# Non-contiguous finished genome sequence and description of *Bacillus massiliogorillae* sp. nov.

**Mamadou Bhoye Keita[1], Seydina M. Diene[1], Catherine Robert[1], Didier Raoult[1,2], Pierre-Edouard Fournier[1], and Fadi Bittar[1]\***

**[1]URMITE, Aix-Marseille Université, Faculté de Médecine, Marseille, France**
**[2]King Fahad Medical Research Center, King Abdul Aziz University, Jeddah, Saudi Arabia**

**\***Correspondence: Fadi Bittar (fadi.bittar@univ-amu.fr)

Strain G2[T] sp. nov. is the type strain of *B. massiliogorillae*, a proposed new species within the genus *Bacillus*. This strain, whose genome is described here, was isolated in France from the fecal sample of a wild western lowland gorilla from Cameroon. *B. massiliogorillae* is a facultative anaerobic, Gram-variable, rod-shaped bacterium. Here we describe the features of this organism, together with the complete genome sequence and annotation. The 5,431,633 bp long genome (1 chromosome but no plasmid) contains 5,179 protein-coding and 98 RNA genes, including 91 tRNA genes.

## Introduction

Strain G2[T] (= CSUR P206 = DSM 26159) is the type strain of *B. massiliogorillae* sp. nov. This bacterium is a Gram-variable, facultatively anaerobic, indole-negative bacillus having rounded-ends. It was isolated from the stool sample of *Gorilla gorilla gorilla* as part of a "culturomics" study aiming at cultivating bacterial species within gorilla feces.

The genus *Bacillus* (Cohn 1872) was created about 140 years ago [1]. To date this genus, comprised mostly of Gram-positive, motile, and spore-forming bacteria, includes 276 species with validly published names [2]. Members of the genus *Bacillus* are ubiquitous bacteria isolated from various environments including soil, fresh and sea water, food, and occasionally from humans and animals in which they are either pathogens, such as *B. anthracis* (the causative agent of anthrax) [3] and *B. cereus* (associated mainly with food poisoning) [4], or saprophytes [5]. *Bacillus* species may also rarely be involved in a variety of human infections, including pneumonia, bacteremia, meningitis, endocarditis, endophthalmitis, osteomyelitis and skin/soft tissue infection [5]. However, in great apes, few data are available about the presence of the genus *Bacillus*. Recent reports have described the isolation of atypical *B. anthracis* (*B. anthracis*-like bacteria) in wild chimpanzees and gorillas from Africa [6-8].

Here we present a summary classification and a set of features for *B. massiliogorillae* sp. nov. strain G2[T] together with the description of the complete genome sequence and annotation. These characteristics support the circumscription of the species *B. massiliogorillae* [9].

## Classification and features

In July 2011, a fecal sample was collected from a wild western lowland gorilla near Messok, a village in the south-eastern part of the DJA FAUNAL Park (Cameroon). The collection of the stool sample was approved by the Ministry of Scientific Research and Innovation of Cameroon. No experimentation was conducted on this gorilla. The fecal specimen was preserved at -80°C after collection and sent to Marseille. Strain G2[T] (Table 1) was isolated in January 2012 by cultivation on *Brucella* agar medium (Oxoid, Dardilly, France). This strain exhibited a 97.3% 16S rRNA nucleotide sequence similarity with *Bacillus simplex*, the phylogenetically closest validly published *Bacillus* species (Figure 1). This value was lower than the 98.7% 16S rRNA gene sequence threshold recommended by Stackebrandtia and Beers to delineate a new species without carrying out DNA-DNA hybridization [23].

