

Paenibacillus phocaensis sp. nov., isolated from the gut microbiota of a healthy infant

M. Tidjani Alou¹, J. Rathored¹, T. T. Nguyen¹, C. Andrieu¹, C. Couderc¹, S. Brah², B. A. Diallo³, P.-E. Fournier¹, D. Raoult^{1,4} and G. Dubourg¹

1) Aix-Marseille Université, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM63, CNRS 7278, IRD 198, INSERM 1095, Marseille, France, 2) Hôpital National de Niamey, BP 238, 3) Laboratoire de microbiologie, département de biologie, Université Abdou Moumouni de Niamey, BP 237, Niamey, Niger and 4) Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract

Paenibacillus phocaensis sp. nov. strain mt24^T (= CSUR P2238 = DSM 101777) is a Gram-negative, facultative anaerobic, spore-forming and motile bacilli. This strain was isolated from the stool sample of a healthy infant from Niger. Its genome was estimated to a size of 5 521 415 bp with a 53.54% GC content. It contains 4835 protein-coding genes and 89 RNAs, among which two were 16S rRNA genes. There were also 101 genes (2.09%) identified as ORFans.

© 2016 The Authors. Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases.

Keywords: Culturomics, genome, gut microbiota, *Paenibacillus phocaensis*, taxonogenomics

Original Submission: 20 September 2016; **Revised Submission:** 28 November 2016; **Accepted:** 1 December 2016

Article published online: 9 December 2016

Corresponding author: G. Dubourg, Aix-Marseille Université, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM63, CNRS 7278, IRD 198, INSERM 1095, 27 Boulevard Jean Moulin, 13385 Marseille, France
E-mail: greg.dubourg@gmail.com

The human gut microbiota is a vast and complex ecosystem that contributes greatly to the human body's health. Among other functions, its involvement in the digestive process, the immune system and the production of some micronutrients has been proven [1]. Several pathologies like obesity, inflammatory bowel disease, irritable bowel syndrome and severe acute malnutrition have been associated with a disruption of the gut microbiota [1,2]. Its establishment begins at the birth when the newborn is exposed to the mother's vaginal microbiota [3]. Great variability has been shown in the composition of the gut microbiota of healthy children [4,5], but it stabilizes and reaches maturity around the age of 2 or 3 years [1].

In order to study the composition of the gut microbiota, culture and metagenomics approaches have been used, although metagenomics have been the method of the choice in the last decade [6]. Recently, a wide-array culture method, the microbial

culturomics strategy, has been developed in our laboratory. It consists in the multiplication of culture conditions with a variation of culture media and physicochemical parameters like temperature, pH and atmosphere in order to characterize a bacterial ecosystem as exhaustively as possible [7]. This method was used to characterize by a culture approach the gut microbiota of healthy African children. A new member of the *Paenibacillus* genus was thus isolated. This genus was created by Ash *et al.* in 1994 [8] and consists of 174 species [9]. The *Paenibacillus* genus consists of Gram-variable, spore-forming, motile bacilli. These aerobic or facultative anaerobic species have a G+C content ranging from 40 to 59%. Its type species is *Paenibacillus polymyxa*. *Paenibacillus* species have not been associated with pathogenicity in humans; nevertheless, *Paenibacillus larvae*, *Paenibacillus lentimorbus* and *Paenibacillus popilliae* are insect pathogens [8].

Bacterial classification is currently based on phenotypic and genotypic characteristics [10–12]. The decreasing cost of genome sequencing has allowed more insights into the properties and capabilities of a species. A new concept of bacterial description, called taxonogenomics, was developed in our laboratory associating phenotypic characteristics, a proteomic description of the strain with the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) profile and a

presentation of the complete sequence of the annotated genome of the described strain [13]. Using this concept, we describe here a new member of the *Paenibacillus* genus called *Paenibacillus phocaensis* (= CSUR P2238 = DSM 101777), which is the 13th previously unknown *Paenibacillus* isolated as a part of a culturomics study.

Materials and Methods

Sample information

A stool sample was collected from a 13-month-old girl in Niamey, Niger. A weight-for-height z score of -0.95 was estimated for this healthy girl, with no growth stunting as showed by the calculated height-for-age z score of -0.88 . No antibiotics were being administered at the time of sample collection. Upon arrival in Marseille where the study was conducted, the stool sample was stored at -80°C . Consent was obtained from the girl's parent, and approval for this study was provided by the Institut Fédératif de Recherche 48 (Faculty of Medicine, Marseille, France) under agreement 09-022.

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

The bacterial diversity of the stool sample was characterized using the 18 culture conditions of standardized culturomics [14]. For each culture condition, a liquid preincubation of the sample was performed, and tenfold serial dilutions of this culture were seeded on 5% sheep's blood-enriched Colombia agar every 3 days for 30 days (bioMérieux, Marcy l'Étoile, France). Colonies were purified and identified using MALDI-TOF MS as previously described [15,16]. The MALDI-TOF MS analysis was carried out using a Microflex Spectrometer (Bruker Daltonics, Leipzig, Germany) with a MTP 96 MALDI-TOF target plate (Bruker). Each colony was tested in duplicate, and the obtained spectra were imported into MALDI BioTyper 2.0 software (Bruker). The spectra were then compared by standard pattern matching (with default parameter settings) to the 7567 references contained in our database, which consists of the Bruker database incremented with data from species that were not already present in the database. A strain is considered identified at the species level for an identification score of ≥ 1.9 . Between identification scores of 1.7 to 1.9, the strain is identified at the genus level. A score of < 1.7 does not allow any identification. The 16S rRNA gene was sequenced in order to obtain an identification as previously described [17].

Growth conditions

Different atmospheres, temperature, pH and NaCl concentrations were tested in order to determine the ideal growth condition of strain mt24^T. Anaerobic and microaerophilic

atmospheres were tested using respectively GENbag anaer and GENbag miroaer systems (bioMérieux). Aerobic growth was tested with or without 5% CO₂. The following temperatures were tested in each atmosphere: 25, 28, 37, 45 and 56°C. Growth was also tested on Colombia agar with different pH (6, 6.5, 7 and 8.5) and on Colombia agar supplement with various NaCl concentration (0.5, 1, 5, 7.5 and 10%).

Morphologic, biochemical and antibiotics susceptibility tests

Phenotypic characteristics like Gram staining, oxidase, catalase, motility and sporulation were determined as described previously [14]. Morphologic observations were also carried out by performing negative staining. Detection formvar-coated grids were deposited on a 40 μL bacterial suspension drop and incubated at 37°C for 30 minutes. Then followed a 10-second incubation on ammonium molybdate 1%. The grids were dried on blotting paper and observed with a Tecnai G20 transmission electron microscope (FEI Company, Limeil-Brevannes, France).

