Noncontiguous finished genome sequence and description of *Bacillus andreraoultii* strain SIT1^T sp. nov.

S. I. Traore[1,2], T. Cimmino[1], J.-C. Lagier[1], S. Khelaifia[1], S. Brah[3], C. Michelle[1], A. Caputo[1], B. A. Diallo[4], P.-E. Fournier[1], D. Raoult[1,5] and J. M. Rolain[1]

1) *Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM 63, CNRS 7278, IRD 198, Inserm U1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine, Aix-Marseille Université, Marseille, France,* 2) *Département d'Epidémiologie des Affections Parasitaires, Faculté de médecine et de pharmacie de Bamako, Mali,* 3) *Hôpital National de Niamey,* 4) *Laboratoire de Microbiologie, Département de Biologie, Université Abdou Moumouni de Niamey, Niamey, Niger and* 5) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Bacillus andreraoultii strain SIT1^T (= CSUR P1162 = DSM 29078) is the type strain of *B. andreraoultii* sp. nov. This bacterium was isolated from the stool of a 2-year-old Nigerian boy with a severe form of kwashiorkor. *Bacillus andreraoultii* is an aerobic, Gram-positive rod. We describe here the features of this bacterium, together with the complete genome sequencing and annotation. The 4 092 130 bp long genome contains 3718 protein-coding and 116 RNA genes.

Keywords: Culturomics, genome
Original Submission: 14 September 2015; **Revised Submission:** 15 December 2015; **Accepted:** 16 December 2015
Article published online: 23 December 2015

Corresponding author: J. M. Rolain, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM 63, CNRS 7278, IRD 198, Inserm U1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine, Aix-Marseille Université, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France
E-mail: jean-marc.rolain@univ-amu.fr

Introduction

Bacillus andreraoultii strain SIT1^T (= CSUR P1162 = DSM 29078) is the type strain of *B. andreraoultii* sp. nov. This bacterium was isolated from the stool of a 2-year-old Nigerian boy with kwashiorkor, a severe form of acute malnutrition, and is part of an effort called culturomics to cultivate all bacterial species from the human gut [1,2]. This is a Gram-positive, aerobic or facultatively anaerobic, motile, spore-forming, indole-negative and rod-shaped bacillus.

The ruling taxonomic classification of prokaryotes is based on a combination of phenotypic and genotypic criteria [3,4]. However, the three essential criteria that are used (16S rRNA gene-based phylogeny [5], G+C content and DNA-DNA hybridization (DDH)) [3,6] exhibit several drawbacks. The number of sequenced bacterial genomes has rapidly increased due to the decrease in cost of sequencing (to date, almost 40 000 bacterial genomes have been sequenced). Therefore, the genomic data, the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) spectrum [7] and phenotypic criteria have been proposed for the description of new bacterial species [8,9].

The genus *Bacillus* was discovered in 1872 (Cohn, 1872) [10]. The genus is composed of 331 species with validly published names (*Bacillus*, http://www.doi.namesforlife.com/). Species of the genus *Bacillus* are ubiquitous bacteria isolated from various environments including soil, gastrointestinal tracts of various insects and animals, vegetation, fresh- and seawater and food [11]. In human beings, the *Bacillus* species may exist in opportunistic form in immunocompromised patients [12] or in pathogenic form, such as *B. cereus* (food poisoning) and *B. anthracis* (anthrax) [13]. Other species may also be found in various human infections, including pneumonia, endocarditis, and ocular, cutaneous, bone or central nervous system infections and bacteraemia [14].

http://dx.doi.org/10.1016/j.nmni.2015.12.005

The following is a summary classification and a set of features for *Bacillus andreraoultii* sp. nov. strain SIT1^T together with the description of the complete genomic sequencing and annotation. These particularities support the circumscription of the species *Bacillus andreraoultii.*

Materials and methods

Sample collection, strain isolation and culture condition

A stool sample was obtained from a 2-year-old Nigerian boy with kwashiorkor, a severe form of acute malnutrition, who was admitted to the emergency room in the national hospital in Niamey, the capital city of Niger, in October 2013 [2]. The study was approved by the local ethics committee of the Institut Fédératif de Recherche IFR48, Faculty of Medicine, Marseille, France, under agreement 09-022. The fecal specimen was preserved at 4°C after collection. Then the stool was sent to Marseille, where it was kept at −80°C until laboratory culture isolation. Strain SIT1^T was isolated in March 2014 by cultivation on 5% sheep's blood–enriched Columbia agar (bioMérieux, Marcy l'Étoile, France) in an anaerobic atmosphere at 37°C after 10 days of stool specimen incubation in a culture bottle containing a blood-enriched Columbia agar liquid medium (bioMérieux). Growth of the strain was tested under anaerobic conditions using GENbag anaer system (bioMérieux), and under aerobic conditions, with or without 5% CO_2. Different growth temperatures (25, 30, 37, 45, 55°C) were also tested.

MALDI-TOF and 16S sequencing

MALDI-TOF protein analysis was carried out as previously described [15] using a Microflex spectrometer (Bruker Daltonics, Leipzig, Germany). In brief, a pipette tip was used to pick one isolated bacterial colony from a culture agar plate and spread it as a thin film on an MSP 96 MALDI-TOF target plate (Bruker). Twenty distinct deposits from 20 isolated colonies were tested for strain SIT1^T. Each smear was overlaid with 2 µL of matrix solution (saturated solution of alpha-cyano-4-hydroxycinnamic acid in 50% acetonitrile and 2.5% trifluoracetic acid) and allowed to dry for 5 minutes. 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 240 shots with variable laser power. The time of acquisition was between 30 seconds and 1 minute per spot. The 20 SIT1^T spectra were imported into MALDI BioTyper 3.0 software (Bruker) and analysed by standard pattern matching (with default parameter settings) against the main spectra of 7379 bacteria. The method of identification included the m/z from 3000 to 15 000 Da. For every spectrum, a maximum of 100 peaks were compared with spectra in the database. The resulting score enabled the identification of species whether tested or not: a score of $\geq$2 with a validly published species enabled identification at the species level, a score of $\geq$1.7 but <2 enabled identification at the genus level and a score of <1.7 did not enable any identification.