**Table 1.** Classification and general features of *Bacillus massiliogorillae* strain G2[T]

| MIGS ID | Property | Term | Evidence code[a] |
|---|---|---|---|
| | | Domain *Bacteria* | TAS [10] |
| | | Phylum *Firmicutes* | TAS [11-13] |
| | | Class *Bacilli* | TAS [14,15] |
| | Current classification | Order *Bacillales* | TAS [16,17] |
| | | Family *Bacillaceae* | TAS [16,18] |
| | | Genus *Bacillus* | TAS [16,19,20] |
| | | Species *Bacillus massiliogorillae* | IDA |
| | | Type strain G2[T] | IDA |
| | Gram stain | Variable | IDA |
| | Cell shape | Rod | IDA |
| | Motility | Motile | IDA |
| | Sporulation | Sporulating | IDA |
| | Temperature range | Mesophile | IDA |
| | Optimum temperature | 37°C | IDA |
| MIGS-6.3 | Salinity | Growth in BHI medium + 2% NaCl | IDA |
| MIGS-22 | Oxygen requirement | Facultative anaerobic | IDA |
| | Carbon source | Unknown | NAS |
| | Energy source | Unknown | NAS |
| MIGS-6 | Habitat | Gorilla gut | IDA |
| MIGS-15 | Biotic relationship | Free living | IDA |
| MIGS-14 | Pathogenicity | Unknown | NAS |
| | Biosafety level | 2 | NAS |
| | Isolation | Gorilla feces | NAS |
| MIGS-4 | Geographic location | Cameroon | IDA |
| MIGS-5 | Sample collection time | July 2011 | IDA |
| MIGS-4.1 | Latitude | Unknown | NAS |
| MIGS-4.1 | Longitude | Unknown | NAS |
| MIGS-4.3 | Depth | Unknown | NAS |
| MIGS-4.4 | Altitude | Unknown | NAS |

Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [21]. If the evidence is IDA, then the property was directly observed for a live isolate by one of the authors or an expert mentioned in the acknowledgements.
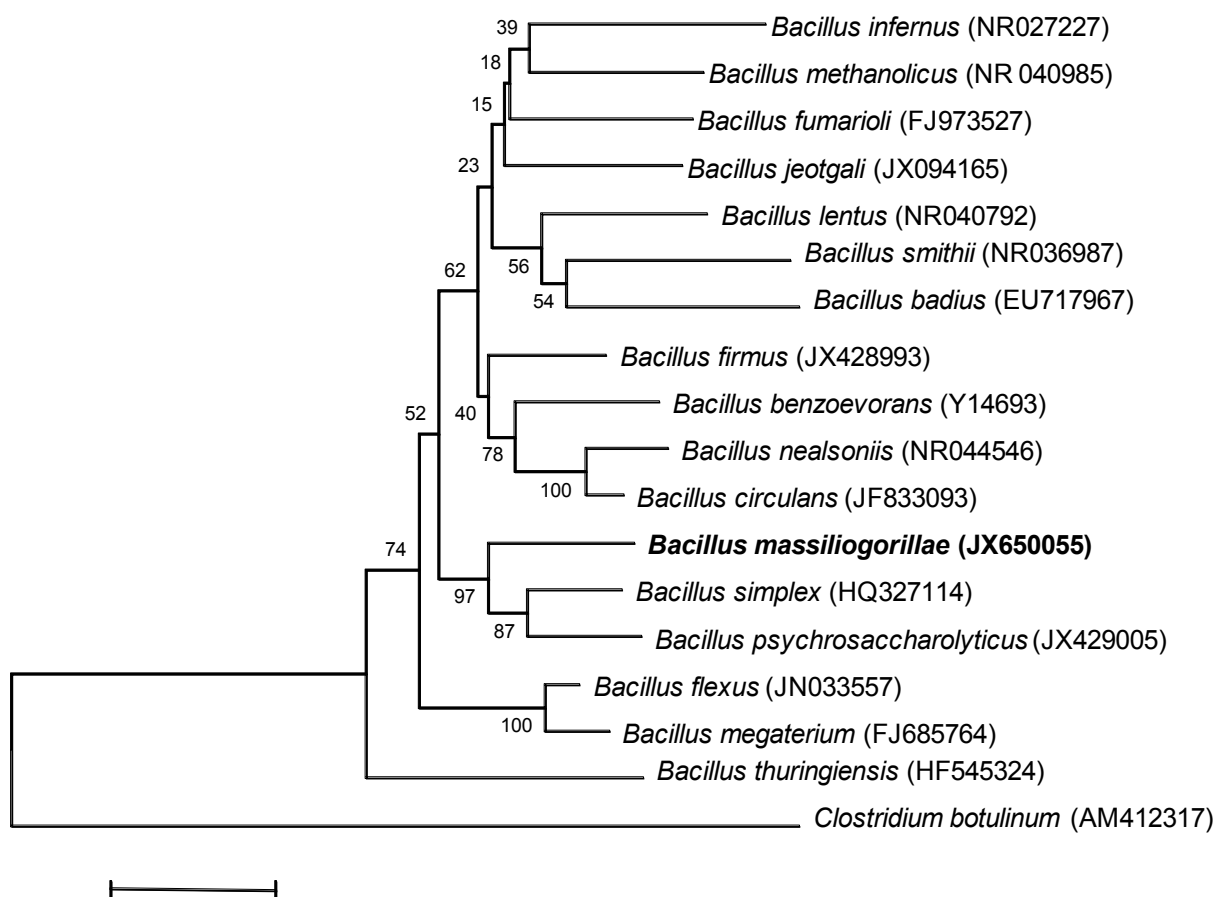
**Figure 1.** Phylogenetic tree highlighting the position of *Bacillus massiliogorillae* strain G2$^T$ relative to other type strains within the *Bacillus* genus. GenBank accession numbers are indicated in parentheses. Sequences were aligned using CLUSTAL X (V2), and phylogenetic inferences obtained using the maximum-likelihood method within the MEGA 5 software [22]. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 1,000 times to generate a majority consensus tree. *Clostridium botulinum* was used as outgroup. The scale bar represents a 2% nucleotide sequence divergence.

