Bacillus rubiinfantis sp. nov. strain mt2[T], a new bacterial species isolated from human gut

M. Tidjiani Alou[1], J. Rathored[1], S. Khelaifia[1], C. Michelle[1], S. Brah[2], B. A. Diallo[3], D. Raoult[1,4] and J.-C. Lagier[1]

1) *Faculté de médecine, Unité des Maladies Infectieuses et Tropicales Emergentes (URMITE), UM63, CNRS7278, IRD198, Inserm 1095, Aix-Marseille Université, Marseille, France,* 2) *Hopital National de Niamey,* 3) *Laboratoire de microbiologie, département de biologie, Université Abdou Moumouni de Niamey, Niamey, Niger* and 4) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Bacillus rubiinfantis sp. nov. strain mt2[T] is the type strain of *B. rubiinfantis* sp. nov., isolated from the fecal flora of a child with kwashiorkor in Niger. It is Gram-positive facultative anaerobic rod belonging to the *Bacillaceae* family. We describe the features of this organism alongside the complete genome sequence and annotation. The 4 311 083 bp long genome (one chromosome but no plasmid) contains 4028 protein-coding gene and 121 RNA genes including nine rRNA genes.

New Microbes and New Infections © 2015 The Authors. Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases.

Corresponding author: J.-C. Lagier, Faculté de Médecine, URMITE, UMR CNRS 7278, IRD 198, INSERM U1095, Aix-Marseille Université, 27 Bd Jean Moulin, 13385 Marseille cedex 5, France
E-mail: jclagier@yahoo.fr

Introduction

Bacterial classification is currently based on phenotypic and genomic characteristics and the building of phylogenetic relationships using a comparison of the 16S ribosomal RNA [1–3]. Culturomics studies are based on the multiplication of culture conditions with a rapid identification method, which has allowed us to recently extend the gut microbiota repertoire [4,5]. In parallel, a new concept of bacterial description, termed taxonogenomics, has recently been developed combining a proteomic description with the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) profile associated with a biochemical and genomic description of the new bacterial species [4]. To date, nine bacterial species have officially been recognized through taxonogenomics [5–15].

The genus *Bacillus* was established in 1872 by Cohn and is composed of strictly aerobic and facultatively anaerobic rod-shaped bacteria that form heat-resisting endospores [16–19]. It includes over 200 described species and subspecies belonging to the *Firmicutes* phylum (http://www.bacterio.net/bacillus.html). These bacteria are widely distributed in the environment (soil, water, air), plants, food and human clinical samples [20]. The most known pathogenic *Bacillus* species is *Bacillus anthracis,* but most *Bacillus* are considered to be harmless to humans [20].

Bacillus rubiinfantis strain mt2[T] (= Collection de souches de l'Unité des Rickettsies (CSUR) P1141 = Deutsche Sammlung von Mikroorganismen (DSM) 28615) is the type strain of *B. rubiinfantis* sp. nov. This bacterium is Gram positive, spore forming, facultative anaerobic and motile. It was isolated from the stool of a child living in Niamey, Niger, with severe acute malnutrition (kwashiorkor).

Material and Methods

Organism information: classification and features

A stool sample was collected from a 2-year-old girl living in Niamey, Niger, with severe acute malnutrition. Consent was

obtained from the child's parents by the National Hospital of Niamey, and the study was approved by the Institut Fédératif de Recherche 48, Faculty of Medicine, Marseille, France (agreement 09-022). The patient did not receive any antibiotics at the time of sample collection; the fecal sample was stored at −80°C.

Strain identification

All the colonies were identified by MALDI-TOF, as described below. In the case of no matching spectra in the database, we further characterized the colonies using 16S rRNA sequencing, as previously described. If the 16S rRNA sequence similarity value was lower than 98.7%, we considered a new species without performing DNA-DNA hybridization as suggested by Stackebrandt and Ebers [21].

Phenotypic characteristics

The main phenotypic characteristics (i.e. Gram staining, motility, sporulation, catalase and oxidase test) were performed as previously described [22]. The chemical characteristics of the newly isolated strain were investigated using API 20NE, API ZYM and API 50CH strips (bioMérieux, Marcy l'Étoile, France). Growth temperatures were tested at 25°C, 30°C, 37°C, 45°C and 55°C. The growth of the strain was tested under anaerobic and microaerophilic conditions using GENbag anaer and GEN-bag microaer systems, respectively (bioMérieux) and under aerobic conditions with or without 5% CO_2. Regarding electron microscopy, detection formvar coated grids were deposited on a 40 μL bacterial suspension drop and incubated during 30 minutes at 37°C. The grids were incubated for 10 seconds on ammonium molybdate 1%, dried on blotting paper and then observed with a Morgani 268D at an operating voltage of 60 kV.

MALDI-TOF protein analysis was carried out as previously described [23] using a Microflex spectrometer (Brüker Daltonics, Leipzig, Germany). Twelve individual colonies were deposited on a MTP 96 MALDI-TOF target plate (Brüker). The 12 spectra were imported into the MALDI BioTyper software (version 2.0; Brüker) and analysed by standard pattern matching (with default parameter settings) against the main spectra of 6.252 bacteria, including 199 spectra from 104 validly named *Bacillus* species used as reference data in the BioTyper database. A score enabled the presumptive identification and discrimination of the tested species from those in a database: a score of >2 with a validated species enabled the identification at the species level, and a score of <1.7 did not enable any identification [23]. No significant score was obtained for strain mt2, thus suggesting that our isolate was not a member of any known species. The obtained reference spectrum for *B. rubiinfantis* strain mt2[T] (Fig. 4) was incremented in our database and compared to other members of the *Bacillus* genus. The observed differences are shown in gel view in Fig. 5.
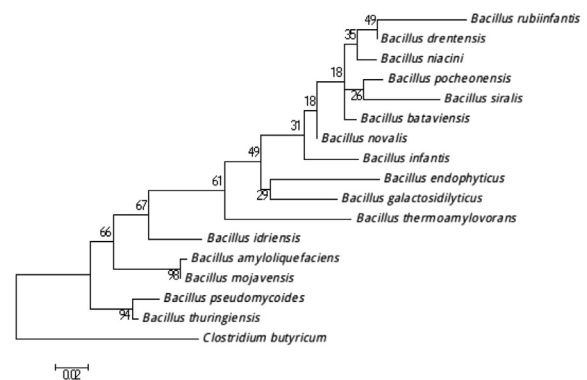