Identification of bacteria continued with a 16S rRNA standard PCR coupled with sequencing. That was performed using GeneAmp PCR System 2720 thermal cyclers (Applied Biosystems, Bedford, MA, USA) and ABI Prism 3130xl Genetic Analyser capillary sequencer (Applied Biosystems) respectively [16]. The 16S rRNA nucleotides sequence was corrected using Chromas Pro 1.34 software (Technelysium, Tewantin, Australia), and the BLASTn searches were performed in the online PubMed National Center for Biotechnology Information (NCBI) database (http://blast.ncbi.nlm.nih.gov.gate1.inist.fr/Blast.cgi).

Morphologic, biochemical and antibiotic susceptibility tests

Bacillus andreraoultii strain SIT1^T was observed, after negative colouration, using a Morgani 268D (Philips, Amsterdam, The Netherlands) transmission electron microscope at an operating voltage of 60 kV. The Gram colouration was performed using Color Gram 2 Kit (bioMérieux) and observed by using the DM1000 photonic microscope (Leica Microsystems, Wetzlar, Germany) with a 100× oil-immersion objective lens. The sporulation test was done doing a thermic shock (80°C during 30 minutes). To evaluate the motility of *Bacillus andreraoultii,* fresh colonies were observed between blades and slats using a DM1000 photonic microscope (Leica) with a 40× objective lens.

API ZYM, API 20 NE and API 50 CH (bioMérieux) gallery systems were used to perform biochemical assays. Oxidase (Becton Dickinson, Franklin Lakes, NJ, USA) and catalase assays (bioMérieux) were done separately. The antibiotic susceptibility was tested using SirScan Discs antibiotics (i2a, Montpellier, France).

Genome sequencing

Genomic DNA extraction of *Bacillus andreraoultii* strain SIT1^T was performed according to the method previously described [17]. The DNA was resuspended in 205 µL TE buffer. The DNA concentration was 401.63 ng/µL as measured by a Qubit fluorometer using the high-sensitivity kit (Life Technologies, Carlsbad, CA, USA).

Genomic DNA of *B. andreraoultii* was sequenced on the MiSeq Technology (Illumina, San Diego, CA, USA) with the

mate pair strategy. The gDNA was barcoded in order to be mixed with 11 other projects with the Nextera Mate Pair sample prep kit (Illumina).

The gDNA was quantified by a Qubit assay with the high-sensitivity kit (Life Technologies) to 87.67 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 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 11 kb, with an optimal size of 4.851 kb. No size selection was performed, and 432.1 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared into small fragments with an optimal size of 684 bp on the Covaris device S2 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 64.49 nmol/L.

The libraries were normalized at 2 nM and pooled. After denaturation 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 a single 39-hour sequencing run were performed at a 2 × 251 bp read length.

All 8.2 Gb of information was obtained from a 947K/mm^2 cluster density, with a cluster passing quality control filters with 99% (18 111 784 clusters). Within this run, the index representation for *Bacillus andreraoultii* was determined to 9.27%. The 1 513 908 paired reads were filtered according to the read qualities. The reads obtained from applications were trimmed, and an optimal assembly of 14 scaffolds and 58 contigs was obtained through the SOAPdenovo software, which generated a genome size of 4.09 Mb.

Genome annotation and comparison

Open reading frames (ORFs) were predicted using Prodigal [18] 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) database using BLASTP (e-value 1e-03, coverage of 0.7, and an identity percentage of 30%). If no hit was found, it search against the NR database using BLASTP with an e-value of 1e-03, coverage of 0.7 and an identity percentage of 30%. If sequence lengths were smaller than 80 amino acids, we used an e-value of 1e-05. The tRNAScanSE tool [19] was used to find tRNA genes, whereas ribosomal RNAs were found by using RNAmmer [20]. Lipoprotein signal peptides and the number of transmembrane helices were predicted using Phobius [21]. ORFans were identified if all the BLASTP performed did not retrieve positive results. Such parameter thresholds have already been used in previous works to define ORFans. Genomes were automatically retrieved from the 16s RNA tree using Phylopattern software [22]. For each selected genome, complete genome sequences, proteome genome sequences and Orfeome genome sequences were retrieved from the FTP site of NCBI. All proteomes were analysed with Proteinortho [23]. Then for each couple of genomes a similarity score was computed. This score is the mean value of nucleotide similarity between all couple of orthologues in the two genomes studied (average genomic identity of orthologous gene sequences, AGIOS) [9]. The resistome was analysed with the ARG-ANNOT (Antibiotic Resistance Gene-ANNOTation) database and BLASTp in GenBank [24]. The exhaustive bacteriocin database available in our laboratories (Bacteriocins from the URMITE database) (http://drissifatima.wix.com/bacteriocins) was performed by collecting all currently available sequences from the databases and from NCBI. Protein sequences from this database allowed putative bacteriocins from human gut microbiota to be identified using BLASTp methodology [25]. PHAST (PHAge search tool) was used to identify phage sequences [26].

An annotation of the entire proteome was performed to define the distribution of functional classes of predicted genes according to the clusters of orthologous groups of proteins using the same method as for the genome annotation.

Results

Strain identification and phylogenetic analyses

No significant MALDI-TOF score was obtained for strain SIT1^T against the Bruker database, suggesting that our isolate was not a member of a known species. We added the spectrum from strain SIT1^T to our database (Figure 1). Then a gel view was performed to show the spectral differences with other members of the genus *Bacillus* (Figure 2).

Using 16S rRNA phylogeny analyses, we demonstrated that strain SIT1^T exhibited a 96% 16S rRNA sequence identity with *Bacillus thermoamylovorans* (GenBank accession no. HM030742), the phylogenetically closest bacterial species with standing in nomenclature (Figure 3). Its 16S rRNA sequence was deposited in GenBank under accession no. LK021120. This value was lower than the 98.7% 16S rRNA gene sequence threshold recommended by Stackebrandt and Ebers [5] to delineate a new species without carrying out DNA-DNA hybridization. Thus, this bacterium was considered as a new species called *Bacillus andreraoultii* strain SIT1^T belonging family *Bacillaceae* (Table 1).

