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

E. H. Seck[1], S. A. Sankar[1], S. Khelaifia[1], O. Croce[1], C. Robert[1], C. Couderc[1], F. Di Pinto[1], C. Sokhna[2,3], P.-E. Fournier[1], D. Raoult[1,4] and J.-C. Lagier[1]

1) *Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM 63, CNRS 7278, IRD 198, Inserm 1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de Médecine,* 2) *Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes IRD 198, CNRS 7278, Aix-Marseille Université, Marseille, France,* 3) *Campus Commun UCAD-IRD of Hann, Dakar, Senegal and* 4) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

We propose the main phenotypic characteristics and the complete genome sequence and annotation of *Planococcus massiliensis* strain ES2^T (= CSUR P1103 = DSM 28915), the type strain of *P. massiliensis* sp. nov., isolated from a faeces sample collected from a healthy Senegalese man. It is an aerobic, Gram-positive, moderately halophilic, motile and rod-shaped bacterium. The 3 357 017 bp long genome exhibits a G+C content of 46.0% and contains 3357 protein-coding genes and 48 RNA genes.

Keywords: Culturomics, genome, human gut microbiota, *Planococcus massiliensis*, taxonogenomics
Original Submission: 25 October 2015; **Revised Submission:** 18 December 2015; **Accepted:** 18 December 2015
Article published online: 31 December 2015

Corresponding author: J.-C. Lagier, URMITE, UMR CNRS 7278, IRD 198, INSERM U1095, Faculté de Médecine, Aix-Marseille Université, 27 Boulevard Jean Moulin, 13385 Marseille Cedex 5, France
E-mail: jean-christophe.lagier@ap-hm.fr

Introduction

Planococcus massiliensis strain ES2^T (= CSUR P1103 = DSM 28915) is the type strain of *P. massiliensis* sp. nov. This bacterium was isolated from a stool sample of a healthy Senegalese man. This isolation is part of a wider culturomics study, with an aim to maximize the culture conditions to explore the human microbiota in depth [1]. For this project, several hundred samples from healthy individuals, antibiotic-treated individuals, or people with, for example, obesity, anorexia nervosa, or malnutrition were analysed by culturomics in order to extend our knowledge of gut microbes [2]. In this case, we created media containing a high salt concentration in order to cultivate halophilic microorganisms [2]. Furthermore, the availability of genomic data for many bacterial species [3] inspired us to propose a new concept for the description of new species of bacteria by adding proteomic information obtained by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) [4] and genomic analyses [5]. This concept changes the current methods of defining a new bacterial species, which are based on the genetic, phenotypic and chemotaxonomic criteria that are not reproducible and cannot be applied to all the bacterial genus [6–8].

Here we present a summary classification and a set of features for the type strain *Planococcus massiliensis* sp. nov. strain ES2^T, together with the description of the complete genomic sequence and its annotation. These characteristics support the circumscription of the species *Planococcus massiliensis.* To our knowledge, *Planococcus massiliensis* is the first representative member of the genus of *Planococcus* isolated from a human subject. To date, 17 recognized species are representatives of *Planococcus* (*P. alkanoclasticus, P. antarcticus, P. citreus, P. columbae, P. donghaensis, P. halocryophilus, P. kocurii, P. maitriensis, P.*

http://dx.doi.org/10.1016/j.nmni.2015.12.006

maritimus, P. mcmeekini, P. okeanokoites, P. plakortidis, P. psychrophylus, P. rifietoensis, P. salinarum, P. stackebrandtii) (http://www.bacterio.net). All these species are Gram-positive, aerobic cocci that are able to grow at moderately low temperatures and high salt concentrations, and have been predominantly isolated from saline environments [9,10].

Material and Methods

Sample collection and culture conditions

Signed informed consent was obtained from each person included in the study. The study and the assent procedure were approved by the National Committee of Senegal and the ethics committee of Federative Research Institute 48 (Faculty of Medicine, Marseille, France) under agreement 09-022. The sample was obtained from a native Senegalese man living in Ndiop, a rural village in the Guinean–Sudanese zone in Senegal. The specimen was collected in sterile plastic containers, formed into aliquots and stored at −80°C until use. For each sample, pH and salinity were systematically determined with a pH meter (ThermoFisher Scientific, Saint Aubin, France) and a digital refractometer (ThermoFisher Scientific) before any analysis was performed. Then it was cultured in a liquid Colombia broth culture medium (Sigma-Aldrich, Saint-Quentin Fallavier, France) modified by adding (per litre): 5 g $MgCl_2$ $6H_2O$; 5 g $MgSO_4$ $7H_2O$; 2 g KCl; 1 g $CaCl_2$ $2H_2O$; 0.5 g NaBr; 0.5 g $NaHCO_3$; 2 g glucose; and 100 g NaCl. pH was adjusted to 7.5 with 10 M NaOH.

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

MALDI-TOF analysis of proteins was used for the identification of bacteria. Each colony was deposited in duplicate on a MALDI-TOF MSP96 target and covered with 1.5 µL of a matrix solution (saturated solution of α-cyano-4-hydroxycinnamic acid in 50% acetonitrile, 2.5% trifluoroacetic acid) to allow crystallisation of molecules. MALDI-TOF was performed using the LT Microflex spectrometer (Bruker Daltonics, Leipzig, Germany). All spectra were recorded in positive linear mode for the mass range from 2000 to 20 000 Da (parameters: ion source 1 (ISI), 20 kV; IS2, 18.5 kV lens, 7 kV). The generated spectra were then compared to the Bruker database, which was supplemented with the new species found through the culturomics project. The resulting score enabled the identification or not of tested species: a score of $\geq$2 with a validly published species enables identification at the species level, a score of $\geq$1.7 but <2 enables identification at the genus level and a score of <1.7 does not enable any identification.

