Microb Ecol (2016) 71:990–998
DOI 10.1007/s00248-015-0722-4

GENES AND GENOMES

Non-contiguous-Finished Genome Sequence and Description of *Paenibacillus camerounensis* sp. nov.

Mamadou Bhoye Keita[1] · Roshan Padhmanabhan[1] · Catherine Robert[1] · Eric Delaporte[2] · Didier Raoult[1,3] · Pierre-Edouard Fournier[1] · Fadi Bittar[1]

Received: 22 May 2015 / Accepted: 15 December 2015 / Published online: 29 December 2015

Abstract Strain G4^T was isolated from the stool sample of a wild gorilla (*Gorilla gorilla gorilla*) from Cameroon. It is a facultative anaerobic, Gram-negative, rod-shaped bacterium. This strain exhibits a 16S rRNA nucleotide sequence similarity of 97.48 % with *Paenibacillus typhae*, the phylogenetically closest species with standing nomenclature. Moreover, the strain G4^T presents some phenotypic differences when compared to other *Paenibacillus* species and shows a low MALDI-TOF Mass Spectrometry score that does not allow any identification. Thus, it is likely that this strain represents a new species. Here, we describe the characteristics of this organism, complete genome sequence, and annotation. The 6,933,847 bp size genome (1 chromosome but no plasmid) contains 5972 protein-coding genes and 54 RNAs genes, including 44 tRNA genes. In addition, digital DNA-DNA hybridization values for the genome of the strain G4^T against the closest *Paenibacillus* genomes range between 19.7 and 22.1, once again confirming its new status as a new species. On the basis of these polyphasic data, consisting of phenotypic and genomic analyses, we propose the creation of *Paenibacillus camerounensis* sp. nov. that contains the strain G4^T.

Keywords *Paenibacillus camerounensis* · Genome · Taxono-genomics · Culturomics

✉ Fadi Bittar
fadi.bittar@univ-amu.fr

[1] Faculté de médecine, URMITE, Aix-Marseille Université, Marseille, France

[2] IRD, University Montpellier 1, Montpellier, France

[3] King Fahad Medical Research Center, King Abdul Aziz University, Jeddah, Saudi Arabia

Abbreviations

URMITE	Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes
CSUR	Collection de Souches de l'Unité des Rickettsies
DSM	Deutsche Sammlung von Mikroorganismen
MALDI-TOF MS	Matrix-assisted laser-desorption/ ionization time-of-flight mass spectrometry
TE buffer	Tris-EDTA buffer
GGDC	Genome-to-Genome Distance Calculator
dDDH	Digital DNA-DNA hybridization

Introduction

The genus *Paenibacillus*, described by Ash et al. [1, 2] about 20 years ago, currently includes 177 species (167 validly and 10 non-validated but published species) [3]. Species of this genus are Gram-positive, negative or variable, frequently motile, and spore-forming bacteria. Many studies have described *Paenibacillus* species in various environments including soil, water, and food. Moreover, *Paenibacillus* species are rarely associated with human diseases, but they may be involved in some infections such as endocarditis, bacteremia, and wound infections [4–9].

Strain G4^T (= CSUR P208 = DSM 26182) is the type strain of *Paenibacillus camerounensis* sp. nov. This bacterium is a Gram-negative, facultative anaerobic, and indole-negative bacillus that has rounded-ends. It was isolated from the feces of western lowland gorilla as part of a culturomics study to describe the bacterial communities of the gorilla gut [10]. Indeed, the use of various culture conditions has allowed

the identification of numerous new bacterial species from gorilla fecal samples [10].

In this study, we present a summary classification, phenotypic features for *P. camerounensis* sp. nov. strain G4^T, together with the description of the complete genome sequence and its annotation. These characteristics support the circumscription of the species *P. camerounensis* [11].

Materials and Methods

Strain Isolation and Phenotypic Tests

Information about the fecal sample collection and conservation are described previously [10]. Strain G4^T was isolated in January 2012 as part of a culturomics study [10] by cultivation on a novel medium which was designed as follows: Mango fruit was crushed and lyophilized and a solution containing 12 mg of mango per ml of sterile water was prepared and filtered using 0.2 μm filters. In addition, a solution of 14 mg of agar per ml of sterile water was prepared. Using these solutions, the medium was prepared (20 ml of filtered mango solution + 80 ml of agar solution). 16S rRNA sequence was performed on this strain [10]. A phylogenetic tree was obtained using the maximum-likelihood method and Kimura 2-parameter model within the MEGA 6 software [12]. Moreover, matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) MS protein analysis was carried out using a Microflex spectrometer (Bruker Daltonics, Leipzig, Germany), and 12 distinct deposits were performed for strain G4^T from 12 isolated colonies. The 12 G4^T spectra were imported into the MALDI BioTyper software (version 2.0, Bruker) and analyzed by standard pattern matching (with default parameter settings) against 6253 bacterial spectra including 124 spectra from 68 *Paenibacillus* strains, used as reference data, in the BioTyper database. Interpretation of scores was as the following: a score ≥2 enabled the identification at the species level; a score between 1.7 and 2 enabled the identification at the genus level; and a score less than 1.7 did not enable any identification (these scores were established by the manufacturer Bruker Daltonics). Different growth temperatures (25, 30, 37, and 45 °C) were tested. Growth of the strain was tested under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems, respectively (BioMérieux, Marcy l'Etoile, France), and under aerobic conditions, with or without 5 % CO_2. API 50CH and API ZYM systems (BioMérieux) were used for carbohydrate metabolism tests and enzyme detection, respectively, as recommended by the manufacturer. The standard disc method was applied for antimicrobial susceptibility testing according to the Société Française de Microbiologie (SFM).

