput sequencing techniques has allowed researchers to make genomic data available for many bacterial species [6–10]. We recently proposed a new method (taxonogenomics) consisting in a polyphasic approach to describe new bacterial species [5]. This strategy combines phenotypic characteristics including MALDI-TOF spectrum and genomic analysis [6–10]. Here we present a summary classification and a set of features for the *Virgibacillus massiliensis* sp. nov., strain Vm-5^T (= CSUR P971 = DSM 28587), including the description of its complete genome sequence and annotation. These characteristics support the circumscription of the species *Virgibacillus massiliensis*.

Virgibacillus massiliensis is the first representative from the *Virgibacillus* genus to be isolated from the human intestinal microbiota. The genus *Virgibacillus* was first described by Heyndrickx *et al.* in 1998 and currently consists of mainly Gram-positive, motile, spore-forming, rod-shaped bacteria that are moderately halophilic [11]. Members of the genus *Virgibacillus* are found in various environments including sediment of a saline lake [12–15], traditional salt-fermented seafood [16], a permafrost core collected from the Canadian high Arctic [17], a marine solar saltern [18–21], biofilm formation on mural paintings [22], seawater [23,24], field soil, a dairy product sample [25], a saline mud sample [26], residual wash water

http://dx.doi.org/10.1016/j.nmni.2015.09.012

produced during the processing of Spanish-style green table olive sewage [27], salt crust [28] and fermented fish [29].

Organism Information

Classification and features

Stool specimens were collected from a 10-year-old Amazonian boy, formed into aliquots and stored at −80°C until use. The child and his parents provided informed consent. The study and the assent procedure were approved by the ethics committees of the Institut Fédératif de Recherche 48, Faculty of Medicine, Marseille, France, under agreement 09-022. The salt concentration of the stool specimen was determined using a digital refractometer (Fisher Scientific, Illkirch, France) and the pH with a pH meter (Table 1).

Strain Vm-5^T (Table 1) was isolated in December 2013 by aerobic culture on a homemade culture medium consisting of a Columbia agar culture medium (Sigma-Aldrich, Saint-Quentin Fallavier, France) modified by adding (per liter) the following: $MgCl_2$ $6H_2O$, 5 g; $MgSO_4$ $7H_2O$, 5 g; KCl, 2 g; $CaCl_2$ 2H2O, 1 g; NaBr, 0.5 g; $NaHCO_3$, 0.5 g, glucose, 2 g and 100 g/L of NaCl. The pH was adjusted to 7.5 with 10 M NaOH before autoclaving. Strain Vm-5^T (GenBank accession number HG931931) exhibited a 16S rRNA sequence identity of 97.3% with *Virgibacillus olivae* strain $E_{30}8^T$ (NR043572), its phylogenetically closest bacterial species with standing in nomenclature (Fig. 1).

TABLE 1. Classification and general features of *Virgibacillus massiliensis* strain Vm-5^T according to MIGS recommendations [30].

MIGS ID	Property	Term	Evidence code[a]
	Current classification	Domain: *Bacteria*	TAS [31]
		Phylum: *Firmicutes*	TAS [32–34]
		Class: *Bacilli*	TAS [35,36]
		Order: *Bacillales*	TAS [37–39]
		Family: *Bacillaceae*	TAS [38–41]
		Genus: *Virgibacillus*	TAS [11]
		Species: *Virgibacillus massiliensis*	IDA
		Type strain: Vm-5^T	IDA
	Gram stain	Positive	IDA
	Cell shape	Rod shaped	IDA
	Motility	Motile by polar flagellum	IDA
	Sporulation	Endospore forming	IDA
	Temperature range	Mesophile	IDA
	Optimum temperature	37°C	IDA
	pH	pH 5 to 9	
	Optimum pH	7.5	
MIGS-6.3	Salinity	0.5–20%	IDA
	Optimum salinity	5%	IDA
MIGS-22	Oxygen requirement	Aerobic	IDA
	Carbon source	Unknown	IDA
	Energy source	Unknown	IDA
MIGS-6	Habitat	Human gut	IDA
MIGS-15	Biotic relationship	Free-living	IDA
	Pathogenicity	Unknown	NAS
	Biosafety level	2	IDA
MIGS-14	Isolation	Human feces	IDA
MIGS-4	Geographic location	France	IDA
MIGS-5	Sample collection time	December 2013	IDA
MIGS-4.1	Latitude	4.916667	IDA
MIGS-4.1	Longitude	−52.316666	IDA
MIGS-4.3	Depth	Surface	IDA
MIGS-4.4	Altitude	0 m above sea level	IDA

MIGS, minimum information about a genome sequence.
[a]Evidence codes are as follows: IDA, inferred from direct assay; TAS, traceable author statement (i.e., direct report exists in the literature); NAS, nontraceable author statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species or on anecdotal evidence). These evidence codes are form the Gene Ontology project (http://www.geneontology.org/GO.evidence.shtml) [42]. If the evidence is IDA, then the property was directly observed for a live isolate by one of the authors or an expert mentioned in the acknowledgements.

Colonies were obtained on our homemade culture medium after 24 hours of incubation in aerobic conditions at 37°C. The colonies of strain Vm-5^T were circular, greyish, shiny and smooth, with a diameter of 2 to 5 mm. Cells stained Gram positive (Fig. 2). They were motile by polar flagella, were terminal spore forming and most commonly occurred as single cells or in pairs. Colonies were not haemolytic on blood-enriched agar.

Strain Vm-5 ^{T} was mesophilic and grew at temperatures ranging from 15 to 45°C, at an optimum temperature of 37°C. The isolate required NaCl for growth and grew at salinity ranging from 5 to 200 g/L of NaCl (optimum at 50 g/L). The optimal pH for growth was 7.5 (pH range 5 to 9). The growth of strain Vm-5^T was tested under aerobic atmosphere, in the presence of 5% CO_2 and in anaerobic and microaerophilic atmospheres created using GENbag anaer and GENbag microaer (bioMérieux, Marcy l'Etoile, France), respectively. The strain was strictly aerobic and grew in the presence of 5% CO_2 but did not grow in microaerophilic or anaerobic atmosphere. The size (2 to 6 µm in length and 0.5 µm in diameter) and ultrastructure of cells were determined by negative staining transmission electron microscopy (Fig. 3).