FIG. 1. Phylogenetic tree highlighting position of *Bacillus rubiinfantis* sp. nov. strain mt2[T] (= CSUR P1141 = DSM 28615) relative to other type strains within *Bacillus* genus. Strains and their corresponding GenBank accession numbers for 16S rRNA genes are (type = [T]): *B. novalis* strain SCTB 113, JN650278; *B. drentensis* DRG 4, HQ436340; *B. niacini* strain TSII-13, JN993716; *B. bataviensis* strain LMG21832, AJ542507; *B. pocheonensis* strain Gsoil 420, AB245377.1; *B. infantis* strain A-49, KC751070; *B. siralis* strain 171544, NR_028709.1; *B. endophyticus* strain 2DT, NR_025122.1; *B. pseudomycoides* strain BIHB 360, FJ859700.1; *B. thuringiensis* strain GMC 108, AB741470; *B. idriensis* strain SMC 4352−2, NR_043268.1; *B. amyloliquefaciens* strain BIHB 35, FJ859496 ; *B. mojavensis* strain BCRC 17124, EF433405.1 ; *B. galactosidilyticus* strain LMG 17892, NR_025580.1 ; *B. thermoamylovorans* strain DKP, NR_029151.1. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using maximum-likelihood method within MEGA6. Numbers at nodes are percentages of bootstrap values obtained by repeating analysis 1000 times to generate majority consensus tree. *Bacteroides thetaiotaomicron* strain ATCC 29148[T] (L16489) was used as outgroup. Scale bar = 1% nucleotide sequence divergence.

Growth conditions and genomic DNA preparation

B. rubiinfantis sp. nov., strain mt2[T] (= CSUR P1141 = DSM 28615), was grown on 5% sheep's blood−enriched Columbia agar (bioMérieux) at 37°C in aerobic atmosphere. Bacteria
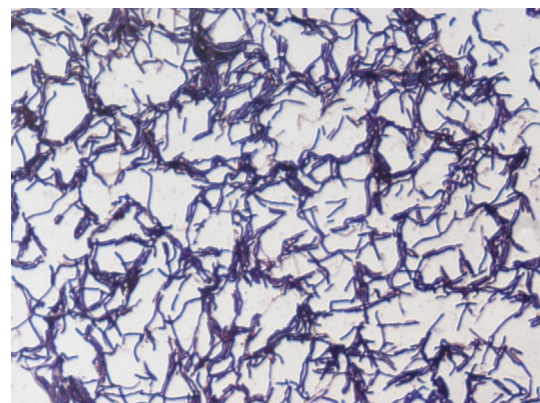
FIG. 2. Gram staining of *B. rubiinfantis* strain mt2[T].
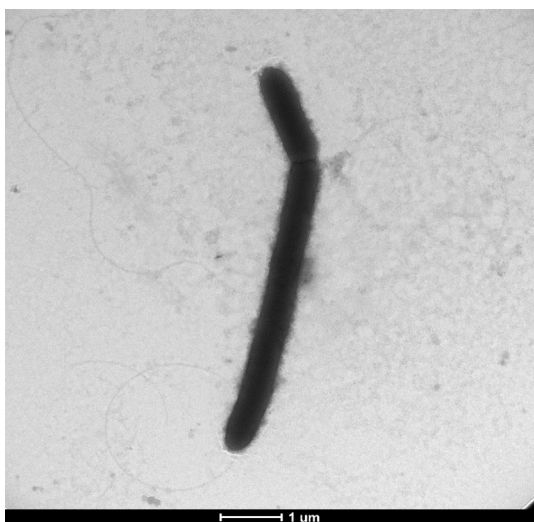
FIG. 3. Transmission electron microscopy of *B. rubiinfantis strain* mt2[T] using Morgani 268D (Philips) at operating voltage of 60 kV. Scale bar = 1 μm

grown on three petri dishes were collected and resuspended in 4 × 100 μL of Tris-EDTA (TE) buffer. Then 200 μL of this suspension was diluted in 1 ml TE buffer for lysis treatment that included a 30- minute incubation with 2.5 μg/μL lysozyme at 37°C, followed by an overnight incubation with 20 μg/μL proteinase K at 37°C. Extracted DNA was then purified using 3 successive phenol−chloroform extractions and ethanol precipitations at −20°C overnight. After centrifugation, the DNA was resuspended in 160 μL TE buffer.