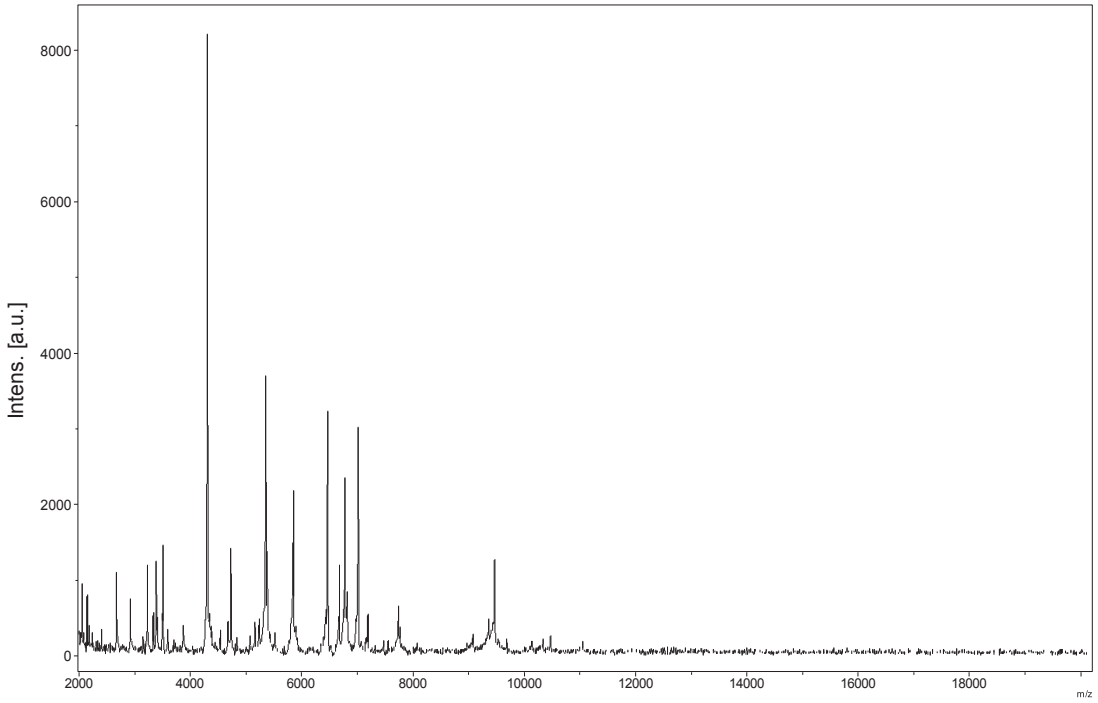

FIG. 1. Reference mass spectrum from *Bacillus andreraoultii* strain SIT1^T. Spectra from 20 individual colonies were compared and reference spectrum generated.

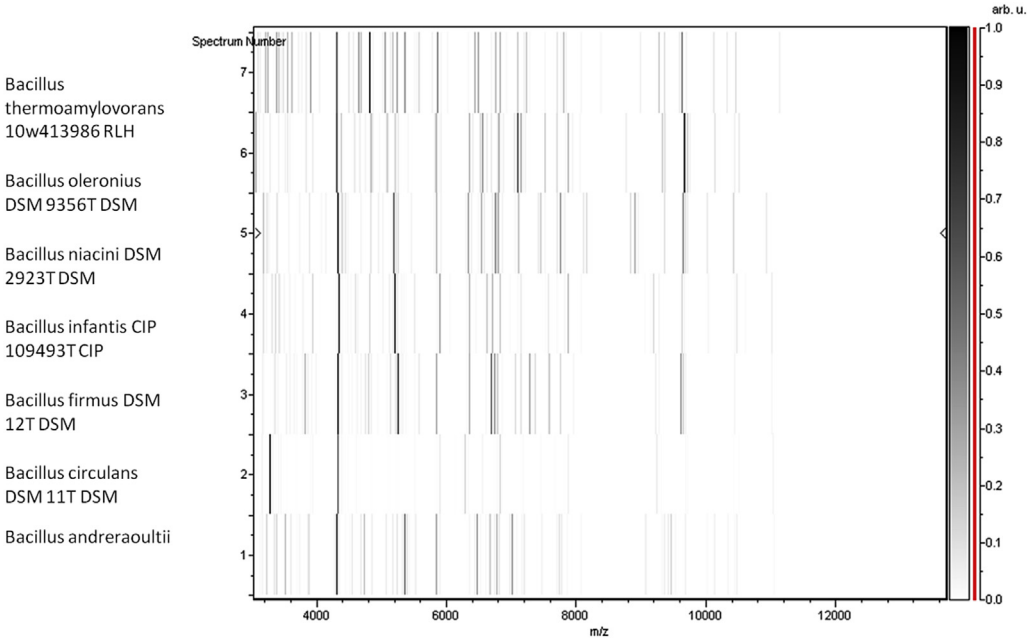

FIG. 2. Gel view comparing *Bacillus andreraoultii* SIT1^T to other *Bacillus* species. Gel view displays raw spectra of loaded spectrum files arranged in pseudo-gel-like look. *x*-axis records m/z value. Left *y*-axis displays running spectrum number originating from subsequent spectra loading. Peak intensity is expressed in arbitrary units by greyscale scheme code, as indicated on right y-axis. Displayed species are indicated at left.