The commercially available Api ZYM, Api 20NE (bioMérieux), was used to characterize the biochemical properties of the strain according to the manufacturer's instructions. The strain was incubated at 37°C for 24 hours. Api 50 CH strips were inoculated with a bacterial suspension in Api 50CHB/E medium supplemented by 10% NaCl (w/v) and incubated at 37° C for 48 hours. *Virgibacillus massiliensis* strain Vm-5^T exhibited catalase and oxidase activities. Negative reactions were observed for alkaline phosphatase, galactosidase, N-acetyl-β-glucosaminidase and urease activities. A positive reaction was observed for nitrate reduction. Substrate oxidation and assimilation were examined using an API 50CH strip (bioMérieux) at 37°C. Negative reactions were obtained for D-lactose, L-arabinose, D-galactose and D-ribose. Positive reactions were obtained for D-glucose, D-fructose, D-mannose, D-mannitol, D-maltose and D-sucrose. Phenotypic characteristics were compared to those of the most closely related species (Table 2). *Virgibacillus massiliensis* differed from other *Virgibacillus* species based on its use of nitrate reductase (+), N-acetyl-glucosamine (−), D-mannose (+), D-sucrose (+) and D-maltose (+).

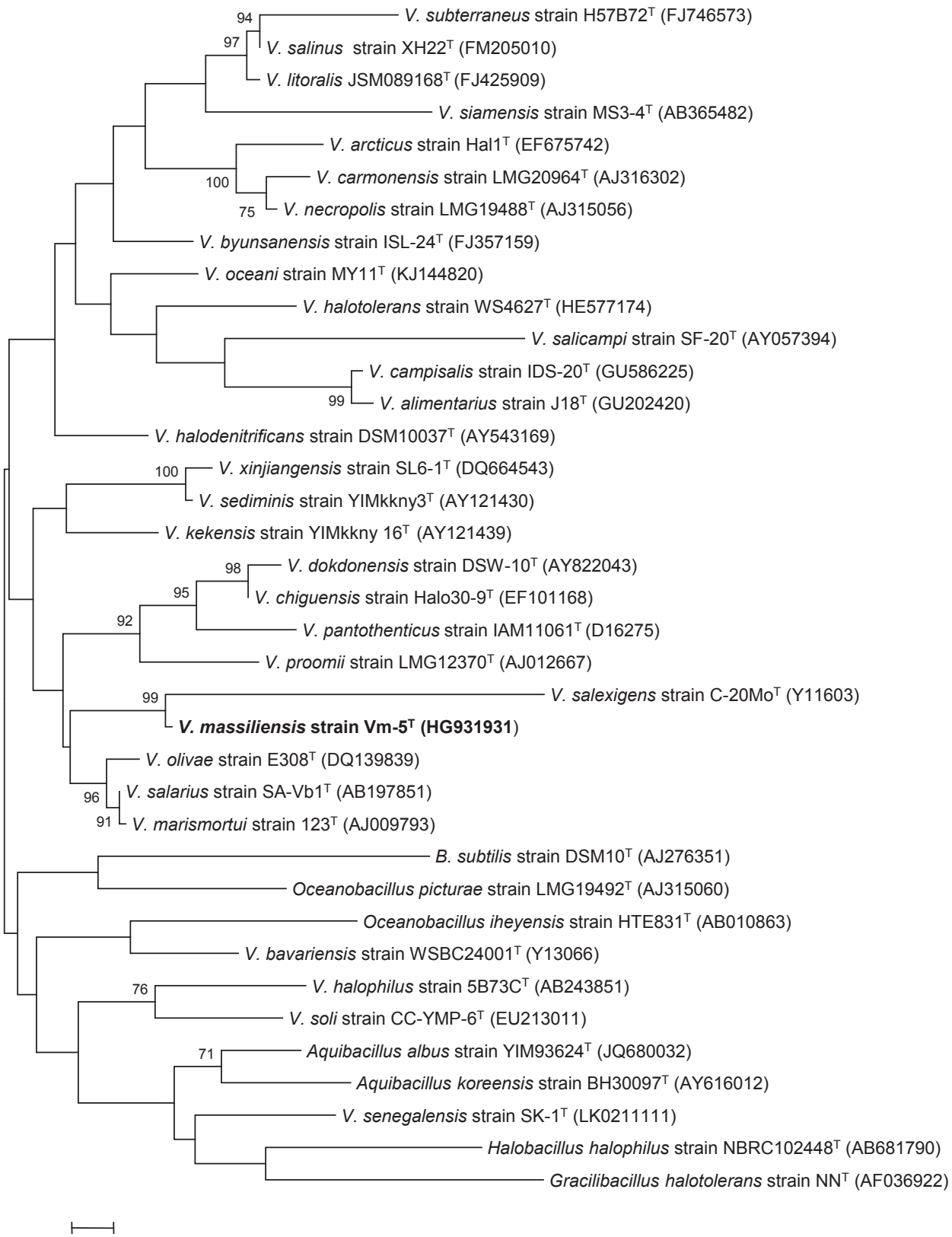

FIG. 1. Unrooted phylogenetic tree based on comparison of 16S rRNA sequences highlighting position of *Virgibacillus massiliensis* strain Vm-5^T relative to other type strains within genus *Virgibacillus* and to type strains of other closely related genera. Sequences were aligned using Clustal W (http://www.clustal.org/clustal2/), and phylogenetic inferences were obtained using maximum-likelihood method within MEGA 6 software (http://www.megasoftware.net/mega.php). Similar phylogenetic organization was obtained using neighbor-joining method. GenBank accession numbers are displayed in parentheses. Scale bar = 0.5% nucleotide sequence divergence. Bootstrap values of 70% or more are indicated at nodes.

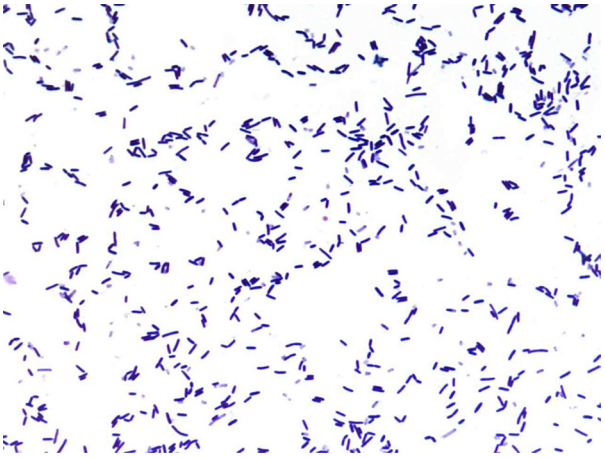

FIG. 2. Gram staining of *Virgibacillus massiliensis* strain Vm-5^T.

Antimicrobial susceptibility testing demonstrated that strain Vm-5^T was susceptible to penicillin, ampicillin, amoxicillin, ceftriaxone, imipenem, doxycycline, rifampicin, vancomycin, nitrofurantoin, erythromycin, ciprofloxacin and gentamicin but was resistant to trimethoprim/sulfamethoxazole and metronidazole.