In order to determine the metabolic features of strain mt24^T, API strips 50CH, ZYM and 20 NE were used according to the manufacturer's instructions (bioMérieux). Starch hydrolysis was determined using the starch agar test. Colonies were grown on a starch agar plate; upon drenching the colonies with an iodine solution, circles of hydrolysis appear when amylase was produced by the tested strain as previously described [18]. Gelatin hydrolysis was tested according to the gelatin stab method [19]. A medium with gelatin as a nutrient and a solidifying agent was prepared, and the strain was inoculated in the appropriate condition until optimal growth. After incubation, in order to obtain growth, the tubs containing the liquefied medium and the grown strain are incubated at 4°C for a few hours. Gelatin hydrolysis induces an elimination the solidifying agent in the medium. Consequently gelatinase production induces a permanent liquefied state of the medium; when a solidification of the medium occurs upon cooling, gelatinase is not produced by the tested strain. Cellular fatty acid methyl ester (FAME) analysis was performed by gas chromatography/mass spectrometry (GC/MS). Two samples were prepared, with approximately 50 mg of bacterial biomass each harvested from several culture plates. FAMES were prepared as described by Sasser [20]. GC/MS analyses were carried out as previously described [21]. Briefly, FAMES were separated using an Elite 5-MS column and monitored by mass spectrometry (Clarus 500-SQ 8 S, PerkinElmer, Courtaboeuf, France). Spectral database search was performed using MS Search 2.0 operated with the Standard Reference Database IA (National Institute of Standards and Technology, Gaithersburg, MD, USA) and the FAMES mass spectral database (Wiley, Chichester, UK).

Antibiotic susceptibility was determined using the disk diffusion method and according to the European Committee on Antimicrobial Susceptibility Testing 2015 recommendations [22].

Genomic DNA preparation

Strain mt24^T was cultured on 5% sheep's blood-enriched Columbia agar (bioMérieux) at 37°C aerobically. Bacteria grown on three petri dishes were resuspended in 4 × 100 µL 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 three 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 (gDNA) of strain mt24^T 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 (Thermo Fisher Scientific, Waltham, MA, USA) to 66.2 ng/µL. The mate pair library was prepared with 1 µg of gDNA using the Nextera mate pair Illumina guide. The gDNA sample was simultaneously fragmented and tagged with a mate pair junction adapter. The pattern of 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 to 11 kb, with an optimal size at 3.927 kb. No size selection was performed, and 505 ng of tagged fragments were circularized. The circularized DNA was mechanically sheared to small fragments with optimum at 597 bp on the Covaris device S2 in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies), 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. An automated cluster generation and sequencing run was performed in a single 39-hour run at a 2 × 251 bp read length.

Genome annotation and analysis

Open reading frames (ORFs) were predicted using Prodigal [23] with default parameters, but the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched against the GenBank [24] and the Clusters of Orthologous Groups (COGs) databases using

BLASTP (*E* value 1e-03, coverage 0.7 and identity percentage 30%). If no hit was found, it was searched against the NR database using BLASTP with an *E* value of 1e-03, a coverage of 0.7 and identity percentage of 30%, and if the sequence length was smaller than 80 aa, we used an *E* value of 1e-05. The tRNAScanSE tool [25] was used to find tRNA genes, whereas ribosomal RNAs were found by using RNAmmer [26]. Lipoprotein signal peptides and the number of transmembrane helices were predicted using Phobius [27]. Mobile genetic elements were predicted using PHAST [28] and RAST [29]. ORFans were identified if all the BLASTP performed did not give positive results (*E* value smaller than 1e-03 for ORFs with sequence size larger than 80 aa or *E* value smaller than 1e-05 for ORFs with sequence length smaller 80 aa). Such parameter thresholds have already been used in previous studies to define ORFans. Artemis [30] and DNA Plotter [31] were used for data management and visualization of genomic features, respectively. The Mauve alignment tool (version 2.3.1) was used for multiple genomic sequence alignment [32].

Comparator species for genomic comparison were identified in the 16S RNA tree using PhyloPattern software [33]. The genome of strain mt24^T was compared to those of *Paenibacillus barengoltzii* strain NBRC 101216, *Paenibacillus macerans* strain ZY18, *Paenibacillus polymyxa* strain CF05, *Paenibacillus sanguinis* strain 2301083, *Paenibacillus sabiniae* strain T27 and *Paenibacillus borealis* strain DSM 13188. For each selected genome, the complete genome sequence, proteome genome sequence and ORFeome genome sequence were retrieved from the FTP site of the National Center for Biotechnology Information. 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). Annotation and comparison processes were performed in the Multi-Agent software system DAGOBAN [34] that include Figenix [35] libraries that provide pipeline analysis. To evaluate the genomic similarity between studied genomes, we determined two parameters, digital DNA-DNA hybridization (DDH) that exhibits a high correlation with DDH [36,37] and average genomic identity of orthologous gene sequences (AGIOS) [38], which was designed to be independent from DDH [38]. The AGIOS score [38] is the mean value of nucleotide similarity between all couples of orthologous proteins between the two studied genomes [38].

Results

Strain identification and phylogenetic analyses

Strain mt24^T was isolated after 15 days' preincubation in aerobic marine broth at 37°C and seeding on 5% sheep's blood-enriched Columbia agar also in aerobic condition at 37°C. A

score under 1.7 was obtained for strain mt24^T after the MALDI-TOF MS analysis and the 16S rRNA gene was sequenced. The sequence available under accession number LN998053 showed a similarity level of 98.2% with *Paenibacillus timonensis* (Fig. 1). This similarity level under the 98.65% threshold defined by Stackebrandt and Ebers [39] to delineate a new species and later confirmed by a recent study [40] allowed us to create a new species within the genus *Paenibacillus* called *Paenibacillus phocaensis* the general characteristics of which are presented in Table 1. The reference spectra for *Paenibacillus phocaensis* (Fig. 2) was incremented in our database. The spectrum of *P. phocaensis* was also compared to other *Paenibacillus* species (Fig. 3).

Phenotypic description

Growth was observed between 25 and 56°C under aerobic, anaerobic and microaerophilic conditions. Strain mt24^T also grew at all the tested pH, but only at a 0.5% NaCl concentration. Ideal growth was observed under aerobic condition at 37°C and pH 7 after an incubation of 48 hours. There was also healthy growth observed at 37°C under anaerobic and microaerophilic conditions and at 28°C under aerobic condition.

TABLE 1. Classification and general features of *Paenibacillus phocaensis* strain mt24^T

Property	Term
Current classification	Domain: <i>Bacteria</i> Phylum: <i>Firmicutes</i> Class: <i>Bacilli</i> Order: <i>Bacillales</i> Family: <i>Paenibacillaceae</i> Genus: <i>Paenibacillus</i> Species: <i>Paenibacillus phocaensis</i> Type strain: mt24
Gram stain	Negative
Cell shape	Bacilli
Motility	Motile
Sporulation	Sporulating
Temperature range	Mesophilic
Optimum temperature	37°C

Cells were Gram negative, motile and sporulating (Fig. 4) and formed translucent concentric colonies. Negative staining of strain mt24^T showed a mean diameter of 0.47 µm and a mean length of 3.44 µm (Fig. 5).