Genomic DNA Preparation

P. camerounensis sp. nov. strain G4^T was cultured aerobically on four Petri dishes (5 % sheep blood-enriched Columbia agar) at 37 °C. Then, the strain was collected from the Petri dishes and suspended in 3 × 500μl of TE buffer and stored at 80 °C. Five hundred microliters of this suspension was thawed, centrifuged 3 min at 10,000 rpm, and resuspended in 3 × 100μl of G2 buffer (EZ1 DNA Tissue kit, Qiagen, Courtaboeuf, France). A mechanical lysis was performed using glass powder on the Fastprep-24 device (Sample Preparation system, MP Biomedicals, USA) twice for 20 s. Then, lysozyme (2.5 μg/μl) was added and the tube was incubated at 37 °C for 30 min. Finally, the extraction was performed using the BioRobot EZ1 Advanced XL (Qiagen). The yield and the concentration were measured by the Quant-it Picogreen kit (Invitrogen, Cergy Pontoise, France) on the Genios Tecan fluorometer at 50 ng/μl.

Genome Sequencing and Assembly

A 3-kb paired end library was sequenced using the 454_Roche_Titanium. This project was loaded on a 1/4 region for each application on PTP Picotiterplate. The library was prepared from 5 μg of bacterial DNA by the DNA fragmentation on the Covaris S-Series (S2) instrument (Woburn, Massachusetts, USA) with an enrichment size at 3.2 kb. The DNA fragmentation was visualized through the Agilent 2100 BioAnalyzer on a DNA labchip 7500. The library was constructed according to the 454 GS FLX Titanium paired-end protocol (Roche). Circularization and nebulization were performed and generated a pattern with an optimum at 606 bp. Following PCR amplification through 17 cycles and double size selection, the single stranded paired-end library was quantified using the Quant-it Ribogreen kit (Invitrogen) on the Genios Tecan fluorometer at 420 pg/μL. The library concentration equivalence was calculated as 1.27E+9 molecules/μL. The library was clonally amplified with 0.5 cpb in 3 emPCR reactions and using the GSTitanium SV emPCR Kit (Lib-L) v2. The yield of the emPCR was 13.88 % between the expected ranges of 5 to 20 % and according to Roche recommendation.

Beads (790,000) for a 1/4 region per application were loaded on the GS Titanium PicoTiterPlate PTP Kit 70 × 75 and sequenced with the GS FLX Titanium Sequencing Kit XLR70 (Roche). The run was performed overnight and then analyzed on the cluster through the gsRunBrowser and Newbler assembler (Roche). A total of 236,286 passed filter wells were obtained and generated 79.84 Mb of sequences with an average length of 337 bp. The passed filter sequences were assembled using Newbler with 90 % identity and 40-bp as overlap. The final assembly identified 153 contigs (>200 bps) generating a genome size of 6.93 Mb, which corresponds to a genome coverage of 52.7×.

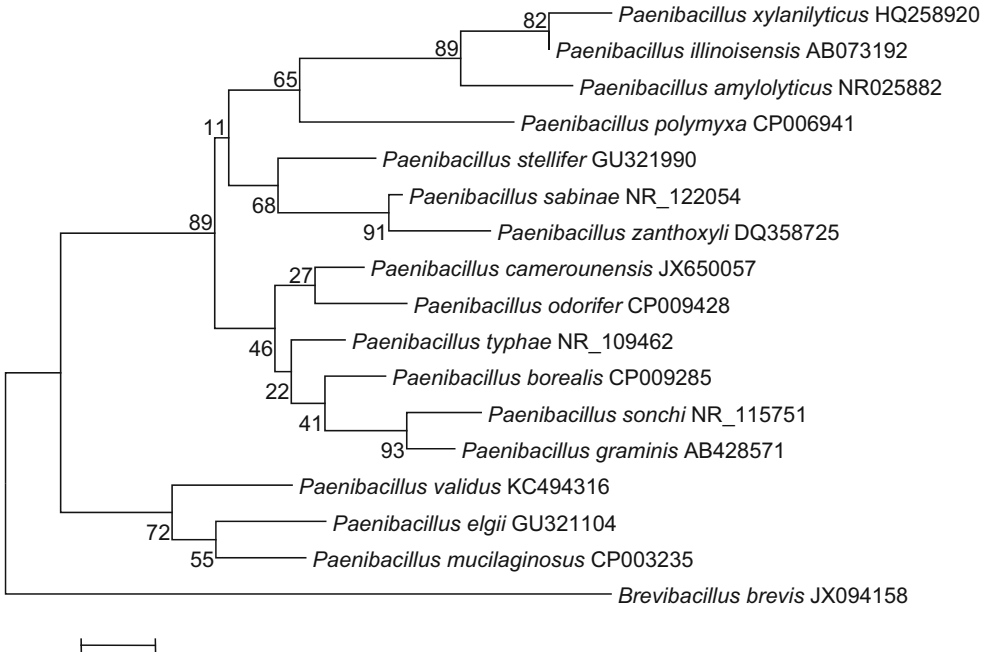

Fig. 1 Phylogenetic tree highlighting the position of *Paenibacillus camerounensis* strain G4^T relative to other type strains within the genus *Paenibacillus*. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 1000 times to generate a majority consensus tree. The *scale bar* represents a rate of substitution per site of 0.005