Different growth temperatures (25, 30, 37, 45°C) were tested. Growth occurred at all tested temperatures, and the optimal growth was observed at 37°C. Colonies were 2-5 mm in diameter on Columbia agar, grey opaque in color. Growth of the strain was tested under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems, respectively (BioMérieux), and in aerobic conditions, with or without 5% $CO_2$. Growth was achieved under aerobic (with and without $CO_2$), microaerophilic and anaerobic conditions. Gram staining showed Gram variable bacilli (Figure 2). A motility test was positive. Cells grown on agar sporulate and the rods have a length ranging from 3.2 to 7.5 μm (mean 5.4 μm) and a diameter ranging from 0.8 to 1.2 μm (mean 1 μm) as determined by negative staining transmission electron microscopy (Figure 3).

Strain G2$^T$ exhibited catalase activity but not oxidase activity. Using the API 50CH system (BioMerieux), a positive reaction was observed for D-glucose, D-fructose, D-ribose, N-acetylglucosamine, amygdalin, arbutin, aesculin, salicin, cellobiose, maltose, D-lactose, D-trehalose, D-saccharose, and hydrolysis of starch. Using the API ZYM system, positive reactions were observed for esterase (C4), esterase lipase (C8), phosphatase acid, α- glucosidase and N-acetyl-β-glucosaminidase. The urease reaction was also positive, but nitrate reduction and indole production were negative. *B. massiliogorillae* is susceptible to amoxicillin, nitrofurantoin, erythromycin, doxycycline, rifampin, vancomycin, gentamycin and imipenem but resistant to trimethoprim-sulfamethoxazole, ciprofloxacin, ceftriaxon and amoxicillin-clavulanic acid.
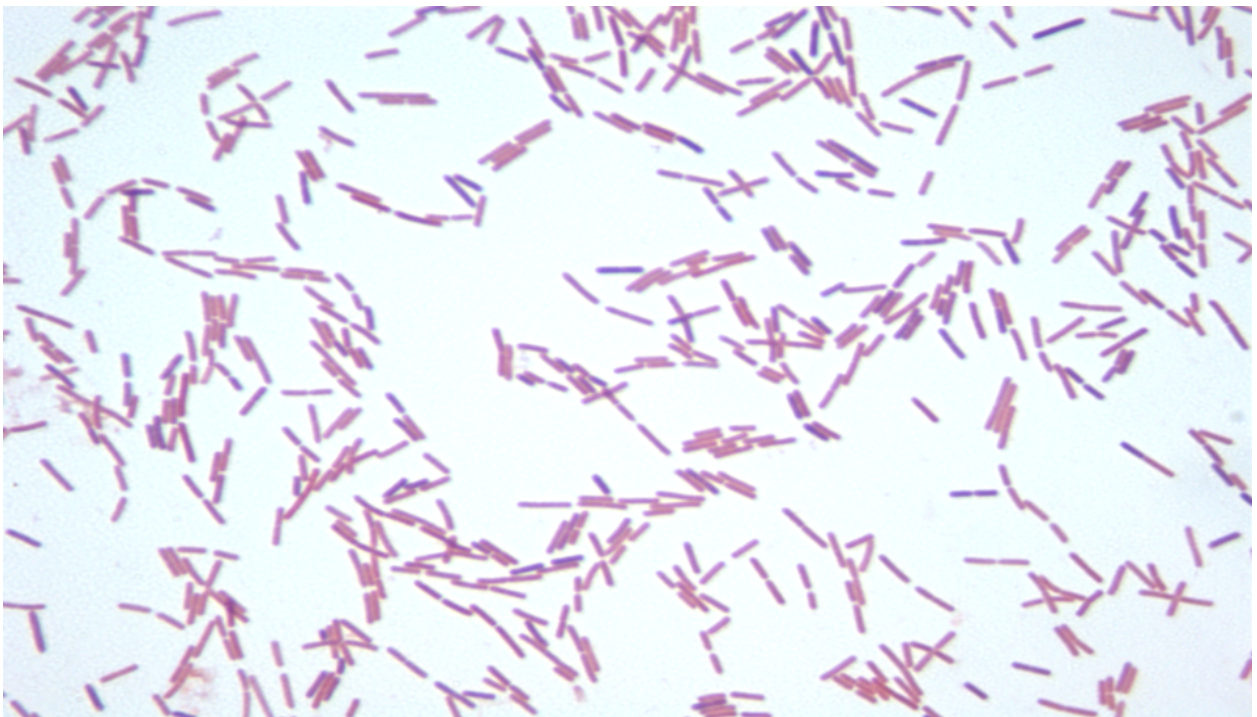
**Figure 2.** Gram staining of *B. massiliogorillae* strain G2$^T$
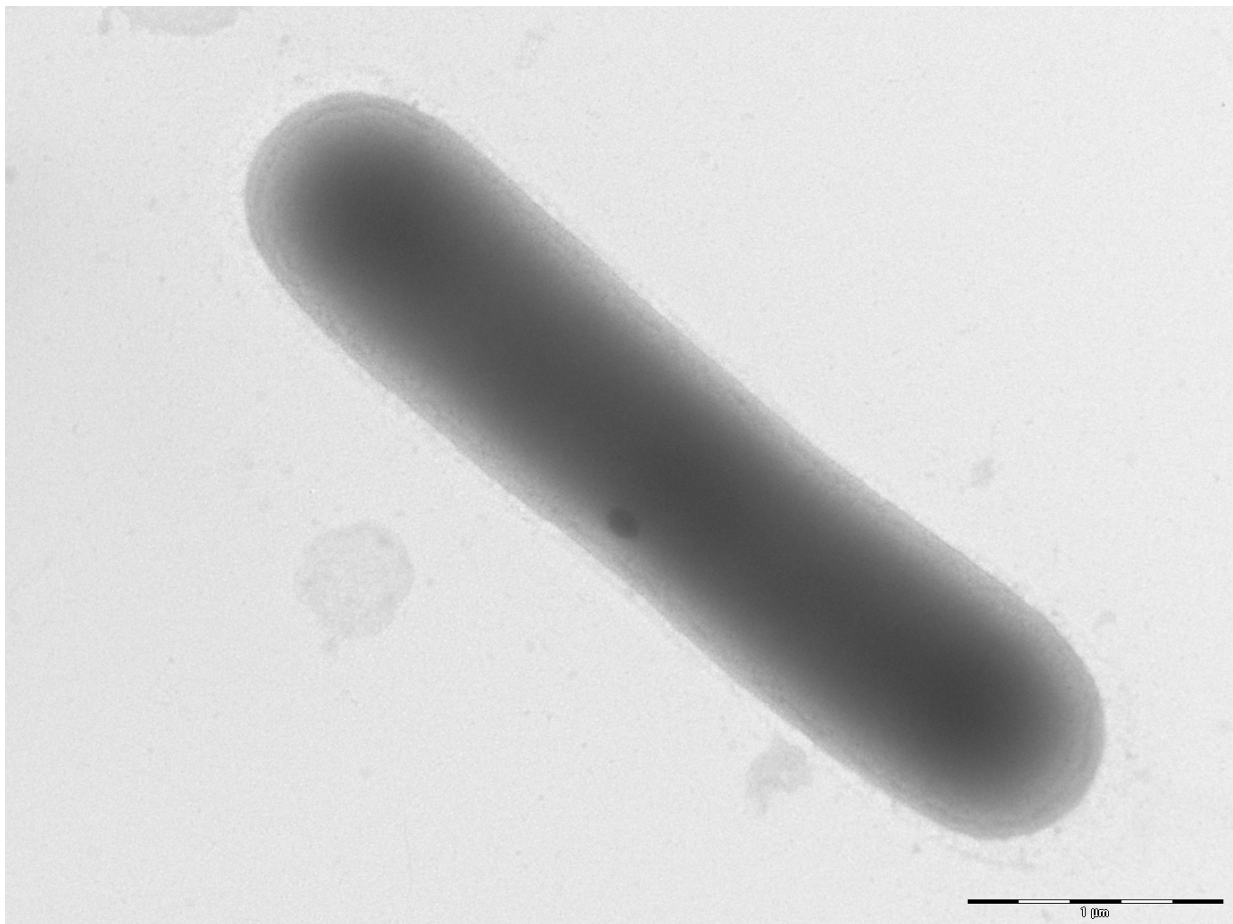