The most abundant fatty acid is 15:0 anteiso (52%). The fatty acids were mainly saturated. Only three unsaturated species were detected with low abundances (<1%) (Table 2). When compared to *Paenibacillus timonensis*, the phylogenetically

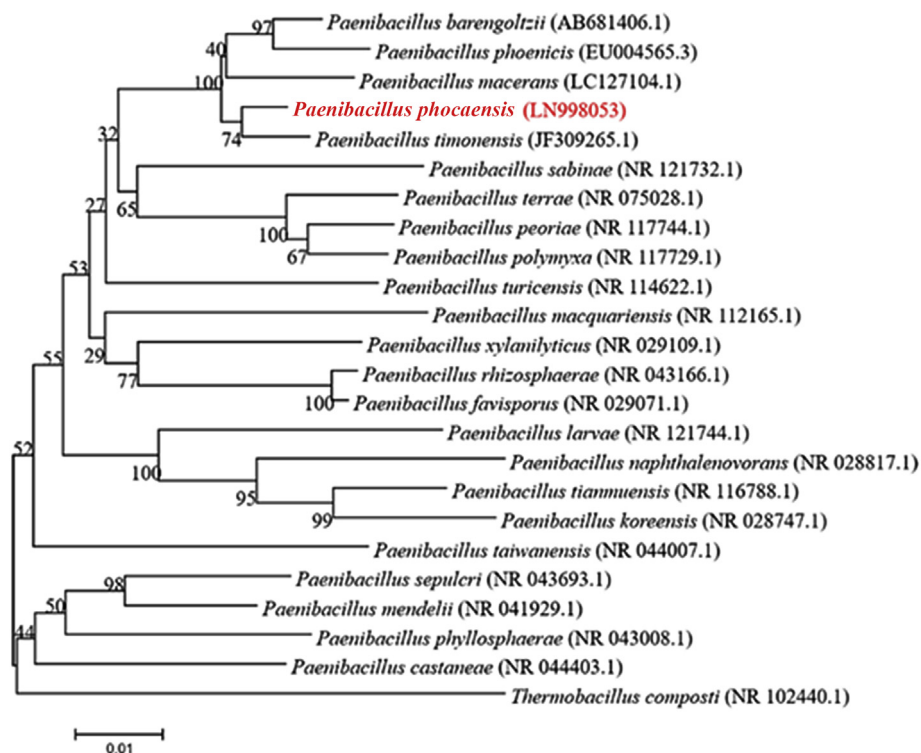


FIG. 1. Phylogenetic tree highlighting position of *Paenibacillus phocaensis* strain mt24^T (= CSUR P2238 = DSM 101777) relative to other close strains. Respective GenBank accession numbers for 16S rRNA genes are indicated in parentheses. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using maximum-likelihood method within MEGA6 software. Numbers at nodes are percentages of bootstrap values obtained by repeating analysis 1000 times to generate majority consensus tree. *Clostridium felsineum* strain NCIMB 10690 was used as outgroup. Scale bar represents 2% nucleotide sequence divergence.

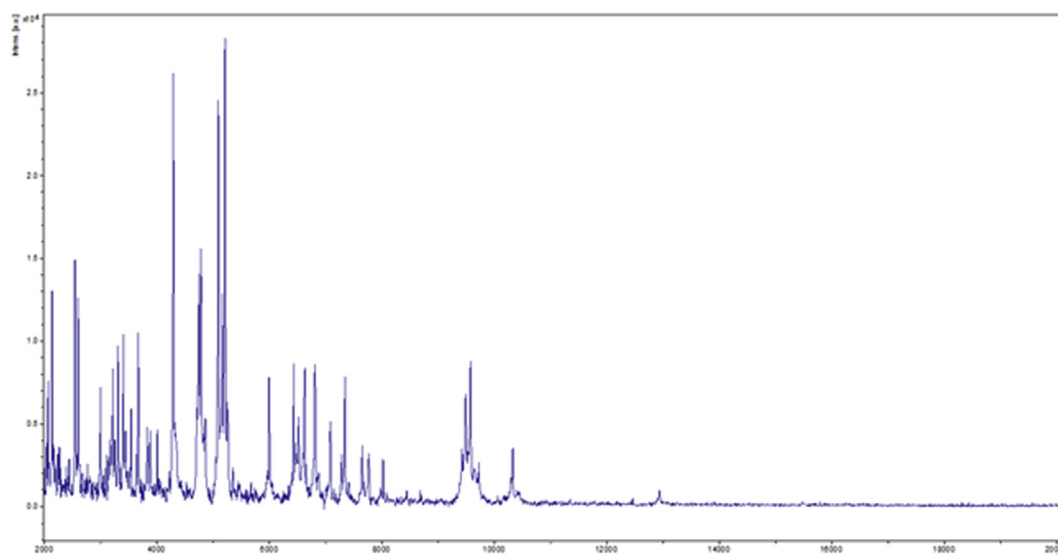


FIG. 2. Reference mass spectrum from *Paenibacillus phocaensis* strain mt24^T (= CSUR P2238 = DSM I01777). Spectra from 12 individual colonies were compared and reference spectrum generated.

closest species, the majority fatty acid was the same. Nevertheless, the other fatty acids found in both species were in different abundances. Only one unsaturated fatty acid was present in *Paenibacillus timonensis*, and three fatty acids present in traces in strain mt24^T were completely absent in *Paenibacillus timonensis* (Table 2).

Strain mt24^T showed oxidase and catalase activities. Activities of the following enzymes were also exhibited using an API ZYM strip: esterase C4, lipase esterase C8, leucine arylamidase, naphthol-AS-BI-phosphohydrolase, α -galactosidase, β -galactosidase, α -glucosidase, β -glucosidase and *N*-acetyl- β -glucosaminidase. Negative reactions were obtained for alkaline phosphatase,