MALDI-TOF analysis

MALDI-TOF protein analysis was used to analyse strain Vm-5^T. Briefly, a pipette tip was used to pick one isolated bacterial colony from a culture agar plate and spread it as a thin film on a MALDI-TOF target plate (Bruker Daltonics, Leipzig, Germany). Twelve distinct deposits were done for strain Vm-5^T from 12 isolated colonies. After air drying, 2 µL of matrix solution (saturated solution of α-cyanohydroxycinnaminic acid in 50% aqueous acetonitrile containing 2.5% trifluoroacetic acid) was applied to each spot. MALDI-TOF was conducted using the Microflex LT spectrometer (Bruker). All spectra were recorded in the positive linear mode for the mass range of 2000 to 20 000 Da (parameter settings: ion source 1 (ISI), 20 kV; IS2, 18.5 kV; lens, 7 kV). A spectrum was obtained after 675 shots with variable laser power. The time of acquisition was between 30 seconds and 1 minute per spot. The 12 spectra of strain Vm-5^T were imported into the MALDI BioTyper software (version 2.0, Bruker) and analysed by standard pattern matching (with default parameter settings) against the main spectra of 7335 bacteria including the spectra from the closely related species *Virgibacillus proomii* strain DSM 13055^T, *V. proomii* strain 10403186, *V. pantothenticus* strain DSM 26^T, *V. halodenitrificans* DSM 10037^T, *Oceanobacillus massiliensis* DSM 24644^T and *O. jeddahmassiliense* strain DSM 28586 ^{T}.

The identification method included m/z from 3000 to 15 000 Da. For every spectrum, a maximum of 100 peaks were compared with spectra in database. The resulting score enabled the identification (or not) of tested species: a score of ≥ 2 with a validly published species enabled identification at the species level. No significant MALDI-TOF score was obtained (< 0.9) for strain Vm-5^T against the Bruker database, suggesting that our isolate was not a member of a known species. We added the spectrum from strain Vm-5^T to our database (Fig. 4). Finally, the gel view showed the spectral differences with other members of the genus *Virgibacillus* (Figs. 4 and 5).

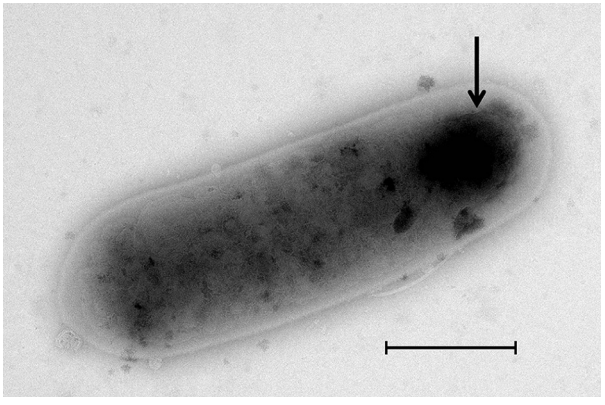

FIG. 3. Transmission electron microscopy of *Virgibacillus massiliensis* strain Vm-5^T, using Morgagni 268D transmission electron microscope (Philips/FEI, Hillsboro, OR, USA) at operating voltage of 60 kV. Scale bar = 500 nm. Arrow indicates terminal spore.

Genome Sequencing Information

Genome project history

The *V. massiliensis* genome was sequenced as part of a culturomics study aiming at isolating all bacterial species colonizing the human gut [1] and because of its potential classification as a new species within the *Virgibacillus* genus. The genome from *V. massiliensis* strain Vm-5^T is the fourth genome of a *Virgibacillus* species and the first genome of *V. massiliensis* sp. nov. This genome consists of seven contigs and was deposited in GenBank under accession numbers CCDP010000001 to CCDP010000007CCDP010000001CCDP010000002CCDP010000003CCDP010000004CCDP010000005CCDP010000006 CCDP010000007. Table 3 shows the project information.

Growth conditions and genomic DNA preparation

Virgibacillus massiliensis sp. nov., strain Vm-5^T (CSUR P971 = DSM 28587), was grown on a homemade culture medium at 37°C in aerobic atmosphere. Bacteria grown on ten 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

TABLE 2. Differential characteristics

Property	*V. massiliensis*	*V. dokdonensis*	*V. halodenitrificans*	*V. kekensis*	*V. marismortui*	*V. olivae*	*V. proomii*	*V. salarius*	*V. sediminis*	*V. senegalensis*	*V. xinjiangensis*
Cell diameter (μm)	0.5–0.8	NA	0.6–0.8	0.3–0.5	NA	0.4–0.6	0.5–0.7	0.6–0.9	0.4–0.7	0.6–0.9	1.4–2.4
Oxygen requirement	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic
Gram stain	+	+	+	+	+	+	+	+	+	+	+
Salt requirement	+	+	+	–	–	+	NA	+	+	+	–
Motility	+	+	+	+	+	+	+	+	+	+	+
Endospore formation	+	+	+	+	+	+	+	+	+	+	+
Indole	–	–	–	–	–	–	–	–	–	–	–
Production of											
Alkaline phosphatase	–	–	NA	–	NA	NA	NA	NA	–	–	NA
Catalase	+	+	+	+	+	+	+	+	+	–	+
Oxidase	+	+	+	+	+	+	NA	+	+	–	–
Nitrate reductase	+	–	+	+	+	+	+	–	+	–	+
Urease	–	–	+	–	NA	NA	NA	–	–	+	–
β-Galactosidase	–	–	+	–	–	NA	+	–	–	+	–
N-acetyl-glucosamine	–	–	NA	–	+	NA	+	+	–	–	NA
Acid from:											
L-Arabinose	–	–	–	–	–	NA	–	–	–	–	–
Ribose	–	+	NA	–	NA	NA	+	NA	+	–	–
D-Mannose	+	+	+	+	+	–	+	+	–	–	–
D-Mannitol	+	–	+	w	–	NA	+	–	–	–	–
D-Sucrose	+	+	–	–	–	–	+	–	–	–	+
D-Glucose	+	+	+	+	+	–	+	+	+	–	+
D-Fructose	+	+	+	–	+	+	+	+	+	–	+
D-Maltose	+	–	+	+	–	–	+	+	+	–	–
D-Lactose	–	+	+	–	–	–	+	–	–	–	–
Habitat	Human gut	Soil	Solar saltern	Salt lake	Mural paintings	Waste wash water	Soil	Salt lake	Salt lake	Human gut	Salt lake

NA, data not available; w, weak reaction.