Genome sequencing and assembly

Genomic DNA of *Bacillus rubiinfantis* was sequenced on the MiSeq Technology (Illumina, San Diego, CA, USA) with the mate-pair strategy. The gDNA was barcoded in order to be mixed with 11 other projects with the Nextera Mate Pair sample prep kit (Illumina). gDNA was quantified by a Qubit assay with the high sensitivity kit (Life Technologies, Carlsbad, CA, USA) to 66.2 ng/μL. The mate-pair library was prepared with 1 μg of genomic DNA using the Nextera mate-pair Illumina guide. The genomic DNA sample was simultaneously fragmented and tagged with a mate-pair junction adapter. The pattern of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) with a DNA 7500 labchip. The DNA fragments ranged in size from 1 kb up to 11 kb, with an optimal size of 3.927 kb. No size selection was performed, and 505 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal at 597 bp on a
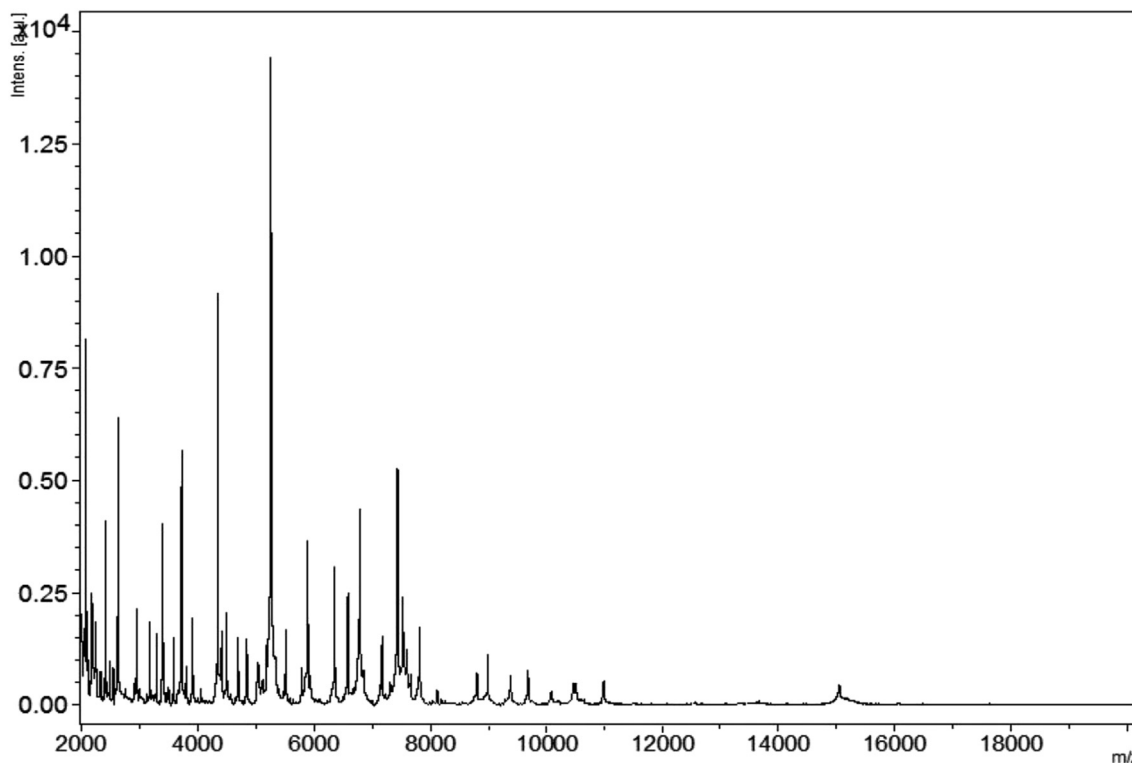
FIG. 4. Reference mass spectrum from *B. rubiinfantis strain* mt2[T]. Spectra from 12 individual colonies were compared and reference spectrum was generated.
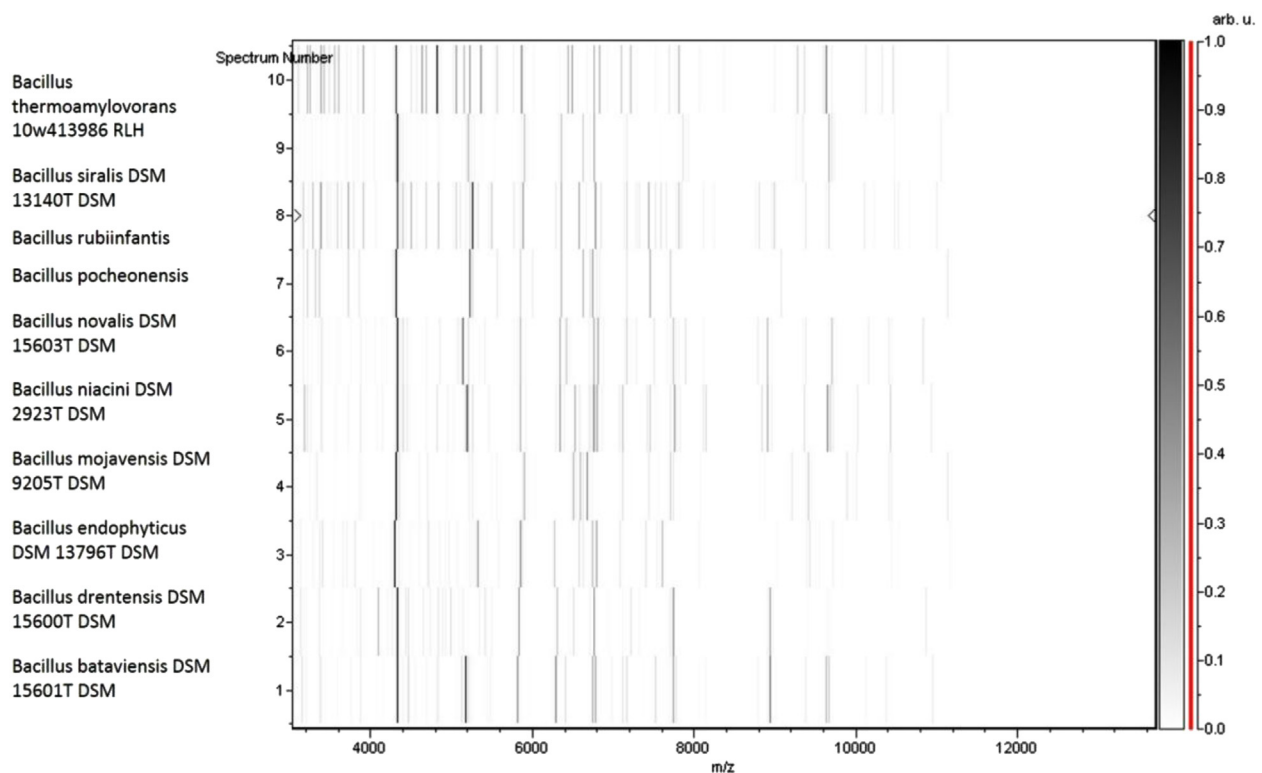
FIG. 5. Gel view comparing *B. rubiinfantis* (= CSUR P1141 = DSM 28615) to other species within genus *Bacillus*. Gel view displays raw spectra of loaded spectrum files arranged in pseudo-gel-like look. X-axis records m/z value. Left y-axis displays running spectrum number originating from subsequent spectra loading. Peak intensity is expressed by greyscale scheme code. Color bar and right y-axis indicate relation between color peak; peak intensity in arbitrary units. Displayed species are indicated on left.