After three assays, unidentified colonies were sequenced using 16S rRNA for formal identification. Isolated colonies were suspended in 200 µL distilled water for DNA extraction using an EZ1 DNA Tissue Kit (Qiagen, Venlo, Netherlands). The amplification of the 16S rRNA was performed by a standard PCR using the universal primer pair FD1 5′-AGAGTTTGATCCTGGCTCAG-3′ and RP2 5′-ACGGCTACCTTGTTACGACTT-3′ [11]. The PCR product was purified and sequenced using the Big Dye Terminator Sequencing kit v.1.1 (Perkin-Elmer, Courtaboeuf, France) with the following internal primers: 536F, 536R, 800F, 800R, 1050F and 1050R; 16S rRNA amplification and sequencing were carried out as previously described by Steven *et al.* [12]. The 16S rRNA nucleotide sequence was corrected using Chromas Pro 1.34 software (Technelysium, Tewantin, Australia), and the BLASTn searches were performed by National Center for Biotechnology Information (http://blast.ncbi.nlm.nih.gov.gate1.inist.fr/Blast.cgi). MEGA6 (Molecular Evolutionary Genetics Analysis) software [13] allowed us to construct a phylogenetic tree. Sequences alignment of the different species was performed by CLUSTALW, and calculation of the evolutionary distance was done with the Kimura two-parameter model [14,15].

Atmospheric, sporulation and microscopy tests

Growth of the strain $ES2^T$ was tested under aerobic atmosphere, in the presence of 5% CO_2, and also in anaerobic and microaerophilic atmospheres created using AnaeroGen and CampyGen respectively (ThermoFisher Scientific).

Spore formation was determined by thermal shock and observed under a microscope.

Gram staining and motility were observed by the use of the DM1000 photonic microscope (Leica Microsystems, Nanterre, France). Cell morphology was examined with a Tecnai G20 (FEI, Limeil-Brevannes, France) transmission electron microscope.

Biochemistry and antimicrobial susceptibility

Biochemical tests were realized by using the commercially available API ZYM, API 50CH and API 20 NE strips (bioMérieux, Marcy l'Étoile, France). Oxidase and catalase reactions were determined by using a BD BBL DrySlide (Becton Dickinson, Franklin Lakes, NJ, USA) according to the manufacturer's instructions.

Sensitivity to antibiotics was determined by using Sircan Discs (i2a, Montpellier, France) on Mueller-Hinton agar in a petri dish (bioMérieux). Doxycycline, rifampicin, vancomycin, nitrofurantoin, amoxicillin, erythromycin, ampicillin, ceftriaxone, ciprofloxacin, gentamicin, penicillin, trimethoprim/sulfamethoxazole, imipenem and metronidazole activity were tested on our strain.

Genomic DNA preparation

After 48 hours of culture, the bacteria were resuspended in sterile water and centrifuged at 4°C at 2000 × *g* for 20 min. Cell pellets were resuspended in 1 mL Tris/EDTA/NaCl solution (10 mM Tris/HCl (pH7.0), 10 mM EDTA (pH8.0) and 300 mM NaCl) and recentrifuged under the same conditions. The pellets were then resuspended in 200 µL of Tris-EDTA (TE) buffer and proteinase K and kept overnight at 37°C for cell lysis. DNA was purified with phenol/chloroform/isoamylalcohol (25:24:1), followed by an overnight precipitation with ethanol at −20°C. It was then resuspended in 205 µL TE buffer and quantified (155 ng/µL) by a Qubit fluorometer using the high-sensitivity kit (ThermoFisher Scientific).

Genome sequencing and assembly

Genomic DNA of *Planococcus massiliensis* 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 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 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 at 4.008 kb. No size selection was performed, and 388.3 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments, with an optimum of 634 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 35.59 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 the sequencing run were performed in a single 39-hour run at a 2 × 251 bp read length. Total information of 10.6 Gb was obtained from a 1326K/mm^2 cluster density with a cluster passing quality control filters of 99.1% (24 492 260 clusters). Within this run, the index representation for *Planococcus massiliensis* was determined to be 7.06%. The 1 481 197 paired reads were filtered according to the read qualities. These reads were trimmed, then assembled using the CLC genomics WB4 software.

Genome annotation and comparison

Open reading frames (ORFs) were predicted using Prodigal [16] with default parameters, but the predicted ORFs were