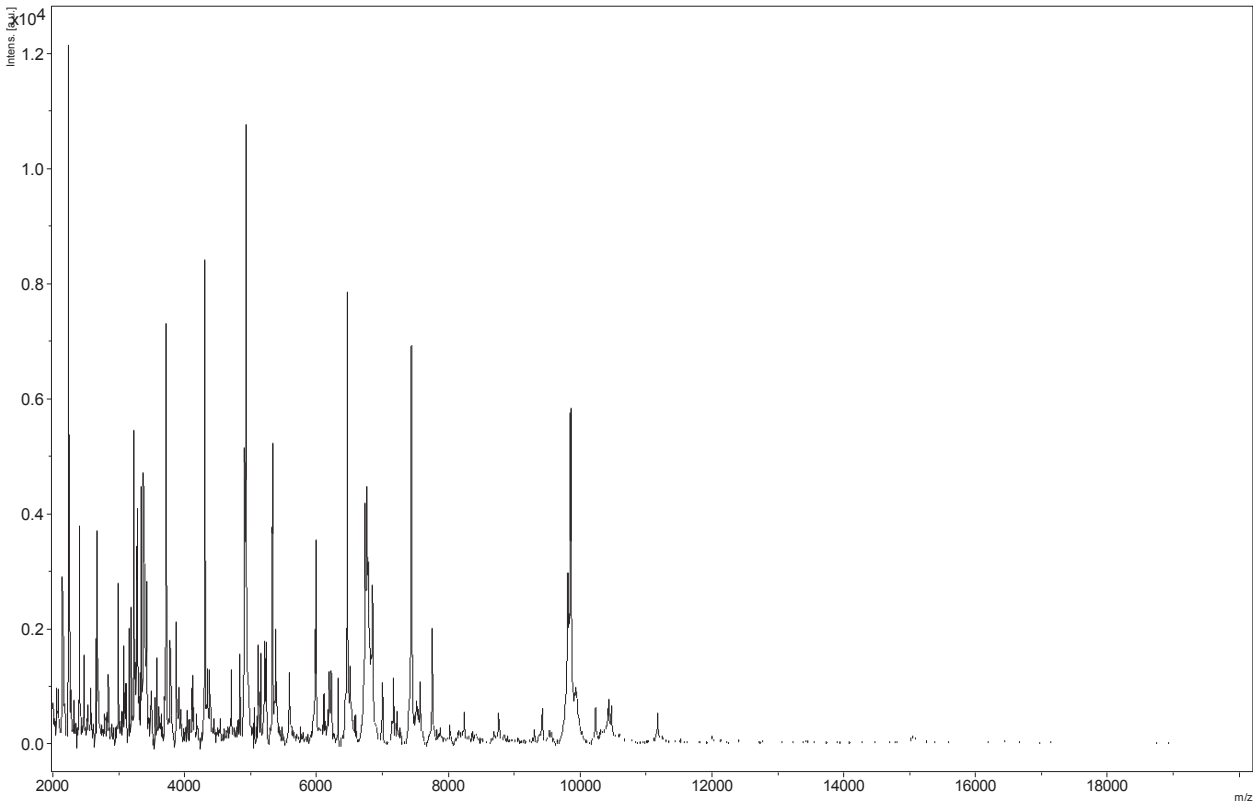

FIG. 4. Reference mass spectrum from *Virgibacillus massiliensis* strain Vm-5^T. Spectra from 10 individual colonies were compared and reference spectrum generated.

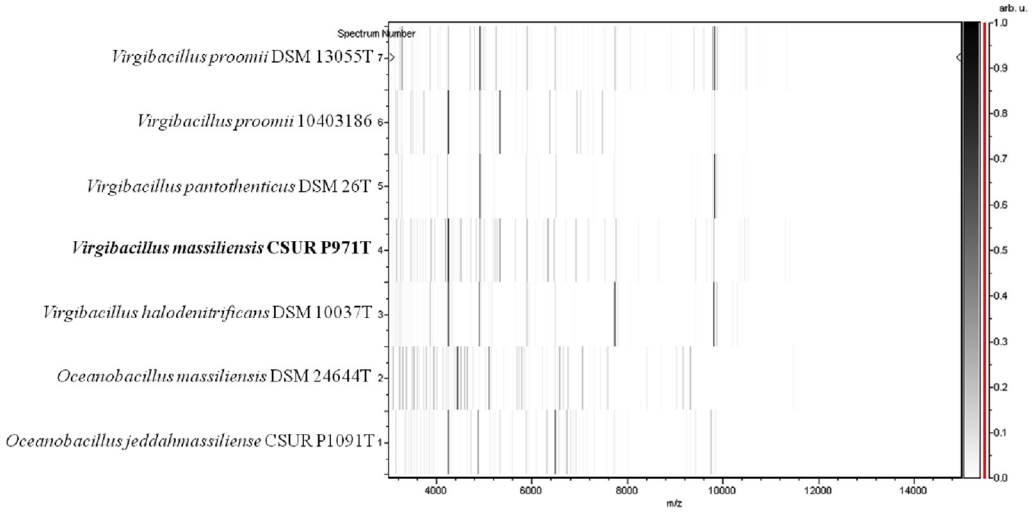

FIG. 5. Gel view comparing *Virgibacillus massiliensis* strain Vm-5^T (= CSURP971^T = DSM 28587) to other species within genus *Virgibacillus* and *Oceanobacillus*. Gel view displays raw spectra of loaded spectrum files arranged in pseudo-gel-like look. X-axis records m/z value. Left y-axis displays running spectrum number originating from subsequent spectra loading. Peak intensity is expressed by greyscale scheme code. Color bar and right y-axis indicate relation between color peak, with peak intensity in arbitrary units. Displayed species are indicated at left.

by an overnight incubation with 20 µg/µL proteinase K at 37°C. Extracted DNA was then purified using 3 successive phenol–chloroform extractions and ethanol precipitations at −20°C overnight. After centrifugation, the DNA was resuspended in 160 µL TE buffer. The yield and concentration was measured by the Quant-it Picogreen kit (Invitrogen, Waltham, MA, USA) on a Genios-Tecan fluorometer at 40.5 ng/µL.

Genome sequencing and assembly

Genomic DNA (gDNA) of *V. massiliensis* Vm-5^T was sequenced on the MiSeq sequencer (Illumina, San Diego, CA, USA) using the mate pair strategy. The gDNA was bar coded in order for it to be mixed with 11 other projects with the Nextera Mate Pair sample prep kit (Illumina). The mate pair library was prepared with 1 µg of 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 profile of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) with a DNA 7500 lab chip. The DNA fragments ranged in size from 1 to 10 kb. No size selection was performed, and only 14 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal at 696 bp on the Covaris device S2 in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies). The libraries were normalized at 2 nM and pooled. After a denaturation step and dilution at 10 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 42-hour run in a 2 × 251 bp. Total information of 4.7 Gb was obtained from a 488K/mm^2 cluster density, with a cluster passing quality control filters of 97.2% (9 590 000 clusters). Within this run, the index representation for *V. massiliensis* strain Vm-5^T was determined to be 11.16%.