Covaris S2 device in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent), and the final concentration library was measured at 59.2 nmol/L. The libraries were normalized at 2 nM and pooled. After a denaturation step and dilution at 15 pM, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and sequencing run were performed in a single 39-hour run at 2 × 251 bp. Total information of 10.7 Gb was obtained from a 1283 K/mm^2 cluster density, with a cluster passing quality control filters of 99% (24 072 208 clusters). Within this run, the index representation for *Bacillus rubiinfantis* was determined to 9.71%. The 2 053 904 paired reads were filtered according to the read qualities. These reads were trimmed, then assembled. A genome with 12 scaffolds and a size of 4.32 Mb was generated. The percentage GC was calculated at 40.0%.

Genome annotation and genome analysis

Open reading frames (ORFs) were predicted using Prodigal (http://prodigal.ornl.gov/) with default parameters, but the predicted ORFs were excluded if they spanned a sequencing gap region (contain N). The predicted bacterial protein sequences were searched against the Clusters of Orthologous Groups (COGs) using BLASTP (E-value 1e^{-03}, coverage 0.7, identity 30%). If no match was found, it searched against the NR database using BLASTP with an E-value of 1e^{-03}, coverage 0.7 and identity 30%, If the sequence lengths were smaller than 80 amino acids, we used an E-value of 1e^{-05}. The tRNAScanSE tool [24] was used to find tRNA genes, whereas ribosomal RNAs were found by RNAmmer [25]. Lipoprotein signal peptides and the number of transmembrane helices were predicted using Phobius [26]. ORFans were identified if all the performed BLASTP did not give positive results (E value smaller than 1e^{-03} for ORFs with sequence size upper than 80 aa or E value smaller than 1e^{-05} for ORFs with sequence length smaller 80 aa). Such parameter thresholds have already been used in previous works to define ORFans.

For the genomic comparison, we compared our type strain *Bacillus rubiinfantis* strain mt2^T to *Bacillus bataviensis* (GenBank accession number AJ542508.1), *Bacillus infantis* (AY904032.1), *Bacillus endophyticus* (AF295302.1), *Bacillus pseudomycoides*

(AF013121.1), *Bacillus thuringiensis* (D16281.1), *Bacillus amyloliquefaciens* (AB006920.1), *Bacillus mojavensis* (AB021191.1), *Bacillus thermoamylovorans* (L27478.1) and *Clostridium butyricum* (AJ58420.1). Genomes were automatically retrieved from the 16S RNA tree using Xegen software (Phylopattern) [27]. For each selected genome, the complete genome sequence, proteome genome sequence and Orfeome genome sequence were retrieved from the File Transfer Protocol (FTP) of National Center for Biotechnology Information. All proteomes were analysed with proteinOrtho [28]. For each couple of genomes, a similarity score was then computed. This score is the mean value of nucleotide similarity between all couples of orthologues between the two genomes studied (AGIOS) [6]. An annotation of the entire proteome was performed to define the distribution of the functional classes of predicted genes according to the clusters of orthologous groups of proteins (using the same method as for the genome annotation).

Results

Phenotypic description

Strain mt2[T] (Table 1) was first isolated in January 2014 by a 21-day anaerobic preincubation in blood culture with sheep's blood and cultivation 5% sheep's blood−enriched Colombia agar (BioMérieux) in anaerobic atmosphere at 37°C.

This strain showed a 96% nucleotide sequence similarity with *B. bataviensis,* the phylogenetically closest *Bacillus* species with a valid published name (Fig. 1). This similarity value is below the 16S rRNA threshold of 98.7% set by Stackebrandt and Ebers to delineate a new species without carrying out DNA-DNA hybridization [21].

Various growth temperatures (25°C, 30°C, 37°C, 45°C and 56°C) were tested. Growth occurred between 25°C and 45°C on blood-enriched Columbia agar (bioMérieux), with the

optimal growth being obtained at 37°C after 48 hours of incubation. Growth of the strain was tested under anaerobic and microaerophilic conditions using GENbag Anaer and GENbag microaer systems, respectively (bioMérieux), and under aerobic conditions, with or without 5% CO_2. Optimal growth was achieved aerobically. Weak cell growth was observed under microaerophilic and anaerobic conditions. The motility test was positive, and the cells were sporulating. Colonies were translucent and 10 mm in diameter on blood-enriched Columbia agar. Cells were Gram-positive rods (Fig. 2). Under electron microscopy, the bacteria grown on agar had a mean diameter and length of 0.54 and 0.62 μm, respectively (Fig. 3).

Strain mt2[T] exhibited catalase activity but was negative for oxidase. Using an API ZYM strip (BioMérieux), positive reactions were observed for esterase (C4), esterase lipase (C8), leucine arylamidase, valine arylamidase, cystine arylamidase, α-chymotrypsin, acid phosphatase, naphtol-AS-BI-phosphohydrolase, α-galactosidase and β-galactosidase. Negative reactions were observed for lipase (C14), trypsin, alkaline phosphatase, β-glucuronidase, α-glucosidase, β-glucosidase, α-fucosidase, α-mannosidase and N-acetyl-β-glucosaminidase. Using an API 20 NE strip, positive reactions were obtained for L-arginine, urea, 4-nitrophenyl- βD-galactopyranoside and capric acid. All other reactions were negative.

Using an API 50 CH strip (BioMérieux), positive reactions were observed for D-galactose, D-glucose, D-fructose, D-mannose, N-acetylglucosamine, esculin ferric citrate, D-maltose, D-lactose, D-melibiose and D-trehalose. Negative reactions were observed glycerol, erythrol, D-arabinose, L-arabinose, D-ribose, L-xylose, D-xylose, D-adonitol, methyl-βD-xylopranoside, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl-αD-mannopyranoside, methyl-αD-glucopyranoside, amygdaline, arbutin, salicin, D-cellobiose, D-maltose, D-saccharose, inulin, D-melezitose, D-raffinose, amidon, glycogen, gentiobiose, D-turanose, D-lyxose, D-tagatose, D-fucose, L-fucose, D-arabitol, L-arabitol, potassium gluconate, potassium 2-ketogluconate and potassium 5-ketogluconate.