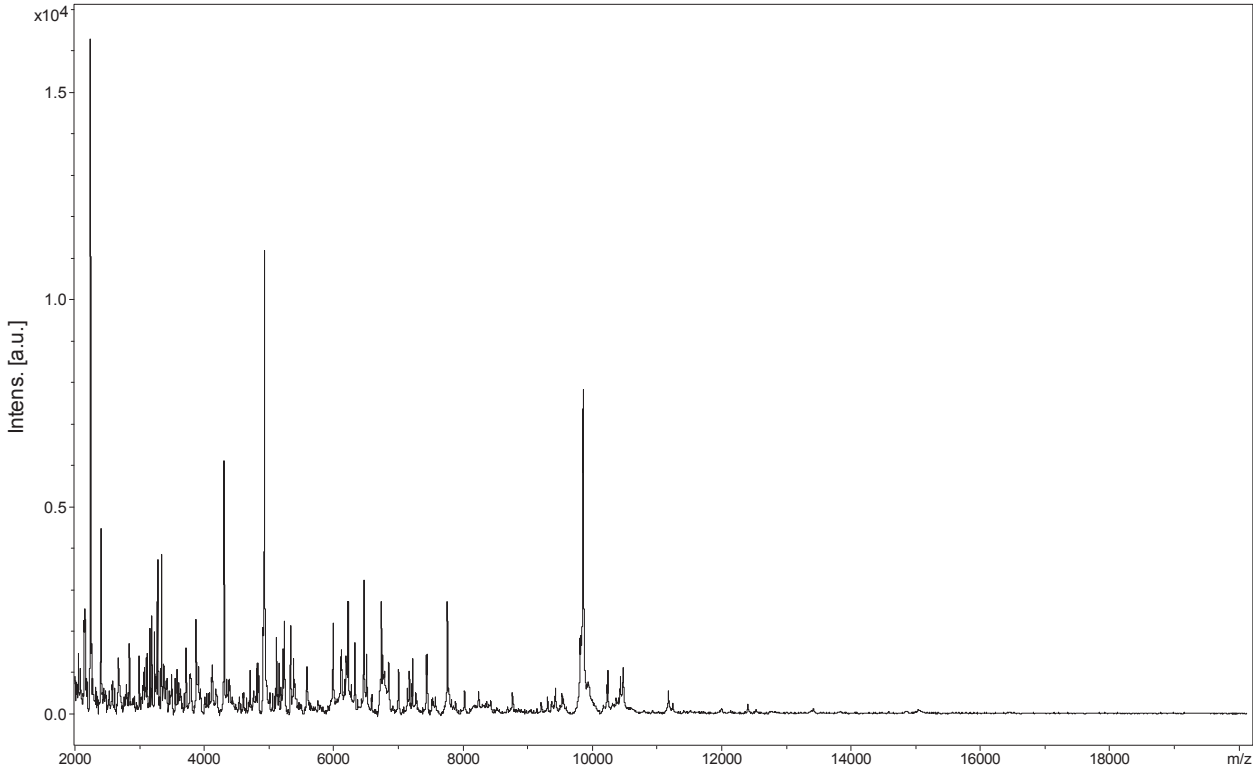

FIG. 1. Reference mass spectrum from Planococcus massiliensis strain ES2^T.

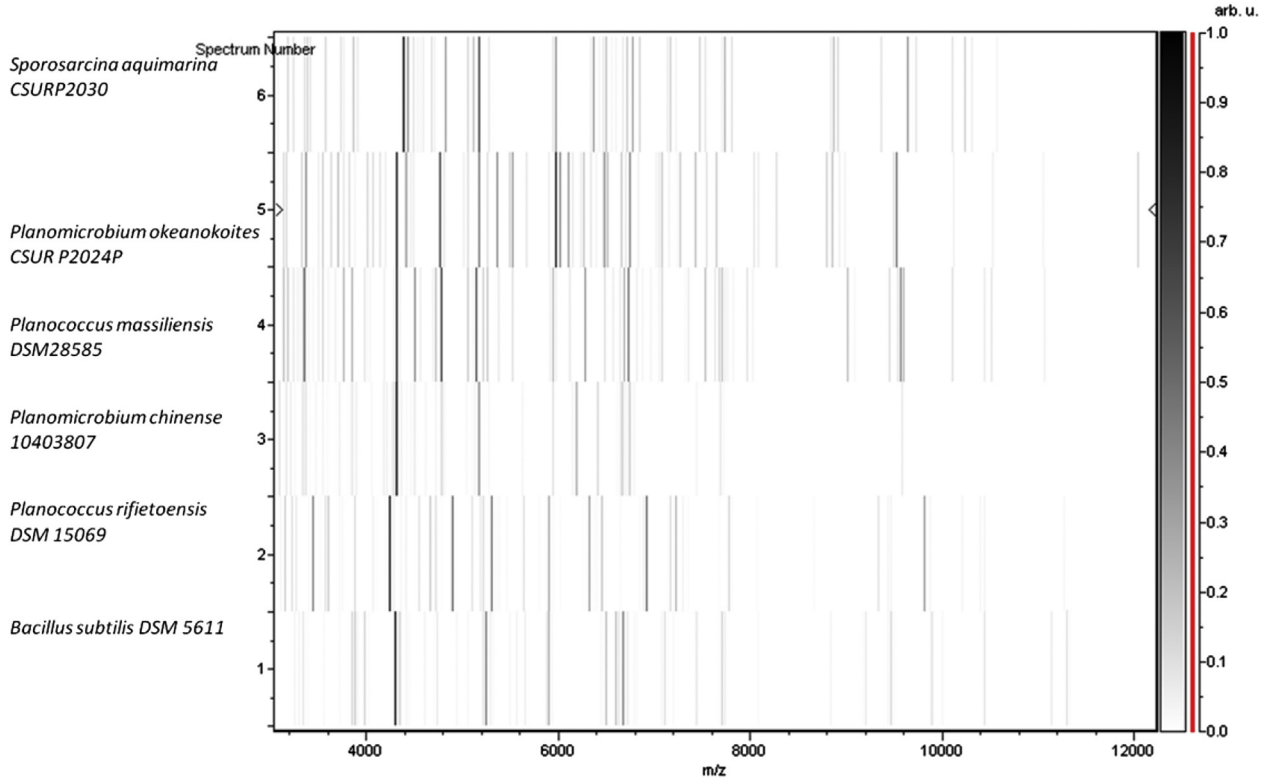

FIG. 2. Gel view comparing Planococcus massiliensis strain ES2^T to other species within genera Planomicrobium, Planococcus and Bacillus. Gel view displays raw spectra of loaded spectrum files arranged in pseudo-gel-like look. x-axis records m/z value. Left y-axis displays running spectrum number originating 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.

excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched against the GenBank database [17] and the Clusters of Orthologous Groups (COGs) database using BLASTP. The tRNAScanSE tool [18] was used to find tRNA genes, whereas ribosomal RNAs were found using RNAmmer [19] and BLASTn against the GenBank database. Lipoprotein signal peptides and the number of transmembrane helices were predicted using SignalP [20] and TMHMM [21] respectively. ORFans were identified if their BLASTP *E* value was lower than 1e-03 for alignment length greater than 80 amino acids. If alignment lengths were smaller than 80 amino acids, we used an *E* value of 1e-05. Such parameter thresholds have already been used in previous works to define ORFans. Artemis [22] was used for data management and DNA Plotter [23] for visualization of genomic features. The Mauve 2.3.1 alignment tool was used for multiple genomic sequence alignment [24]. To estimate the mean level of nucleotide sequence similarity at the genome level, we used the MAGI homemade software to calculate the average genomic identity of gene sequences (AGIOS) among compared genomes. Briefly, this software combines the Proteinortho software [25] for detecting orthologous proteins in pairwise genomic comparisons, then retrieves the corresponding genes and determines the mean percentage of nucleotide sequence identity among orthologous ORFs using the Needleman-Wunsch global alignment algorithm. Genomes from the genus *Planococcus* and closely related genera were used for the calculation of AGIOS values.