Illumina reads where trimmed using Trimmomatic [43], then assembled through Spades software [44,45]. Contigs obtained were combined together by SSpace [46] and Opera software [47] helped by GapFiller [48] to reduce the set. Some manual refinements using CLC Genomics v7 software (CLC bio, Aarhus, Denmark) and homemade tools in Python improved the genome. Finally, the draft genome of *V. massiliensis* strain Vm-5^T consists of seven contigs.

TABLE 3. **Project information**

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality draft
MIGS-28	Libraries used	1 mate paired
MIGS-29	Sequencing platforms	MiSeq Illumina
MIGS-31.2	Sequencing coverage	62
MIGS-30	Assemblers	Spades
MIGS-32	Gene calling method	Prodigal
	GenBank ID	CCDP010000001–CCDP010000007
	GenBank date of release	September 2014
MIGS-13	Source material identifier	Vm-5^T
	Project relevance	Stool samples from healthy Amazonian boy

MIGS, minimum information about a genome sequence.

Genome annotation

Noncoding genes and miscellaneous features were predicted using RNAmmer [49], ARAGORN [50], Rfam [51], PFAM [52]

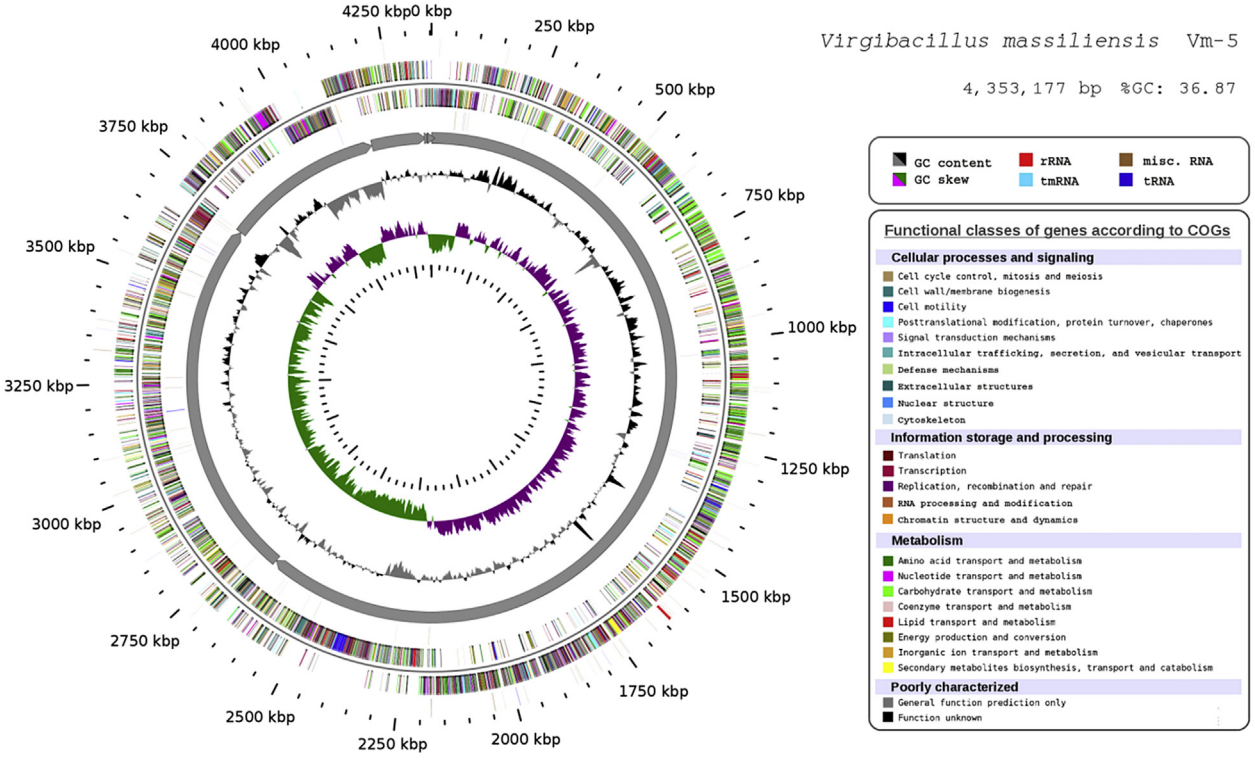

FIG. 6. Circular representation of *Virgibacillus massiliensis* Vm-5^T genome. Circles from center to outside: GC screw (green/purple), GC content (grey/black), scaffolds in grey arrows, tRNA (blue), rRNA (red), tmRNA (light blue), miscellaneous RNA (brown) on forward strand, genes on forward strand colored by COGs categories, genes on reverse strand colored by COGs categories, tRNA (blue), rRNA (red), tmRNA (light blue), miscellaneous RNA (brown) on reverse strand. COGs, Clusters of Orthologous Groups database.

and Infernal [53]. Coding DNA sequences (CDSs) were predicted using Prodigal [54], and functional annotation was achieved using BLAST+ [55] and HMMER3 [56] against the UniProtKB database [57].

Genome properties

The genome of *V. massiliensis* strain Vm-5^T contains 4 353 177 bp with a G + C content of 36.87% (Fig. 6, Table 4). One hundred twenty-five RNAs were detected, including five rRNAs (one 16S rRNA, one 23S rRNA, three 5S rRNA), 42 tRNAs and 78 miscellaneous RNAs. Overall, 4394 genes were identified, representing a coding capacity of 3 754 518 bp (coding percentage, 86.25%). Among these genes, 322 (7.33%) were identified as putative proteins and 1107 (25.19%) were annotated as hypothetical proteins. Moreover, 4291 genes matched a least one sequence in the Clusters of Orthologous Groups (COGs) database [58,59] with BLASTP default parameters. The properties and the statistics of the genome are summarized in Table 4. The distribution of genes into COGs functional categories is presented in Table 5.

Insights into the genome sequence

We compared the genome of *V. massiliensis* strain Vm-5^T to that of *V. halodenitrificans* strain 1806 (ALEF01000001), which is currently the closest available sequenced genome based on 16S rRNA comparison (Table 6).

