High-quality genome sequence and description of *Bacillus ndiopicus* strain FF3^T sp. nov.

C. I. Lo[1,2], R. Padhmanabhan[1], O. Mediannikov[1,2], A. Caputo[1], C. Michelle[1], N. Faye[3], C. Sokhna[2], D. Raoult[1,2,4], P.-E. Fournier[1] and F. Fenollar[1,2]

1) *Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, Marseille, France,* 2) *Campus International UCAD-IRD, Dakar,* 3) *Université Cheikh Anta Diop de Dakar, Laboratoire de Parasitologie générale, Fann, Senegal and* 4) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Strain FF3^T was isolated from the skin-flora of a 39-year-old healthy Senegalese man. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry did not allow any identification. This strain exhibited a 16S rRNA sequence similarity of 96.8% with *Bacillus massiliensis*, the phylogenetically closest species with standing nomenclature. Using a polyphasic study made of phenotypic and genomic analyses, strain FF3^T was Gram-positive, aeroanaerobic and rod shaped and exhibited a genome of 4 068 720 bp with a G+C content of 37.03% that coded 3982 protein-coding and 67 RNA genes (including four rRNA operons). On the basis of these data, we propose the creation of *Bacillus ndiopicus* sp. nov.

Keywords: *Bacillus ndiopicus*, genome, Senegal, skin, taxonogenomics

Original Submission: 11 August 2015; **Revised Submission:** 7 October 2015; **Accepted:** 9 October 2015

Article published online: 19 October 2015

Corresponding author: F. Fenollar, Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille Cedex 05, France

E-mail: florence.fenollar@univ-amu.fr

Introduction

Bacillus subtilis was the first type species described in the genus *Bacillus* (Cohn 1872) [1]. Currently there are 301 species and seven subspecies with validly published names [2]. Generally members of this genus are environmental bacteria present in soil, food, and fresh and sea water. In humans, some strains can be pathogenic, such as *Bacillus cereus* (associated mainly with food poisoning) and *Bacillus anthracis* (the causative agent of anthrax) [3–5]. Other strains are saprophytes [6]. Several *Bacillus* species are also isolated from different plants in which they are endophytes [7].

Recently high-throughput genome sequencing and mass spectrometry analyses of bacteria have given unprecedented access to an abundance of genetic and proteomic information [8–10]. Currently a polyphasic approach is performed to describe new bacterial taxa, including their genome sequence, matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) spectrum, and major phenotypic characteristics such as Gram staining, culture, metabolic characteristics, habitat and (if applicable) pathogenicity [9,10].

Bacillus ndiopicus strain FF3^T (= CSUR P3025 = DSM 27837) is designated as the type strain of *Bacillus ndiopicus.* This bacterium is a Gram-positive rod that is aeroanaerobic. This bacterium was isolated from the skin of a healthy Senegalese man as part of a culturomics [11] study aiming at cultivating bacterial species from skin flora.

Here we provide a summary classification and set of features for *B. ndiopicus* sp. nov. strain FF3^T, together with the description of the complete genomic sequencing and annotation. These characteristics support the circumscription of the species *B. ndiopicus.*

Organism information

Classification and features

In December 2012, a skin specimen was sampled with a swab from a healthy Senegalese volunteer living in Ndiop, a rural village in the Guinean–Sudanian area in Senegal (Table 1). This 39-year-old man was included in a research project approved by the National Ethic Committee for health research (CNERS) in Senegal and the ethics committee of the Institut Fédératif de Recherche IFR48, Faculty of Medicine, Marseille, France (agreements 09-022 and 11-017) [12].

Strain FF3^T (Table 1) was isolated by cultivation on 5% blood's sheep enriched Columbia agar (bioMérieux, Marcy l'Etoile, France), under aerobic conditions, in December 2012.

http://dx.doi.org/10.1016/j.nmni.2015.10.009

TABLE 1. Classification and general features of *Bacillus ndiopicus* strain FF3^T [15]

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain: *Bacteria*	TAS [27]
		Phylum: *Firmicutes*	TAS [28,29]
		Class: *Bacilli*	TAS [30,31]
		Order: *Bacillales*	TAS [32]
		Family: *Bacillaceae*	TAS [33]
		Genus: *Bacillus*	TAS [34,35]
		Species: *Bacillus ndiopicus*	IDA
		(Type) strain: FF3^T	IDA
	Gram stain	Positive	IDA
	Cell shape	Rods	IDA
	Motility	Motile	IDA
	Sporulation	Sporulating	NAS
	Temperature range	Mesophile	IDA
	Optimum temperature	37°C	IDA
	pH range; optimum	5.6–8.4; 7.0	IDA
	Carbon source	Unknown	
MIGS-6	Habitat	Human skin	IDA
MIGS-6	Salinity	Unknown	
MIGS-22	Oxygen requirement	Aeroanaerobic	IDA
MIGS-15	Biotic relationship	Free-living	IDA
MIGS-14	Pathogenicity	Unknown	
MIGS-4	Geographic location	Ndiop, Senegal	TAS
MIGS-5	Sample collection	December 2012	TAS
MIGS-4.1	Latitude	14.5333	TAS
MIGS-4.1	Longitude	−16.2667	TAS
MIGS-4.4	Altitude	5 m above sea level	TAS

MIGS, minimum information about a genome sequence.
[a]Evidence codes are as follows: IDA, inferred from direct assay; TAS, traceable author statement (i.e., a 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 anecdotal evidence). These evidence codes are from the Gene Ontology project (http://www.geneontology.org/GO.evidence.shtml) [36]. If the evidence code is IDA, then the property should have been directly observed, for the purpose of this specific publication, for a live isolate by one of the authors, or an expert or reputable institution mentioned in the acknowledgements.