Cells are susceptible to vancomycin, rifampicin, doxycycline, erythromycin, amoxicillin, nitroflurantoin, ciprofloxacin, ceftriaxone, penicillin G, gentamicin and imipenem but resistant to trimethoprim/ sulfamethoxazole and metronidazole. The differences exhibited by the comparison with other representatives of the genus *Bacillus* are detailed in Table 2.

Genome properties

The genome of *B. rubiinfantis* strain mt2[T] is 4 311 083 bp long with a 40.0% G+C content (Fig. 6). Of the 4149 predicted genes, 4028 were protein-coding genes and 121 were RNAs. Out of these 121 rRNA genes, 11 are 16S rRNA genes, 9 are 23S rRNA, 11 are 5S rRNA and 90 predicted tRNA genes were identified in

TABLE 1. Classification and general features of *Bacillus rubiinfantis* strain mt2

Property	Term
Current classification	Domain: *Bacteria*
	Phylum: *Firmicutes*
	Class: *Bacilli*
	Order: *Bacillales*
	Family: *Bacillaceae*
	Genus: *Bacillus*
	Species: *Bacillus rubiinfantis*
	Type strain: mt2
Gram stain	Positive
Cell shape	Rod
Motility	Motile
Sporulation	sporulating
Temperature range	Mesophilic
Optimum temperature	37°C

TABLE 2. Differential characteristics of *Bacillus rubiinfantis* strain mt2^T, *Bacillus novalis* LMG 21837, *Bacillus bataviensis* LMG 21833, *Bacillus drentensis* LMG 21831, *Bacillus niacini* DSM 2923, *Bacillus endophyticus* UCM 5715, *Bacillus siralis* NCIMB 13601, *Bacillus thermoamylovorans* CNCM I 1378, *Bacillus pocheonensis* DSM 18135, *Bacillus mojavensis* NRRL B 14698 [31–39]

Property	B. rubiinfantis	B. nolis	B. bataviensis	B. drentensis	B. niacini	B. endophyticus	B. siralis	B. thermo-amylovorans	B. pocheonensis	B. mojavensis
Cell diameter (μm)	0.5–0.6	0.6–1.2	0.7–1.2	0.6–1.2	0.9–1.4	0.5–1.5	0.5–0.8	0.45–0.5	1.5–3	1–2
Oxygen requirement	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic
Gram stain	+	+	+	+	+	+	+	+	+	+
Salt requirement	NA	NA	NA	NA	NA	+	+	NA	+	NA
Motility	+	+	+	+	−	−	NA	+	−	+
Endospore formation	+	+	+	+	+	+	+	−	+	+
Indole	−	−	−	−	+	−	NA	−	−	−
Production of:										
Alkaline phosphatase	−	NA	NA	NA	NA	NA	NA	NA	NA	NA
Catalase	+	NA	NA	NA	+	+	+	+	+	+
Oxidase	−	NA	NA	NA	+	+	+	+	+	+
Nitrate reductase	−	+	+	+/−	+	−	+	−	+	+
Urease	+	−	−	−	−	−	NA	NA	−	−
β-Galactosidase	+	+	+	+	+	−	−	+	+	+
N-Acetyl-glucosamine	−	+	+	+	+	NA	NA	+	+	NA
Acid from:										
L-Arabinose	−	−	−	−	−	+	−	−	+	+
Ribose	−	+	+	+	−	+	−	+	−	−
Mannose	+	+	+	+	+	+	−	+	+	+
Mannitol	−	+	+	−	−	+	−	−	−	+
Sucrose	−+	−	+	+	+	+	−	+	−	+
D-Glucose	+	+	+	+	+	+	−	+	−	+
D-Fructose	+	+	+	+	+	NA	−	+	+	+
D-Maltose	−	+	+	+	+	+	−	+	+	+
D-Lactose	+	−	+	+	+	+	+	+	−	−
Habitat	Human gut	Soil	Soil	Soil	soil	Cotton plants	Grassland soil	Palm vine	Soil	soil

NA: non-available data
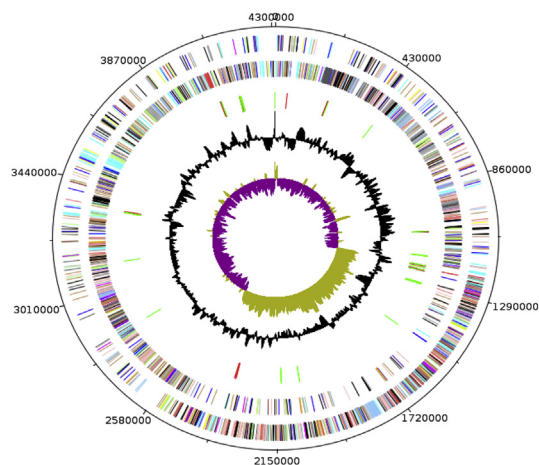
FIG. 6. Graphical circular map of chromosome. From outside to center: genes on forward strand colored by COGs categories (only gene assigned to COGs), RNA genes (Trnas green, rRNAs red), GC content and GC skew. COGs, Clusters of Orthologous Groups.

the genome. A total of 3026 genes (72.93%) were assigned a putative function. Two hundred two genes were identified as ORFans (5.04%). The remaining genes were annotated as hypothetical proteins. The properties and statistics of the genome are summarized in Tables 3 and 4. The distribution of genes into COGs functional categories is presented in Table 4.

It was the 70th organism identified within the *Bacillus* genera. The genome's GenBank accession number is CCFE01000000 and consists of 12 scaffolds and 33 contigs.