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

Attribute	Value	% of total[a]
Genome size (bp)	4 353 177	100
DNA coding region (bp)	3 754 518	86.25
DNA G + C content (bp)	1 605 022	36.87
Total genes	4519	100
rRNA	5	0.11
tRNA	42	0.93
tmRNA	0	0
miscRNA	78	1.73
Protein-coding genes	4394	86.25
Genes with function prediction	3287	74.81
Genes assigned to COGs	4291	97.66

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

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

Code	Value	% of total[a]	Description
J	208	4.85	Translation, ribosomal structure and biogenesis
A	3	0.07	RNA processing and modification
K	340	7.92	Transcription
L	207	4.82	Replication, recombination and repair
B	9	0.21	Chromatin structure and dynamics
D	58	1.35	Cell cycle control, cell division, chromosome partitioning
Y	0	0.0	Nuclear structure
V	65	1.51	Defense mechanisms
T	141	3.29	Signal transduction mechanisms
M	222	5.17	Cell wall/membrane biogenesis
N	75	1.75	Cell motility
Z	4	0.09	Cytoskeleton
W	0	0	Extracellular structures
U	51	1.19	Intracellular trafficking and secretion, and vesicular transport
O	149	3.47	Posttranslational modification, protein turnover, chaperones
C	269	6.27	Energy production and conversion
G	360	8.39	Carbohydrate transport and metabolism
E	371	8.65	Amino acid transport and metabolism
F	122	2.84	Nucleotide transport and metabolism
H	134	3.12	Coenzyme transport and metabolism
I	153	3.57	Lipid transport and metabolism
P	234	5.45	Inorganic ion transport and metabolism
Q	47	1.1	Secondary metabolites biosynthesis, transport and catabolism
R	510	11.89	General function prediction only
S	662	15.43	Function unknown

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

The draft genome sequence of *V. halodenitrificans* strain 1806 had a smaller size compared to *V. massiliensis* strain Vm-5^T (3.9 Mb vs. 4.3 Mb, respectively), a smaller total number of genes (3886 and 4394 genes, respectively), and a lower ratio of genes per Mb (996.4 genes/Mb vs. 1010, respectively), but a higher G + C content (37.41% and 36.87, respectively).

Genome Comparison

At the time of analysis, only four whole genome sequences of *Virgibacillus* were available at the National Center for Biotechnology Information. Therefore, whole genome comparison was done between *V. massiliensis, V. alimentarius* (GenBank accession number NZ_JFBD00000000), *V. halodenitrificans* (NZ_ALEF01000000) and *V. senegalensis* (NZ_CCXU01000000) (Table 7). Among *Virgibacillus* genomes, that of *V. massiliensis* (4.35 Mb) is the largest, followed by *V. halodenitrificans* (3.92 Mb), *V. senegalensis* (3.92) and *V. alimentarius* (3.05 Mb).

To estimate the mean level of nucleotide sequence similarity at the genome level between *V. massiliensis* and the other four *Virgibacillus* genomes, we calculated the average genomic identity of orthologous gene sequences (AGIOS) values using an in-lab pipeline named Marseille Average Genomic Identity (MAGi). Briefly, this pipeline combines Proteinortho 4 software (with the following parameters: e-value 1e^{-05}, 30% of identity, 50% coverage and algebraic connectivity of 50%) for detecting orthologous proteins between genomes compared pairwise, then retrieves the corresponding gene nucleotide sequences and determines the mean percentage of nucleotide sequence identity among orthologous open reading frames using the Needleman-Wunsch global alignment algorithm [5]. Similarity values at the genome level were also calculated using genome-to-genome distance (GGDC) calculator software [60].

The number of orthologous genes is indicated in Table 7. The AGIOS and GGDC values obtained using the MAGi

TABLE 6. Percentage of genes associated with the 25 general COGs functional categories for *Virgibacillus massiliensis* Vm-5^T and *Virgibacillus halodenitrificans* 1806

Code	Description	*V. massiliensis* Vm-5^T (% of total)	*V. halodenitrificans* 1806 (% of total)	Difference (%)
J	Translation, ribosomal structure and biogenesis	4.85	4.16	0.69
A	RNA processing and modification	0.07	0.0	0.07
K	Transcription	7.92	6.2	1.72
L	Replication, recombination and repair	4.82	3.12	1.7
B	Chromatin structure and dynamics	0.21	0.02	0.19
D	Cell cycle control, cell division, chromosome partitioning	1.35	0.83	0.52
Y	Nuclear structure	0.0	0.0	0.0
V	Defense mechanisms	1.51	1.25	0.26
T	Signal transduction mechanisms	3.29	3.64	-0.35
M	Cell wall/membrane biogenesis	5.17	3.9	1.27
N	Cell motility	1.75	1.54	0.21
Z	Cytoskeleton	0.09	0.0	0.09
W	Extracellular structures	0.0	0.0	0.0
U	Intracellular trafficking and secretion, and vesicular transport	1.19	1.23	-0.04
O	Posttranslational modification, protein turnover, chaperones	3.47	2.44	1.03
C	Energy production and conversion	6.27	4.26	2.01
G	Carbohydrate transport and metabolism	8.39	5.46	2.93
E	Amino acid transport and metabolism	8.65	7.9	0.75
F	Nucleotide transport and metabolism	2.84	2.29	0.55
H	Coenzyme transport and metabolism	3.12	2.96	0.16
I	Lipid transport and metabolism	3.57	2.41	1.16
P	Inorganic ion transport and metabolism	5.45	4.49	0.96
Q	Secondary metabolites biosynthesis, transport and catabolism	1.1	1.44	-0.34
R	General function prediction only	11.89	9.79	2.1
S	Function unknown	15.43	30.65	-15.22

COGs, Clusters of Orthologous Groups database.

TABLE 7. Numbers of orthologous proteins shared between genomes[a]

	V. massiliensis	*V. alimentarius*	*V. halodenitrificans*	*V. senegalensis*
V. massiliensis	**4394**	1811	2283	1903
V. alimentarius		**3070**	1920	2036
V. halodenitrificans			**7088**	1611
V. senegalensis				**3824**

Bold numbers indicate numbers of proteins of each genome.
[a]Shown are differences of gene number (in percentage) related to each Clusters of Orthologous Groups (COGs) database category between *Virgibacillus massiliensis* Vm-5^T and *Virgibacillus halodenitrificans* 1806. The proportion of COGs is similar between the two species. The maximum difference is related to "Function unknown" with 15.22% and "Carbohydrate transport and metabolism" with 2.93%.

pipeline and GGDC software, respectively, among the different studied genomes are summarized in Table 8. The AGIOS values among *Virgibacillus* genomes ranged from 65.54 between *V. alimentarius* and *V. senegalensis* to 69.66% between *V. alimentarius* and *V. halodenitrificans*. The AGIOS values obtained between *V. massiliensis* and other compared species were within this range (65.78% with *V. senegalensis* to 66.88% with *V. halodenitrificans*). The GGDC values among *Virgibacillus* genomes ranged from 19.1% between *V. alimentarius* and *V. halodenitrificans* to 27.7% between *V. senegalensis* and *V. halodenitrificans*. The GGDC values obtained between *V. massiliensis* and other compared species were also within a similar range (17.8% with *V. alimentarius* to 26.8% with *V. senegalensis*). These values are consistent with the status of new species of *V. massiliensis*.

Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Virgibacillus massiliensis* sp. nov., represented here by strain Vm-5^T. This strain was isolated from a stool specimen from a healthy Amazonian boy. This description was based on a single isolate, similarly to the descriptions of *V. halotolerans* (Seiler and Wenning, 2013) and *V. oceani* (Yin *et al.*, 2015).

Taxonomic and nomenclatural proposals Description of *Virgibacillus massiliensis* sp. nov.

Virgibacillus massiliensis (mas.si.li.en'sis. L. masc. adj. *massiliensis*, from Massilia, the Roman name for Marseille, France, where the type strain was isolated). Growth occurred between 15 and 45°C on a salt-enriched culture medium. Strain Vm-5^T required NaCl for growth and grew at salinity ranging from 5 to 200 g/L of NaCl (optimum, 100 g/L). The optimal growth was observed at 37°C in aerobic atmosphere. The optimal pH for growth was 7.5 (pH range 5 to 9). Strain Vm-5^T grew in presence of 5% CO_2 but not in a microaerophilic or anaerobic atmosphere. Colonies were circular, greyish, shiny and smooth, with a diameter of 2 to 5 mm. Cells stained Gram positive. Cells were motile by polar flagella, spore forming (2 to 6 μm in length and 0.5 μm in diameter) and generally occurred as single cells or in pairs. Strain Vm-5^T exhibited catalase and oxidase activities. Strain Vm-5^T was positive for nitrate reduction but negative for phosphatase alkaline activity, β-galactosidase, αN-acetyl-β-glucosaminidase and urease. Strain Vm-5^T was negative for ribose, L-arabinose and D-lactose assimilation and positive for D-glucose, D-fructose, D-mannose, D-mannitol, D-maltose and D-sucrose. Strain Vm-5^T was susceptible to penicillin, ampicillin, amoxicillin, ceftriaxone, imipenem doxycycline, rifampicin, vancomycin, nitrofurantoin, erythromycin, ciprofloxacin and gentamicin but was resistant to trimethoprim/sulfamethoxazole and metronidazole.

The percentage of G + C content of the genome is 36.87%. The 16S rRNA and genome sequences are deposited in GenBank under accession numbers HG931931 and CCDP010000001 to CCDP010000007, respectively. The habi tat of the microorganism is the human digestive tract. The type strain Vm-5^T (= CSUR P971 = DSM 28587) was isolated from the stool specimen of a healthy Amazonian boy.

TABLE 8. AGIOS (upper right) and GGDC[a] values (lower left) shared between *Virgibacillus* genomes

	V. massiliensis	*V. alimentarius*	*V. halodenitrificans*	*V. senegalensis*
V. massiliensis	100	69.73	69.88	65.78
V. alimentarius	17.8 ± 2.5	100	69.66	65.54
V. halodenitrificans	19.1 ± 2.5	19.1 ± 2.5	100	65.91
V. senegalensis	26.8 ± 2.5	26.5 ± 2.5	27.7 ± 2.5	100

AGIOS, average genomic identity of orthologous gene sequences; GGDC, genome-to-genome distance.
[a]GGDC values were calculated by formula 2. Standard deviations are provided for each GGDC value.

Acknowledgement

Funded by the Méditerranée Infection Foundation.

Conflict of Interest

None declared.