B. ndiopicus strain FF3^T exhibited a 96.8% nucleotide sequence similarity with *Bacillus massiliensis* (Glazunova *et al.*, 2006), the phylogenetically closest *Bacillus* species (Fig. 1). These values were lower than the 98.7% 16S rRNA gene sequence threshold recommended by Meier-Kolthoff *et al.* [13] to delineate a new species within the phylum *Firmicutes* without carrying out DNA-DNA hybridization. Different growth temperatures (25, 30, 37, 45 and 56°C) were tested. Optimal growth was observed at 37 and 45°C after 24 hours of incubation; weak growth was noticed at 30°C. Colonies were 1 mm in diameter and transparent on 5% blood-enriched Columbia agar. Growth of the strain was tested under anaerobic and microaerophilic conditions using the GENbag anaer and GENbag microaer systems, respectively (bioMérieux), and under aerobic conditions, with or without 5% CO_2. Optimal growth was obtained under aerobic condition with 5% CO_2 and under microaerophilic condition at 37 and 45°C.

Gram staining showed Gram-positive rods (Fig. 2). The motility test was positive by means of peritrichous flagella. Cells grown on agar have a mean diameter of 1.2 µm (ranging from 0.8 to 1.6 µm) and a mean length of 2.5 µm (ranging from 1.8 to 3.2 µm) (Fig. 3).

Strain FF3^T exhibited catalase and oxidase activities. Using the API ZYM strip (bioMérieux), positive reactions were observed with alkaline phosphatase, esterase, α-chymotrypsin and lipase. Negative reactions were observed for leucine arylamidase, valine arylamidase, cystine arylamidase, phosphatase acid, trypsin, naphthol-AS-BI-phosphohydrolase, β-glucuronidase, α-glucosidase, β-glucosidase, N-acetyl-β-glucosaminidase, α-mannosidase and α-fucosidase. Using the API 20E strip (bioMérieux), only the citrate test was positive; all others tests were negative, including indole, β-galactosidase, urease, ornithine decarboxylase, mannitol, sorbitol and rhamnose fermentation. Using the API 50CH strip (bioMérieux), no positive reaction was observed, including for glycerol, D-arabinose, D-xylose, L-rhamnose, amygdalin, D-cellobiose, D-fucose, potassium 5-ketogluconate, L-arabitol, starch, D-maltose and D-mannose. *B. ndiopicus* was susceptible *in vitro* to penicillin, amoxicillin, amoxicillin−clavulanic acid, ceftriaxone, imipenem, gentamicin, ciprofloxacin, erythromycin, doxycycline, rifampicin and vancomycin, but resistant to nitrofurantoin and metronidazole. When compared with representative species from the genus *Bacillus*, *B. ndiopicus* strain FF3^T exhibited several phenotypic differences, which are summarized in Table 2.

MALDI-TOF protein analysis was performed using a Microflex LT (Bruker Daltonics, Leipzig, Germany), as previously reported [14]. The scores previously established by Bruker allowing validating (or not) the identification of species compared to the database of the instrument were applied. Briefly, a score of $\geq$2.000 with a species with a validly published name provided allows the identification at the species level; a score of $\geq$1.700 and <2.000 allows the identification at the genus level; and a score of <1.700 does not allow any identification. We performed 12 distinct deposits from 12 isolated colonies of strain FF3^T. Two microliters of matrix solution (saturated solution of alpha-cyano-4-hydroxycinnamic acid) in 50% acetonitrile and 2.5% trifluoroacetic acid were distributed on each smear and submitted at air drying for 5 minutes. Then the spectra from the 12 different colonies were imported into MALDI Biotyper 2.0 software (Bruker) and analysed by standard pattern matching (with default parameter settings) against the main spectra of 6252 bacterial spectra including 199 spectra from 104 *Bacillus* species. Scores ranging from 1.2 to 1.4 were obtained for strain FF3^T, suggesting that this isolate was not a member of any known species. The reference mass spectrum from strain FF3^T was incremented in our database (Fig. 4). The gel view highlighted spectrum differences with other *Bacillus* species (Fig. 5).

Genome sequencing information

Genome project history

The organism was selected for sequencing on the basis of its 16S rRNA similarity, phylogenetic position and phenotypic