Genome comparison

The draft genome sequence of *Bacillus rubiinfantis* is larger than those of *Bacillus amyloliquefaciens*, *Bacillus mojavensis* and *Bacillus thermoamylovorans* (4.01, 3.96 and 3.71) but smaller than those of *Bacillus pseudomycoides*, *Bacillus bataviensis*, *Bacillus*

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

Attribute	Genome (total)	
	Value	% of total[a]
Size (bp)	4 311 083	100
G+C content (bp)	1 724 433	40.0
Coding region (bp)	3 672 471	84.88
Total genes	4149	100
RNA genes	121	2.80
Protein-coding genes	4028	97.08
Genes with function prediction	3026	72.93
Genes assigned to COGs	2666	64.25
Genes with peptide signals	244	5.88
Genes with transmembrane helices	1045	25.18
Genes with Pfam domains	1405	34.02
CRISPR repeats	01	0.02

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

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

Code	Value	% of total[a]	Description
J	163	3.97	Translation
A	0	0	RNA processing and modification
K	265	5.34	Transcription
L	235	4.62	Replication, recombination, repair
B	1	0.02	Chromatin structure and dynamics
D	40	0.92	Cell cycle control, mitosis, meiosis
Y	0	0	Nuclear structure
V	65	1.46	Defense mechanisms
T	197	3.57	Signal transduction mechanisms
M	184	3.60	Cell wall/membrane biogenesis
N	60	1.24	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	46	0.97	Intracellular trafficking and secretion
O	122	2.73	Posttranslational modification, protein turnover, chaperones
C	211	4079	Energy production and conversion
G	183	3.85	Carbohydrate transport and metabolism
E	397	9.14	Amino acid transport and metabolism
F	91	2.09	Nucleotide transport and metabolism
H	107	2.46	Coenzyme transport and metabolism
I	133	3.15	Lipid transport and metabolism
P	218	4.87	Inorganic ion transport and metabolism
Q	81	1.79	Secondary metabolites biosynthesis, transport, catabolism
R	428	9.41	General function prediction only
S	291	6.41	Function unknown
—	3518	33.47	Not in COGs

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

thuringiensis, *Bacillus infantis*, *Bacillus endophyticus*, *Bacillus infantis* and *Clostridium butyricum* (5.75, 5.37, 5.24, 5.1, 4.88 and 4.62 MB, respectively). The G+C content of *Bacillus rubiinfantis* is smaller than those of *Bacillus amyloliquefaciens*, *Bacillus infantis* and *Bacillus mojavensis* (46.1, 46.0 and 43.7) but larger than those of *Bacillus bataviensis*, *Bacillus thermoamylovorans*, *Bacillus endophyticus*, *Bacillus pseudomycoides*, *Bacillus thuringiensis* and *Clostridium butyricum* (39.6%, 37.3%, 36.4%, 35.4%, 35.4% and 28.8%, respectively). The gene content of *Bacillus rubiinfantis* is larger than those of *Bacillus mojavensis*, *Bacillus amyloliquefaciens* and *Bacillus thermoamylovorans* (4131, 4005 and 3458) but lower than that of *Bacillus pseudomycoides*, *Bacillus thuringiensis*, *Bacillus bataviensis*, *Bacillus endophyticus*, *Bacillus infantis* and *Clostridium butyricum* (5941, 5263, 5238, 4996, 4837 and 4231, respectively). To evaluate the genomic similarity among studied *Bacillus* strains, we determined two parameters, dDDH, Which exhibits a high correlation with DDH [29,30], and AGIOS [6], which was designed to be independent from DDH (Tables 5 and 6).

however, the distribution of genes into COGs categories was similar in all compared genomes (Fig. 7(. In addition, *B. rubiinfantis* shared 4022, 5265, 5077, 4078, 2152, 5863, 3468, 5271 and 4148 orthologous genes with *B. amyloliquefaciens*, *B. bataviensis*, *B. endophyticus*, *B. mojavensis*, *B. infantis*, *B. pseudomycoides*, *B. thuringiensis*, *B. thermoamylovorans* and *Clostridium butyricum* (Table 6). Among species with standing in nomenclature, AGIOS values ranged from 57.74 between *Clostridium*

TABLE 5. Pairwise comparison of *Bacillus rubiinfantis* (upper right) with eight other species using GGDC, formula 2 (DDH estimates based on identities/HSP length)[a]

	B. rubiinfantis	B. amyloliquefaciens	B. bataviensis	B. endophyticus	B. mojavensis	B. infantis	B. pseudomycoides	B. thermoamylovorans	B. thuringiensis	C. butyricum
B. rubiinfantis	100% ± 00	2.54% ± 0.14	2.66% ± 0.23	2.54% ± 0.16	2.55% ± 0.14	2.57% ± 0.20	2.54% ± 0.16	2.54% ± 0.23	2.55% ± 0.14	2.53% ± 0.16
B. amyloliquefaciens		100% ± 00	2.54% ± 0.22	2.54% ± 0.20	3.01% ± 0.22	2.54% ± 0.20	2.53% ± 0.19	2.53% ± 0.24	2.54% ± 0.13	2.53% ± 0.18
B. bataveinsis			100% ± 00	2.54% ± 0.21	2.54% ± 0.19	2.61% ± 0.20	2.54% ± 0.21	2.53% ± 0.24	2.55% ± 0.18	2.53% ± 0.17
B. endophyticus				100% ± 00	2.55% ± 0.18	2.54% ± 0.23	2.55% ± 0.22	2.53% ± 0.22	2.56% ± 0.20	2.53% ± 0.19
B. mojavensis					100% ± 00	2.54% ± 0.16	2.54% ± 0.16	2.53% ± 0.23	2.55% ± 0.12	2.53% ± 0.17
B. infantis						100% ± 00	2.53% ± 0.19	2.53% ± 0.22	2.55% ± 0.18	2.52% ± 0.11
B. pseudomycoides							100% ± 00	2.53% ± 0.19	3.00% ± 0.16	2.53% ± 0.18
B. thermoamylovorans								100% ± 00	2.53% ± 0.17	2.52% ± 0.21
B. thuringiensis									100% ± 00	2.54% ± 0.17
C. butyricum										100% ± 00

[a]Confidence intervals indicate inherent uncertainty in estimating DDH values from intergenomic distances based on models derived from empirical test data sets (which are always limited in size) These results are in accordance with 16S rRNA (Fig. 1) and phylogenomic analyses as well as GGDC results.
GGDC, Genome-to-Genome Distance Calculator; DDH, DNA-DNA Hybridization; HSP, High-Scoring Pair.