References

[1] 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.
[2] Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, Krichevsky MI, et al. Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. Int J Syst Bacteriol 1987;37:463–4.
[3] Rossello-Mora 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.
[4] Welker M, Moore ER. Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. Syst Appl Microbiol 2011;34:2–11.
[5] 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 new bacterial species. Int J Syst Evol Microbiol 2014;64:384–91.
[6] Kokcha S, Mishra AK, Lagier JC, Million M, Leroy Q, Raoult D, et al. Non-contiguous finished genome sequence and description of *Bacillus timonensis* sp. nov. Stand Genomic Sci 2012;6:346–55.
[7] Ramasamy D, Lagier JC, Gorlas A, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Bacillus massiliosenegalensis* sp. nov. Stand Genomic Sci 2013;8:264–78.
[8] Mishra AK, Pfleiderer A, Lagier JC, Robert C, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Bacillus massilioanorexius* sp. nov. Stand Genomic Sci 2013;8:465–79.
[9] Keita MB, Diene SM, Robert C, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Bacillus massiliogorillae* sp. nov. Stand Genomic Sci 2013;9:93–105.
[10] Roux V, Million M, Robert C, Magne A, Raoult D. Non-contiguous finished genome sequence and description of *Oceanobacillus massiliensis* sp. nov. Stand Genomic Sci 2013;9:370–84.
[11] Heyndrickx M, Lebbe L, Kersters K, De Vos P, Frosyth H, Logan NA. Virgibacillus: a new genus to accommodate *Bacillus pantothenticus* (Proom and Knight 1950). Emended description of *Virgibacillus pantothenticus*. Int J Syst Bacteriol 1998;48:99–106.
[12] Carrasco IJ, Marquez MC, Ventosa A. *Virgibacillus salinus* sp. nov., a moderately halophilic bacterium from sediment of a saline lake. Int J Syst Evol Microbiol 2009;59:3068–73.
[13] Chen YG, Cui XL, Wang YX, Zhang YQ, Tang SK, Li WJ, et al. *Virgibacillus sediminis* sp. nov., a moderately halophilic bacterium isolated from a salt lake in China. Int J Syst Evol Microbiol 2009;59:2058–63.
[14] Jeon C, Kim J, Park DJ, Xu LH, Jiang CL, Kim CJ. *Virgibacillus xinjiangensis* sp. nov., isolated from a Salt Lake of Xin-jiang Province in China. J Microbiol 2009;47:705–9.
[15] Zhang YJ, Zhou Y, Ja M, Shi R, Chun-Yu WX, Yang LL, et al. *Virgibacillus albus* sp. nov., a novel moderately halophilic bacterium isolated from Lop Nur salt lake in Xinjiang Province, China. Antonie Van Leeuwenhoek 2012;102:553–60.
[16] Kim J, Jung MJ, Roh SW, Nam YD, Shin KS, Bae JW. *Virgibacillus alimentarius* sp. nov., isolated from a traditional Korean food. Int J Syst Evol Microbiol 2011;61:2851–5.
[17] Niederberger TD, Steven B, Charvet S, Barbier B, Whyte LG. *Virgibacillus arcticus* sp. nov., a moderately halophilic, endospore-forming bacterium from permafrost in the Canadian high Arctic. Int J Syst Evol Microbiol 2009;59:2219–25.
[18] Lee SY, Kang CH, Oh TK, Yoon JH. *Virgibacillus campisalis* sp. nov., from a marine solar saltern. Int J Syst Evol Microbiol 2012;62:347–51.
[19] Wang CY, Chang CC, Ng CC, Chen TW, Shyu YT. *Virgibacillus chiguensis* sp. nov., a novel halophilic bacterium isolated from Chigu, a previously commercial saltern located in southern Taiwan. Int J Syst Evol Microbiol 2008;58:341–5.
[20] Yoon JH, Oh TK, Park YH. Transfer of *Bacillus halodenitrificans* Denariaz et al. 1989 to the genus *Virgibacillus* as *Virgibacillus halodenitrificans* comb. nov. Int J Syst Evol Microbiol 2004;54:2163–7.
[21] Yoon JH, Kang SJ, Jung YT, Lee KC, Oh HW, Oh TK. *Virgibacillus byunsanensis* sp. nov., isolated from a marine solar saltern. Int J Syst Evol Microbiol 2010;60:291–5.
[22] Heyrman J, Logan NA, Busse HJ, Balcaen A, Lebbe L, Rodriguez-Diaz M, et al. *Virgibacillus carmonensis* sp. nov., *Virgibacillus necropolis* sp. nov. and *Virgibacillus picturae* sp. nov., three novel species isolated from deteriorated mural paintings, transfer of the species of the genus *Salibacillus* to *Virgibacillus*, as *Virgibacillus marismortui* comb. nov. and *Virgibacillus salexigens* comb. nov., and emended description of the genus *Virgibacillus*. Int J Syst Evol Microbiol 2003;53:501–11.
[23] Waino M, Tindall BJ, Schumann P, Ingvorsen K. *Gracilibacillus* gen. nov., with description of *Gracilibacillus halotolerans* gen. nov., sp. nov.; transfer of *Bacillus dipsosauri* to *Gracilibacillus dipsosauri* comb. nov., and *Bacillus salexigens* to the genus *Salibacillus* gen. nov., as *Salibacillus salexigens* comb. nov. Int J Syst Evol Bacteriol 1999;49:821–31.
[24] 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.
[25] Seiler H, Wenning M. *Virgibacillus halotolerans* sp. nov., isolated from a dairy product. Int J Syst Evol Microbiol 2013;63:3358–63.
[26] Chen YG, Cui XL, Fritze D, Chai LH, Schumann P, Wen ML, et al. *Virgibacillus kekensis* sp. nov., a moderately halophilic bacterium isolated from a salt lake in China. Int J Syst Evol Microbiol 2008;58:647–53.
[27] Quesada T, Aguilera M, Morillo JA, Ramos-Cormenzana A, Monteoliva-Sanchez M. *Virgibacillus olivae* sp. nov., isolated from waste wash-water from processing of Spanish-style green olives. Int J Syst Evol Microbiol 2007;57:906–10.
[28] Hua NP, Hamza-Chaffai A, Vreeland RH, Isoda H, Naganuma T. *Virgibacillus salarius* sp. nov., a halophilic bacterium isolated from a Saharan salt lake. Int J Syst Evol Microbiol 2008;58:2409–14.
[29] Tanasupawat S, Chamroensaksri N, Kudo T, Itoh T. Identification of moderately halophilic bacteria from Thai fermented fish and proposal of *Virgibacillus siamensis* sp. nov. J Gen Appl Microbiol 2010;56:369–79.
[30] Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol 2008;26:541–7.
[31] Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea, Bacteria*, and *Eucarya*. Proc Natl Acad Sci U S A 1990;87:4576–9.
[32] Murray RGE. The higher taxa, or, A place for everything...? In: Holt JG, editor. Bergey's manual of systematic bacteriology. Baltimore: Williams & Wilkins; 1984. p. 1:31–34.
[33] Garrity GM, Holt JG. The road map to the manual. In: Garrity GM, Boone DR, Castenholz RW, et al., editors. Bergey's manual of systematic bacteriology. 2nd ed. New York: Springer; 2001. p. 1:119–169.
[34] Gibbons NE, Murray RGE. Proposals concerning the higher taxa of bacteria. Int J Syst Bacteriol 1978;28:1–6.

[35] Ludwig W, Schleifer KH, Whitman WB. Class I. *Bacilli* class nov. In: De Vos P, Garrity G, Jones D, et al., editors. Bergey's manual of systematic bacteriology. 2nd ed. New York: Springer-Verlag; 2009. p. 3:19–20.
[36] Euzéby J. List of new names and new combinations previously effectively, but not validly, published. List no. 132. Int J Syst Evol Microbiol 2010;60:469–72.
[37] Prévot AR. Clostridiales. In: Hauduroy P, Ehringer G, Guillot G, Magrou J, Prevot AR, Rosset, et al., editors. Dictionnaire des bactéries pathogènes. 2nd ed. Paris: Masson et Cie; 1953. p. 142–65.
[38] Skerman VBD, McGowan V, Sneath PHA. Approved lists of bacterial names. Int J Syst Bacteriol 1980;30:225–420.
[39] Fischer A. Untersuchungen über bakterien. Jahrbücher für Wissenschaftliche Botanik 1895;27:1–163.
[40] Hirota K, Aino K, Nodasaka Y, Yumoto I. *Oceanobacillus indicireducens* sp. nov., a facultative alkaliphile that reduces an indigo dye. Int J Syst Evol Microbiol 2013;63:1437–42.
[41] Yumoto I, Hirota K, Nodasaka Y, Nakajima K. *Oceanobacillus oncorhynchi* sp. nov., a halotolerant obligate alkaliphile isolated from the skin of a rainbow trout (*Oncorhynchus mykiss*), and emended description of the genus *Oceanobacillus*. Int J Syst Evol Microbiol 2005;55:1521–4.
[42] 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* as *Oceanobacillus picturae* comb. nov. with emended descriptions. Int J Syst Evol Microbiol 2006;56:251–7.
[43] Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, et al. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res 2012;40(Web Server issue):W622–7.
[44] Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol 2013;20:714–37.
[45] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 2012;19:455–77.
[46] Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 2011;27:578–9.
[47] Gao S, Sung WK, Nagarajan N. Opera: reconstructing optimal genomic scaffolds with high-throughput paired-end sequences. J Comput Biol 2011;18:1681–91.
[48] Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. Genome Biol 2012;13:R56.
[49] 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.
[50] Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 2004;32:11–6.
[51] Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. Nucleic Acids Res 2003;31:439–41.
[52] Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. Nucleic Acids Res 2012;40:D290–301.
[53] Nawrocki EP, Kolbe DL, Eddy SR. Infernal 1.0: inference of RNA alignments. Bioinformatics 2009;25:1335–7.
[54] 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.
[55] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics 2009;10:421.
[56] Eddy SR. Accelerated profile HMM searches. PLoS Comp Biol 2011;7:e1002195.
[57] UniProt Consortium. Ongoing and future developments at the Universal Protein Resource. Nucleic Acids Res 2011;39:D214–9.
[58] Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genomoe-scale analysis of protein functions and evolution. Nucleic Acids Res 2000;28:33–6.
[59] Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. Science 1997;278:631–7.
[60] 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.