**Figure 3.** Transmission electron microscopy of *B. massiliogorillae* strain G2$^T$, using a Morgani 268D (Philips) at an operating voltage of 60kV. The scale bar represents 1 μm.

**Table 2.** Differential phenotypic characteristics between *B. massiliogorillae* sp. nov. strain G2$^T$ and phylogenetically close *Bacillus* species.

| Characteristic | *B. massiliogorillae* sp. nov. | *B. simplex* | *B. psychrosaccharolyticus* | *B. circulans* |
|---|---|---|---|---|
| Cell diameter (µm) | 0.87-1.2 | 0.7-0.9 | 0.9-1 | 0.5-0.8 |
| Oxygen requirement | aerobic | aerobic | facultative anaerobic | aerobic |
| Gram stain | var | var | var | var |
| Salt requirement | < 5% | <7% | <10% | <7% |
| Motility | + | v | + | + |
| Endospore formation | + | + | + | + |
| | | | | |
| **Production of** | | | | |
| Alkaline phosphatase | + | na | na | na |
| Acid phosphatase | + | na | na | na |
| Catalase | + | + | + | + |
| Oxidase | - | - | na | na |
| Nitrate reductase | - | na | + | na |
| Urease | + | na | na | w |
| α-galactosidase | - | na | na | na |
| β-galactosidase | - | na | + | + |
| β -glucuronidase | - | na | na | na |
| α -glucosidase | + | na | na | na |
| N-acetyl- βglucosaminidase | + | na | na | na |
| Indole | - | na | na | na |
| Esterase | + | na | na | na |
| Esterase lipase | + | na | na | na |
| Naphthyl-AS-BI-phosphohydrolase | - | na | na | na |
| Phenylalanine arylamidase | - | na | na | na |
| Leucine arylamidase | - | na | na | na |
| Cystine arylamidase | - | na | na | na |
| Valine arylamidase | - | na | na | na |
| Glycine arylamidase | - | na | na | na |
| | | | | |
| **Utilization of** | | | | |
| D-mannose | - | - | na | + |
| Amygdalin | + | - | na | v |
| L-Arabinose | - | - | + | + |
| Cellobiose | + | - | na | + |
| Lactose | + | - | + | + |
| D-xylose | - | -/w | na | + |
| Glucose | + | na | + | + |
| Mannitol | - | na | + | + |
| Arabinose | - | na | + | + |
| Xylose | - | na | + | + |
| Glycerol | - | na | + | + |
| D-Galactose | - | na | na | + |
| Starch | + | na | na | + |
| **Habitat** | gorilla gut | soil | soil and lowland marshes | environment and fish gut |

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

When compared to other *Bacillus* species, *B. massiliogorillae* differed from *B. simplex* [24] for the utilization of amygdalin, cellobiose, lactose and glucose (Table 2). It also differed from *B. psychrosaccharolyticus* [25] in nitrate reductase and β-galactosidase production, and in the utilization of L-arabinose, mannitol, xylose and glycerol (Table 2). Differences were also observed with *B. circulans* [26] in β-galactosidase production and the utilization of D-mannose, L-arabinose, D-xylose, mannitol, arabinose, xylose, glycerol and D-galactose (Table 2).

Matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) MS protein analysis was carried out as previously described [27,28]. Deposits were done for strain G2$^T$ from 12 isolated colonies. Each smear was overlaid with 2µL of matrix solution (saturated solution of alpha-cyano-4-hydroxycinnamic acid) in 50% acetonitrile, 2.5% tri-fluoracetic-acid, and allowed to dry for five minutes. Measurements were performed with a Microflex spectrometer (Bruker Daltonics, Leipzig, Germany). Spectra were recorded in the positive linear mode for the mass range of 2,000 to 20,000 Da (parameter settings: ion source 1 (IS1), 20 kV; IS2, 18.5 kV; lens, 7 kV). A spectrum was obtained after 675 shots at a variable laser power. The time of acquisition was between 30 seconds and 1 minute per spot. The 12 G2$^T$ spectra were imported into the MALDI BioTyper software (version 2.0, Bruker) and analyzed by standard pattern matching (with default parameter settings) against 6,252 bacterial spectra including 199 spectra from 104 *Bacillus* species, used as reference data, in the BioTyper database. The method of identification included the m/z from 3,000 to 15,000 Da. For every spectrum, 100 peaks at most were taken into account and compared with spectra in the database. A score enabled the identification, or not, from the tested species: a score > 2 with a validated species enabled the identification at the species level, a score > 1.7 but < 2 enabled identification at the genus level; and a score < 1.7 did not enable any identification. For strain G2$^T$, the scores obtained ranged from 1.177 to 1.343, thus suggesting that our isolate was not a member of a known species. We incremented our database with the spectrum from strain G2$^T$ (Figure 4). Spectrum differences with other of *Bacillus* species are shown in Figure 5.