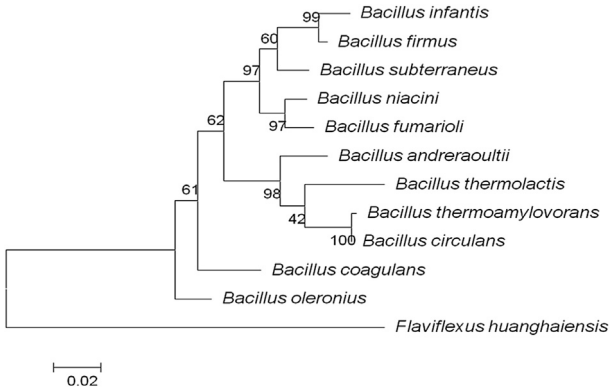

FIG. 3. Phylogenetic tree highlighting position of *Bacillus andreraoultii* sp. nov. strain SIT1^T (= CSUR P1162 = DSM 29078) relative to other type strains within *Bacillus* genus. Strains and their corresponding GenBank accession numbers for 16S rRNA genes are (type = T): *B. infantis* strain NRRL B-14911, NR121756; *B. firmus* strain 5695m-D2, AJ509007; *B. subterraneus* strain COOI3B, NR104749; *B. niacini* strain Et9/1, KJ722425; *B. fumarioli* strain R-14705, AJ581126; *B. thermolactis* strain R-33520, AM910339; *B. thermoamylovorans* strain N12-2, HM030742; *B. circulans* strain WSBC20059, Y13063; *B. coagulans* strain 36D1, DQ297926; *B. oleronius* strain ATCC700005, NR043325; *Flaviflexus huanghaiensis* strain H5, JN815236. 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. *Flaviflexus huanghaiensis* strain H5 (JN815236) was used as outgroup. Scale bar = 2% nucleotide sequence divergence.

TABLE 1. Classification and general features of *Bacillus andreraoultii* strain SIT1^T according to MIGS recommendations [27].

MIGS ID	Property	Term	Evidence code[a]
	Current classification	Domain: *Bacteria*	TAS [28]
		Phylum: *Firmicutes*	TAS [29–31]
		Class: *Bacilli*	TAS [32,33]
		Order: *Bacillales*	TAS [34,35]
		Family: *Bacillaceae*	TAS [35,36]
		Genus: *Bacillus*	TAS [10,35]
		Species: *Bacillus andreraoultii*	IDA
		Type strain SIT1^T	IDA
	Gram stain	Positive	IDA
	Cell shape	Rod-shaped	IDA
	Motility	Motile	IDA
	Sporulation	Sporulating	IDA
	Temperature range	Mesophile	IDA
	Optimum temperature	37–45°C	IDA
MIGS-6.3	Salinity	0	
MIGS-22	Oxygen requirement	Aerobic or facultative anaerobic	IDA
	Carbon source	Unknown	IDA
	Energy source	Unknown	IDA
MIGS-6	Habitat	Human gut	IDA
MIGS-15	Biotic relationship	Free-living	IDA
MIGS-14	Pathogenicity	Unknown	
	Biosafety level	2	
	Isolation	Human faeces	
MIGS-4	Geographic location	Marseille, France	IDA
MIGS-5	Sample collection time	March 2013	IDA
MIGS-4.1	Latitude	43.296482	IDA
MIGS-4.1	Longitude	5.36978	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. 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). 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 by an expert or reputable institution mentioned in the acknowledgements.

Phenotypic and biochemical characteristics

B. andreraoultii growth occurred at all temperatures tested; however, optimal growth was observed between 37 and 45°C after 24 hours of incubation. Colonies were 0.1 to 0.3 µm diameter on blood-enriched Columbia agar. Growth was achieved aerobically and weak growth anaerobically. Gram staining showed rod-shaped, Gram-positive bacilli (Figure 4). Cells were grown on agar sporulate. A motility test was positive. Cells grown on agar are smooth and greyish after 24 hours of incubation, and they have an average width and length of 0.5 µm and 3 µm, respectively, and exhibited flagella (Figure 5).

Strain SIT1^T showed catalase activity but was negative for oxidase. Using an API ZYM strip, positive reactions were observed for alkaline phosphatase, esterase (C4), esterase lipase (C8), acid phosphatase, naphthol-AS-BI-phosphohydrolase, α-galactosidase, β-galactosidase, α-glucosidase and β-glucosidase. Negative reactions were observed for lipase (C14), leucine arylamidase, valine arylamidase, cystine arylamidase, trypsin, α-chymotrypsin, β-glucuronidase, N-acetyl- β-glucosaminidase, α-mannosidase and α-fucosidase. Using an API 20 NE strip, the nitrate reductase–hydrolysis reaction (β-glucosidase, esculine), and β-galactosidase and assimilation reaction (potassium gluconate) were also positive. Negative reactions were found for urease, indole, arginine dihydrolase and fermentation (glucose). Using an API 50 CH strip, positive reactions were recorded for L-arabinose, D-ribose, D-xylose, D-glucose, D-fructose, D-mannose, N-acetylglucosamine, arbutine, esculine, salicine, D-celiobiose, D-maltose, D-saccharose, D-trehalose and amidon.

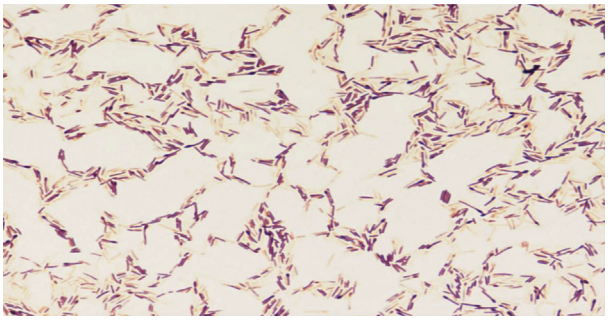

FIG. 4. Gram staining of *B. andreraoultii* strain SIT1^T.

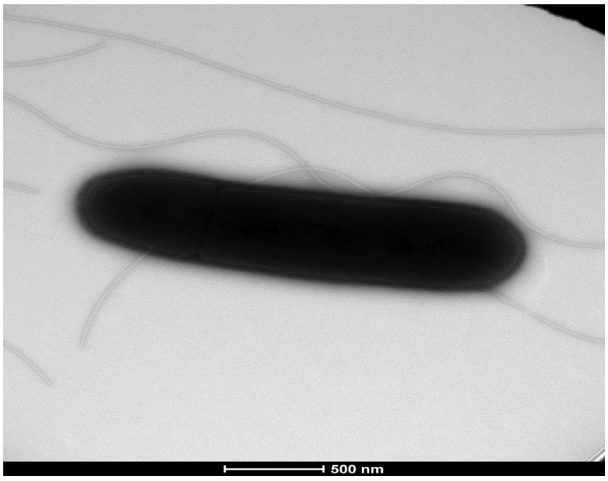

FIG. 5. Transmission electron microscopy of *B. andreraoultii* strain SIT1^T using Morgani 268D (Philips) at operating voltage of 60kV. Scale bar = 500 nm.