Genome Annotation

Open reading frames (ORFs) were predicted using Prodigal [13] with default parameters, but the predicted ORFs were excluded if they spanned a sequencing region gap. The predicted bacterial protein sequences were searched against the GenBank database [14] and the Clusters of Orthologous Groups (COG) databases using BLASTP. The tRNAScanSE tool [15] was used to find tRNA genes, whereas ribosomal RNAs were found by using RNAmmer [16] and BLASTn against the GenBank

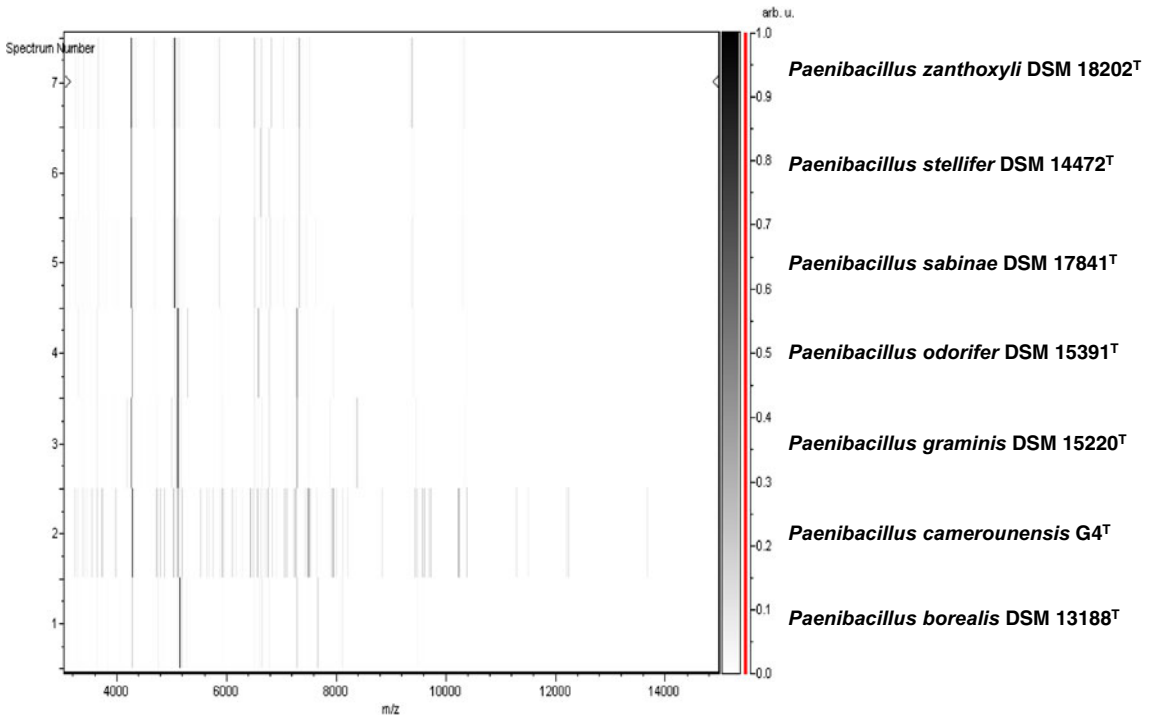

Fig. 2 Gel view comparing *Paenibacillus camerounensis* G4^T spectra with other members of the *Paenibacillus* genus. The Gel View displays the raw spectra of all loaded spectrum files arranged in a pseudo-gel-like look. The *x*-axis records the *m*/*z* value. The *left y*-axis displays the running spectrum number originating from subsequent spectra loading. The peak intensity is expressed by a *grayscale* scheme code. The *color bar* and the *right y*-axis indicate the relation between the color a peak is displayed with and the peak intensity in arbitrary units

Table 1 Differential phenotypic characteristics between *P. camerounensis* sp. nov. strain G4^T and phylogenetically close *Paenibacillus* species

