Noncontiguous finished genome sequence and description of *Kallipyga gabonensis* sp. nov.

G. Mourembou[1,2], J. Rathored[1], J. B. Lekana-Douki[3,4], A. Ndjoyi-Mbiguino[5], F. Fenollar[1], C. Michelle[1], P.-E. Fournier[1], D. Raoult[1,6] and J.-C. Lagier[1]

1) *Aix Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, INSERM 1095, Marseille, France,* 2) *Ecole Doctorale Régionale d'Afrique Centrale* 3) *Unité de Parasitologie Médicale (UPARAM), CIRMF, Franceville, Gabon,* 4) *Département de Parasitologie Mycologie et de Médecine Tropicale* 5) *Département de Microbiologie, Laboratoire national de référence IST/sida, Faculté de Médecine, Université des Sciences de la Santé, Libreville, Gabon and* 6) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Taxonogenomics coupled with culturomics promotes the isolation and characterization of bacteria. *Kallipyga gabonensis* sp. nov. strain GM4 is a strictly anaerobic, Gram-positive, and non motile coccus isolated from the stool of a Gabonese male teenager. The genome is 1,621,211 bp long with 50.01% G+C content and two scaffolds. Of the 1,536 predicted genes, 1,475 were protein-coding genes and 61 were RNA genes. A total of 931 genes were assigned a putative function, and 79 genes were identified as ORFans.
New Microbes and New Infections © 2015 The Authors. Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases.

Keywords: Culturomics, Gabon, genome, *Kallipyga gabonensis*sp. nov., taxonogenomics
Original Submission: 5 October 2015; **Revised Submission:** 4 November 2015; **Accepted:** 5 November 2015
Article published online: 14 November 2015

Corresponding author: J.-C. Lagier, Aix Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, INSERM 1095, Marseille, France
E-mail: jclagier@yahoo.fr

Introduction

Bacteraemia represents an important cause of infectious diseases throughout the world [1]. It is mainly the consequence of the destruction of the gastrointestinal barrier by some factor, principally infectious factors [2]. In 1997 Wilairatana *et al.* [2] reported that gastrointestinal permeability is increased during *Plasmodium falciparum* malaria but reverts to normal during convalescence. This knowledge highlighted the need to perform an exhaustive study of the microorganisms existing in the human gut. Overall, 1 g of human stool contains between 10^{11} and 10^{12} bacteria [3]; unfortunately, only approximately 2000 different bacteria are known of the human microbiota. This number remains insufficient as a result of the lack of comprehensive culture methods. To enhance the isolation of bacteria, our laboratory performed an efficient method called culturomics [4]. Since then, several new bacteria, including new genera and species, have been reported in the human gut [5–8]. To describe these bacteria, 16S ribosomal RNA sequencing coupled with a taxonogenomic strategy was recently developed. This strategy is based on a systematic comparison of genomic and phenotypic characteristics, especially the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) spectrum, with the phylogenetically closest species existing in databases [5–8].

Using the culturomics method, a new bacterium was isolated from a healthy Gabonese male teenager. The strain exhibited 97.9% of similarity with the 16S nucleotide sequence of *Kallipyga massiliensis* (JN837487). However, this value was lower than the 98.7% 16S rRNA gene sequence threshold recommended by Stackebrandt and Ebers [9] to delineate a new species without carrying out DNA-DNA hybridization.

Here we present methods of isolation and characterization of *Kallipyga gabonensis* sp. nov. strain GM4 (= CSUR P1915 = DSM 100575), followed by its classification and a set of

http://dx.doi.org/10.1016/j.nmni.2015.11.007

features including a description of the complete genome sequencing and annotation.

Material and Methods

Ethics and sample collection

After receiving signed informed consent, a stool specimen was collected at Lebamba, Gabon, in January 2015. This study had been previously approved by the National Ethic Committee of Gabon and IFR48, Marseille, France (0023/2013/SG/CNE and 09-022, respectively). The specimen was from a healthy Gabonese male 16-year-old (body mass index, 19.03 kg/m^2) who belonged to the Nzebi tribe of Gabon. The stool sample was stored at −80°C until it was sent to URMITE (Marseille, France) for analysis.

Strain isolation

The stool sample was diluted in phosphate-buffered saline (Life Technologies, Carlsbad, CA, USA) in April 2015. The inoculum was obtained and preincubated in anaerobic conditions at 37°C in a culture bottle containing a blood-enriched Columbia agar liquid medium (bioMérieux, Marcy l'Étoile, France). After pre-incubation, 100 µL of cultured media was recovered, diluted and incubated 24 or 48 hours on a 5% sheep's blood–enriched Columbia agar solid medium (bioMérieux) at 37°C in the anaerobic conditions of the generator (GENbag anaer; bioMérieux). The colonies of bacteria obtained were isolated on 5% sheep's blood–enriched Columbia agar (bioMérieux). They were then identified using MALDI-TOF and 16S rRNA sequencing [5–8].