Negative reactions were recorded for glycerol, erythritol, D-arabinose, L-xylose, D-adonitol, methyl-βD-xylopyranoside, D-galactose, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl-αD-mannopyranoside, methyl-αD-glucopyranoside, amygdalin, D-lactose, D-melibiose, inulin, D-melezitose, D-raffinose, glycogen, xylitol, gentiobiose, D-turanose, D-lyxose, D-tagatose, D-fucose, L-fucose, D-arabitol, L-arabitol, potassium gluconate, potassium 2-ketogluconate and potassium 5-ketogluconate.

Bacillus andreraoultii SIT^1 was resistant to trimethroprim-sulphamethoxazole, macrolide (erythromycin), vancomycin and the third generation of cephalosporin (ceftriaxone) but was susceptible to fosfomycin, imipenem, penicillin, amoxicillin, gentamicin, ciprofloxacin, doxycycline and rifampicin. Five species with validly published names in the *Bacillus* genus were selected to make a phenotypic comparison with *B. andreraoultii* (Table 2).

Genomic characteristics and genome comparison

Bacillus andreraoultii SIT1^T was selected for sequencing on the basis of its phenotypic differences, phylogenetic position and

TABLE 2. **Differential characteristics of *Bacillus andreraoultii* SIT1^T; *B. thermoamylovorans* strain LMG 18084^T; *B. thermolactis* strain R-6488^T; *B. circulans* strain LMG 13261^T; *B. subterraneus* strain COO13B^T; and *B. niacin* 2923^T**

Property	*B. andreraoultii*	*B. thermoamylovorans*	*B. thermolactis*	*B. circulans*	*B. subterraneus*	*B. niacini*
Cell diameter (μm)	0.1–0.3	0.5–4	1–4	1–3	0.5–12	3–5
Mean length (μm)	3	5	10	4.2	25	5.6
Oxygen requirement	F/anaerobic	F/anaerobic	F/anaerobic	F/anaerobic	F/anaerobic	Aerobic
Gram stain	+	+	+	−	−	+
Motility	+	+	−	+	+	−
Flagella	+	NA	+	+	+	−
Endospore formation	+	+	+	+	−	+
Production of:						
Alkaline phosphatase	+	NA	NA	NA	NA	NA
Acid phosphatase	+	NA	NA	NA	NA	NA
Catalase	+	+	+	NA	+	NA
Oxidase	−	+	+	−	−	+
Nitrate reductase	+	+	+	v	+	+
Indole	−	−	−	−	−	−
Urease	−	−	−	−	−	−
α-Galactosidase	+	NA	NA	NA	NA	NA
β-Galactosidase	+	NA	NA	NA	−	NA
β-Glucuronidase	−	NA	NA	NA	−	NA
α-Glucosidase	+	NA	NA	NA	+	NA
β-Glucosidase	+	NA	NA	NA	+	NA
Esterase	+	NA	NA	NA	NA	NA
Esterase lipase	+	NA	NA	NA	NA	NA
Naphthol-AS-BI-phosphohydrolase	+	NA	NA	NA	NA	NA
N-acetyl-β-glucosaminidase	−	NA	NA	NA	NA	NA
Pyrazinamidase	NA	NA	NA	NA	NA	NA
α-Mannosidase	−	NA	NA	NA	NA	−
α-Fucosidase	−	NA	NA	NA	NA	NA
Leucine arylamidase	−	NA	NA	NA	NA	NA
Valine arylamidase	−	NA	NA	NA	NA	NA
Cystine arylamidase	−	NA	NA	NA	NA	NA
α-Chemotrypsin	−	NA	NA	NA	NA	NA
Trypsin	−	NA	NA	NA	NA	NA
Utilization of:						
5-Keto-gluconate	−	−	−	v	NA	NA
D-Xylose	+	+	+	+	+	−
D-Fructose	+	+	+	+	+	+
D-Glucose	+	+	+	+	+	+
D-Mannose	+	+	−	+	−	NA
Habitat	Human gut	Wine, grass . . .	Raw milk	Bee larvae	Subterranean water	Soil

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

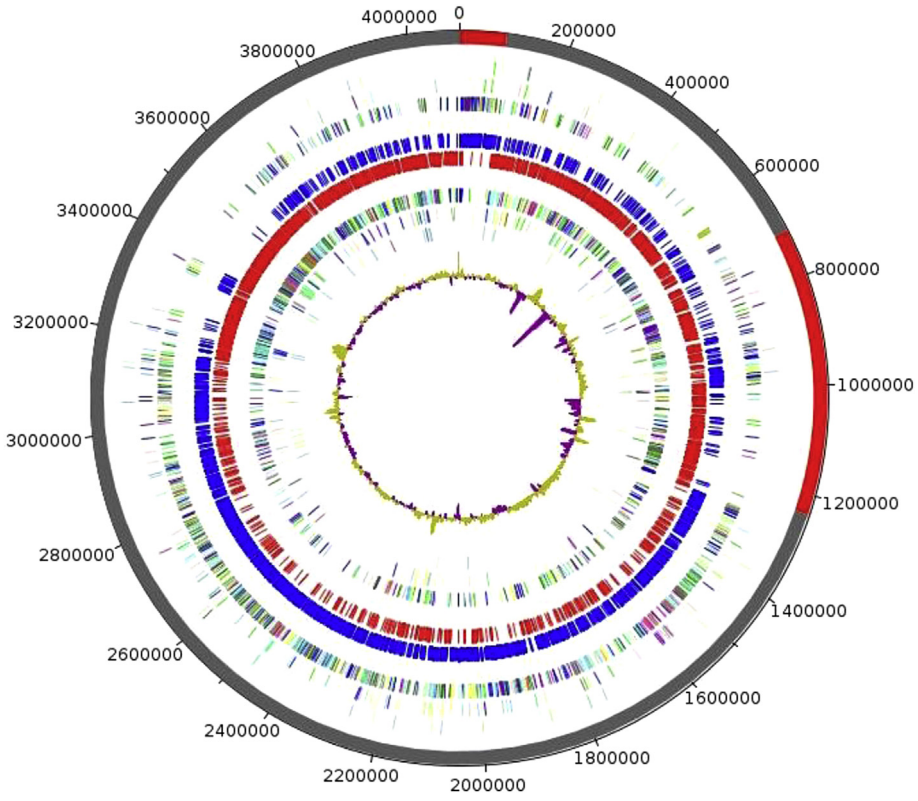

FIG. 6. Graphical circular map of genome. From outside to center: Contigs (red/grey), COGs category of genes on forward strand (three circles), genes on forward strand (blue circle), genes on reverse strand (red circle), COGs category on reverse strand (three circles), GC content.