Characteristic	1	2	3	4	5	6	7	8	9
Gram stain	−	+	+	+	−	+	+	+	+
Salt requirement	<5 %	na	<5 %	na	na	na	<5 %	<5 %	<5 %
Production of									
Catalase	+	+	−	+	+	+	+	+	+
Oxidase	−	−	−	−	−	−	−	−	−
Nitrate reductase	−	+	−	+	−	−	+	+	+
Indole	−	na	na	na	−	na	−	na	−
Gelatin hydrolysis	+	na	na	na	−	na	−	−	−
Utilization of									
L-Arabinose	+	+	na	+	+	+	na	na	+
D-Ribose	−	−	na	+	+	−	na	na	−
D-Xylose	+	+	−	+	+	+	−	−	+
L-Xylose	−	−	−	−	na	+	−	−	−
D-Adonitol	−	−	na	−	−	−	na	na	−
D-Galactose	+	+	na	+	+	+	na	na	+
D-Glucose	+	+	−	+	+	+	−	−	+
D-Fructose	+	+	+	+	+	+	−	−	+
D-Mannose	+	+	na	var	+	+	na	na	+
L-Sorbose	−	−	na	−	na	na	na	na	−
L-Rhamnose	−	−	na	−	−	−	na	na	+
Dulcitol	−	−	na	−	na	na	na	na	−
Inositol	−	−	na	−	−	−	na	na	−
D-Mannitol	+	+	na	−	var	−	na	na	+
D-Sorbitol	−	−	na	−	var	−	−	−	−
N-Acetylglucosamine	+	+	na	+	+	−	na	na	+
Amygdalin	+	+	na	+	+	+	na	na	+
Arbutin	+	+	na	+	−	−	na	na	+
Aesculin	+	+	na	+	+	+	na	na	+
Salicin	+	+	na	+	+	+	na	na	+
D-Cellobiose	+	+	na	+	+	+	na	na	+
D-Maltose	+	+	−	+	+	+	−	−	+
D-Lactose	+	+	na	+	+	+	−	−	+
D-Melibiose	+	+	na	+	var	+	na	na	+
D-Saccharose	+	na	na	na	+	na	na	na	na
D-Trehalose	+	+	na	+	+	+	na	na	+
Inulin	+	var	na	+	+	−	na	na	+
D-Melezitose	+	+	na	−	+	−	na	na	+
D-Raffinose	+	+	na	+	na	+	na	na	+
Starch	+	+	+	+	−	+	−	−	+
Glycogen	+	+	na	+	−	+	na	na	+
Xylitol	+	−	na	−	var	−	na	na	+
Gentiobiose	+	+	na	+	na	+	na	na	+
D-Turanose	+	+	na	+	na	+	na	na	+
D-Tagatose	−	−	na	−	var	−	na	na	−
D-Fucose	−	var	na	−	+	−	na	na	−
L-Fucose	−	−	na	var	+	−	na	na	−
D-Arabitol	−	−	na	−	+	−	na	na	−
L-Arabitol	−	−	na	−	na	−	na	na	−
Potassium gluconate	−	var	na	−	na	−	na	na	−
Habitat	Gorilla gut	Soil and plant roots	Rhizosphere soil	Plant roots and food	spruce forest humus	Food-packaging paperboard	Rhizosphere soil	Rhizosphere soil	Plant roots

Strains: 1, G4^T; 2, *P. graminis* DSM 15220^T; 3, *P. sonchi* X19-5^T; 4, *P. odorifer* DSM 15391^T; 5, *P. borealis* DSM 13188^T; 6, *P. stellifer* DSM 14472^T; 7, *P. sabinae* T27^T; 8, *P. zanthoxyli* JH29^T; 9, *P. typhae* xj7^T

var variable, +: positive result, −: negative result, *na* data not available, *w* weak positive result

database. 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. To estimate the mean level of nucleotide sequence similarity at the genome level between *P. camerounensis* sp. nov. strain G4^T and other *Paenibacillus* species, we use the Average Genomic Identity of orthologous gene Sequences (AGIOS) homemade software. Briefly, this software combines the Proteinortho software [17] for detecting orthologous proteins between genomes compared two by two, then retrieves the corresponding genes and determines the mean percentage of nucleotide sequence identity among orthologous ORFs using the Needleman-Wunsch global alignment algorithm. Moreover, we used the Genome-to-Genome Distance Calculator (GGDC) web server available at (http://ggdc.dsmz.de) to estimate the overall similarity among the compared genomes and to replace the wet-lab DNA-DNA hybridization (DDH) by a digital DDH (dDDH) [18, 19]. GGDC 2.0 BLAST+ was chosen as an alignment method and the recommended formula 2 was taken into account to interpret the results.

Strain and Sequences Deposition

Strain G4T was deposited in two microbial culture collections; the German collection of microorganisms (Deutsche Sammlung von Mikroorganismen, DSM) under the accession number DSM 26182 and the French culture collection (Collection de Souches de l'Unité des Rickettsies, CSUR) under the accession number CSUR P208. The 16S rRNA and genome sequences are available in GenBank database under accession numbers JX650057 and CCDG000000000, respectively.

Results and Discussion

Classification and Phenotypic Features

Strain G4^T had a 97.48 % 16S rRNA nucleotide sequence similarity with *Paenibacillus typhae*, the phylogenetically closest validly published *Paenibacillus* species (Fig. 1), when it was compared against the NCBI database and Ribosomal Database Project (RDP). This value was lower than the percentage of 16S rRNA gene sequence threshold recommended by Meier-Kolthoff et al. for *Firmicutes* to delineate a new species without carrying out DNA-DNA hybridization with maximum error probability of 0.01 % [20]. Moreover, for strain G4^T, a poor MALDI-TOF-MS score (<1.4) was obtained that did not allow any identification, suggesting it was not a member of any known species. We added the spectrum from strain G4^T to our database. Spectrum differences with other *Paenibacillus* species are presented in Fig. 2.