MALDI-TOF and 16S rRNA identification

In order to identify bacteria, MALDI-TOF was used following the same protocol as previously described [5–8]. Identification of bacteria continued with a 16S rRNA standard PCR coupled with sequencing. The apparatuses used were GeneAmp PCR System 2720 thermal cyclers (Applied Biosystems, Foster City, CA, USA) and the ABI Prism 3130xl Genetic Analyzer capillary sequencer (Applied Biosystems), respectively [10]. The 16S rRNA nucleotide sequence obtained after amplification and sequencing was corrected using Chromas Pro 1.34 software (Technelysium, Tewantin, Australia). Then a BLASTN was systematically performed in the online PubMed National Center for Biotechnology Information (NCBI) database (http://blast.ncbi.nlm.nih.gov.gate1.inist.fr/Blast.cgi). The spectrum of *K. gabonensis* strain GM4 was entered in the Bruker database, and its 16S rRNA sequence was deposited in GenBank under accession number LN849790. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using the maximum-likelihood method within MEGA software [5–8].

Growth conditions

The growth temperature of *K. gabonensis* strain GM4 was assessed by cultivating this bacterium on 5% sheep's blood–enriched Colombia agar (bioMérieux) in anaerobic conditions. Four growth temperatures were tested: 28, 37, 45 and 55°C. The GENbag anaer and GENbag microaer systems (bioMérieux) were used to assess the ability of the bacterium to grow anaerobically or in microaerophilic conditions, respectively, at 37°C. The limit of salinity's acceptance of *K. gabonensis* strain GM4 was tested using 0, 5, 15 and 45% NaCl concentrations. Accordingly, three pHs were tested: 5, 7 and 8.5.

Biochemical, sporulation and motility assays

API ZYM, API 20A and API 50CH (bioMérieux) gallery systems were used to perform biochemical assays. The sporulation test was done with a thermic shock (80°C during 10 minutes). To evaluate the motility of *K. gabonensis* strain GM4, fresh colonies were observed between blades and slats using a DM1000 photonic microscope (Leica Microsystems, Nanterre, France) with a 40× objective lens.

Antibiotic susceptibility

Sixteen antibiotics were used to test antibiotic susceptibility of the bacterium. They included amoxicillin, doxycycline, nitrofurantoin, vancomycin, amoxicillin/clavulanic acid, clindamycin, rifampicin, gentamicin, tobramycin, erythromycin, metronidazole, trimethoprim/sulfamethoxazole, amikacin, ciprofloxacin, imipenem/cilastatin and oxacillin (i2a, Montpellier, France).

Microscopy

Images of *K. gabonensis* strain GM4 were obtained using a Tecnai G20 transmission electron microscope (FEI, Limeil-Brevannes, France) at an operating voltage of 60 kV. Gram staining was observed using a DM1000 photonic microscope (Leica Microsystems) with a 100× oil-immersion objective lens.

Genome sequencing and assembly

A MiSeq sequencer (Illumina, San Diego, CA, USA) added to the mate-pair strategy was used to sequence the genomic DNA (gDNA) of *K. gabonensis* strain GM4. gDNA was barcoded to allow mixing with 11 other projects by using 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 fragmented and tagged using a mate-pair junction adapter. Validation of the pattern of fragmentation was done with a DNA 7500 labchip on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA). The size of the DNA fragments ranged from 1 to 10 kb, with an optimal size of 4.08 kb. There was no size selection, and only 464 ng of tagmented fragments were circularized. The

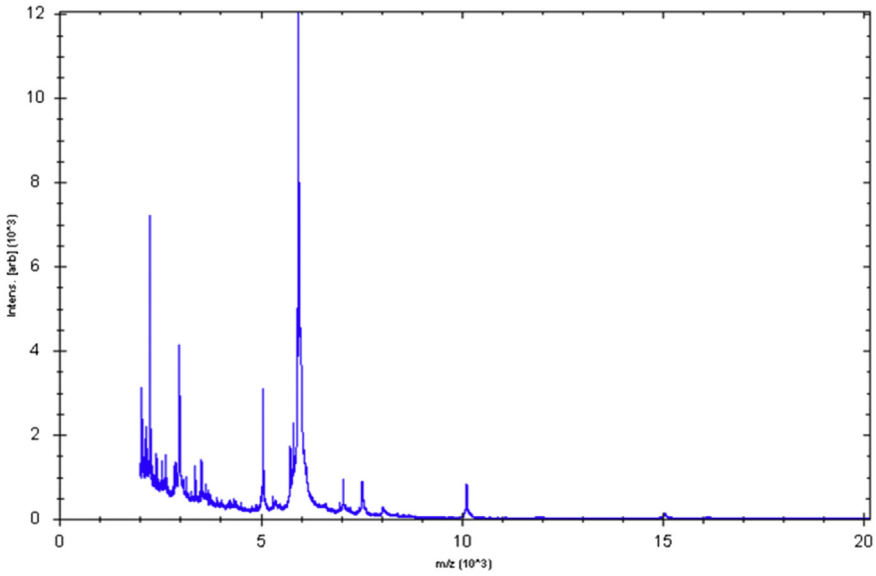

FIG. 1. Reference mass spectrum (matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, MALDI-TOF) from *Kallipyga gabonensis* strain GM4. Spectra from 12 individual colonies were compared and reference spectrum generated.

circularized DNA was mechanically sheared to small fragments (optimal at 569 bp) in microtubes by using the Covaris S2 ultrasonicator (Covaris, Woburn, MA, USA). The profile of the library was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies). The final library concentration was measured at 24.4 nmol/L, normalized at 2 nM and pooled. The library pool was loaded onto the reagent cartridge and then onto the instrument along with the flow cell after a denaturation step and dilution at 15 pM. Cluster generation and the sequencing run were automatically performed in a single 39-hour run in 2 × 251 bp. From a 1189K/mm^2 cluster density and a cluster passing quality control filters of 99.1% (22 579 000 clusters), 10.1 GB of information was obtained. The reads obtained were trimmed; assemblage was then performed using the CLC genomicsWB4 software [5–8].