16S rRNA sequence similarity to other members of the *Bacillus* genus. It was part of a culturomics study aimed at isolating all bacterial species from human digestive flora in patients with kwashiorkor, an acute form of malnutrition. It is the first sequenced genome from *B. andreraoultii* sp. nov. The European Molecular Biology Laboratory (EMBL) accession number of *B. andreraoultii* genome is CCFJ00000000 and consists of 14 scaffolds and 58 contigs (Figure 6). Table 3 shows the project information and its association with minimum information about a genome sequence (MIGS) version 2.0 compliance [27].

The genome is 4 092 130 bp long with 35.42% G+C content. On the 3843 predicted genes, 3718 were protein-coding genes and 116 were RNAs genes. The remaining genes were annotated as hypothetical proteins (712 genes, >19.15%). The properties and statistics of the genome are summarized in Table 4. The distribution of genes into functional COGs categories is presented in Table 5. The draft genome sequence of *Bacillus andreraoultii* was smaller than those of *Bacillus halodurans* C-125, *Bacillus pseudofirmus* OF4 and *Lysinibacillus sphaericus*

TABLE 3. Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality draft
MIGS-28	Libraries used	Mate pair
MIGS-29	Sequencing platform	Illumina MiSeq
MIGS-31.2	Fold coverage	185×
MIGS-30	Assemblers	SOAPdenovo
MIGS-32	Gene calling method	Prodigal
	GenBank date of release is 8-07-2015	
	NCBI project ID	PRJEB6477
	EMBL accession	CCFJ00000000
MIGS-13	Project relevance	Study of human gut microbiome

EMBL, European Molecular Biology Laboratory; MIGS, minimum information about a genome sequence; NCBI, National Center for Biotechnology Information.

TABLE 4. Nucleotide content and gene count levels of genome

Attribute	Value	% of total[a]
Size (bp)	4 092 130	100
G+C content (bp)	1 434 521	35.42
Coding region (bp)	3 248 339	79.38
Total genes	3834	100
RNA genes	116	3.0
Protein-coding genes	3718	100
Genes with function prediction	2505	67.37
Genes assigned to COGs	2420	65.08
Genes with peptide signals	308	8.26
Genes with transmembrane helices	918	24.69

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

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

Code	Value	% Value	Description
J	163	4.3840775	Translation
A	0	0	RNA processing and modification
K	192	5.164067	Transcription
L	233	6.26681	Replication, recombination and repair
B	1	0.02689618	Chromatin structure and dynamics
D	31	0.8337816	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	62	1.6675632	Defense mechanisms
T	118	3.1737492	Signal transduction mechanisms
M	107	2.8778913	Cell wall/membrane biogenesis
N	65	1.7482517	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	41	1.1027434	Intracellular trafficking and secretion
O	105	2.824099	Posttranslational modification, protein turnover, chaperones
C	145	3.899946	Energy production and conversion
G	172	4.626143	Carbohydrate transport and metabolism
E	255	6.858526	Amino acid transport and metabolism
F	71	1.9096289	Nucleotide transport and metabolism
H	95	2.5551372	Coenzyme transport and metabolism
I	116	3.119957	Lipid transport and metabolism
P	181	4.8682084	Inorganic ion transport and metabolism
Q	47	1.2641205	Secondary metabolites biosynthesis, transport and catabolism
R	344	9.252286	General function prediction only
S	245	6.589564	Function unknown
—	1298	34.911243	Not in COGs

COGs, Clusters of Orthologous Groups database.

C3-41 (4.09, 4.20, 4.25 and 4.82 MB respectively) but larger than those of *Solibacillus silvestris* StLB046, *Bacillus pumilus* SAFR-032 and *Bacillus coagulans* 2-6 WK1 (3.98, 3.70 and 3.07 MB, respectively). The G+C content of *Bacillus andreraoultii* was smaller than those of *Bacillus pumilus* SAFR-032, *Bacillus coagulans* 2-6, *Bacillus halodurans* C-125, *Bacillus pseudofirmus* OF4 and *Lysinibacillus sphaericus* C3-41 (35.42, 41.29, 47.29, 43.69, 39.86 and 37.13% respectively). The gene content of *Bacillus andreraoultii* was smaller than those of *Bacillus halodurans* C-125, *Bacillus pseudofirmus* OF4 and *Lysinibacillus sphaericus* C3-41 (3718, 4052, 4335 and 4771, respectively) but larger than those of *Bacillus pumilus* SAFR-032 and *Bacillus coagulans* 2-6 (3681 and 2971 respectively) (Table 6). The distribution of genes into COGs categories in the genomes from all seven compared *Bacillus* species, *Anoxybacillus flavithermus, Lysinibacillus sphaericus* and *solibacillus silvestris* (Figure 7).

Bacillus andreraoultii contained a bacteriocin (colicin) consisting of 175 amino acids harboured with the bacteriocins of *Paenibacillus* and *Lactobacillus* (Figure 8) and sharing 62% of the homology of *Bacillus vireti.* The results did not indicate the presence of nonribosomal peptide synthetases and polyketide synthases and phage. We performed the analysis of the resistome of *Bacillus andreraoultii* SIT1^T antibiotic classes of the macrolide–lincosamide–streptogramin B (MLSB) antibiotics, such as ATP-binding transporters (ABC) *lsaB* and *msrD,* major facilitator transporters *mefA,* transferases (*vatA*) and two-component system vancomycin resistance *vanS*/*vanR*, *norA* (Table 7).

Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses (taxonogenomics), we formally propose the creation of *Bacillus andreraoultii* sp. nov. that contains the strain SIT1^T. This bacterium was isolated from the stool of a 2-year-old Nigerian boy with kwashiorkor, a severe form of acute malnutrition.

Description of *Bacillus andreraoultii* sp. nov.

The *Bacillus andreraoultii* name come from André Raoult, who was a military doctor who worked with malnutrition and kwashiorkor in Senegal and who described its specific features [37].

Strain SIT1^T is aerobic, Gram positive, endospore forming, motile and rod shaped. Growth was achieved aerobically between 25 and 55°C (optimum 37 to 45°C). After 24 hours of growth on 5% sheep's blood–enriched Columbia agar at 37°C, bacterial colonies were smooth and greyish with a diameter of 0.1 to 0.3 mm. The cells had a mean width and length of 0.5 µm and 3 µm, respectively, and exhibited flagella. They were catalase positive and oxidase negative.

Bacillus andreraoultii SIT1^T was resistant to trimethroprim-sulphamethoxazole, macrolide (erythromycin), vancomycin

TABLE 6. Numbers of orthologous proteins shared between genomes (upper right) and AGIOS values obtained (lower left)

	Bacillus halodurans	*Bacillus pseudofirmus*	*Bacillus coagulans*	*Lysinibacillus sphaericus*	*Bacillus pumilus*	*Bacillus andreraoultii*
B. halodurans	4052[a]	1819	1293	1360	1603	1429
B. pseudofirmus	68.7%	4335[a]	1300	1388	1577	1452
B. coagulans	63.5%	63.2%	2971[a]	1166	1366	1362
L. sphaericus	63.2%	63.9%	62.4%	4771[a]	1409	1317
B. pumilus	64.8%	65.4%	64.4%	64%	3681[a]	1472
B. andreraoultii	64.4%	65.2%	64.1%	65.6%	65.6%	3718[a]

AGIOS, average genomic identity of orthologous gene sequences.
[a]Numbers of proteins per genome. Percentage 0.660 = 6.

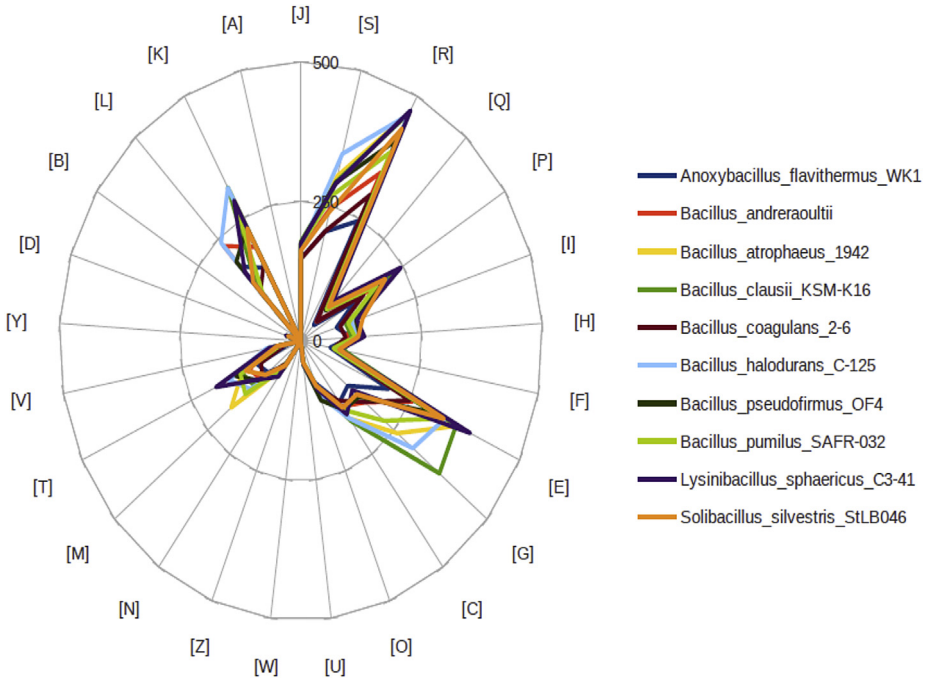

FIG. 7. Distribution of functional classes of predicted genes in genomes from *Bacillus andreraoultii, Anoxybacillus flavithermus, Bacillus atrophaeus, Bacillus clausii, Bacillus coagulans, Bacillus halodurans, Bacillus pseudofirmus, Bacillus pumilus, Lysinibacillus sphaericus* and *Solibacillus silvestris* chromosomes according to clusters of orthologous groups of proteins.

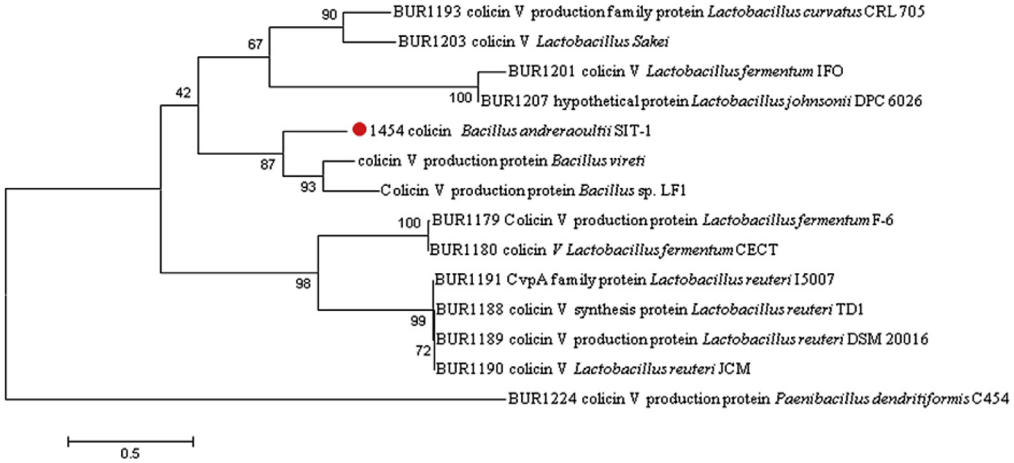

FIG. 8. Molecular phylogenetic analysis by maximum likelihood method of representatives of genus *Bacillus andreraoultii* SIT1^{T} inferred from 16S rRNA gene sequence. Tree with highest log likelihood (−2930.4905) is shown. Percentage of trees in which associated taxa clustered together is shown next to branches. Initial trees for heuristic search were obtained automatically by applying neighbour-joining and BioNJ algorithms to matrix of pairwise distances estimated using JTT model, then selecting topology with superior log likelihood value. Tree is drawn to scale, with branch lengths measured in number of substitutions per site. Analysis involved 14 amino acid sequences. There 155 positions in final data set. Evolutionary analyses were conducted in MEGA5.