TABLE 6. Numbers of orthologous protein shared between genomes (upper right) average percentage similarity of nucleotides corresponding to orthologous protein shared between genomes (lower left)

	B. amyloliquefaciens	B. bataviensis	B. endophyticus	B. mojavensis	B. infantis	B. pseudomycoides	B. rubiinfantis	B. thermoamylovorans	B. thuringiensis	C. butyricum
B. amyloliquefaciens	4022[a]	1674	1864	2028	686	1642	1486	1442	1707	798
B. bataviensis	65.44	5265[a]	1765	1653	878	1785	1850	1599	1807	854
B. endophyticus	65.43	66.90	5077[a]	1849	749	1769	1547	1481	1839	833
B. mojavensis	77.91	65.86	66.25	4078[a]	693	1637	1466	1413	1694	776
B. infantis	65.39	74.32	67.37	65.97	2152[a]	744	740	633	757	391
B. pseudomycoides	64.79	67.00	68.06	65.55	67.10	5863[a]	1556	1416	2212	868
B. rubiinfantis	65.31	73.95	66.82	65.78	72.33	67.19	4028[a]	1532	1571	778
B. thermoamylovorans	64.64	67.30	66.27	65.10	67.67	66.64	67.41	3468[a]	1445	785
B. thuringiensis	64.42	66.65	67.70	65.23	66.92	83.41	66.80	66.33	5271[a]	879
C. butyricum	57.74	60.66	61.81	59.14	61.44	62.07	60.72	61.49	62.11	4148[a]

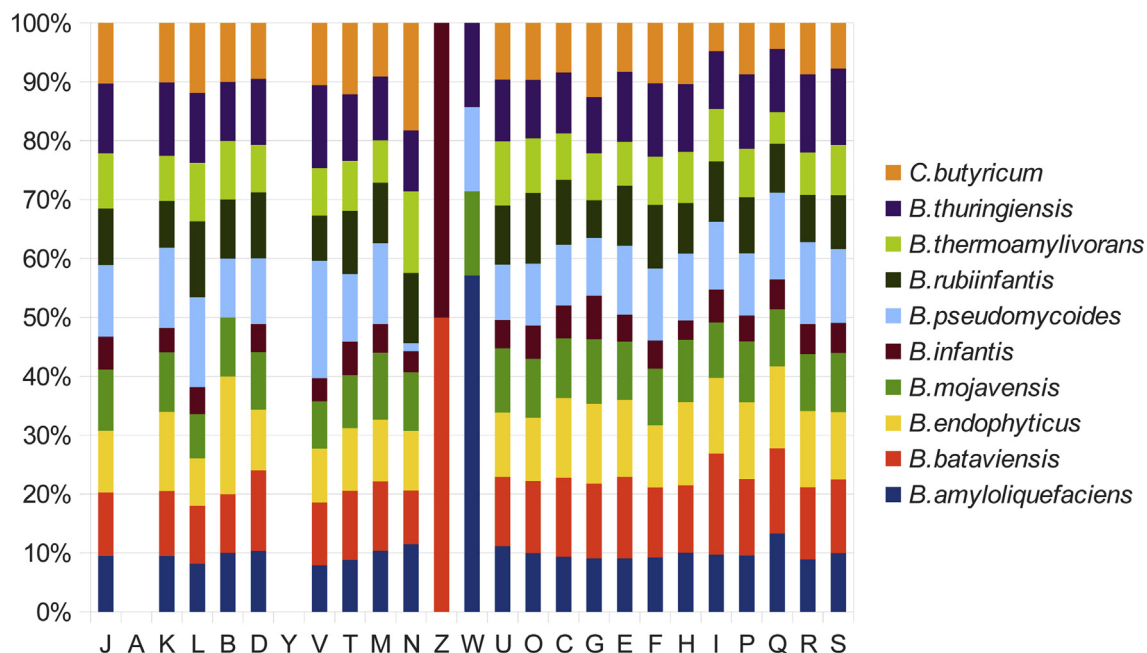[a]Number of proteins per genome.

FIG. 7. Distribution of functional classes of predicted genes according to COGs. COGs, Clusters of Orthologous Groups.

butyricum and *B. amyloliquefaciens* to 83.41% between *B. thuringiensis* and *B. pseudomycoides*.

Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Bacillus rubiinfantis* sp. nov. that contains the strain mt2. This bacterial strain has been isolated from the fecal flora of a 2-year-old girl from Niamey, Niger, with kwashiorkor.

Taxonomic and nomenclatural proposals: description of *Bacillus rubiinfantis* strain mt2^T sp. nov.
The cells are Gram-positive, sporulating and rod-shaped bacilli with a diameter of 0.5 μm. Colonies are translucent and 10 mm in diameter on 5% sheep's blood–enriched Columbia agar (bioMérieux) and are catalase positive and oxidase negative. Positive reactions are observed for esterase (C4), esterase lipase (C8), leucine arylamidase, valine arylamidase, cystine arylamidase, α-chymotrypsin, acid phosphatase, naphtol-AS-BI-phosphohydrolase, α-galactosidase and β-galactosidase. Negative reactions are observed for lipase (C14), trypsin, nitrate reduction, alkaline phosphatase, β-glucuronidase, α-glucosidase, β-glucosidase, α-fucosidase, α-mannosidase and N-acetyl-β-glucosaminidase.

Cells are susceptible to vancomycin, rifampicin, doxycycline, erythromycin, amoxicillin, nitroflurantoin, ciprofloxacin,

ceftriaxone, penicillin G, gentamicin and imipenem but resistant to trimethoprim/sulfamethoxazole and metronidazole.

The G+C content of the genome is 40.0%. The 16S rRNA gene sequence and whole-genome shotgun sequence of *B. rubiinfantis* strain mt2^T are deposited in GenBank under accession numbers LK021113 and CCFE01000000, respectively. The type strain mt2^T (= CSUR P1141 = DSM 28615) was isolated from the stool of a child living in Niamey, Niger, with kwashiorkor.