Genome annotation and comparison

Prodigal software (http://prodigal.ornl.gov/) adjusted with default parameters was used to predict open reading frames (ORFs), and the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched against the GenBank [11] and Clusters

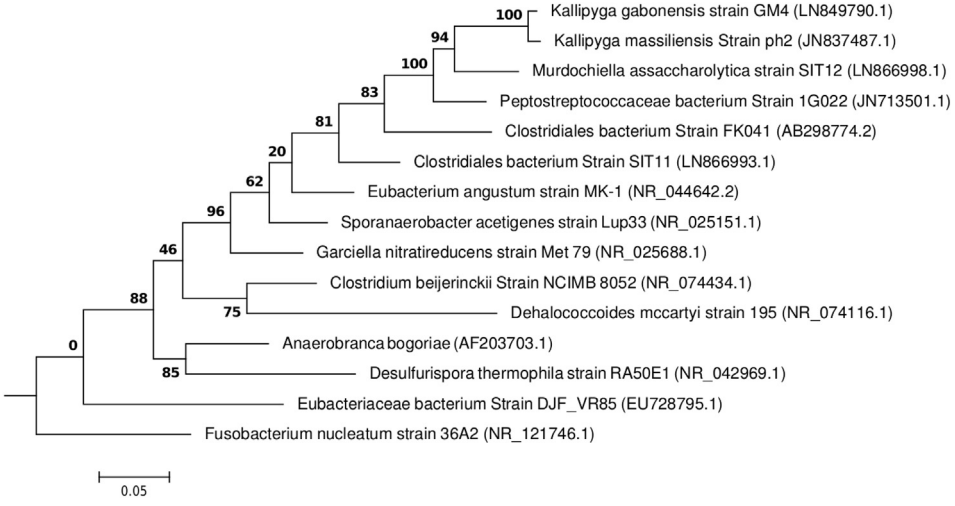

FIG. 2. Phylogenetic tree highlighting position of *Kallipyga gabonensis* strain GM4 compared to other *Clostridiales* bacteria. Sequences were aligned using CLUSTALW and phylogenetic inferences obtained using maximum-likelihood method within MEGA software. Scale bar = 0.05% nucleotide sequence divergence.

TABLE 1. Classification and general features of *Kallipyga gabonensis* strain GM4

Property	Term
Current classification	Domain: *Bacteria* Phylum: *Firmicutes* Class: *Clostridia* Order: *Clostridiales* Family: *Clostridiales Incertae Sedis XI* Genus: *Kallipyga* Species: *gabonensis* Type strain: GM4
Gram stain	Positive
Cell shape	Cocci
Motility	Not motile
Sporulation	Nonsporulating
Temperature range	Mesophilic
Optimum temperature	37°C

of Orthologous Groups (COGs) databases using BLASTP on the NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The tRNAScan-SE and RNAmmer tools allowed us to predict respectively the tRNAs and rRNAs [12,13]. SignalP software and the TMHMM program helped to predict, respectively, signal peptides and numbers of transmembrane helices [14,15]. Mobile genetic elements were predicted using PHAST and RAST tools [16,17]. ORFans were identified if the BLASTP *E* value was lower than 1e-03 for alignment lengths greater than 80 amino acids. An *E* value of 1e-05 was used if alignment lengths were smaller than 80 amino acids. These threshold parameters were already used in previous works to define ORFans [5–8]. For the data management and genomic features visualization, Artemis and DNAPlotter were respectively used [18,19]. The multiple genomic sequence alignment was performed with the Mauve alignment tool (version 2.3.1) [20].

Homemade average genomic identity of orthologous gene sequences (AGIOS) software was used to estimate the mean level of nucleotide sequence similarity between *K. gabonensis* and other *Clostridiales* [21]. By combining AGIOS and Proteinortho software [22], we detected orthologous proteins between genomes compared two by two. The corresponding genes were then retrieved, and the mean percentage of nucleotide sequence identity among orthologous ORFs was determined using the Needleman-Wunsch global alignment algorithm. To make a genome comparison, *K. gabonensis* strain GM4 was compared with *Desulfitobacterium hafniense* strain DP7, *Peptostreptococcaceae bacterium* strain VA2, *Fusobacterium nucleatum* strain 36A2, *Kallipyga massiliensis* strain ph2 and *Clostridiales bacterium* strain 9403326.

All annotation and comparison processes were done *via* the Multi-Agent Software System DAGOBAH [23] that includes Figenix [24].