Among the different growth temperatures tested, the strain G4^T grew at two temperatures (25 and 37 °C), but the optimal growth occurred at 37 °C. Colonies were 1–2.5 mm in diameter on Columbia agar, appearing as a brown color. Growth was achieved under aerobic (with and without CO_2), microaerophilic, and anaerobic conditions. Gram staining showed Gram-negative

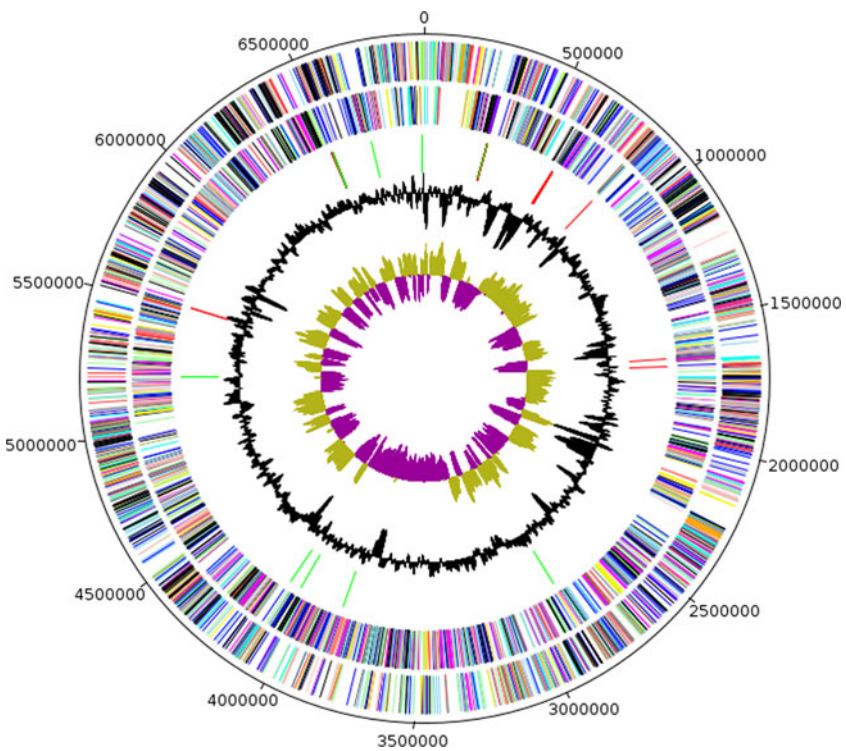

Fig. 3 Graphical circular map of the chromosome. *From outside to the center*: Genes on the forward strand colored by COG categories (only genes assigned to COG), genes on the reverse strand colored by COG categories (only gene assigned to COG), RNA genes (tRNAs *green*, rRNAs *red*), G+C content and GC skew. *Purple* and *olive* indicating negative and positive values, respectively

Table 2 Number of genes associated with the 25 general COG functional categories

Code	Value	% of total[a]	Description
J	189	3.17	Translation
A	0	0.00	RNA processing and modification
K	433	7.26	Transcription
L	134	2.25	Replication, recombination and repair
B	0	0.00	Chromatin structure and dynamics
D	32	0.54	Cell cycle control, mitosis and meiosis
Y	0	0.00	Nuclear structure
V	140	2.35	Defense mechanisms
T	250	4.19	Signal transduction mechanisms
M	207	3.47	Cell wall/membrane biogenesis
N	31	0.52	Cell motility
Z	0	0.00	Cytoskeleton
W	0	0.00	Extracellular structures
U	29	0.49	Intracellular trafficking and secretion
O	112	1.88	Posttranslational modification, protein turnover, chaperones
C	163	2.73	Energy production and conversion
G	555	9.30	Carbohydrate transport and metabolism
E	265	4.44	Amino acid transport and metabolism
F	96	1.61	Nucleotide transport and metabolism
H	116	1.94	Coenzyme transport and metabolism
I	60	1.01	Lipid transport and metabolism
P	252	4.22	Inorganic ion transport and metabolism
Q	40	0.67	Secondary metabolites biosynthesis, transport and catabolism
R	473	7.93	General function prediction only
S	379	6.35	Function unknown
–	535	8.96	Not in COGs

[a] The total is based on the total number of protein coding genes in the annotated genome

bacilli. A motility test gave a positive result. The strain grown on agar sporulate and the rods have a length of about 14 μm and a diameter of about 0.73 μm, as determined by negative staining transmission electron microscopy.

Strain G4^T exhibited catalase activity but not oxidase activity. Using API 50CH system, after 24 h of incubation at 37 °C, a positive reaction was observed for glycerol, methyl-βD-xylopyranoside, D-mannose, amygdalin, L-arabinose, D-cellobiose, D-lactose, xylitol, D-xylose, D-glucose, inulin, D-melezitose, glycogen, D-mannitol, D-galactose, *N*-acetylglucosamine, arbutin, aesculin, gentiobiose, D-turanose, D-maltose, D-saccharose, D-trehalose, salicin, D-melibiose, D-raffinose, D-fructose, and hydrolysis of starch. By contrast, negative reactions were observed for D- arabinose, erythritol, L-xylose, D-adonitol, L-rhamnose, dulcitol, inositol, D-sorbitol, D-tagatose, potassium gluconate, potassium 2-cetogluconate, D-ribose,

Table 3 Genomic comparison (sequence size and C+G contents) of *P. camerounensis* sp. Nov., strain G4^T with seven other species of the genus *Paenibacillus*