# Genome sequencing information
## Genome project history

The organism was selected for sequencing on the basis of its phylogenetic position and 16S rRNA similarity to other members of the genus *Bacillus*, and is part of a "culturomics" study of the gorilla flora aiming at isolating all bacterial species within gorilla feces. It was the 61$^{st}$ genome of a *Bacillus* species and the first genome of *Bacillus massiliogorillae* sp. nov. A summary of the project information is shown in Table 2. The Genbank accession number is CAVL000000000 and consists of 66 large contigs. Table 3 shows the project information and its association with MIGS version 2.0 compliance [29].

## Growth conditions and DNA isolation

*B. massiliogorillae* sp. nov. strain G2$^T$, CSUR P206, DSM 26159, was grown aerobically on 5% sheep blood-enriched Columbia agar at 37°C. Four petri dishes were spread and resuspended in 3x500µl of TE buffer and stored at 80°C. Then, 500µl of this suspension were thawed, centrifuged 3 minutes at 10,000 rpm and resuspended in 3x100µL of G2 buffer (EZ1 DNA Tissue kit, Qiagen). A first mechanical lysis was performed by glass powder on the Fastprep-24 device (Sample Preparation system, MP Biomedicals, USA) using 2x20 seconds cycles. DNA was then treated with 2.5µg/µL lysozyme (30 minutes at 37°C) and extracted using the BioRobot EZ1 Advanced XL (Qiagen). The DNA was then concentrated and purified using the Qiamp kit (Qiagen). The yield and the concentration was measured by the Quant-it Picogreen kit (Invitrogen) on the Genios Tecan fluorometer at 50ng/µl.

## Genome sequencing and assembly

The paired-end library was prepared with 5 µg of bacterial DNA using the DNA fragmentation on the Covaris S-Series (S1, S2) instrument (Woburn, Massachusetts, USA) with an enrichment size at 3-5-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 500 bp. After PCR amplification through 15 cycles followed by double size selection, the single stranded paired-end library was quantified using the Quant-it Ribogreen kit (Invitrogen) on the Genios Tecan fluorometer at 339 pg/µL. The library concentration equivalence was calculated as 1.00E+08 molecules/µL. The library was stored at -20°C until further use.
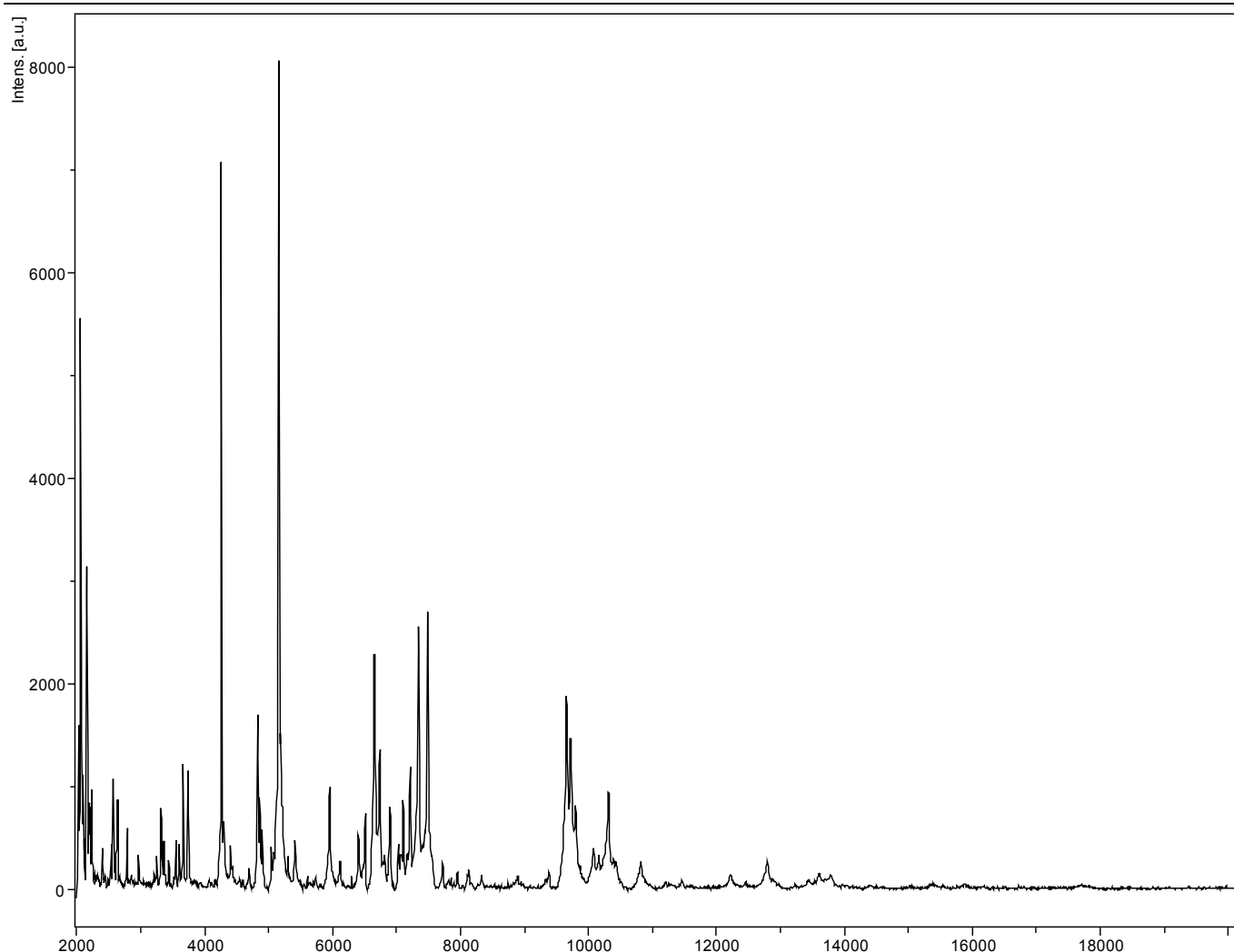
**Figure 4.** Reference mass spectrum from *B. massiliogorillae* strain G2$^T$. Spectra from 12 individual colonies were compared and a reference spectrum was generated.