Acknowledgements

Conflict of Interest

None declared.

References

[1] Viale AM, Arakaki AK, Soncini FC, Ferreyra RG. Evolutionary relationships among eubacterial groups as inferred from GroEL

(Chaperonin) sequences comparison. Int J Syst Bacteriol 1944;44(3): 527–33.

[2] Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposals for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 1990;87(12):4576–9.

[3] Wolf M, Muller T, Dandekar T, Pollack JD. Phylogeny of Firmicutes with special reference to Mycoplasma (Mollicutes) as inferred from phosphoglycerate kinase amino acid sequence data. Inst J Syst Evol Microbiol 2004;54(Pt3):871–5.

[4] Lagier JC, Hugon P, Khelaifia S, Fournier PE, La Scola B, Raoult D. The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. Clin Microbiol Rev 2015 Jan;28(1): 237–64.

[5] Lagier J-C, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 2012;18:1185–93.

[6] Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi-Tamisier 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.

[7] Lagier JC, El Karkouri K, Nguyen TT, Armougom F, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Anaerococcus senegalensis* sp. nov. Stand Genomic Sci 2012;6: 116–25.

[8] Lagier JC, Armougom F, Mishra AK, Nguyen TT, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Alistipes timonensis* sp. nov. Stand Genomic Sci 2012;6:315–24.

[9] Roux V, El Karkouri K, Lagier JC, Robert C, Raoult D. Non-contiguous finished genome sequence and description of *Kurthia massiliensis* sp. nov. Stand Genomic Sci 2012;7:221–32.

[10] Kokcha S, Ramasamy D, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Brevibacterium senegalense* sp. nov. Stand Genomic Sci 2012;7:233–45.

[11] Ramasamy D, Kokcha S, Lagier JC, Nguyen T-T, Raoult D, Fournier PE. Genome sequence and description of *Aeromicrobium massiliense* sp. nov. Stand Genomic Sci 2012;7:246–57.

[12] Lagier JC, Elkarkouri K, Rivet R, Couderc C, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Senegalemassilia anaerobia* gen. nov., sp. nov. Stand Genomic Sci 2013;7: 343–56.

[13] 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.

[14] Hugon P, Mishra AK, Lagier JC, Nguyen TT, Couderc C, Raoult D, et al. Non-contiguous finished genome sequence and description of *Brevibacillus massiliensis* sp. nov. Stand Genomic Sci 2013;8:1–14.

[15] Lagier JC, Gimenez G, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of Herbaspirillum massiliense sp. nov. Stand Genomic Sci 2012 Dec 19;7(2):200–9.

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

[17] Fritze D. Taxonomy of the genus Bacillus and related genera: the aerobic endospore-forming bacteria. Physiopathology 2004;94(11): 1245–8.

[18] Gordon RE, Haynes WC, Pang CH. The genus Bacillus. US Dep Agric Handh 1973;427.

[19] Goto K, Omura T, Hara Y, Sadaie Y. Application of the 16s rDNA sequence as an index for rapid identification of species in the genus Bacillus. J Gen Appl Microbiol 2000;46(1):1–8.

[20] http://www.bacterio.net/bacillus.html.

[21] Prevot AR. Dictionnaire des bactéries pathogens. In: Hauduroy P, Ehringer G, Guillot G, Magrou J, Prevot AR, Rosset, Urbain A, editors. Paris: Masson; 1953. p.1–692.

[22] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152–5.

[23] 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 Jan;28(1):208–36.

[24] Seng P, Abat C, Rolain JM, Colson P, Lagier JC, Gouriet F, et al. Identification of rare pathogenic bacteria in a clinical microbiology laboratory: impact of matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol 2013;51:2182–94.

[25] Prodigal (http://prodigal.ornl.gov/).

[26] Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucl Acids Res 2007;35:3100–8.

[27] 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.

[28] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340:783–95.

[29] Gouret P, Thompson JD, Pontarotti P. PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinformatics 2009;10:298. http://dx.doi.org/10.1186/1471-2105-10-298.

[30] Lechner M, Findeib S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics 2011;12:124.

[31] Auch AF, Von Jan M, Klenk HP, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci 2010;2:117–34.

[32] Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 2013;14:60.

[33] Heyram J, Vanparys B, Logan NA, Balcaen A, Rodriguez-Diaz M, Felske A, 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(1):47–57.

[34] Nagel M, Andreesen JR. *Bacillus niacini* sp. nov., a nicotinate- metabolizing mesophile isolated from soil. Int J Syst Evol Microbiol 1991;41(1): 134–9.

[35] Reva ON, Smirnov VV, Petterson B, Priest FG. *Bacillus endophyticus* sp. nov., isolated from the inner tissues of cotton plants (Gossypium sp.). Int J Syst Evol Microbiol 2002;52(1):101–7.

[36] Petterson B, De Silva SK, Uhlén M, Priest FG. *Bacillus siralis* sp. nov., a novel species from silage with a higher order structural attribute in the 16S rRNA genes. Int J Syst Evol Microbiol 2000;50(6):2181–7.

[37] Combet-Blanc Y, Olivier B, Streicher C, Patel BK, Dwivedi PP, Pot B, et al. *Bacillus thermoamylovorans sp. nov.*, a moderately thermophilic and amylolytic bacterium. Int J Syst Evol Microbiol 1995;45(1):9–16.

[38] Ten LN, Baek SH, Im WT, Larina LL, Lee JS, Oh HM, et al. *Bacillus pocheonensis* sp. nov., a moderately halotolerant, aerobic bacterium isolated from soil of a ginseng field. Int J Syst Evol Microbiol 2007;57(11):2532–7.

[39] Robert MS, Nakamura LK, Cohan FM. *Bacillus mojavensis* sp. nov., distinguishable from Bacillus subtilis by sexual isolation, divergence in DNA sequence, and differences in fatty acid composition. Int J Syst Evol Microbiol 1994;44(2):256–64.