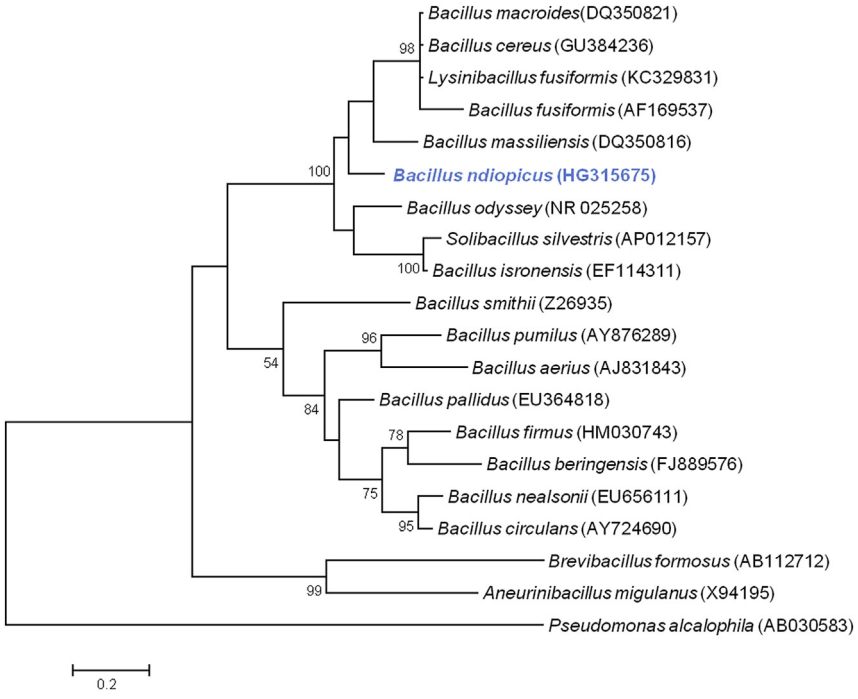

FIG. 1. Phylogenetic tree highlighting the position of *Bacillus ndiopicus* strain FF3^T relative to the most closely related type strains within the genus *Bacillus*. The strains and their corresponding GenBank accession numbers for 16S rRNA genes are provided (type = ^{T}), and in parentheses we indicate GA if the genome is available or GNA if the genome is not available at the National Center for Biotechnology Information website: *Bacillus macroides* strain LMG 18474 (GNA), *Bacillus cereus* strain ZQN6, *Lysinibacillus fusiformis* strain H1k (GA: AYMK00000000), *Bacillus massiliensis* strain 4400831 (GA: JPVQ00000000), *Bacillus ndiopicus* strain FF3^T (GA: CCAP000000000), *Bacillus odyssey* strain NBRC 100172 (GA: JPVP00000000), *Solibacillus silvestris* strain StLB046 (GA: AP012157), *Bacillus isronensis* (GA: AMCK00000000), *Bacillus smithii* strain 7_3_47FAA (GA: ACWF00000000), *Bacillus pumilus* strain BA06 (GA: AMDH00000000), *Bacillus aerius* strain 24K, *Bacillus pallidus* strain CW 7, *Bacillus firmus* strain DS1 (GA: APVL00000000), *Bacillus beringensis* strain BR035 (GNA), *Bacillus nealsonii* strain AAU1 (GA: ASRU00000000), *Bacillus circulans* NBRC 13626 (GNA), *Brevibacillus formosus* strain F12 (GNA), *Aneurinibacillus migulanus* (GA: GCA_000878905), and *Pseudomonas alcaliphila* strain JAB1 (GNA). Sequences were aligned using MUSCLE [40] and the phylogenetic tree inferred by the maximum likelihood method with Kimura two-parameter model from MEGA6 software [41]. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 1000 times to generate a majority consensus tree. *P. alcaliphila* was used as outgroup. Scale bar = rate of substitution per site of 0.2.

differences with other members of the genus *Bacillus,* which support that *Bacillus ndiopicus* strain FF3^T likely represents a new bacterial species. This strain is part of a study aiming to characterize the skin flora of healthy Senegalese people. Currently there are more of 270 sequenced genomes of *Bacillus* species [8]. Strain FF3^T is the first genome of *B. ndiopicus* sp. nov., and its GenBank accession number is CCAP000000000. The genome consists of 23 large contigs. Table 3 shows the project information and its association with minimum information about a genome sequence (MIGS)

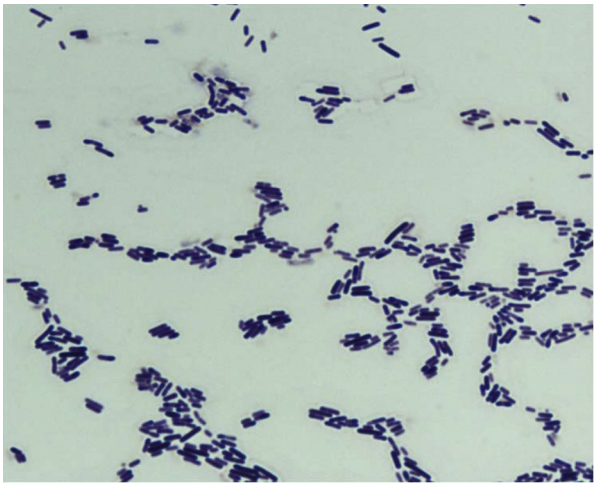

FIG. 2. Gram staining of *Bacillus ndiopicus* strain FF3^T.

2.0 compliance [15]; associated MIGS records are summarized.

Growth conditions and DNA isolation

Bacillus ndiopicus strain FF3^T (= CSUR P3025 = DSM 27837) was grown aerobically on 5% sheep's blood–enriched Columbia agar (bioMérieux) at 37°C. Then we suspended all bacterial colonies in 500 µL of Tris-EDTA (TE) buffer 10×. We remove 100 µL of this solution. This volume is completed by 400 µL TE buffer 10×, 25 µL proteinase K and 50 µL sodium dodecyl sulfate and then incubated overnight at 56°C for complete cells lysis. The next day this lysate is purified by washing with a phenol–chloroform solution three times. It is precipitated in absolute ethanol and incubated at −20°C for at least 2 hours. After a first centrifugation at 4°C for 30 minutes at 8000 rpm, the pellet is taken up in 70% ethanol kept at −20°C. A second centrifugation in the same conditions for 20 minutes is performed. After drying the tube in an oven at 37°C for 5 minutes, the DNA is taken up with 65 µL with buffer EB. The genomic DNA concentration was measured at 47.7 ng/µL by the Qubit assay with the high sensitivity kit (Life Technologies, Carlsbad, CA, USA).

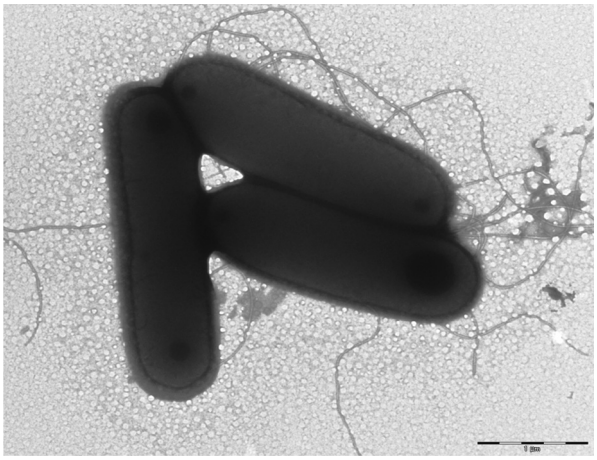

FIG. 3. Transmission electron microscopy of *Bacillus ndiopicus* strain FF3^T. Cells were observed on a Tecnai G20 device operated at 200 keV. Scale bar = 1 µm.