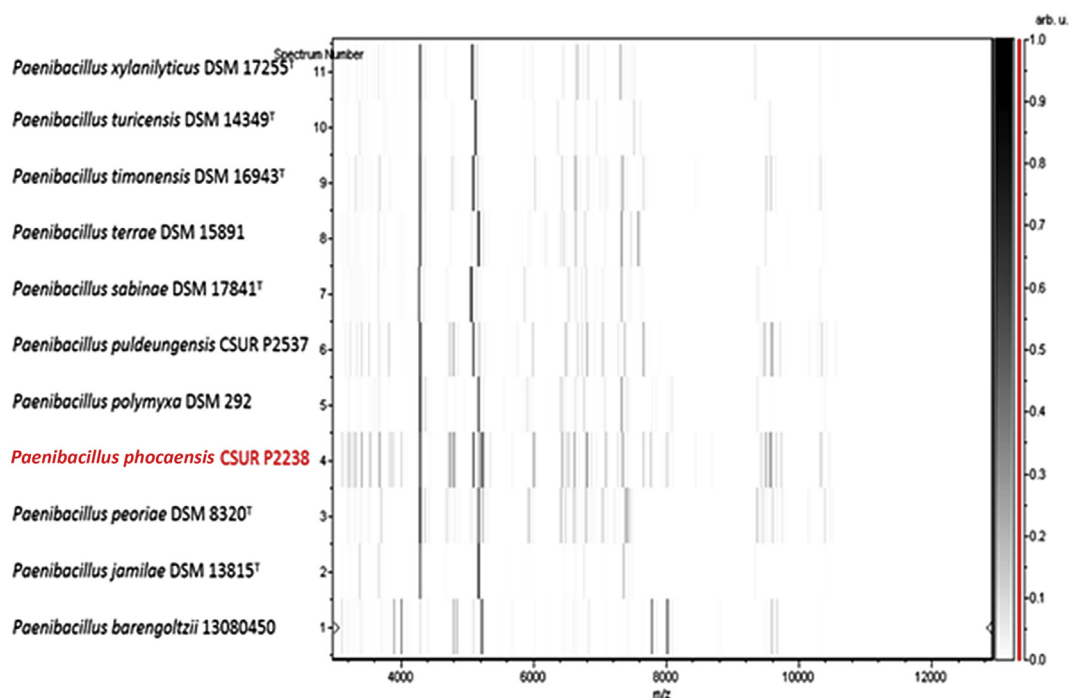


FIG. 3. Reference mass spectrum from *Paenibacillus phocaensis* strain mt24^T (= CSUR P2238 = DSM I01777). Spectra from 12 individual colonies were compared and reference spectrum generated.



FIG. 4. Gram staining of *Paenibacillus phocaensis* strain mt24^T (= CSUR P2238 = DSM 101777).

lipase C14, valine arylamidase, cysteine arylamidase, trypsin, α -chymotrypsin, acid phosphatase, β -glucuronidase, α -mannosidase and α -fucosidase. Using an API 20 NE strip, positive reactions were observed for nitrate reduction, β -glucosidase and β -galactosidase. Acid was produced from the fermentation of glucose, mannitol and *N*-acetyl-glucosamine. Using an API 50CH strip, L-arabinose, D-ribose, D-galactose, D-glucose, D-fructose, D-mannose, L-rhamnose, D-mannitol, methyl- α -D-glucopyranoside, *N*-acetyl-glucosamine, amygdaline, esculin, salicin, D-cellobiose, D-maltose, D-lactose, D-melibiose, D-sucrose, D-trehalose, D-melezitose, D-raffinose, starch, glycogen, xylitol, gentiobiose, L-fucose, potassium gluconate and potassium 2-

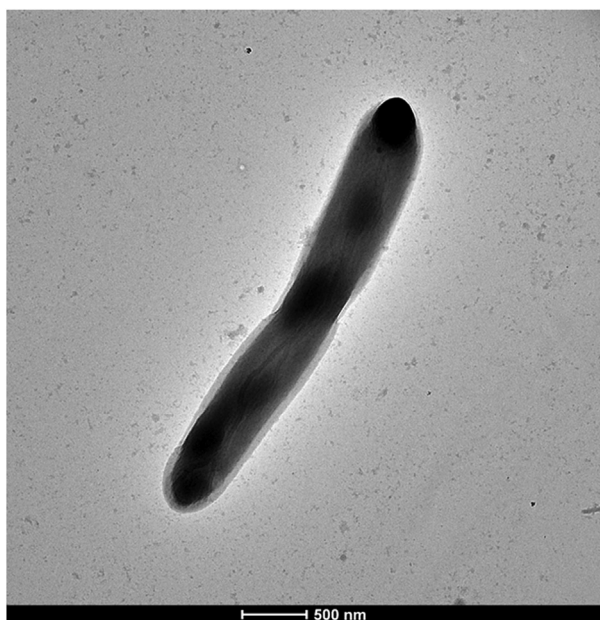


FIG. 5. Transmission electron microscopy of *Paenibacillus phocaensis* strain mt24^T (= CSUR P2238 = DSM 101777) using Tecnai G20 transmission electron microscope (FEI Company) at operating voltage of 60 kV. Scale bar = 500 nm.

TABLE 2. Cellular fatty acid composition (%) of *Paenibacillus phocaensis* strain mt24^T

Fatty acid	IUPAC name	<i>Paenibacillus phocaensis</i> (mean relative %) ^a
15:0 anteiso	12-methyl-Tetradecanoic acid	52.0 \pm 1.1
16:00	Hexadecanoic acid	10.6 \pm 0.4
16:0 iso	14-methyl-Pentadecanoic acid	8.7 \pm 0.5
15:0 iso	13-methyl-Tetradecanoic acid	8.5 \pm 0.1
17:0 anteiso	14-methyl-Hexadecanoic acid	7.5 \pm 0.3
15:00	Pentadecanoic acid	3.9 \pm 0.3
17:0 iso	15-methyl-Hexadecanoic acid	3.0 \pm 0.1
14:0 iso	12-methyl-Tridecanoic acid	1.8 \pm 0.1
17:00	Heptadecanoic acid	1.0 \pm 1.1
18:1n9	9-Octadecenoic acid	TR
14:00	Tetradecanoic acid	TR
18:00	Octadecanoic acid	TR
18:2n6	9,12-Octadecadienoic acid	TR
5:0 anteiso	2-methyl-Butanoic acid	TR
18:1n6	12-Octadecenoic acid	TR
13:00	Tridecanoic acid	TR

IUPAC, International Union of Pure and Applied Chemistry; TR, trace amounts <1%.

^aMean peak area percentage \pm standard deviation.

ketogluconate were metabolized. *Paenibacillus timonensis* was not able to metabolize L-rhamnose, L-fucose, glycogen, xylitol and potassium 2-ketogluconate. Negative reactions were observed for glycerol, erythrol, D-arabinose, D-ribose, D-xylose, D-adonitol, methyl- β -D-xylopyranoside, L-sorbose, dulcitol, inositol, D-sorbitol, methyl- α -D-mannopyranoside, arbutin, D-turanose, D-lyxose, D-tagatose, D-fucose, D-arabitol, L-arabitol and potassium 5-ketogluconate. Though negative reactions were observed for D-sorbitol, arbutin and D-turanose, they were found to be metabolized by *Paenibacillus timonensis*. Strain mt24^T and *Paenibacillus timonensis* both showed negative reactions for gelatinase and positive reactions for starch hydrolysis.

Cells were sensitive to doxycycline, ceftriaxone, penicillin, gentamycin, trimethoprim/sulfamethoxazole, imipenem, vancomycin, rifamycin and teicoplanin but resistant to colistin, ciprofloxacin, erythromycin, oxacillin, nitrofurantoin and metronidazole.

Phenotypic features of *P. phocaensis* were compared to those of close relatives (Table 3).