The genome of *Planococcus massiliensis* strain ES2^T (GenBank accession no. CCXS00000000) was compared to the one of *Planomicrobium glaciei* strain CHR43 (NTS) (GenBank accession no. AUYR00000000), *Planococcus halocryophilus* strain Or1 (GenBank accession no. ANBV01000000), *Planococcus donghaensis* strain MPA1U2 (NTS) (GenBank accession no. AEPB01000000) and *Bacillus subtilis* subsp. *spizizenii* strain TU-B-10 (GenBank accession no. CP002905).

To estimate the overall similarity between the genomes, the Genome-to-Genome Distance Calculator (GGDC) was used [26,27]. The system calculates the distances by comparing the genomes to obtain high-scoring segment pairs (HSP) and interfering distances from a set of formulas: 1, HSP length/total length; 2, identities/HSP length; and 3, identities/total length.

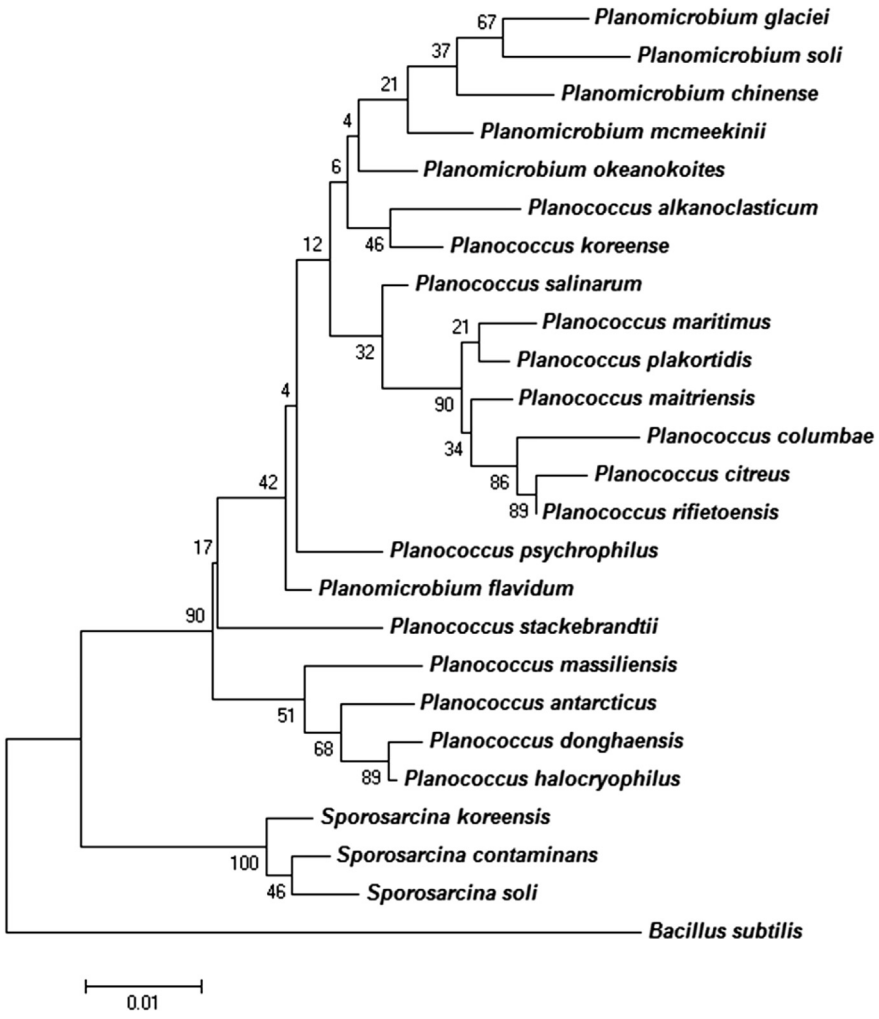

FIG. 3. Phylogenetic tree highlighting position of Planococcus massiliensis sp. nov. strain ES2^T (1536 bp) relative to other type strains within genus. Planomicrobium glaciei strain 0423 (EU036220), Planomicrobium soli strain XN13 (JQ772482), Planomicrobium chinense strain DX3-12 (AJ697862), Planomicrobium mcmeekinii strain S23F2 (AF041791), Planomicrobium okeanokoites strain ATCC 33414 (D55729), Planomicrobium alkanoclasticum strain MAE2 (AF029364), Planomicrobium koreense strain JG07 (AF144750), Planococcus salinarum strain ISL-16 (FJ765415), Planococcus maritimus strain TF-9 (AF500007), Planococcus plakortidis strain MTCC 8491 (JF775504), Planococcus maitriensis strain S1 (AJ544622), Planococcus columbae strain PgEx11 (AJ966515), Planococcus citreus strain ATCC 14404 (X62172), Planococcus rifietoensis strain M8 (AJ493659), Planomicrobium psychrophilus strain CMS 53°r (AJ314746), Planomicrobium flavidum strain ISL-41 (FJ265708), Planococcus stackebrandtii strain K22−03 (AY437845), Planococcus massiliensis (LK021122−1516 bp), Planococcus antarcticus strain CMS 26or (AJ314745), Planococcus donghaensis strain JH1 (EF079063), Planococcus halocryophilus strain Or1 (JF742665), Sporosarcina koreensis strain F73 (DQ073393), Sporosarcina contaminans strain CCUG 53915 (FN298444), Sporosarcina soli strain I80 (DQ073394). GenBank accession numbers are indicated in parentheses. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using maximum likelihood method within MEGA software. Bacillus subtilis subsp. spizizenii strain TU-B-10 (AF074970) was used as outgroup. Scale bar = 0.005% nucleotide sequence divergence.