Genome sequencing and assembly

Genomic DNA of *Bacillus ndiopicus* was sequenced on the MiSeq Technology (Illumina, San Diego, CA, USA) with two applications, paired end and mate pair. The paired-end and the mate-pair strategies were barcoded in order to be mixed with 11 other genomic projects prepared with the Nextera XT DNA sample prep kit (Illumina) and 11 other projects with the Nextera Mate-Pair sample prep kit (Illumina).

The genomic DNA was diluted to 1 ng/µL to prepare the paired-end library. The tagmentation step fragmented and tagged the DNA with an optimal size distribution at 0.95 kb. Then limited-cycle PCR amplification (12 cycles) completed the tag adapters and introduced dual-index barcodes. After purification on AMPure XP beads (Beckman Coulter, Fullerton, CA, USA), the libraries were then normalized on specific beads according to the Nextera XT protocol (Illumina). Normalized libraries were pooled into a single library for sequencing on the MiSeq. The pooled single strand library was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and paired-end sequencing with dual index reads were performed in a single 39-hour run in 2 × 250 bp.

Total information of 6.8 Gb was obtained from a 807K/mm^2 cluster density, with a cluster passing quality control filters of 90.88% (14 553 000 clusters). Within this run, the index representation for *Bacillus ndiopicus* was determined to 17.96% and present 2 375 297 reads filtered according to the read qualities.

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 labchip. The DNA fragments were ranged in size from 1.5 to 13 kb, with an optimal size at 8 kb. No size selection was performed, and 600 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments on a Covaris device S2 in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent). 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

TABLE 2. Differential characteristics of *Bacillus ndiopicus* strain FF3^T with *B. kribbensis* [37], *B. massiliensis* [38], *B. vireti* [39], *B. soli* [39]

Property	*B. ndiopicus*	*B. kribbensis*	*B. massiliensis*	*B. vireti*	*B. soli*
Cell diameter (μm)	0.8–1.6	1.4–2.0	0.3–0.5	0.6–0.9	0.6–1.2
Oxygen requirement	Aeroanaerobic	Aerobic	Aerobic	Facultative anaerobic	Facultative anaerobic
Gram stain	+	+	–	–	Variable
Motility	+	+	+	+	+
Endospore formation	+	+	+	+	+
Production of:					
Alkaline phosphatase	+	NA	NA	NA	NA
Acid phosphatase	–	NA	NA	NA	NA
Catalase	+	+	+	NA	NA
Oxidase	–	–	+	NA	NA
Nitrate reductase	–	–	–	+	+
Urease	–	NA	+	–	–
α-Galactosidase	–	NA	NA	NA	NA
β-Galactosidase	–	NA	NA	NA	NA
β-Glucuronidase	–	+	NA	NA	NA
α-Glucosidase	–	+	NA	NA	NA
β-Glucosidase	–	+	NA	NA	NA
Esterase	+	+	NA	NA	NA
Esterase lipase	+	+	NA	NA	NA
Naphthol-AS-BI-phosphohydrolase	–	+	NA	NA	NA
N-acetyl-β-glucosaminidase	–	NA	NA	+	+
Utilization of:					
5-Keto-gluconate	–	NA	–	–	–
D-Xylose	–	+	–	–	–
D-Fructose	–	+	–	+	+
D-Glucose	–	+	–	+	+
D-Mannose	–	–	–	+	+
Habitat	Human skin	Soil	Human CSF	Soil	Soil

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

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 2 × 250 bp. *Bacillus ndiopicus* was determined to 8.09%. The 1 023 790 reads were filtered according to the read qualities. CLC Genomics Workbench 8.5.x was used for genome assembly.

Genome annotation

Open reading frames (ORFs) prediction was carried out using Prodigal [16] with default parameters. We removed the predicted ORFs if they spanned a sequencing gap region. Functional assessment of protein sequences was performed by comparing them with sequences in the GenBank [17] and Clusters of

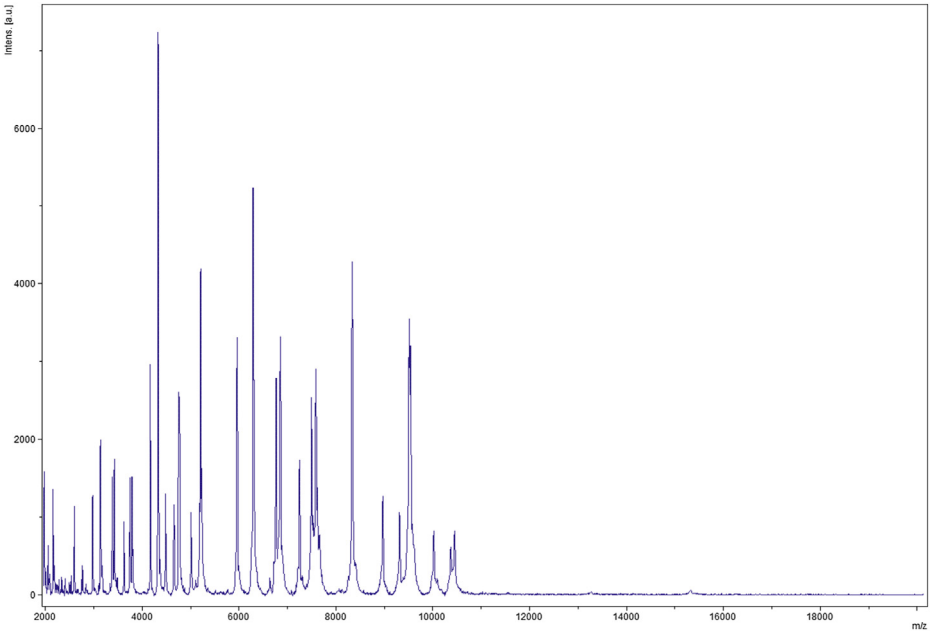

FIG. 4. Reference mass spectrum from *Bacillus ndiopicus* strain FF3^T. Spectra from 12 individual colonies were compared and reference spectrum generated.