Genome properties

The genome is 5 521 412 bp long with 53.54% GC content (Fig. 6). It is composed of seven scaffolds (composed of 39 contigs). Of the 4924 predicted genes, 4835 were protein-coding genes and 89 were RNAs (eight genes are 5S rRNA, two genes are 16S rRNA, three genes are 23S rRNA and 76 genes are tRNA genes). A total of 3560 genes (73.63%) were assigned as putative function (by COGs or by NR BLAST). A total of 101 genes were identified as ORFans (2.09%). The remaining 902 genes (20.52%) were annotated as hypothetical proteins. The properties and statistics of the genome are summarized in Table 4, while the distribution of genes into COGs functional categories is presented in Table 5.

TABLE 3. Differential characteristics of *Penicillus phocaensis* strain mt24^T DSM 101777, *Paenibacillus timonensis* strain 2301032^T CCUG 48216, *Paenibacillus sanguinis* strain 2301083^T CCUG 48214, *Paenibacillus barengoltzii* strain SAFN-016^T ATCC BAA-1209, *Paenibacillus xylanilyticus* strain XIL14^T LMG 21957, *Paenibacillus phoenicis* strain 3PO2SA^T NRRL B-59348, *Paenibacillus sabiniae* strain T27^T DSM 17841, *Paenibacillus puldeungensis* CAU 9324^T CCUG 59189, *Paenibacillus terrae* strain AMI41^T JCM 11466 and *Paenibacillus jamilae* strain B.3^T DSM 13815 [41–47]

Property	<i>Paenibacillus phocaensis</i>	<i>Paenibacillus timonensis</i>	<i>Paenibacillus sanguinis</i>	<i>Paenibacillus barengoltzii</i>	<i>Paenibacillus xylanilyticus</i>	<i>Paenibacillus phoenicis</i>	<i>Paenibacillus sabiniae</i>	<i>Paenibacillus puldeungensis</i>	<i>Paenibacillus terrae</i>	<i>Paenibacillus jamilae</i>
Cell diameter (µm)	0.5–0.4	0.5	0.5	0.5–0.8	1.5–1.55	1.0–1.5	0.7–0.8	0.3–0.4	1.3–1.8	0.5–1.2
Oxygen requirement	+	+	+	+	+	+	+	+	+	+
Gram stain	–	+	+	+	+	+	+	+	+/-	+/-
Salt requirement	–	–	–	–	+	+	+	+	+	–
Motility	+	+	+	+	+	+	+	–	+	+
Endospore formation	–	+	+	+	+	+	+	+	+	+
Indole	–	NA	NA	–	–	+	–	NA	NA	–
Production of:										
Catalase	+	+	–	+	+	+	+	–	+	+
Oxidase	+	–	–	+	–	NA	–	+	–	–
Nitrate reductase	–	w	+	–	+	+	+	+	+	+
Urease	–	NA	NA	+	–	–	NA	NA	–	NA
β-Galactosidase	+	NA	NA	+	+	+	NA	NA	NA	+
N-Acetyl-glucosamine	+	+	+	+	+	+	NA	–	w	–
Acid from:										
L-Arabinose	–	+	+	+	+	+	NA	+	+	+
Ribose	+	–	–	+	NA	+	NA	+	+	+
Mannose	+	+	–	+	+	+	NA	–	+	+
Mannitol	+	–	+	+	+	+	NA	–	+	+
Sucrose	+	–	+	–	+	+	–	+	+	+
D-Glucose	+	+	–	–	+	+	–	+	+	+
D-Fructose	+	+	+	+	NA	+	–	–	+	+
D-Maltose	+	+	+	+	+	+	–	+	+	+
D-Lactose	+	+	+	NA	NA	+	–	+	+	+
Habitat	Human gut	Blood	Blood	Clean room floor	Air	Floor	Soil	Soil	Soil	Olive oil

+, positive result; –, negative result; w, weakly positive result; NA, data not available.

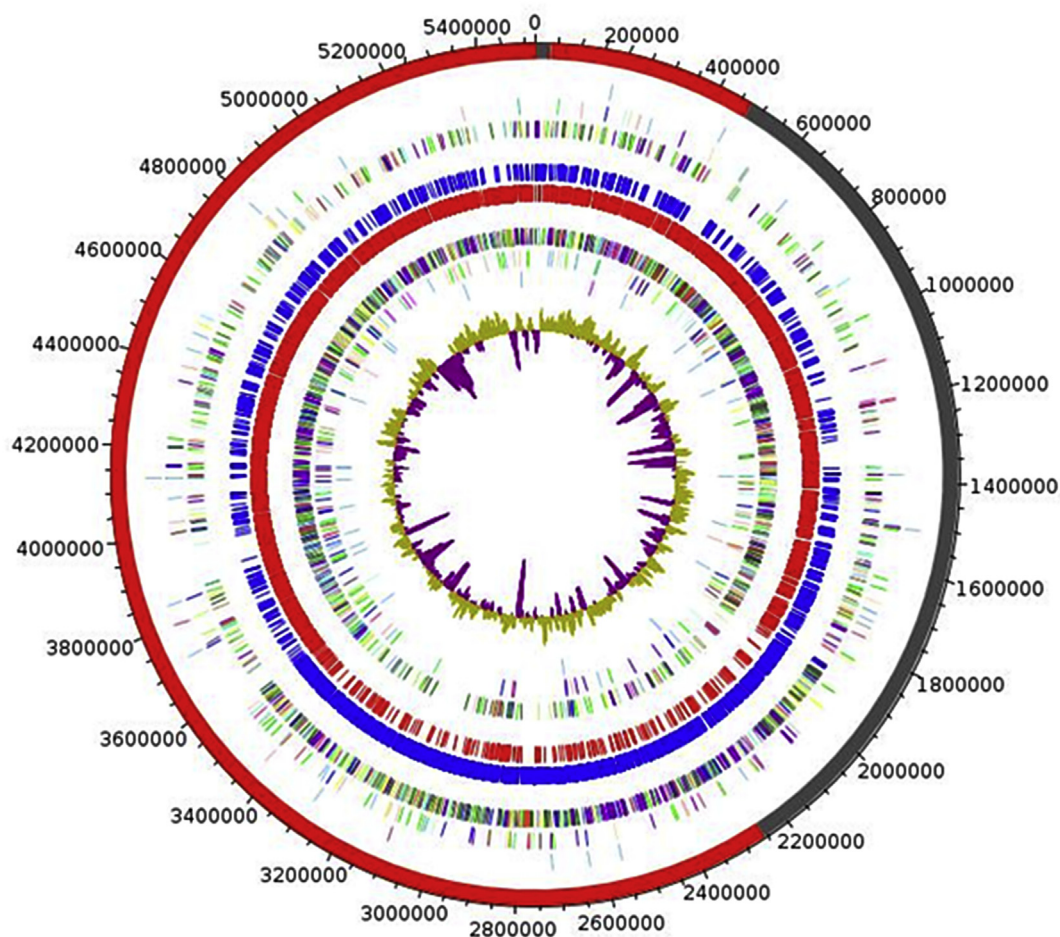


FIG. 6. Graphical circular map of chromosome. From outside to centre: genes on forward strain colored by COGs categories (only gene assigned to COGs), RNA genes (tRNAs green, rRNAs red), G+C content and G+C skew. COGs, Clusters of Orthologous Groups database.