Results

Strain identification

We did not obtain a significant MALDI-TOF score for strain ES2^T against the Bruker database, suggesting that our isolate was not a known species. Its spectrum was added to our database (Figure 1). The gel view highlighted the spectral differences with other members of the genus *Planococcus* (Figure 2). PCR-based identification of the 16S rRNA of our new isolate (GenBank accession no. LK021122) revealed 1516 bp long sequences. This indicated a 97.95% 16S rRNA sequence similarity with *Planococcus halocryophilus* (GenBank accession no. AJ314745), the phylogenetically closest validated *Planococcus* species (Figure 3). The other closest species were *P. donghaensis* (97.72%), *Planococcus glaciei* (97.06%) and *B. subtilis* (91.92%). The species *P. massiliensis, P. halocryophilus* and *P. donghaensis* shared a single cluster, whereas *P. glaciei* was present in a distant clade in the phylogenetic tree (Figure 3). This value of similarity remains lower than the 98.7% 16S rRNA gene sequence threshold recommended by Stackebrandt and Ebers to delineate a new species without carrying out DNA-DNA hybridization [6]. Thus, this bacterium was considered to be a new species called *Planococcus massiliensis* strain ES2^T sp. nov.

Physiologic and biochemical characteristics

Strain ES2^T is able to grow at temperatures between 25 and 40° C (optimum 37°C) and pH 6–9 (optimum pH 7.0–8.0), and it tolerates NaCl concentrations between 5 to 200 g/L (optimum 75 g/L). We tested the *Planococcus massiliensis* growth on 5% sheep's blood–enriched Columbia agar (bioMérieux, Marcy l'Étoile, France, at 37°C) but we observed weak growth with colonies measuring about 0.3 to 0.6 mm after 48 hours of growth. It is a motile, non-spore-forming and Gram-positive bacterium (Figure 4). Atmospheric testing demonstrated that *Planococcus massiliensis* was strictly aerobic and grew in the presence of 5% CO_2 but did not grow in an anaerobic atmosphere. Colonies that grow on our homemade culture medium were orange, circular, entire, smooth and convex, and they had a diameter of 1.0 to 2.0 mm after 48 hours. Individual cells exhibited a diameter of 0.6 to 0.9 µm and had a slightly curved form with a flagellum under electron microscopy (Figure 5).

Using API galleries, we observed positive reactions for esterase, lipase, trypsin, naphthol-AS-BI-phosphohydrolase, α-glucosidase, D-glucose, D-fructose, D-mannose, D-ribose and D-arabinose. Negative reactions were observed for leucine arylamidase, valine arylamidase, β-galactosidase, alkaline phosphatase, cystine arylamidase, α-chymotrypsin, acid phosphatase, α-galactosidase, β-glucuronidase, β-glucosidase, α-mannosidase, α-fucosidase, arginine dihydrolase, N-acetyl-β-glucosaminidase, nitrate, D-galactose, D-mannitol and urease. The strain was also oxidase positive but catalase negative. Phenotypic characteristics were compared to those of the most closely related species (Table 1).

Antimicrobial susceptibility testing demonstrate that strain ES2^T was susceptible to doxycycline, rifampicin, vancomycin, nitrofurantoin, amoxicillin, erythromycin, ampicillin,

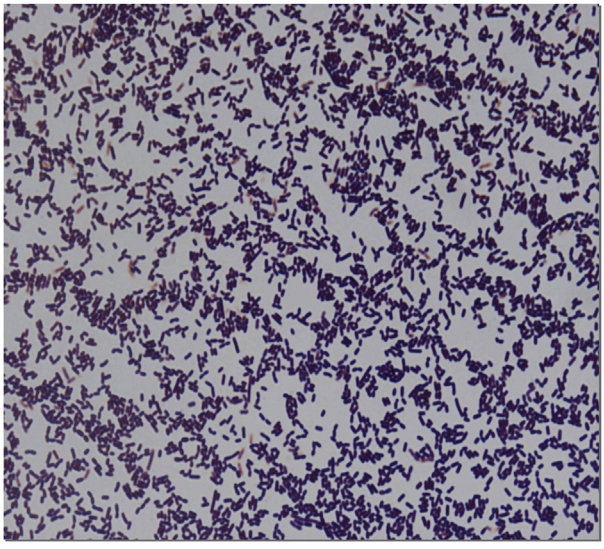

FIG. 4. Gram staining of Planococcus massiliensis strain ES2^T.

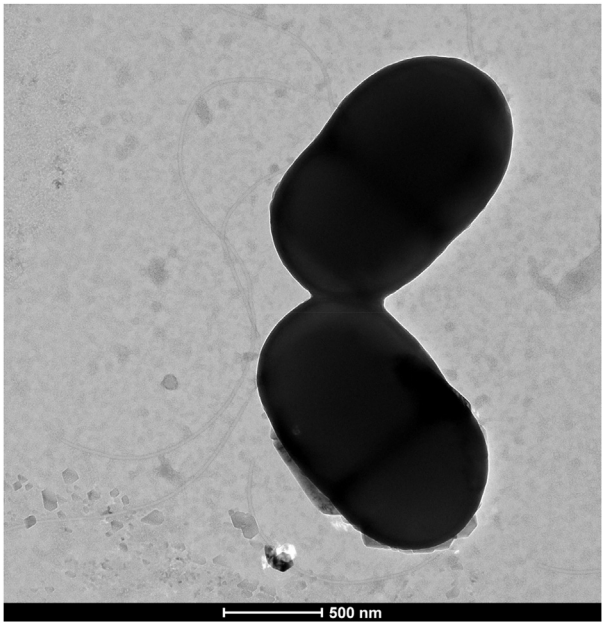

FIG. 5. Transmission electron microscopy of Planococcus massiliensis strain ES2^T. Cells are observed on Tecnai G20 transmission electron microscope operated at 200 keV. Scale bar = 500 nm.