Results and Discussion

MALDI-TOF and phylogenic analysis

The spectrum generated from clean *Kallipyga gabonensis* strain GM4 spots was unable to match those of the Bruker database (Fig. 1). Using 16S rRNA phylogeny analysis, we demonstrated that *K. gabonensis* strain GM4 exhibited 97.9% similarity with the 16S nucleotide sequence of *K. massiliensis* (JN837487) (Fig. 2). This value of similarity remains lower than the 98.7% 16S rRNA gene sequence threshold recommended by Stackebrandt and Ebers [9] to delineate a new species without

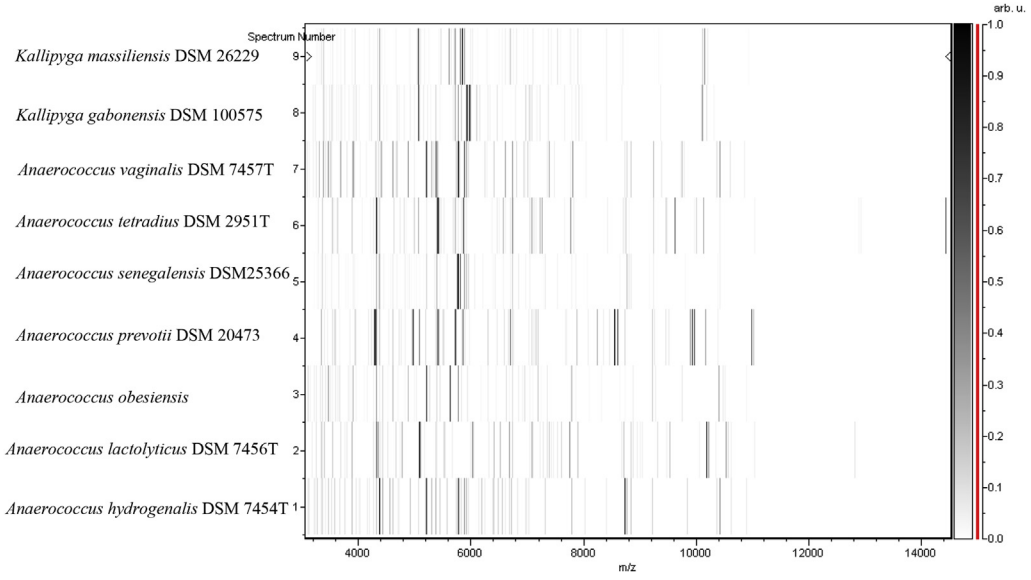

FIG. 3. Gel view comparing *Kallipyga gabonensis* strain GM4 with other members of *Clostridiales* family.

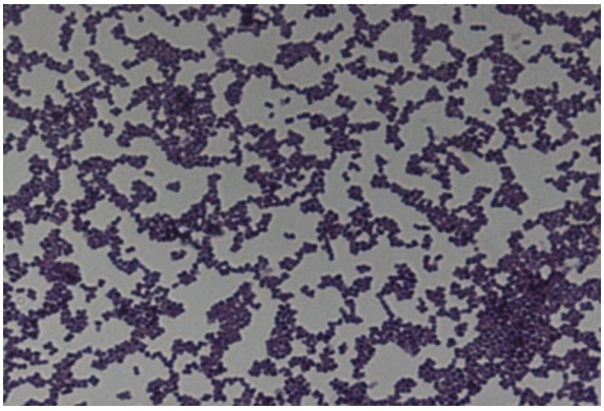

FIG. 4. Gram staining of *Kallipyga gabonensis* strain GM4.

carrying out DNA-DNA hybridization. Thus, this bacterium was considered to be a new species called *K. gabonensis* sp. nov. strain GM4 belonging the *Clostridiales Incertae Sedis XI* family created in 2009 [25] (Table 1). 16S rRNA sequence of *Kallipyga gabonensis* was 1,486 bp long (LN 849790). The 16S rRNA gene was amplified using PCR with the universal primer pair fD1 and rP2. The sequences reactions were performed with the 536F, 536R, 1050F, 1050R, 800F and 800R primers as previously described [4]. *Clostridiales Incertae Sedis XI* is a heterogeneous group that includes anaerobic and morphologically variable bacteria. It currently comprises 11 genera, including *Anaerococcus, Peptoniphilus* and *Tissierella*. The new genus *Kallipyga*, belonging to this group, was recently created in 2013 and included to date only one species, *Kallipyga massiliensis* [26]. The definition of *Clostridiales Incertae Sedis XI* is mainly based on phylogenetic analyses of 16S rRNA sequences, and its members have no precise taxonomic or phylogenetic affiliation. A gel view was performed in order to see the spectra differences of *Kallipyga gabonensis* with other close bacteria, including the second species of the *Kallipyga* genus (*K. massiliensis*) (Fig. 3). The 16S rRNA sequence of *K. gabonensis* was deposited in GenBank under accession number LN849790.