TABLE 4. Nucleotide content and gene count levels of genome

Attribute	Genome (total)	
	Value	% of total ^a
Size (bp)	5 521 412	100
G+C content (%)	2 952 929	53.54
Coding region (bp)	4 699 807	85.11
Total genes	4924	100
RNA genes	89	1.80
Protein-coding genes	4835	98.19
Genes with function prediction	3755	76.25
Genes assigned to COGs	3338	67.79
Genes with peptide signals	707	14.35
CRISPR repeats	04	0.08
ORFans genes	101	2.05
Genes with transmembrane helices	1222	24.81
Genes associated with PKS or NRPS	09	0.10
No of antibiotic resistance genes	0	0
No. of genes associated Pfam-A domains	4484	91.06

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

Genome comparison

Compared to other close species reported in Table 6, the genome size of *Paenibacillus phocaensis* (5.52 Mb) is larger than the genome sizes of *Paenibacillus barengoltzii* (4.78 Mb) and *Paenibacillus sanguinis* (4.8 Mb) but smaller than *Paenibacillus borealis* (8.61 Mb), *Paenibacillus macerans* (7.34 Mb), *Paenibacillus polymyxa* (5.76 Mb) and *Paenibacillus sabiniae* (5.27 Mb) genome sizes. The G+C content of *P. phocaensis* (53.54%) is also higher than the G+C content of *Paenibacillus macerans*, *Paenibacillus sabiniae*, *Paenibacillus barengoltzii*, *Paenibacillus borealis*, *Paenibacillus sanguinis* and *Paenibacillus polymyxa* (52.6, 52.6, 51.9, 51.4, 49.3 and 45.5% respectively). Similarly, the number of protein-coding genes in *P. phocaensis* (4835) is higher than *P. polymyxa* (4715), *P. sabiniae* (4634), *P. sanguinis* (4238) and *P. barengoltzii* (4145) and lower than *P. macerans* (6166) and *P. borealis* (6698). Likewise, the total number of genes (4924) is higher than *P. barengoltzii* (4284), *P. sanguinis* (4388) and *P. sabiniae* (4857) and lower than that of *P. polymyxa* (4928), *P. macerans* (6474)

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

Code	Value	% of total ^a	Description
J	174	2.98	Translation
A	0	0.0	RNA processing and modification
K	476	8.16	Transcription
L	178	3.05	Replication, recombination and repair
B	1	0.02	Chromatin structure and dynamics
D	33	0.57	Cell cycle control, mitosis and meiosis
Y	0	0.0	Nuclear structure
V	113	1.94	Defense mechanisms
T	324	5.55	Signal transduction mechanisms
M	207	3.55	Cell wall/membrane biogenesis
N	74	1.27	Cell motility
Z	3	0.05	Cytoskeleton
W	0	0.0	Extracellular structures
U	56	0.96	Intracellular trafficking and secretion
O	106	1.82	Posttranslational modification, protein turnover, chaperones
C	161	2.76	Energy production and conversion
G	619	10.61	Carbohydrate transport and metabolism
E	309	5.30	Amino acid transport and metabolism
F	81	1.39	Nucleotide transport and metabolism
H	116	1.99	Coenzyme transport and metabolism
I	90	1.54	Lipid transport and metabolism
P	251	4.30	Inorganic ion transport and metabolism
Q	83	1.42	Secondary metabolites biosynthesis, transport and catabolism
R	550	9.43	General function prediction only
S	316	5.42	Function unknown
—	417	7.14	Not in COGs

COGs, Clusters of Orthologous Groups database.
^aTotal is based on total number of protein-coding genes in annotated genome.

and *P. borealis* (6945). Nevertheless, Fig. 7 highlights the fact that the repartition of the genes into COGs categories is similar in all the genomes besides the COGs category coding for extracellular structures, which was only represented in *P. borealis*.

P. phocaensis shared 2192, 2217, 2526, 2651, 2922 and 2946 orthologous proteins with *P. sabinae*, *P. polymyxa*, *P. sanguinis*, *P. borealis*, *P. barengoltzii* and *P. macerans* respectively. Among species with standing in nomenclature, AGIOS ranged from 67.09% between *P. macerans* and *P. polymyxa* and 76.40% between *P. macerans* and *P. barengoltzii*. When compared to the other species, AGIOS values ranged from 67.18% between *P. phocaensis*

TABLE 6. Genome comparison of closely related species to *Paenibacillus phocaensis* strain mt24^T

Organism	INSDC	Size (Mb)	G+C (%)	Protein-coding genes	Total genes
<i>Paenibacillus phocaensis</i> strain mt24 ^T	FCOQ00000000	5.52	53.54	4835	4924
<i>Paenibacillus barengoltzii</i> strain NBRC 101216	ASSZ00000000.1	4.78	51.9	4145	4284
<i>Paenibacillus macerans</i> strain ZY18	JMQA00000000.1	7.34	52.6	6166	6474
<i>Paenibacillus polymyxa</i> strain CF05	CP009909.1	5.76	45.5	4715	4928
<i>Paenibacillus sanguinis</i> strain 2301083	ARGO00000000.1	4.8	49.3	4238	4388
<i>Paenibacillus sabinae</i> strain T27	CP004078.1	5.27	52.6	4634	4857
<i>Paenibacillus borealis</i> strain DSM 13188	CP009285.1	8.61	51.4	6698	6945

INSDC, International Nucleotide Sequence Database Collaboration.

and *P. polymyxa* to 83.07% between *P. phocaensis* and *P. barengoltzii* (Table 7). Values between 20.3% (similarity between *P. phocaensis* and *P. sanguinis*) and 25.6% (*P. phocaensis* and *P. barengoltzii*) were obtained for digital DDH (Table 8). These values are under the 70% threshold used to delineate a new species with DDH [36,37], thus establishing the status of *P. phocaensis* as a new species.

Conclusion

Given the 98.2% similarity level of the 16S r RNA sequence of strain mt24^T with *Paenibacillus timonensis*, its MALDI-TOF MS profile and its annotated genome comparison with close species, we hereby create a new species within the genus *Paenibacillus* called *Paenibacillus phocaensis* (= CSUR P2238 = DSM 101777).

Description of *Paenibacillus phocaensis* sp. nov. strain mt24^T

Paenibacillus phocaensis (pho.ce'ensis, 'from Phocée,' the city after which Marseille was named; strain mt24^T was isolated in Marseille).