TABLE 1. Differential characteristics of Planococcus massiliensis strain ES2^T compared to other closely related Planococcus species

Property	*P. massiliensis*	*P. okeanokoites*	*P. koreense*	*P. mcmeekinii*	*P. donghaensis*	*P. halocryophilus*	*P. glaciei*	*P. salinarum*	*P. columbae*	*P. alkanoclasticum*	*P. soli*
Cell diameter (μm)	0.6–0.9	0.4–0.8	0.4–0.8	0.6–0.9	0.8–1.2	0.8–1.2	0.4–0.8	0.4–0.8	0.8–1.0	0.4–0.8	0.8–1.0
Oxygen requirement	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic
Gram stain	+	+ to v	+ to v	+	+	+	+	+	+	+ to v	+
Salt requirement	+	+	+	+	+	+	+	–	+	+	+
Motility	+	+	+	+	–	+	+	–	+	+	+
Endospore formation	–	+	+	+	–	–	+	+	–	+	+
Indole	–	–	–	–	–	+	–	NA	–	–	–
Production of:											
Alkaline phosphatase	–	NA	NA	NA	–	NA	+	NA	NA	+	–
Catalase	–	+	+	+	NA	NA	+	NA	NA	+	+
Oxidase	+	w	–	–	+	+	–	+	–	–	–
Nitrate reductase	–	–	–	+	–	–	+	–	–	–	–
Urease	–	–	–	–	NA	–	–	NA	NA	NA	NA
Arginine dihydrolase	–	NA	NA	NA	NA	–	–	NA	NA	NA	NA
B-Galactosidase	–	NA	NA	NA	+	NA	–	–	NA	–	–
N-acetyl-β-glucosaminidase	–	NA	–	NA	+	NA	NA	NA	NA	–	–
Acid from:											
L-Arabinose	–	–	–	–	–	–	–	NA	–	–	NA
D-Ribose	+	+	–	–	+	+	+	NA	NA	–	NA
D-Mannose	+	–	–	–	–	+	NA	–	NA	–	–
D-Mannitol	–	–	–	–	–	+	NA	–	–	–	NA
D-Sucrose	–	–	–	–	+	+	NA	–	+	–	–
D-Glucose	+	–	w	+	+	+	NA	–	–	+	–
D-Fructose	+	+	–	+	–	+	NA	+	+	+	–
D-Maltose	–	–	+	w	+	+	NA	–	NA	–	–
D-Lactose	–	–	+	–	–	+	NA	–	+	–	–
Habitat	Human gut	Fermented seafood	Fermented seafood	Fermented seafood	Sea	Soil	Glacier	Coastal sediment	Coastal sediment	Coastal sediment	Soil

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

TABLE 2. Nucleotide content and gene count levels of genome

Attribute	Value	% of total[a]
Size (bp)	3 357 017	100
G+C content (bp)	1 544 227	46.0
Coding region (bp)	2 972 253	88.53
Total genes	3405	100
RNA genes	48	1.40
Protein-coding genes	3357	98.59
Genes with function prediction	2319	68.10
Genes assigned to COGs	2405	70.63
Genes with peptide signals	188	5.52
Genes with transmembrane helices	776	22.79

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

ceftriaxone, ciprofloxacin, gentamycin, penicillin, trimethoprim/sulfamethoxazole and imipenem, but it was resistant to metronidazole.

Genome properties

The GenBank Bioproject number is PRJEB6479 and consists of 192 large contigs. The draft genome of *P. massiliensis* ES2^T consists of six scaffolds with 32 contigs and generated a genome size of 3 357 017 bp with a 46.0% G+C content (Table 2, Figure 6). Of the 3405 predicted genes, 3357 are protein-coding genes and 48 are RNAs (eight genes are 5S rRNA, two are 16S rRNA, three are 23S rRNA and 35 are tRNA). A total of 2601 genes (66.90%) were assigned a putative function. A total of 75 genes (1.93%) were identified as ORFans. The remaining genes were annotated as hypothetical proteins. The properties and statistics of the genome are summarized in Table 2.

The distribution of genes into COGs functional categories is presented in Table 3.

Genome comparison

The draft genome of *Planococcus massiliensis* strain ES2^T is smaller than those of *Planomicrobium glaciei, Planococcus halocryophilus* and *Bacillus subtilis* subsp. *spizizenii* (3.35, 3.9, 3.43 and 4.21 Mb respectively) but larger than those of *Planococcus donghaensis* (3.30 Mb). The G+C content of *Planococcus massiliensis* is smaller than those of *P. glaciei* (46.0 and 47.0% respectively) but larger than those of *Planococcus halocryophilus, P. donghaensis* and *Bacillus subtilis* (39.9, 39.7 and 43.8% respectively). The gene content of *P. massiliensis* is smaller than those of *P. glaciei, P. halocryophilus* and *B. subtilis* (3405, 3967, 3429 and 4307 respectively) but larger than that of *P. donghaensis* (3251). The number of rRNA genes varied from four for *P. donghaensis,* 13 for *P. massiliensis,* 30 for *B. subtilis,* 60