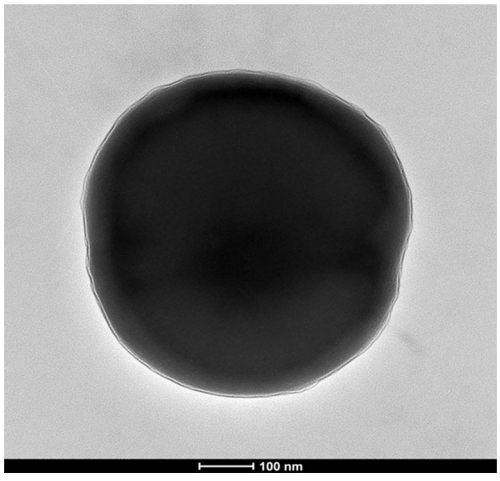

FIG. 5. Transmission electron microscopy of *Kallipyga gabonensis* strain GM4, using Tecnai G20, at operating voltage of 60 kV. Scale bar = 100 nm.

Phenotypic and biochemical characterization

K. gabonensis sp. nov. strain GM4 is Gram positive, comprises non motile cocci, and is non–spore forming (Fig. 4). This bacterium is catalase positive and oxidase negative. It is able to grow under temperatures ranging from 28 to 45°C, with an optimum growth at 37°C, after 48 hours of culture. pH test results showed that *K. gabonensis* can survive under pH conditions ranging between 5 and 8.5, but has optimal growth at pH 5. It grows better with salinity concentrations of around 0. Individual colonies of *K. gabonensis* are whitish and exhibit a diameter of 2 mm on 5% sheep's blood–enriched Colombia agar. Its individual cell exhibits a diameter of 0.6 µm in electron microscopy (Fig. 5). Furthermore, this bacterium is strictly anaerobic; it is unable to grow in aerobic and microaerophilic atmospheres.

Using the API 20A gallery, positive reactions were observed for D-glucose and esculin ferric citrate. Using API 50CH, we concluded that *K. gabonensis* is able to ferment esculin ferric

TABLE 2. **Differential characteristics of *Kallipyga gabonensis* strain GM4 (data from this study) with *Kallipyga massiliensis* strain ph2 [26], *Fenollaria massiliensis* strain 9401234(T) [29] and *Anaerococcus senegalensis* strain JC48^T [5]**

Property	*K. gabonensis*	*K. massiliensis*	*F. massiliensis*	*A. senegalensis*
Cell diameter	0.6 µm	0.6 µm	0.6–1 µm	0.8 µm
Oxygen requirement	–	–	–	–
Gram stain	+	+	–	+
Mobility	–	–	–	–
Endospore formation	–	–	–	–
Indole	–	–	+	+
Production of:				
Alkaline phosphatase	–	–	–	+
Catalase	+	–	–	+
Oxydase	–	–	–	–
Nitrate reductase	NA	–	NA	NA
Urease	–	–	–	+
β-Galactosidase	+	–	–	–
N-acetyl-glucosaminidase	–	–	–	–
Acid from:				
L-Arabinose	–	–	–	NA
Ribose	–	+	NA	NA
Mannose	–	–	–	+
Mannitol	–	–	–	NA
Sucrose	+	NA	NA	NA
D-Glucose	+	+	–	NA
D-Fructose	–	+	NA	NA
D-Maltose	–	–	–	NA
D-Lactose	–	–	–	NA
G+C content (%)	50.0	51.4	34.5	28.4
Habitat	Human	Human	Human	Human

NA, data not available.

TABLE 3. Nucleotide content and gene count levels of chromosome

Attribute	Genome (total)	
	Value	% of total
Size (bp)	1,621,211	100
G+C content (bp)	810,790	50.0
Coding region (bp)	1,380,371	91.3
Extrachromosomal elements	0	0
Total genes	1,536	100
RNA genes	61	3.9
Protein-coding genes	1,475	100
Genes with function prediction	931	63.1
Genes assigned to COGs	913	61.9
Genes with peptide signals	153	10.4
Genes with transmembrane helices	347	23.5

COGs, Clusters of Orthologous Groups database.

citrate (as found with the API 20A gallery), D-melibiose, sucrose, inulin and D-raffinose. The result of the API ZYM gallery shows that *K. gabonensis* possesses esterase (C4), leucine arylamidase, naphthol-AS-BI-phosphohydrolase, slightly esterase lipase (C8), acid phosphatase and β-glucosidase activities. All biochemical reactions not listed here but present in the three galleries we used remained negative. These biochemical characteristics were compared with those of other close bacteria (Table 2). Among the antibiotics tested, *K. gabonensis* was sensitive only to metronidazole, imipenem and oxacillin. No sensitivity was recorded with amoxicillin, doxycycline, nitrofurantoin, vancomycin, amoxicillin/clavulanic acid, clindamycin, erythromycin, tobramycin, trimethoprim, amikacin, ciprofloxacin 5 and gentamicin.

Genome properties

The genome is 1,621,211 bp long with 50.01% G+C content (Table 3). It contains two scaffolds (composed of two contigs) (Fig. 6). Of the 1,536 predicted genes, 1,475 were protein-coding genes and 61 were RNA genes (four are 5S rRNA, four are 16S rRNA, four are 23S rRNA, 49 are tRNA genes). A total of 931 genes (63.12%) were assigned a putative function (by COGs or by nrBLAST), and 79 genes were identified as ORFans (5.36%). The remaining genes were annotated as encoding hypothetical proteins (394 genes = 26.71%). Table 4 summarizes the distribution of genes into COGs functional categories of the *K. gabonensis* genome. The genome sequence has been deposited in GenBank under accession number CXYV00000000.