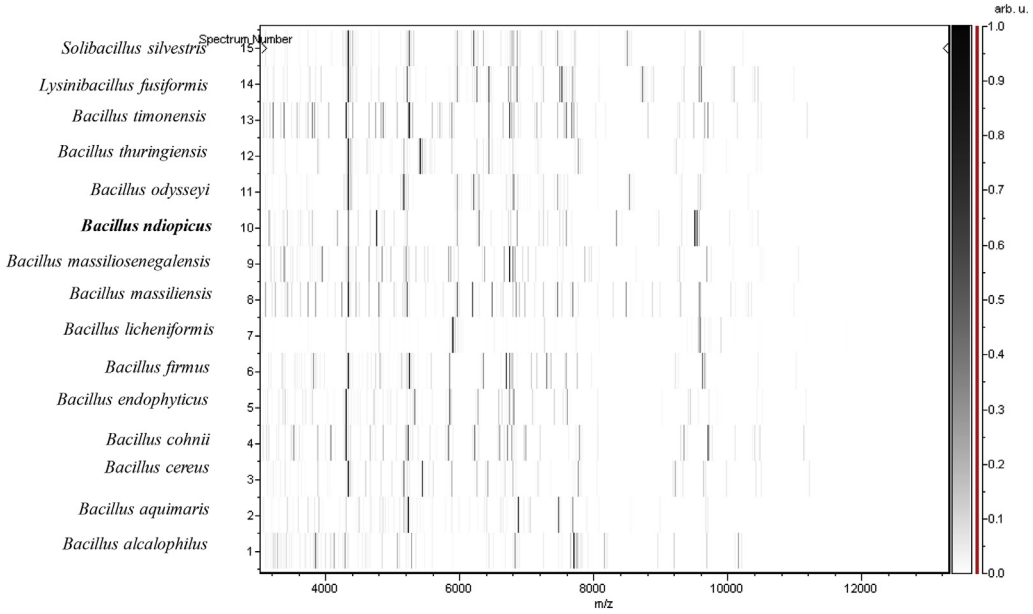

FIG. 5. Gel view comparing *Bacillus ndiopicus* strain FF3T spectrum to other members of family *Bacillaceae*. Gel view displays raw spectra of all loaded spectrum files arranged in pseudo-gel-like look. The *x*-axis records *m/z* value. Left *y*-axis displays running spectrum number originating from subsequent spectra loading. Peak intensity is expressed by greyscale scheme code. Color bar and right *y*-axis indicating relation between color peak is displayed, with peak intensity in arbitrary units. Displayed species are indicated at left.

Orthologous Groups (COGs) databases using BLASTP. tRNAs, rRNAs, signal peptides and transmembrane helices were identified using tRNAscan-SE 1.21 [18], RNAmmer [19], SignalP [20] and TMHMM [21], respectively. Artemis [22] was used for data management, and DNA Plotter [23] was used for visualization of genomic features. In-house Perl and bash scripts were used to automate these routine tasks. ORFans were sequences which have no homology in a given database—that is, nonredundant (nr) or identified if their BLASTP *E* value was lower than 1e-03 for alignment lengths greater than 80 aa. PHAST was used to identify, annotate and graphically display prophage sequences within bacterial genomes or plasmids [24].

TABLE 3. **Project information**

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality draft
MIGS-28	Libraries used	Paired end and mate pair
MIGS-29	Sequencing platforms	MiSeq
MIGS-31.2	Fold coverage	52×
MIGS-30	Assemblers	CLC genomics workbench
MIGS-32	Gene calling method	Prodigal
	Locus tag	Not reported
	GenBank ID	CCAP000000000
	GenBank date of release	March 18, 2014
	GOLD ID	Gp0101144
	BIOPROJECT	PRJNA224116
MIGS-13	Source material identifier	DSM 27837
	Project relevance	Study of human skin flora

MIGS, minimum information about a genome sequence.

To estimate the nucleotide sequence similarity at the genome level between *B. ndiopicus* and other members of *Bacillaceae* family, orthologous proteins were detected by Proteinortho software [25] (with the following parameters: *E* value 1e-5, 30% percentage of identity, 50% coverage and algebraic connectivity of 50%) and genomes compared two by two. After fetching the corresponding nucleotide sequences of orthologous proteins for each pair of genomes, we determined the mean percentage of nucleotide sequence identity using the Needleman-Wunsch global alignment algorithm. The script created to calculate AGIOS (average genomic identity of orthologous gene sequences) values was named MAGi (Marseille Average genomic identity) and is written in Perl and Bioperl modules.

Genome properties

The genome of *B. ndiopicus* strain FF3T is 4 068 720 bp long (one chromosome, no plasmid) with a 37.03% G+C content (Fig. 6). Of note, we acknowledge the fact that because the genome of *Bacillus ndiopicus* is a draft sequence, its exact size might be slightly different from that of our sequence, but given the fold coverage (52×), we are confident that the missing fragments are probably small and do not significantly influence the genome size. Of the 3982 predicted genes, 3915 were protein-coding genes and 67 were RNAs. A total of 1697 genes (43.34%) were assigned a putative function. The properties of the genome are presented in Table 4. Using PHAST software,