Species	Strain	Genome accession number	Genome size (Mb)	G+C content
Paenibacillus camerounensis	G4^T	CCDG000000000	6.93	51.40
Paenibacillus graminis	DSM 15220^T	CP009287	7.17	50.60
Paenibacillus sonchi	X19-5^T	AJTY00000000	7.51	50.40
Paenibacillus odorifer	DSM 15391^T	CP009428	6.81	44.20
Paenibacillus borealis	DSM 13188^T	CP009285	8.16	51.40
Paenibacillus stellifer	DSM 14472^T	CP009286	5.66	53.50
Paenibacillus sabinae	T27^T	CP004078	5.27	52.60
Paenibacillus zanthoxyli	JH29^T	ASSD00000000	5.05	50.90

Table 4 Numbers of orthologous genes shared between genomes (lower left triangle), average percentage similarity of nucleotides corresponding to orthologous proteins shared between genomes (upper right triangle)

	P. camerounensis	*P. sonchi*	*P. zanthoxyli*	*P. sabinae*	*P. borealis*	*P. stellifer*	*P. graminis*	*P. odorifer*
P. camerounensis	*5972*	74.16	69.35	69.92	75.58	69.21	74.93	70.90
P. sonchi	3445	*6705*	69.53	69.84	75.67	68.76	91.06	71.29
P. zanthoxyli	2494	2464	*4907*	81.38	69.27	73.56	69.91	68.09
P. sabinae	2851	2696	2800	*4865*	69.87	74.7	70.46	68.25
P. borealis	4016	3724	2655	3014	*6967*	69.04	76.64	71.72
P. stellifer	2956	2788	2626	2984	3152	*5161*	69.25	66.79
P. graminis	3743	3942	2611	2944	4042	3067	*6211*	72.06
P. odorifer	3664	3357	2456	2796	3867	2900	3683	*5960*

Italicized numbers indicate numbers of proteins per genome

potassium 5-cetogluconate, methyl- αD-mannopyranoside, and methyl-αD-glucopyranoside. In assays with API ZYM, positive reactions were observed for esterase (C4), esterase lipase (C8), alkaline phosphatase, α-glucosidase, leucine arylamidase, and acid phosphatase activities, but negative reactions were observed for lipase (C14), trypsin, α-chymotrypsin, naphthyl-AS-BI-phosphohydrolase, β-glucuronidase, cystine arylamidase, valine arylamidase, glycine arylamidase, α-galactosidase, α-mannosidase, α-fucosidase, *N*-acetyl-β-glucosaminidase, and β-glucosidase. The urease and esculin reactions were positive, but nitrate reduction and indole production were negative. *P. camerounensis* is susceptible to amoxicillin-clavulanic acid, penicillin, gentamycin 15, gentamycin 500, ciprofloxacin, ceftriaxone, imipenem, nitrofurantoin, amoxicillin, erythromycin, doxycycline, rifampicin, and vancomycin, but resistant to trimethoprim/sulfamethoxazole and metronidazole.

When compared to other *Paenibacillus* species [21–27], *P. camerounensis* sp. nov. strain G4^T exhibited the phenotypic differences detailed in Table 1.

Genome Sequencing Information and Genome Properties

On the basis of phenotypic characteristics and MALDI-TOF results of this strain and because of the low16S rRNA similarity to other members of the genus *Paenibacillus*, it is likely that the strain represents a new species and thus it was chosen for genome sequencing. It was the 45th genome of a *Paenibacillus* species (Genomes Online Database) and the first genome of *P. camerounensis* sp. nov.

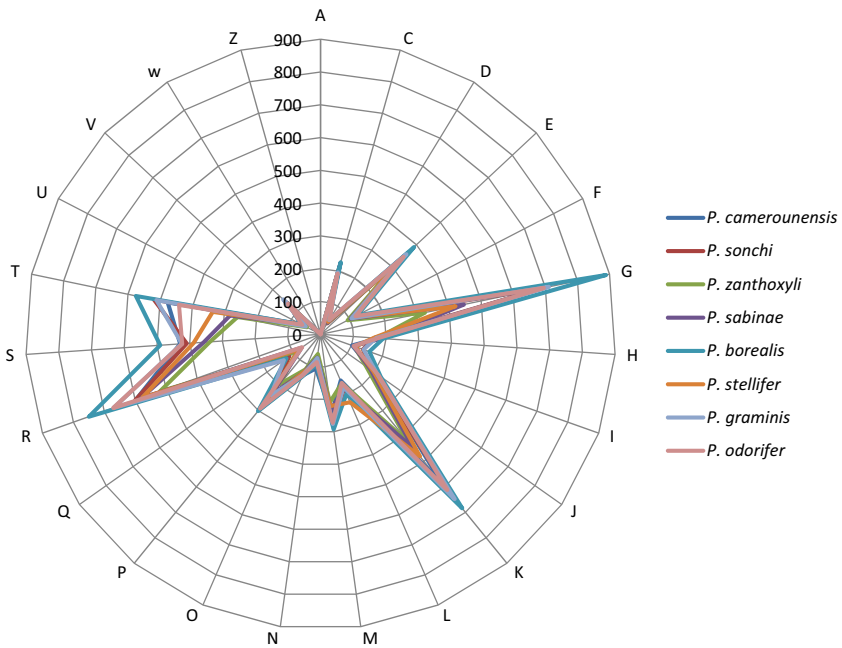

Fig. 4 Distribution of predicted genes of *P. camerounensis* and seven other *Paenibacillus* species into COG categories