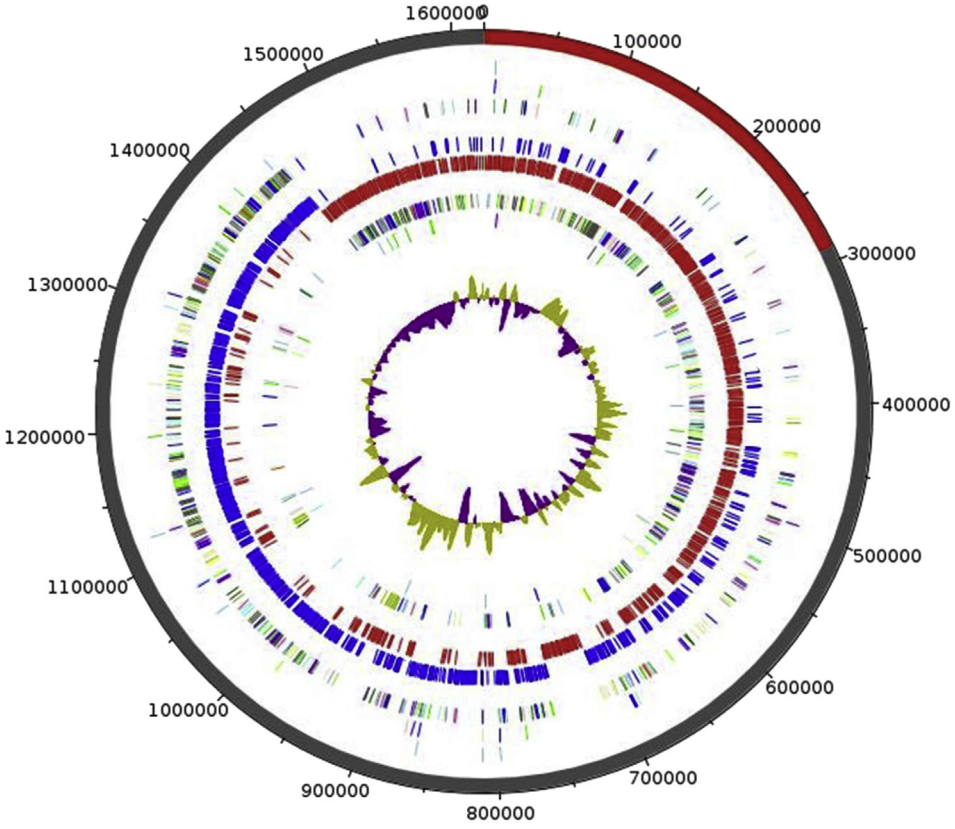

FIG. 6. Graphical circular map of genome of *Kallipyga gabonensis* strain GM4. From outside to center, contigs (red/grey), COGs category of genes on forward strand (three circles), genes on forward strand (blue circle), genes on reverse strand (red circle), COGs category on reverse strand (three circles), G+C content. COGs, Clusters of Orthologous Groups database.

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

Code	Value	% of total	Description
J	130	8.8	Translation
A	0	0	RNA processing and modification
K	59	4.0	Transcription
L	85	5.8	Replication, recombination and repair
B	0	0	Chromatin structure and dynamics
D	12	0.8	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	27	1.8	Defence mechanisms
T	24	1.6	Signal transduction mechanisms
M	43	2.9	Cell wall/membrane biogenesis
N	0	0	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	11	0.7	Intracellular trafficking and secretion
O	39	2.6	Posttranslational modification, protein turnover, chaperones
C	59	4.0	Energy production and conversion
G	76	5.2	Carbohydrate transport and metabolism
E	70	4.7	Amino acid transport and metabolism
F	45	3.0	Nucleotide transport and metabolism
H	31	2.1	Coenzyme transport and metabolism
I	31	2.1	Lipid transport and metabolism
P	42	2.8	Inorganic ion transport and metabolism
Q	7	0.5	Secondary metabolites biosynthesis, transport and catabolism
R	106	7.2	General function prediction only
S	71	4.8	Function unknown
—	562	38.1	Not in COGs

COGs, Clusters of Orthologous Groups database.

Genome comparison

The draft genome sequence of *Kallipyga gabonensis* strain GM4 (1.62 Mb) is lower than those of *Desulfitobacterium hafniense* strain DP7 (5.73 Mb), *Peptostreptococcaceae bacterium* strain VA2 (3.58 Mb), *Fusobacterium nucleatum* strain 36A2 (2.17 Mb), *Kallipyga massiliensis* strain ph2 (1.77 Mb) and *Clostridiales bacterium* strain 9403326 (1.71 Mb). The G+C percentage of *K. gabonensis* strain GM4 (50.0%) is lower than that of *Kallipyga massiliensis* strain ph2 (51.4%) but higher than those of *Clostridiales bacterium* strain 9403326, *Desulfitobacterium hafniense* strain DP7, *Peptostreptococcaceae bacterium* stain VA2, and *Fusobacterium nucleatum* (48.8, 47.4, 28.6 and 27.2% respectively). The gene content of *Kallipyga gabonensis* strain GM4 (1536) is lower than those of *Desulfitobacterium hafniense* strain DP7, *Peptostreptococcaceae bacterium* strain VA2, *Fusobacterium nucleatum* strain 36A2, *Kallipyga massiliensis* strain ph2 and *Clostridiales bacterium* strain 9403326 (5,365, 3,567, 2,062, 1,616 and 1,538, respectively). The protein-coding genes of *Kallipyga gabonensis* strain GM4 (1475) are more numerous than in *Clostridiales bacterium* strain 9403326 (1,463) and less numerous than those of *Desulfitobacterium hafniense* strain DP7, *Peptostreptococcaceae bacterium* strain VA2, *Fusobacterium nucleatum* strain 36A2 and *Kallipyga massiliensis* strain ph2 (5,203, 3,193, 1,983 and 1,546, respectively). Specifically in comparison with *Kallipyga massiliensis*, the genome of *K. gabonensis* includes a larger number of genes assigned with peptide signals (153 vs. 90), a larger number of ORFans (79 vs. 42) and a larger number of RNA genes (61 vs. 50) but a lower number of genes assigned to COGs (913 vs. 1165).