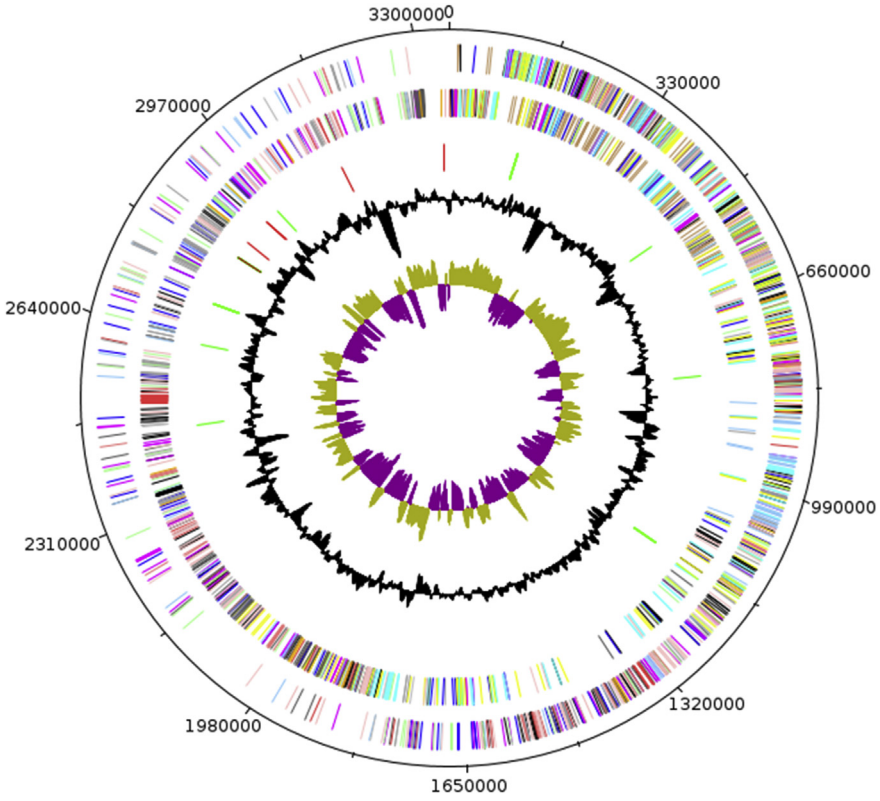

FIG. 6. Graphical circular map of genome of Planococcus massiliensis strain ES2^T. From outside to center: contigs (red/grey), COGs category of genes on forward strand (three circles), genes on forward strand (blue circle), genes on reverse strand (red circle), COGs category on reverse strand (three circles), GC content.

TABLE 3. Number of genes associated with 25 general COGs functional categories

Code	Value	% Value	Description
J	174	5.18	Translation
A	0	0	RNA processing and modification
K	248	7.38	Transcription
L	133	3.96	Replication, recombination and repair
B	1	0.02	Chromatin structure and dynamics
D	34	1.01	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	64	1.90	Defense mechanisms
T	152	4.52	Signal transduction mechanisms
M	132	3.93	Cell wall/membrane biogenesis
N	41	1.22	Cell motility
Z	1	0.02	Cytoskeleton
W	0	0	Extracellular structures
U	48	1.42	Intracellular trafficking and secretion
O	95	2.82	Posttranslational modification, protein turnover, chaperones
X	14	0.41	Phages, Prophages, Transposable elements, Plasmids
C	163	4.85	Energy production and conversion
G	216	6.84	Carbohydrate transport and metabolism
E	317	9.44	Amino acid transport and metabolism
F	80	2.38	Nucleotide transport and metabolism
H	95	2.82	Coenzyme transport and metabolism
I	131	3.90	Lipid transport and metabolism
P	169	5.03	Inorganic ion transport and metabolism
Q	85	2.53	Secondary metabolites biosynthesis, transport and catabolism
R	487	14.50	General function prediction only
S	269	8.01	Function unknown
—	1183	35.23	Not in COGs

COGs, Clusters of Orthologous Groups database.

for *P. halocryophilus* and 62 for *P. glaciei* respectively. A large number of genes assigned to COGs functional categories for amino acid transport and metabolism, transcription, carbohydrate transport and metabolism and translation were identified. Nevertheless, we observed a relative lower number of genes assigned for amino acid transport and metabolism in *P. massiliensis* compared to other species (Figure 7). The genes for RNA processing and modification, nuclear structure and extracellular structures were absent in all the genomes. Finally, the genes coding for COGs category cytoskeleton were present only in *P. massiliensis* and *P. glaciei* (Figure 7). In addition, *P. massiliensis* shared 3880, 2775, 3146 and 4099 orthologous genes with *P. glaciei, P. halocryophilus, P. donghaensis* and *B. subtilis* (Table 4). The average nucleotide sequence identity ranged from 85.84% between *P. donghaensis* and *P. halocryophilus* to 56.69% between *P. halocryophilus* and *B. subtilis* (Table 4). The genomic similarity level between strain ES2^T and closely related *Planomicrobium* and *Planococcus* species was also estimated using the GGDC (Table 5). This comparison of the genomes using

TABLE 4. Numbers of orthologous proteins shared between genomes (upper right) and AGIOS values obtained (lower left)

	PM	PD	PG	PH	BS
PM	3357[a]	2255	2275	2195	1533
PD	70.55	3146[a]	2318	2340	1544
PG	74.92	70.50	3880[a]	2256	1562
PH	67.78	85.84	67.70	2775[a]	1497
BS	59.21	58.76	59.25	56.69	4099[a]

AGIOS, average genomic identity of orthologous gene sequences; BS, *Bacillus subtilis;* PD, *Planococcus donghaensis;* PG, *Planomicrobium glaciei;* PH, *Planococcus halocryophilus;* PM, *Planococcus massiliensis.*
[a]Number of proteins per genome.

TABLE 5. Pairwise comparisons of Planomicrobium species using GGDC, formula 2 (DDH estimates based on identities/HSP length)[a]

	PM	PD	PG	PH	BS
PM	100.00%	19.1% ± 2.76	20.9% ± 2.91	18.9% ± 2.76	27.4% ± 2.54
PD		100.00%	19.2% ± 2.73	39.2% ± 3.34	27.7% ± 2.54
PG			100.00%	19.2% ± 2.73	28.6% ± 2.54
PH				100.00%	29.7% ± 2.54
BS					100.00%

BS, *Bacillus subtilis;* DDH, DNA-DNA hybridization; GGDC, Genome-to-Genome Distance Calculator; PD, *Planococcus donghaensis;* PG, *Planomicrobium glaciei;* PH, *Planococcus halocryophilus;* PM, *Planococcus massiliensis.*
[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) [27]. Distance formulas are explained in Auch *et al.* [26]. Formula 2 is recommended, particularly for draft genomes.