The organism is facultative anaerobic. The major fatty acid is 15:0 anteiso. It is oxidase and catalase positive. Nitrates were reduced into nitrites. Strain mt24^T showed oxidase and catalase activities. Urease and alkaline phosphatase were negative. Positive reactions were obtained for esterase C4, lipase esterase C8, leucine arylamidase, naphthol-AS-BI-phosphohydrolase, α -galactosidase, β -galactosidase, α -glucosidase, β -glucosidase and *N*-acetyl- β -glucosaminidase. Negative reactions were obtained for lipase C14, valine arylamidase, cysteine arylamidase, trypsin, α -chymotrypsin, acid phosphatase, β -glucuronidase, α -mannosidase and α -fucosidase. Using an API 50CH strip, L-arabinose, D-ribose, D-galactose, D-glucose, D-fructose, D-mannose, L-rhamnose, D-mannitol, methyl- α -D-glucopyranoside, *N*-acetylglucosamine, amygdaline, esculin, salicin, D-cellobiose, D-maltose, D-lactose, D-melibiose, D-sucrose, D-trehalose, D-melezitose, D-raffinose, starch, glycogen, xylitol, gentiobiose, L-fucose, potassium gluconate and potassium 2-ketogluconate were metabolized. Negative reactions were observed for glycerol, erythrol, D-arabinose, D-ribose, D-xylose, D-adonitol, methyl- β -D-xylopyranoside, L-sorbose, dulcitol, inositol, D-sorbitol, methyl- α -D-mannopyranoside, arbutin, D-turanose, D-lyxose, D-tagatose, D-fucose, D-arabitol, L-arabitol and potassium 5-ketogluconate.

Cells were sensitive to doxycycline, ceftriaxone, penicillin, gentamycin, trimethoprim/sulfamethoxazole, imipenem, vancomycin, rifamycin and teicoplanin but resistant to colistin, ciprofloxacin, erythromycin, oxacillin, nitrofurantoin and metronidazole.

The G+C content is 53.54%. The whole genome shotgun sequence and 16S r RNA sequence are deposited in GenBank

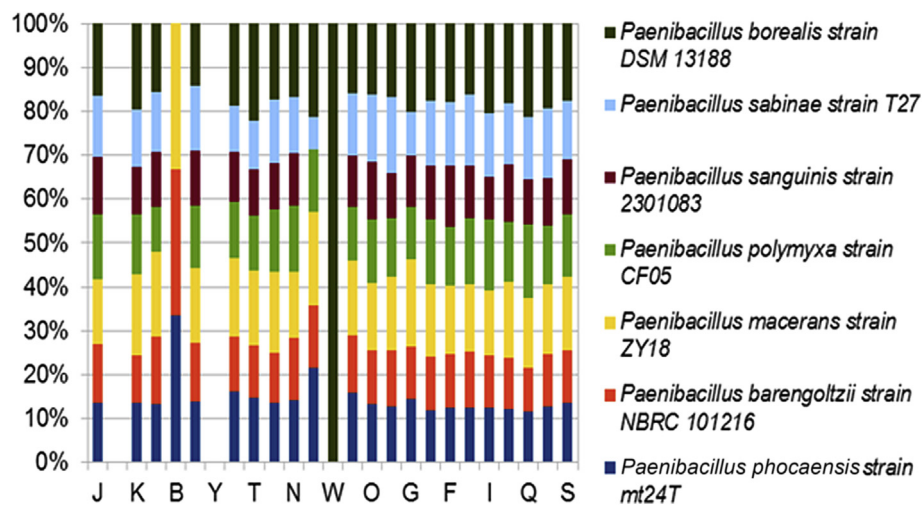


FIG. 7. Distribution of functional classes of predicted genes according to COGs of protein. COGs, Clusters of Orthologous Groups database.

TABLE 7. Numbers of orthologous protein shared between genomes (upper right)^a

	<i>Paenibacillus phocaensis</i>	<i>Paenibacillus barengoltzii</i>	<i>Paenibacillus borealis</i>	<i>Paenibacillus macerans</i>	<i>Paenibacillus polymyxa</i>	<i>Paenibacillus sanguinis</i>	<i>Paenibacillus sabiniae</i>
<i>P. phocaensis</i>	4916	2922	2651	2946	2217	2526	2192
<i>P. baragnostici</i>	83.07	4305	2514	2793	2170	2421	2139
<i>P. borealis</i>	68.73	68.48	6968	2747	2453	2394	2481
<i>P. macerans</i>	77.60	76.40	68.68	6476	2414	2568	2254
<i>P. polymyxa</i>	67.18	67.44	67.79	67.09	4929	2093	2167
<i>P. sanguinis</i>	75.32	74.88	68.44	74.05	67.62	4395	1981
<i>P. sabiniae</i>	69.32	68.85	74.05	69.49	67.57	68.31	4866

^aAverage percentage similarity of nucleotides corresponding to orthologous protein shared between genomes (lower left) and numbers of proteins per genome (bold).

TABLE 8. Pairwise comparison of *Paenibacillus phocaensis* strain mt24^T with other species using GGDC, formula 2 (DDH estimates based on identities/HSP length),^a upper right

	<i>Paenibacillus phocaensis</i>	<i>Paenibacillus barengoltzii</i>	<i>Paenibacillus borealis</i>	<i>Paenibacillus macerans</i>	<i>Paenibacillus polymyxa</i>	<i>Paenibacillus sanguinis</i>	<i>Paenibacillus sabiniae</i>
<i>P. phocaensis</i>	100 ± 00	25.6 ± 2.41	22.3 ± 2.36	22.7 ± 2.37	25.8 ± 2.41	20.3 ± 2.23	22.2 ± 2.36
<i>P. barengoltzii</i>		100 ± 00	22.6 ± 2.36	21.1 ± 2.33	26.5 ± 2.42	19.3 ± 2.29	22.8 ± 2.37
<i>P. borealis</i>			100 ± 00	22.8 ± 2.37	25.7 ± 2.41	19.6 ± 2.30	20.4 ± 2.32
<i>P. macerans</i>				100 ± 00	27.4 ± 2.43	19.9 ± 2.31	23.1 ± 2.37
<i>P. polymyxa</i>					100 ± 00	19.6 ± 2.30	30.8 ± 2.45
<i>P. sanguinis</i>						100 ± 00	19.9 ± 2.30
<i>P. sabiniae</i>							100 ± 00

DDH, DNA-DNA hybridization; GGDC, Genome-to-Genome Distance Calculator; HSP, high-scoring segment pairs.

^aConfidence intervals indicate inherent uncertainty in estimating DDH values from intergenomic distances based on models derived from empirical test data sets (which are always limited in size). These results are in accordance with the 16S rRNA (Fig. 1) and phylogenomic analyses as well as GGDC results.

under accession numbers FCOQ00000000 and LN998053 respectively. The type strain mt24 (= CSUR P2238 = DSM 101777) was isolated from the faeces of a 13-month-old girl from Niger.