Among the species with standing in nomenclature, AGIOS values ranged from 85.78 between *Kallipyga gabonensis* strain GM4 and *Kallipyga massiliensis* strain ph2 to 51.41 between *Kallipyga massiliensis* strain ph2 and *Fusobacterium nucleatum* strain 36A2. The AGIOS strategy involves the first step of performing proteinOrtho analysis to identify the group of orthologous genes. Then for each couple of orthologous genes in each couple of genomes, an alignment at the nucleotide level is done, and a similarity score is calculated for each couple of orthologous genes. For each couple of genomes, a mean similarity score is defined.

To evaluate the genomic similarity among the studied strains, we determined two parameters, digital DNA-DNA hybridization (DDH), which exhibits a high correlation with DDH [27,28] and AGIOS [21], which was designed to be independent from DDH. As a result of recent innovations, the area of genome sequencing called for bioinformatics methods to replace the wet-lab DDH by *in silico* genome-to-genome comparison. This is the method for inferring whole-genome distances which are well able to mimic DDH. These distance functions can also cope with heavily reduced genomes and repetitive sequence regions. Some of them are also very robust against missing fractions of genomic information (which may result from incomplete genome sequencing). This digitally derived genome-to-genome distance shows a better correlation with 16S rRNA gene sequence distances than DDH values.

TABLE 5. Numbers of orthologous proteins shared between genomes (upper right)[a]

	Clostridiales bacterium	*Desulfitobacterium hafniense*	*Kallipyga gabonensis*	*Kallipyga massiliensis*	*Peptostreptococcaceae bacterium*	*Fusobacterium nucleatum*
Clostridiales bacterium	**969**	359	551	562	371	309
Desulfitobacterium hafniense	58.79	**5,416**	482	507	515	447
Kallipyga gabonensis	63.38	58.54	**1,476**	760	500	424
Kallipyga massiliensis	63.58	58.44	85.78	**1,564**	512	441
Peptostreptococcaceae bacterium	55.38	57.07	54.00	52.37	**3,458**	447
Fusobacterium nucleatum	54.19	55.65	53.01	51.41	65.69	**2,161**

[a]Average percentage similarity of nucleotides corresponding to orthologous proteins shared between genomes (lower left) and numbers of proteins per genome (bold).

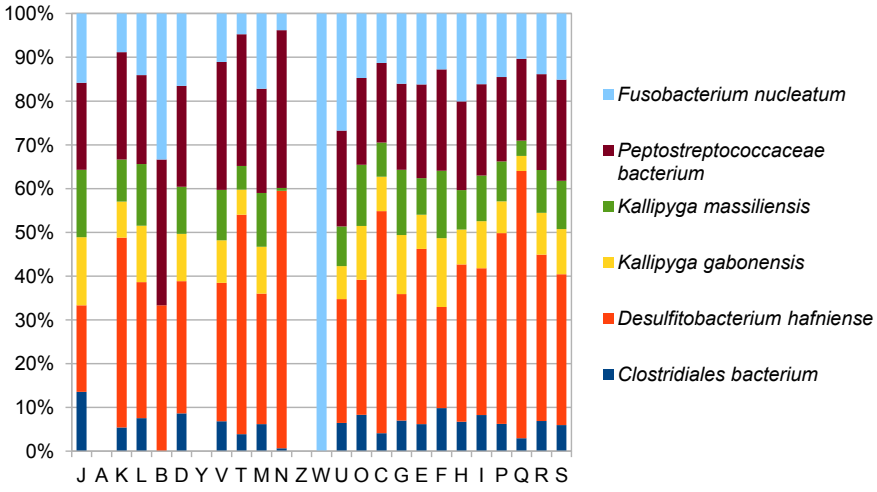

FIG. 7. Distribution of functional classes of predicted genes according to COGs proteins. COGs, Clusters of Orthologous Groups database.

The distribution of functional classes of predicted genes according to the clusters of orthologous groups of proteins is summarized in Table 5 and Fig. 7. Table 6 summarizes the pairwise comparison of *Kallipyga gabonensis* with other species using the Genome-to-Genome Distance Calculator.

Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Kallipyga gabonensis* sp. nov., which contains the strain GM4. This bacterium was isolated in Marseille (France) from a stool specimen of a healthy Gabonese male youth.