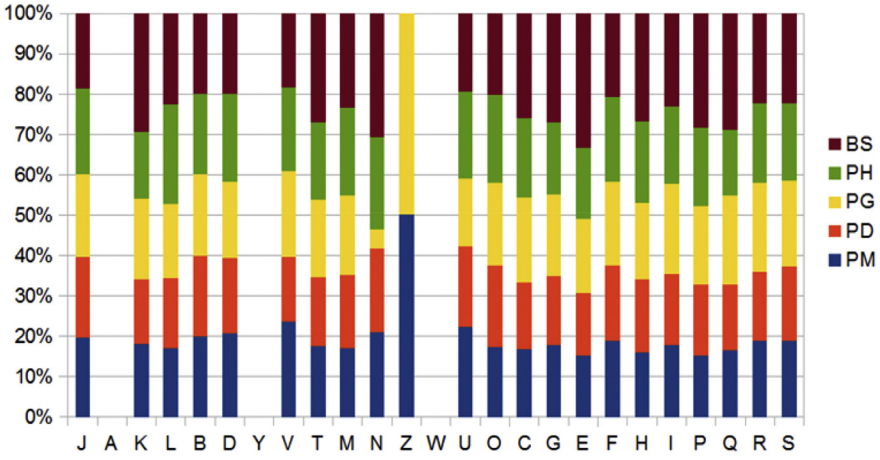

FIG. 7. Distribution of functional classes of predicted genes according to clusters of orthologous groups of proteins. BS, Bacillus subtilis; PD, Planococcus donghaensis; PG, Planomicrobium glaciei; PH, Planococcus halocryophilus; PM, Planococcus massiliensis.

GGDC revealed that *P. massiliensis* shows a slightly higher DNA-DNA hybridization (DDH) estimate with *P. glaciei* compared to those with *P. halocryophilus* and *B. subtilis* (Table 5). For *B. subtilis,* a higher DDH value was estimated with *P. halocryophilus* but did not vary much to the other genomes. These results are in accordance with the 16S rRNA (Figure 1). However, given the confidence intervals (Table 5), the DDH estimates do not show significant differences.

Conclusion

In the context of culturomics studies, several new bacterial species are isolated and then characterized. It is in this context that we studied the phenotypic and phylogenetic characteristics and conducted genomic analyses on strain ES2^T. Results allowed us to formally propose the creation of *Planococcus massiliensis* sp. nov., represented by the strain ES2^T. *P. massiliensis* represents the eighth halophilic bacterium isolated from human stool. Because the colon is not a high-salinity environment, it would be interesting to examine the role of salt or salty products as a potential source of any unusual taxon, such as halophilic.

Taxonomic and nomenclatural proposals

Description of *Planococcus massiliensis* sp. nov.

Planococcus massiliensis (mas.si.li.en'sis, L., masc. adj., *massiliensis* for Massilia, the old Roman name for Marseille, where the strain was isolated).

Strain ES2^T grows at an optimum temperature of 37°C, at pH 7.0–8.0 and at NaCl concentration of 75 g/L. Cells are Gram-positive, strictly aerobic, straight or curved rods (0.6–0.9 μm), motile, and nonendospore forming. Colonies are orange, circular, entire, smooth and convex, 1.0–2.2 mm in diameter.

P. massiliensis shows positive reactions for esterase, lipase, trypsin, naphthol-AS-BI-phosphohydrolase, α-glucosidase, D-glucose, D-fructose, D-mannose, D-ribose and D-arabinose. The strain is also oxidase positive but catalase negative.

Strain ES2^T is susceptible to doxycycline, rifampicin, vancomycin, nitrofurantoin, amoxicillin, erythromycin, ampicillin, ceftriaxone, ciprofloxacin, gentamicin, penicillin, trimethoprim/sulfamethoxazole and imipenem.

P. massiliensis ES2^T (= CSUR P1103, = DSM 28915) was isolated from a stool sample of a healthy Senegalese man. It exhibited a genome size of 3 357 017 bp with a 46.0% G+C content. The 16S rRNA sequence was deposited in GenBank under accession number LK021122, and the whole genome shotgun sequence has been deposited in GenBank under accession number CCXS00000000.

Acknowledgements

The authors thank the Xegen Company (http://www.xegen.fr/) for automating the genomic annotation process and C. Andrieu for administrative assistance. This study was funded by the Fondation Méditerranée Infection.

Conflict of Interest

None declared.

References

[1] Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 2012;18:1185–93.
[2] 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;1:237–64.
[3] Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, Krichevsky MI, et al. Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. Int J Syst Bacteriol 1987;37:463–4.
[4] Welker M, Moore ERB. Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. Syst Appl Microbiol 2011;34:2–11.
[5] Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi M, Sentausa E, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol 2011;64:384–91.
[6] Stackebrandt E. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152–5.
[7] Tindall BJ, Rosselló-Móra R, Busse HJ, Ludwig W, Kämpfer P. Notes on the characterization of prokaryote strains for taxonomic purposes. Int J Syst Evol Microbiol 2000;60:249–66.
[8] Vandamme P, Pot B, Gillis M, de Vos P, Kersters K, Swings J. Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiol Rev 1996;60:407–38.
[9] Nadia C, Mykytczuk S, Wilhelm RC, Whyte LG. *Planococcus halocryophilus* sp. nov., an extreme sub-zero species from high Arctic permafrost. Int J Syst Evol Microbiol 2012;62:1937–44.
[10] Yoon JH, Kang SJ, Lee SY, Oh KH, Oh TK. *Planococcus salinarum* sp. nov., isolated from a marine solar saltern, and emended description of the genus *Planococcus*. Int J Syst Evol Microbiol 2010;60:754–8.
[11] Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol 1991;173:697–703.
[12] Steven B, Briggs G, McKay CP, Pollard WH, Greer CW, Whyte LG. Characterization of the microbial diversity in a permafrost sample from the Canadian high Arctic using culture dependent and culture-independent methods. FEMS Microbiol Ecol 2007;59:513–23.

[13] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. c: Molecular Evolutionary Genetics Analysis, version 6.0. Mol Biol Evol 2013;30: 2725–9.
[14] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–80.
[15] Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 1980;16:111–20.
[16] 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.
[17] Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2012;40:D48–53.
[18] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955–64.
[19] Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 2007;35:3100–8.
[20] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340:783–95.
[21] Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001;305:567–80.
[22] Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. Bioinformatics 2000;16:944–5.
[23] Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 2009;25:119–20.
[24] Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004;14:1394–403.
[25] 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.
[26] 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.
[27] 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.