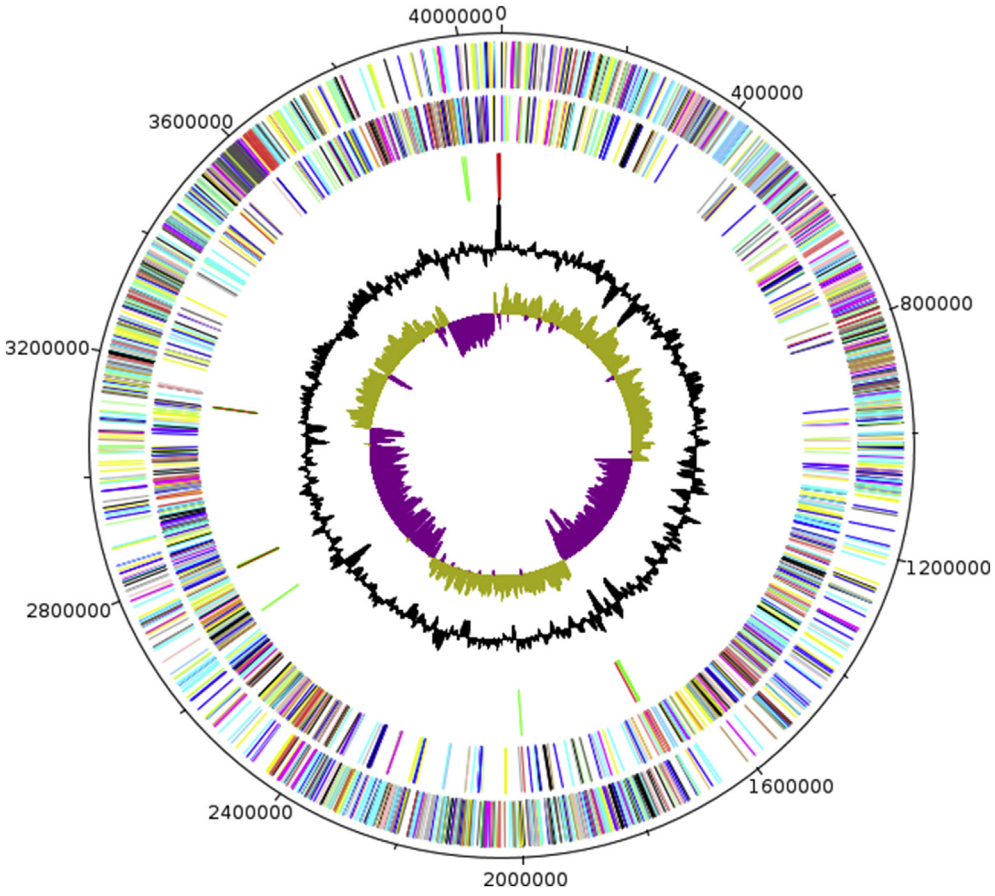

FIG. 6. Graphical circular map of *Bacillus ndiopicus* strain FF3^T chromosome. From outside in, outer two circles show ORFs oriented in forward (colored by COGs categories) and reverse (colored by COGs categories) directions, respectively. Third circle marks rRNA gene operon (red) and tRNA genes (green). Fourth circle shows G+C% content plot. Innermost circle shows GC skew; purple and olive indicate negative and positive values, respectively.

three prophage regions were identified, including one complete and two incomplete prophages (Table 5). A total of 167 were identified as ORFans (42.65%). The distribution of genes into COGs functional categories is presented in Table 6.

Genomic comparative

Today there are more than 277 sequenced genomes of *Bacillus* species (finished and draft) available in Genomes Online Database [3]. Here we compared *B. ndiopicus* genome sequence against other members of genus *Bacillus*, including *Bacillus coagulans* strain 2-6, *B. coagulans* strain 36D1, *Lysinibacillus sphaericus* strain C3-41, *Bacillus bataviensis* stain LMG 21833, and *Bacillus isronensis* strain B3W22. Table 7 shows a comparison of genome size, G+C% content, and number of proteins for each genome selected for taxonogenomic study. Indeed, *Bacillus ndiopicus* has a genome size of 4.06 Mb higher than those of *B. coagulans* 2–6 (3.07 Mb), *B. coagulans* 36D1 (3.55 Mb) and *B. isronensis* B3W22 (4.02 Mb) but lower than those of *B. bataviensis* LMG 21833 (5.37 Mb) and *Lysinibacillus sphaericus* C3-41 (4.82 Mb).

Bacillus ndiopicus strain FF3^T has a G+C content (37.03%) lower than those of all the compared species such as *B. coagulans* strain 2-6 (47.3%), *B. coagulans* strain 36D1 (46.5%), *B. bataviensis* strain LMG 21833 (39.6%), *B. isronensis* strain

TABLE 4. **Genome information**

Attribute	Value	% of total[a]
Genome size (bp)	4 068 720	
DNA coding (bp)	3 460 992	85.0
DNA G+C (bp)	1 506 586	37.03
DNA scaffolds	8	
Total genes	3982	100
Protein coding genes	3915	98.31
RNA genes	67	
Pseudo genes	51	1.18
Genes in internal clusters	208	4.82
Genes with function prediction	1697	43.34
Genes assigned to COGs	1892	48.32
Genes with Pfam domains	3235	75.45
Genes with peptide signals	60	1.53
Genes with transmembrane helices	530	13.5
CRISPR	4	

COGs, Clusters of Orthologous Groups database; CRISPR, clustered regularly interspaced short palindromic repeat.
[a]Total is based on total number of protein-coding genes in annotated genome.

TABLE 5. Identified prophage regions of *Bacillus ndiopicus*[a]

Region	Region length (kb)	Completeness	No. of coding sequence	Region position	Phage	GC%
1	15.6	Incomplete	16	269 940–285 579	PHAGE_Geobac_virus_E2_NC_009552	36.36
2	62.1	Complete	82	1 127 027–1 189 204	PHAGE_Thermu_OH2_NC_021784	37.40
3	18.7	Incomplete	25	1 843 157–1 861 873	PHAGE_Clostr_phiC2_NC_009231	36.67

[a]Region indicates number assigned to region; region length, length of sequence of that region (in bp); completeness, prediction of whether region contains a complete or incomplete prophage; region position, start and end positions of region on bacterial chromosome; phage, phage with highest number of proteins most similar to those in region; and GC%, percentage of GC nucleotides of region.

B3W22 (38.8%) and *L. sphaericus* strain C3-41 (37.1%). As it has been suggested in the literature that the G+C content deviation is at most 1% within species, these data are an additional argument for the creation of a new taxon [26].