Description of *Kallipyga gabonensis* sp. nov.

Kallipyga gabonensis (ga.bon.en'.sis. L. masc. n. *gabonensis*, for Gabon, where the specimen was collected). A non motile, Gram-positive cocci, *K. gabonensis* is catalase positive and oxidase negative, non–spore forming, strictly anaerobic, and has an exhibited individual cell with a diameter of 0.6 µm. Colonies of *K. gabonensis* are whitish and exhibit a diameter of 2 mm. Maximal growth was recorded at 37°C, with a pH of 5 and a salinity concentration around 0. *K. gabonensis* was sensitive to metronidazole, imipenem and oxacillin. Using API gallery systems, *K. gabonensis* strain GM4 induced positive reactions with D-glucose, esculin ferric citrate, D-melibiose, sucrose, inulin and D-raffinose. It also possesses esterase (C4), leucine arylamidase, naphthol-AS-BI-phosphohydrolase, slightly esterase lipase (C8), acid phosphatase and β-glucosidase activities.

The potential pathogenicity of the type of strain GM4 (= CSUR P1915 = DSM 100575) is unknown. This strain exhibited a G+C content of 50.01% and a genome length of 1.6 Mb. The 16S rRNA sequence of *K. gabonensis* was deposited in GenBank under accession number LN849790. The whole genome shotgun sequence of *K. gabonensis* has been deposited in GenBank under accession number CXYV00000000.

TABLE 6. Pairwise comparison of *Kallipyga gabonensis* with other species[a]

	Clostridiales bacterium	*Desulfitobacterium hafniense*	*Kallipyga gabonensis*	*Kallipyga massiliensis*	*Peptostreptococcaceae bacterium*	*Fusobacterium nucleatum*
Clostridiales bacterium	100% ± 00	20.8% ± 2.52	21.2% ± 2.53	29.7% ± 2.56	22.9% ± 2.53	20% ± 2.52
Desulfitobacterium hafniense		100% ± 00	21.7% ± 2.52	21.4% ± 2.52	23.4% ± 2.53	18.9% ± 2.52
Kallipyga gabonensis			100% ± 00	30.1% ± 3.12	21.4% ± 2.54	17.9% ± 2.52
Kallipyga massiliensis				100% ± 00	21.2% ± 2.53	17.3% ± 2.52
Peptostreptococcaceae bacterium					100% ± 00	17.7% ± 2.54
Fusobacterium nucleatum						100% ± 00

DDH, DNA-DNA hybridization; GGDC, Genome-to-Genome Distance Calculator; HSP, High-scoring Segment Pair.
[a]Data were generated by GGDC, formula 2 (DDH estimates based on identities/HSP length). Confidence intervals indicate inherent uncertainty in estimating DDH values from intergenomic distances based on models derived from empirical test data sets (which are always limited in size). These results are in accordance with 16S rRNA (Fig. 2) and phylogenomic analyses as well as GGDC results.

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] Reddy EA, Shaw AV, Crump JA. Community-acquired bloodstream infections in Africa: a systematic review and meta-analysis. Lancet Infect Dis 2010;10:417–32.

[2] Wilairatana P, Meddings JB, Ho M, Vannaphan S, Looareesuwan S. Increased gastrointestinal permeability in patients with *Plasmodium falciparum* malaria. Clin Infect Dis 1997;24:430–5.

[3] Raoult D, Henrissat B. Are stool samples suitable for studying the link between gut microbiota and obesity? Eur J Epidemiol 2014;29:307–9.

[4] 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.

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

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

[7] Mishra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous-finished genome sequence and description of *Clostridium senegalense* sp. nov. Stand Genomic Sci 2012;6:386–95.

[8] Michra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous-finished genome sequence and description of *Peptinophilus timonensis* sp. nov. Stand Genomic Sci 2012;7:1–11.

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

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

[11] Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucl Acids Res 2012;40:D48–53.

[12] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucl Acids Res 1997;25:955–64.

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

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

[15] 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.

[16] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucl Acids Res 2011;39:W347–52.

[17] 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.

[18] Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. Bioinformatics 2000;16:944–5.

[19] Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 2009;25:119–20.

[20] Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004;14:1394–403.

[21] 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 new bacterial species. Int J Syst Evol Microbiol 2014;64:384–91.

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

[23] 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 DAGOBAH. New York: Springer; p. 71–87.

[24] 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.

[25] Ludwig W, Schleifer KH, Whitman WB. Revised road map to the phylum *Firmicutes*. In: De Vos P, Garrity G, et al., editors. Bergey's manual of systematic bacteriology. 2nd ed. New York: Springer-Verlag; 2009. p. 3:1–13.

[26] Hugon P, Ramasamy D, Robert C, Couderc C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Kallipyga massiliensis* gen. nov., sp. nov., a new member of the family *Clostridiales Incertae Sedis XI*. Stand Genomic Sci 2013;8:500–15.

[27] 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.

[28] 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.

[29] Pagnier I, Croce O, Robert C, Raoult D, La Scola B. Non-contiguous finished genome sequence and description of *Fenollaria massiliensis* gen. nov., sp. nov., a new genus of anaerobic bacterium. Stand Genomic Sci 2014;9:704–17.