The genome is 6,933,847 bp long (one chromosome, but no plasmid) (Fig. 3) with a 51.4 % G+C content. It is composed of 153 contigs. Of the 6022 predicted genes, 5972 were protein-coding genes, 54 were RNAs (one gene is 16S rRNA, one gene is 23S rRNA, eight are 5S rRNA, and 44 genes whose two pseudogenes of tRNA) and 133 (2.22 %) were annotated as peptide signals. A total of 4491 genes (75.25 %) were assigned to COGs, Genes (3956) (66.8 %) with function prediction and 1750 genes (29.32 %) as transmembrane helices. In addition, 1418 genes were assigned as hypothetical proteins and the number of Orfans found was 1406. The distribution of genes into COGs functional categories is presented in Table 2.

Comparison with Other *Paenibacillus* Species Genomes

The genome of *P. camerounensis* strain G4^T was compared to those of seven close *Paenibacillus* species (Table 3). The draft genome of *P. camerounensis* is larger in size than those of *Paenibacillus odorifer*, *Paenibacillus stellifer*, *Paenibacillus sabinae*, and *Paenibacillus zanthoxyli* (6.93 vs 6.81, 5.66, 5.27, and 5.05 Mb, respectively), but smaller in size than that of *Paenibacillus graminis*, *Paenibacillus sonchi*, and *Paenibacillus borealis* (6.93 vs 7.17, 7.51, and 8.16 Mb). *P. camerounensis* has a higher G+C content than those observed in *P. graminis*, *P. sonchi*, *P. odorifer*, and *P. zanthoxyli* (51.40 vs 50.60 %, 50.40, 44.20, and 50.90 %, respectively) but lower than those of *P. stellifer* and *P. sabinae* (51.40 vs 53.50 and 52.60 %, respectively) and equal to that of *P. borealis* (Table 3). The protein content of *P. camerounensis* is lower than those of *P. sonchi*, *P. borealis*, and *P. graminis* (5972 vs 6705, 6967, and 6211, respectively) but higher than those of *P. zanthoxyli*, *P. sabinae*, *P. stellifer*, and *P. odorifer* (5972 vs 4907, 4865, 5161, and 5960, respectively) (Table 4). The distribution of genes into COG categories was similar in all the six compared genomes (Fig. 4). In addition, *P. camerounensis* shares 3445, 2494, 2851, 4016, 2956, 3743, and 3664 orthologous genes with *P. sonchi*, *P. zanthoxyli*, *P. sabinae*, *P. borealis*, *P. stellifer*, *P. graminis*, and *P. odorifer*, respectively (Table 4). Based on the analysis of MAGi, the Average Genomic Identity of Orthologus Gene Sequence [AGIOS] ranged from 66.79 to 91.06 % among *Paenibacillus* species. The range of AGIOS calculated using MAGi varies from 69.21 to 75.58 between *P. camerounensis* and other compared *Paenibacillus* species. Strain G4^T is closer to *P. borealis* with 75.58 % genomic identity, with over 4016 orthologus genes shared between them. dDDH estimation of the strain G4^T against the compared genomes ranged between 19.7 and 22.1. These values are very low and below the cutoff of 70 %, thus confirming again the new species status of the strain G4^T. Tables 3 and 4 summarize the number of orthologous genes and the average percentage of nucleotide sequence identity between the different genomes studied.

Conclusions

On the basis of phenotypic characteristics (Table 1), phylogenetic position (Fig. 1), MALDI-TOF analyses, genomic analyses (taxonogenomics) (Tables 3 and 4), and GGDC results, we formally propose the creation of *P. camerounensis* (ca.me.rou.ne'n.sis. L. gen. masc. n. *camerounensis* of Cameroun the French name of Cameroon where the gorilla fecal sample was collected) sp. nov. that contains the strain G4^T.

P. camerounensis is a facultative anaerobic, rod-shaped, endospore-forming, motile, and Gram-negative bacterium. Optimal growth occurs at 37 °C. Bacterial cell has a diameter of 0.73 μm and a length of 14 μm. Colonies are brown and 1 to 2.5 mm in diameter on blood-enriched Columbia agar. The G+C content of the genome is 51.4 %. The GenBank accession numbers for 16S rRNA and genome sequences are JX650057 and CCDG000000000, respectively. The type strain G4^T (= CSUR P208 = DSM 26182) was isolated from the fecal sample of a western lowland gorilla from Cameroon.

Acknowledgments Fadi Bittar was supported by a Chair of Excellence IRD provided by the Institut de Recherche pour le Développement / Méditerranée-Infection foundation. Mamadou Bhoye Keita was funded by the Méditerranée-Infection foundation. The authors thank the Xegen company for automating the genome annotation process.