TABLE 7. List of genes associated with antibiotic resistance in *Bacillus andreraoultii* SIT-1

Gene function	ORF	Gene name	GC	Size (aa)	Function	Best BLAST hit, GenBank	% Coverage	% Identity
MLSB	328	*lsaB*	32.7	494	ABC transporter	*Bacilli*	100	94
Glycopeptide	395	*van S*	34.7	378	Sensor histidine kinase VanS	*Bacillus* sp. J37	99	82
Glycopeptide	396	*vanR*	36.6	232	Vancomycin response regulator VanR	*Clostridium clariflavum*	100	96
MLSB	431	*vatA*	35	213	Chloramphenicol acetyltransferase	*Paucisalibacillus globulus*	100	99
MLSB	572	*msrd*	38.7	487	ABC transporters	*Clostridium kluyveri* NBRC 12016	100	100
MLSB	573	*mefA*	37.2	408	Macrolide-efflux protein	*Clostridium kluyveri*	100	100
MLS	583	*msrd*	35.8	177	ABC-F type ribosomal protection protein	*Bacteroides*	100	91

ORF, open reading frame; MLSB, macrolide–lincosamide–streptogramin B.

and the third generation of cephalosporin (ceftriaxone). It contained a bacteriocin.

The genome is 4 092 130 bp long, and the G+C content is 35.42%. The 16S rRNA gene sequence and whole-genome shotgun sequence of *B. andreraoultii* strain SIT1^T are deposited in GenBank under accession nos. LK021120 and CCFJ00000000, respectively. The type strain SIT1^T (= CSUR P1162 = DSM 29078) was isolated from the stool of a 2-year-old Nigerien boy with kwashiorkor, a severe form of acute malnutrition.

Conflict of Interest

None declared.

References

[1] Dubourg G, Lagier JC, Robert C, Armougom F, Hugon P, Metidji S, et al. Culturomics and pyrosequencing evidence of the reduction in gut microbiota diversity in patients with broad-spectrum antibiotics. Int J Antimicrob Agents 2014;44:117–24.

[2] 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.

[3] Tindall BJ, Rossello-Mora R, Busse HJ, Ludwig W, Kampfer P. Notes on the characterization of prokaryote strains for taxonomic purposes. Int J Syst Evol Microbiol 2010;60(pt 1):249–66.

[4] Stackebrandt E, Frederiksen W, Garrity GM, Grimont PA, Kämpfer P, Maiden MC, et al. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Microbiol 2002;52(pt 3):1043–7.

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

[6] 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.

[7] 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.

[8] 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.

[9] Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi M, Sentausa E, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol 2014;64(pt 2):384–91.

[10] Cohn F. Untersuchungen über Bakterien. Beitrage zur Biologie der Pflanzen Heft 1872;1:127–224.

[11] Nicholson WL. Roles of *Bacillus* endospores in the environment. Cell Mol Life Sci 2002;59:410–6.

[12] Logan NA. Bacillus species of medical and veterinary importance. J Med Microbiol 1988;25:157–65.

[13] Jernigan JA, Stephens DS, Ashford DA, Omenaca C, Topiel MS, Galbraith M, et al. Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States. Emerg Infect Dis 2001;7:933–44.

[14] Bottone EJ. Bacillus cereus, a volatile human pathogen. Clin Microbiol Rev 2010;23:382–98.

[15] Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 2009;49:543–51.

[16] Nkamga VD, Huynh HT, Aboudharam G, Ruimy R, Drancourt M. Diversity of human-associated *Methanobrevibacter smithii* isolates revealed by multispacer sequence typing. Curr Microbiol 2015;70:810–5.

[17] Padmanabhan R, Dubourg G, Lagier JC, Couderc C, Michelle C, Raoult D, et al. Genome sequence and description of *Corynebacterium ihumii* sp. nov. Stand Genomic Sci 2014;9:1128–43.

[18] 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.

[19] 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.

[20] 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.

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

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

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

[24] Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 2008;9:75.

[25] Drissi F, Buffet S, Raoult D, Merhej V. Common occurrence of antibacterial agents in human intestinal microbiota. Front Microbiol 2015;6:441.

[26] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res 2011;39(Web Server issue): W347–52.
[27] 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.
[28] 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.
[29] Gibbons NE, Murray RGE. Proposals concerning the higher taxa of bacteria. Int J Syst Bacteriol 1978;28:1–6.
[30] Garrity GM, Holt JG. 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.
[31] Murray RGE. The higher taxa, or A place for everything... ? In: Holt JG, editor. Bergey's manual of systematic bacteriologyVol. 1. Baltimore: Williams & Wilkins; 1984. p. 31–4.
[32] Oren A, Garrity GM. List of new names and new combinations previously effectively, but not validly, published. Int J Syst Evol Microbiol 2015;65:3763–7.
[33] 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.
[34] Prévot AR. Dictionnaire des bactéries pathogens. In: Hauduroy P, Ehringer G, Guillot G, et al., editors. Paris: Masson; 1953. p. 1–692.
[35] Skerman VBD, McGowan V, Sneath PHA. Approved list of bacterial names. Int J Syst Bacteriol 1980;30:225–420.
[36] Fischer A. Untersuchungen über bakterien. Jahrbücher für Wissenschaftliche Botanik 1895;27:1–163.
[37] Raoult A. Aspects of malnutrition in a big child in French East Africa; data from liver puncture biopsy. I. Bull Soc Pathol Exot Filiales 1959;52: 97–114.