Acknowledgements

The authors thank the Xegen Company (www.xegen.fr/) for automating the genomic annotation process. This study was funded by the Fondation Méditerranée Infection. We thank K. Griffiths for English-language review.

Conflict of Interest

None declared.

References

- [1] Salazar N, Arbolea S, Valdés L, Stanton C, Ross P, Ruiz L, et al. The human intestinal microbiome at extreme ages of life. Dietary intervention as a way to counteract alterations. *Front Genet* 2014;5:406.
- [2] Smith MI, Yatsunenkov T, Manary MJ, Trehan I, Mkakosya R, Cheng J, et al. Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. *Science* 2013;339:548–54.

- [3] Gritz EC, Bhandari V. The human neonatal gut microbiome: a brief review. *Front Pediatr* 2015;3:17. <http://dx.doi.org/10.3389/fped.2015.00017>.
- [4] Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. *PLoS Biol* 2007;5:e177.
- [5] Yatsunenkov T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature* 2012;486:222–7.
- [6] Sankar SA, Lagier JC, Pontarotti P, Raoult D, Fournier PE. The human gut microbiome, a taxonomic conundrum. *Syst Appl Microbiol* 2015;38:276–86.
- [7] Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect* 2012;18:1185–93.
- [8] De Vos P, Garrity GM, Jones D, Krieg NR, Ludwig W, Rainey FA, et al., editors. *Systematic bacteriology*. New York: Springer; 2009.
- [9] Parte AC. LPSN—list of prokaryotic names with standing in nomenclature. *Nucleic Acids Res* 2014;42:D613–6.
- [10] Viale AM, Arakaki AK, Soncini FC, Ferreyra RG. Evolutionary relationships among eubacterial groups as inferred from GroEL (chaperonin) sequence comparisons. *Int J Syst Bacteriol* 1994;44:527–33.
- [11] 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.
- [12] Wolf M, Müller T, Dandekar T, Pollack JD. Phylogeny of *Firmicutes* with special reference to *Mycoplasmata* (*Mollicutes*) as inferred from phosphoglycerate kinase amino acid sequence data. *Int J Syst Evol Microbiol* 2004;54:871–5.
- [13] Fournier PE, Lagier JC, Dubourg G, Raoult D. From culturomics to taxonomogenomics: a need to change the taxonomy of prokaryotes in clinical microbiology. *Anaerobe* 2015;36:73–8.
- [14] 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;28:237–64.
- [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] 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.
- [17] Drancourt M, Bollet C, Carlioz A, Martelin R, Gayral JP, Raoult D. 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. *J Clin Microbiol* 2000;38:3623–30.
- [18] Stark E, Wellerson R, Tetrault PA, Kossack CF. Bacterial alpha amylase paper disc tests on starch agar. *Appl Microbiol* 1953;1:236–43.
- [19] Frazier WC. A method for the detection of changes in gelatin due to bacteria. *J Infect Dis* 1926;39:302–9.
- [20] Sasser M. *Bacterial identification by gas chromatographic analysis of fatty acids methyl esters (GC-FAME)*. Newark, NY: Microbial ID; 2006.
- [21] Dione N, Sankar SA, Lagier JC, Khelaifia S, Michele C, Armstrong N, et al. Genome sequence and description of *Anaerosalibacter massiliensis* sp. nov. *New Microb New Infect* 2016;10:66–76.
- [22] Matuschek E, Brown DFJ, Kahlmeter G. Development of the EUCAST disk diffusion antimicrobial susceptibility testing method and its implementation in routine microbiology laboratories. *Clin Microbiol Infect* 2014;20:O255–66.
- [23] 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.
- [24] Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res* 2012;40:D48–53.
- [25] 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.
- [26] Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007;35:3100–8.
- [27] Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* 2004;338:1027–36.
- [28] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. *Nucleic Acids Res* 2011;39:W347–52.
- [29] 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.
- [30] Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. *Bioinformatics* 2000;16:944–5.
- [31] Carver T, Thomson N, Blesby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 2009;25:119–20.
- [32] Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004;14:1394–403.
- [33] Gouret P, Thompson JD, Pontarotti P. PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. *BMC Bioinformatics* 2009;10:298.
- [34] Gouret P, Paganini J, Dainat J, Louati D, Darbo E, Pontarotti P, et al. Integration of evolutionary biology concepts for functional annotation and automation of complex research in evolution: the multi-agent software system DAGOBAN. In: Pontarotti P, editor. *Evolutionary biology—concepts, biodiversity, macroevolution and genome evolution*. Berlin: Springer; 2011. p. 71–87.
- [35] Gouret P, Vitiello V, Balandraud N, Gilles A, Pontarotti P, Danchin EG. FIGENIX: intelligent automation of genomic annotation: expertise integration in a new software platform. *BMC Bioinformatics* 2005;6:198.
- [36] 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.
- [37] 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.
- [38] Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi M, Sentausa E, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. *Int J Syst Evol Microbiol* 2014;64:384–91.
- [39] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* 2006;33:152.
- [40] Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol* 2014;64:346–51.
- [41] Roux V, Raoult D. *Paenibacillus massiliensis* sp. nov., *Paenibacillus sanguinis* sp. nov. and *Paenibacillus timonensis* sp. nov., isolated from blood cultures. *Int J Syst Evol Microbiol* 2004;54:1049–54.
- [42] Osman S, Satomi M, Venkateswaran K. *Paenibacillus pasadenensis* sp. nov. and *Paenibacillus barengoltzii* sp. nov., isolated from a spacecraft assembly facility. *Int J Syst Evol Microbiol* 2006;56:1509–14.
- [43] Rivas R, Mateos PF, Martínez-Molina E, Velázquez E. *Paenibacillus xylanilyticus* sp. nov., an airborne xylanolytic bacterium. *Int J Syst Evol Microbiol* 2005;55:405–8.

- [44] Benardini JN, Vaishampayan PA, Schwendner P, Swanner E, Fukui Y, Osman S, et al. *Paenibacillus phoenicis* sp. nov., isolated from the Phoenix Lander assembly facility and a subsurface molybdenum mine. *Int J Syst Evol Microbiol* 2011;61:1338–43.
- [45] Ma Y, Xia Z, Liu X, Chen S. *Paenibacillus sabiniae* sp. nov., a nitrogen-fixing species isolated from the rhizosphere soils of shrubs. *Int J Syst Evol Microbiol* 2007;57:6–11.
- [46] Traiwan J, Park MH, Kim W. *Paenibacillus puldeungensis* sp. nov., isolated from a grassy sandbank. *Int J Syst Evol Microbiol* 2011;61:670–3.
- [47] Yoon JH, Oh HM, Yoon BD, Kang KH, Park YH. *Paenibacillus kribbensis* sp. nov. and *Paenibacillus terrae* sp. nov., bioflocculants for efficient harvesting of algal cells. *Int J Syst Evol Microbiol* 2003;53:295–301.