References

1. Ash C, Priest FG, Collins MD (1993) Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Proposal for the creation of a new genus *Paenibacillus*. Antonie Van Leeuwenhoek 64:253–260
2. Trüper HG (2005) The type species of the genus *Paenibacillus* Ash et al. 1994 is *Paenibacillus polymyxa*. Opinion 77. Judicial Commission of the International Committee on Systematics of Prokaryotes. Int J Syst Evol Microbiol 55:513
3. Abstract for the genus *Paenibacillus*. NamesforLife, LLC. Retrieved April 25, 2015
4. Glaeser SP, Falsen E, Busse HJ, Kämpfer P (2013) *Paenibacillus vulneris* sp. nov., isolated from a necrotic wound. Int J Syst Evol Microbiol 63:777–782
5. Anikpeh YF, Keller P, Bloemberg GV, Grünenfelder J, Zinkernagel AS (2010) Spacecraft bacterium, *Paenibacillus pasadenensis*, causing wound infection in humans. BMJ Case Rep. doi:10.1136/bcr.06.2010.3058
6. Rieg S, Martin Bauer T, Peyerl-Hoffmann G, Held J, Ritter W, Wagner D, Kern WV, Serr A (2010) *Paenibacillus larvae* bacteremia in injection drug users. Emerg Infect Dis 16:487–489
7. Ouyang J, Pei Z, Lutwick L, Dalal S, Yang L, Cassai N, Sandhu K, Hanna B, Wieczorek RL, Bluth M, Pincus MR (2008) Case report: *Paenibacillus thiaminolyticus*: a new cause of human infection, inducing bacteremia in a patient on hemodialysis. Ann Clin Lab Sci 38:393–400

8. Teng JL, Woo PC, Leung KW, Lau SK, Wong MK, Yuen KY (2003) Pseudobacteraemia in a patient with neutropenic fever caused by a novel paenibacillus species: *Paenibacillus hongkongensis* sp. nov. Mol Pathol 56:29–35
9. Ferrand J, Hadou T, Selton-Suty C, Goehringer F, Sadoul N, Alauzet C, Lozniewski A (2013) Cardiac device-related endocarditis caused by *Paenibacillus glucanolyticus*. J Clin Microbiol 51:3439–3442
10. Bittar F, Keita MB, Lagier JC, Peeters M, Delaporte E, Raoult D (2014) *Gorilla gorilla gorilla* gut: a potential reservoir of pathogenic bacteria as revealed using culturomics and molecular tools. Sci Rep 4:7174
11. Sentausa E, Fournier PE (2013) Advantages and limitations of genomics in prokaryotic taxonomy. Clin Microbiol Infect 19:790–795
12. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739
13. Prodigal. http://prodigal.ornl.gov
14. GenBank database. http://www.ncbi.nlm.nih.gov/genbank
15. Lowe TM, Eddy SR (1997) t-RNAscan-SE: a program for improved detection of transfer RNA gene in genomic sequence. Nucleic Acids Res 25:955–964
16. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108
17. Lechner M, Findeib S, Steiner L, Marz M, Stadler PF, Prohaska SJ (2011) Proteinortho: detection of (Co-)orthologs in large-scale analysis. BMC Bioinforma 12:124
18. Auch AF, von Jan M, Klenk HP, Göker M (2010) Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci 2: 117–134
19. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinforma 14:60
20. Meier-Kolthoff JP, Göker M, Spröer C, Klenk HP (2013) When should a DDH experiment be mandatory in microbial taxonomy? Arch Microbiol 6:413–418
21. Berge O, Guinebretière MH, Achouak W, Normand P, Heulin T (2002) *Paenibacillus graminis* sp. nov. and *Paenibacillus odorifer* sp. nov., isolated from plant roots, soil and food. Int J Syst Evol Microbiol 52(Pt 2):607–616
22. Hong YY, Ma YC, Zhou YG, Gao F, Liu HC, Chen SF (2009) *Paenibacillus sonchi* sp. nov., a nitrogen-fixing species isolated from the rhizosphere of *Sonchus oleraceus*. Int J Syst Evol Microbiol 59(Pt 11):2656–2661
23. Elo S, Suominen I, Kämpfer P, Juhanoja J, Salkinoja-Salonen M, Haahtela K (2001) *Paenibacillus borealis* sp. nov., a nitrogen fixing species isolated from spruce forest humus in Finland. Int J Syst Evol Microbiol 51(Pt 2):535–545
24. Kong BH, Liu QF, Liu M, Liu Y, Liu L, Li CL, Yu R, Li YH (2013) *Paenibacillus typhae* sp. nov., isolated from roots of *Typha angustifolia* L. Int J Syst Evol Microbiol 63(Pt 3):1037–1044
25. Suominen I, Spröer C, Kämpfer P, Rainey FA, Lounatmaa K, Salkinoja-Salonen M (2003) *Paenibacillus stellifer* sp. nov., a cyclodextrin-producing species isolated from paperboard. Int J Syst Evol Microbiol 53(Pt 5):1369–1374
26. Ma Y, Xia Z, Liu X, Chen S (2007) *Paenibacillus sabinae* sp. nov., a nitrogen-fixing species isolated from the rhizosphere soils of shrubs. Int J Syst Evol Microbiol 57(Pt 1):6–11
27. Ma Y, Zhang J, Chen S (2007) *Paenibacillus zanthoxyli* sp. nov., a novel nitrogen-fixing species isolated from the rhizosphere of *Zanthoxylum simulans*. Int J Syst Evol Microbiol 57(Pt 4): 873–877