The number of orthologous genes shared between *B. ndiopicus* and other *Bacillus* species as well as the average percentage nucleotide identity calculated using the MAGi method is tabulated in Table 8. On the basis of the analysis of MAGi, the AGIOS ranged from 61.79 to 95.94% among the studied members. The range of AGIOS calculated using MAGi varies from 61.79 to 70.95% between *B. ndiopicus* and other compared *Bacillus* species. Antibiotic resistance genes were detected within the genome using the ARDB website (Table 9).

Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses (taxonogenomics), we formally propose the creation of *Bacillus ndiopicus* sp. nov. that contains strain FF3^T as the type strain. The strain was isolated from the skin of a 39-year-old healthy Senegalese man living in Ndiop, Senegal.

TABLE 6. Number of genes associated with general COGs functional categories

Code	Value	%	Description
J	166	4.24	Translation, ribosome structure and biogenesis
A	0	0.00	RNA processing and modification
K	231	5.90	Transcription
L	127	3.24	Replication, recombination and repair
B	0	0.00	Chromatin structure and dynamics
D	33	0.84	Cell cycle control, cell division, chromosome partitioning
V	76	1.94	Defense mechanisms
T	126	3.21	Signal transduction mechanisms
M	112	2.86	Cell wall/membrane biogenesis
N	23	0.58	Cell motility
U	21	0.53	Intracellular trafficking and secretion
O	65	1.66	Posttranslational modification, protein turnover, chaperones
C	105	2.68	Energy production and conversion
G	98	2.50	Carbohydrate transport and metabolism
E	231	5.90	Amino acid transport and metabolism
F	75	1.91	Nucleotide transport and metabolism
H	89	2.27	Coenzyme transport and metabolism
I	70	1.78	Lipid transport and metabolism
P	155	3.95	Inorganic ion transport and metabolism
Q	24	0.61	Secondary metabolites biosynthesis, transport and metabolism
R	348	8.88	General function prediction only
S	303	7.73	Function unknown
—	195	4.98	Not in COGs

COGs, Clusters of Orthologous Groups database.

Description of *Bacillus ndiopicus* strain FF3^T sp. nov.

B. ndiopicus (n.dio.pi.cus. L. gen. masc. n. *ndiopicus*, of Ndiop, the name of the Senegalese village where the man from whom strain FF3^T was cultivated lives).

Cells stain Gram positive, are rod shaped and endospore forming, motile and have a mean diameter of 1.2 µm and a mean length of 2.5 µm. Peritrichous flagellae were observed. Colonies are 1 mm in diameter and transparent on 5% sheep's blood. Optimal growth is achieved at 37°C in an aerobic atmosphere supplemented with 5% CO_2. Catalase and oxidase activities are positive. Positive reactions were obtained with citrate, alkaline phosphatase, esterase, lipase and α-chymotrypsin. Negative reactions were observed for leucine arylamidase, valine arylamidase, cystine arylamidase, phosphatase acid, trypsin, naphthol-AS-BI-phosphohydrolase, β-glucuronidase, α-glucosidase, β-glucosidase, N-acetyl-β-glucosaminidase, α-mannosidase and α-fucosidase. *B. ndiopicus* is susceptible *in vitro* to penicillin, amoxicillin, amoxicillin–clavulanic acid, ceftriaxone, imipenem, gentamicin, ciprofloxacin, erythromycin, doxycycline, rifampicin and vancomycin, but resistant to nitrofurantoin and metronidazole.

TABLE 7. Genome comparison of *Bacillus ndiopicus* strain FF3^T with other *Bacillus* species

No.	Organism	Accession	Size (Mb)	No. of proteins	GC %
1	*Bacillus coagulans* 2-6	NC_015634	3.07	2971	47.3
2	*Bacillus coagulans* 36D1	NC_016023	3.55	3289	46.5
3	*Lysinibacillus sphaericus* C3-41	CP000817	4,82	4584	37.1
4	*Bacillus bataviensis* LMG 21833	NZ_AJLS00000000	5.37	5207	39.6
5	*Bacillus isronensis* B3W22	NZ_AMCK01000000	4.02	3883	38.8
6	*Bacillus ndiopicus* strain FF3^T	CCAP000000000	4.06	3915	37.03

TABLE 8. Orthologous gene comparison and average nucleotide identity of Bacillus ndiopicus strain FF3^T with other compared genomes

	Bacillus ndiopicus	*Bacillus bataviensis*	*Bacillus coagulans* 2-6	*Bacillus coagulans* 36D1	*Bacillus isronensis*	*Lysinibacillus sphaericus*
Bacillus ndiopicus	3915	63.67	61.87	61.79	70.95	70.78
Bacillus bataviensis	1623	5207	64.71	64.49	63.31	63.61
Bacillus coagulans 2-6	1281	1617	2971	95.94	62.21	61.78
Bacillus coagulans 36D1	1359	1737	1824	3289	62.11	61.76
Bacillus isronensis	1934	1681	1332	1434	3883	69.18
Lysinibacillus sphaericus	1981	1669	1321	1413	1965	4584

TABLE 9. Antibiotic resistance genes in *Bacillus ndiopicus* strain FF3 genome

Gene	Size (aa)	Function	E-value	Antibiotic	GenBank ID
baca	275	Undecaprenyl pyrophosphate phosphatase	3e-66	Bacitracin	NC_009832
lmrb	465	ABC transporter system, macrolide-lincosamide-streptogramin B efflux pump	1e-128	Lincomycin	AB000617
vanA	266	D-Alanyl-D-alanine carboxypeptidase	1e-65	Vancomycin	AM410096
vatb	168	Virginiamycin A acetyltransferase	6e-13	Streptogramin A	U19459
str	282	Streptomycin resistance protein	2e-94	Streptomycin	P12055
bmr	390	Major facilitator superfamily transporter; multidrug resistance efflux pump	1e-128	Chloramphenicol/fluoroquinolone	D84432

The G+C content of the genome is 37.03%. The 16S rRNA and genome sequences are deposited in GenBank under accession numbers HG315675 and CCAP000000000, respectively. The type strain FF3^T (= CSUR P3025 = DSM 27837) was isolated from the skin of a healthy 39-year-old Senegalese man living in Ndiop, Senegal.

Acknowledgements

We thank C. Couderc for help in performing the MALDI-TOF analysis. We also thank F. Di Pinto for creating the electron microscope photos. This study was funded by the Méditerranée Infection Foundation.

Conflict of Interest

None declared.

References

[1] Cohn F. Untersuchungen über Bakterien. Beitrage zur Biologie der Pflanzen Heft 1872;1:127–224.
[2] Parte AC. LPSN—list of prokaryotic names with standing in nomenclature. Nucleic Acids Res 2014;42(Database issue):D613–6.
[3] Castagnola E, Fioredda F, Barretta MA, et al. *Bacillus sphaericus* bacteraemia in children with cancer: case reports and literature review. J Hosp Infect 2001;48:142–5.
[4] Keita MB, Diene SM, Robert C, Raoult D, Fournier PE, Bittar F. Non-contiguous finished genome sequence and description of *Bacillus massiliogorillae* sp. nov. Stand Genomic Sci 2013;9:93–105.
[5] Bottone EJ. Bacillus cereus, a volatile human pathogen. Clin Microbiol Rev 2010;23:382–98.
[6] Mandell GL, Bennett JE, Dolin R. Principles and practice of infectious diseases. Amsterdam: Elsevier; 2010.
[7] Zhang YZ, Chen WF, Li M, et al. *Bacillus endoradicis* sp. nov., an endophytic bacterium isolated from soybean root. Int J Syst Evol Microbiol 2012;62:359–63.
[8] Pagani I, Liolios K, Jansson J, et al. The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res 2012;40:D571–9.
[9] Ramasamy D, Mishra AK, Lagier JC, 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.
[10] Sentausa E, Fournier PE. Advantages and limitations of genomics in prokaryotic taxonomy. Clin Microbiol Infect 2013;19:790–5.
[11] Lagier JC, Armougom F, Million M, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 2012;18:1185–93.
[12] Trape JF, Tall A, Diagne N, et al. Malaria morbidity and pyrethroid resistance after the introduction of insecticide-treated bednets and artemisinin-based combination therapies: a longitudinal study. Lancet Infect Dis 2011;11:925–32.
[13] Meier-Kolthoff JP, Göker M, Spröer C, Klenk HP. When should a DDH experiment be mandatory in microbial taxonomy? Arch Microbiol 2013;195:413–8.
[14] Seng P, Drancourt M, Gouriet F, 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.
[15] Field D, Garrity G, Gray T, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol 2008;26:541–7.
[16] Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform 2010;11:119.

[17] Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2012;40:48–53.
[18] 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.
[19] 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.
[20] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340:783–95.
[21] Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001;305:567–80.
[22] Rutherford K, Parkhill J, Crook J, 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] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res 2011;39:347–52.
[25] Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinform 2011;12:124.
[26] Meier-Kolthoff JP, Klenk HP, Göker M. Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. Int J Syst Evol Microbiol 2014;64:352–6.
[27] Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea, Bacteria*, and *Eukarya*. Proc Natl Acad Sci U S A 1990;87:4576–9.
[28] Skerman VBD, Sneath PHA. Approved list of bacterial names. Int J Syst Bacteriol 1980;30:225–420.
[29] Garrity GM, Holt J. The road map to the manual. In: Garrity GM, Boone DR, Castenholz RW, editors. Bergey's manual of systematic bacteriology. 2nd ed., vol. 1. New York: Springer; 2001. p. 119–69.
[30] 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.
[31] 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., vol. 3. New York: Springer-Verlag; 2009. p. 19–20.
[32] Prevot AR. In: Hauduroy P, Ehringer G, Guillot G, et al., editors. Dictionnaire des bactéries pathogènes. Paris: Masson; 1953. p. 692. Type genus: Bacillus Cohn 1872.
[33] Fischer A. Untersuchungen über bakterien. Jahrbücher für Wissenschaftliche Botanik 1895;27:1–163.
[34] Gibson T, Gordon RE. Genus I. *Bacillus* Cohn 1872, 174; Nom. gen. cons. Nomencl. Comm. Intern. Soc. Microbiol 1937, 28; Opin. A. Jud. Comm 1955, 39. In: Buchanan RE, Gibbons NE, editors. Bergey's manual of determinative bacteriology. 8th ed. Baltimore: Williams & Wilkins; 1974. p. 529–50.
[35] Mathews WC, Caperna J, Toerner JG, Barber RE, Morgenstern H. Neutropenia is a risk factor for Gram-negative *Bacillus* bacteremia in human immunodeficiency virus–infected patients: results of a nested case–control study. Am J Epidemiol 1998;148:1175–83.
[36] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–9.
[37] Jee-Min L, Che OJ, Jung RL, Dong-Jin P, Chang-Jin K. *Bacillus* kribbensis sp. nov., isolated from a soil sample in Jeju, Korea. Int J Syst Evol Microbiol 2007;57:2912–6.
[38] Glazunova OO, Raoult D, Roux V. *Bacillus massiliensis* sp. nov., isolated from cerebrospinal fluid. Int J Syst Evol Microbiol 2006;56:1485–8.
[39] Jeroen H, Bram V, Niall AL, et al. *Bacillus novalis* sp. nov., *Bacillus vireti* sp. nov., *Bacillus soli* sp. nov., *Bacillus bataviensis* sp. nov. and *Bacillus drentensis* sp. nov., from the Drentse A grasslands. Int J Syst Evol Microbiol 2004;54:47–57.
[40] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32:1792–7.
[41] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 2013;30:2725–9.