The paired-end library was clonally amplified with 0.5 cpb and 1 cpb in 2 emPCR reactions with the GS Titanium SV emPCR Kit (Lib-L) v2 (Roche). The yield of the emPCR was 19.4%, slightly above the expected yield ranging from 5 to 20% recommended by the Roche procedure.

Approximately 790,000 beads for a ¼ region were loaded on the GS Titanium PicoTiterPlate PTP Kit 70x75 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 322,962 passed filter wells were obtained and generated 64.2 Mb of sequences with a length average of 310 bp. The passed filter sequences were assembled using Newbler with 90% identity and 40 bp as overlap. The final assembly identified 60 scaffolds generating a genome size of 4.6 Mb.

## Genome annotation

Open Reading Frames (ORFs) were predicted using Prodigal [30] with default parameters but the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched against the GenBank database [31] and the Clusters of Orthologous Groups (COG) databases using BLASTP. The tRNAScanSE tool [32] was used to find tRNA genes, whereas ribosomal RNAs were found by using RNAmmer [33] and BLASTn against the GenBank 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 *B. massiliogorillae* sp nov. strain G2$^T$ and another 3 *Bacillus* species (Table 6), we compared genomes pairwise and determined the mean percentage of nucleotide sequence identity among orthologous ORFs using BLASTn. Orthologous genes were detected using the Proteinortho software [34].

**Figure 5**. Gel view comparing *Bacillus massiliogorillae* G2$^T$ spectra with other members of the *Bacillus* genus (*B. thuringiensis*, *B. smithii*, *B. simplex*, *B. psychrosaccharolyticus*, *B. nealsonii*, *B.* *megaterium*, *B. lentus*, *B. flexus*, *B. firmus*, *B. circulans* and *B. benzoevorans*). 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 Gray scale 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.
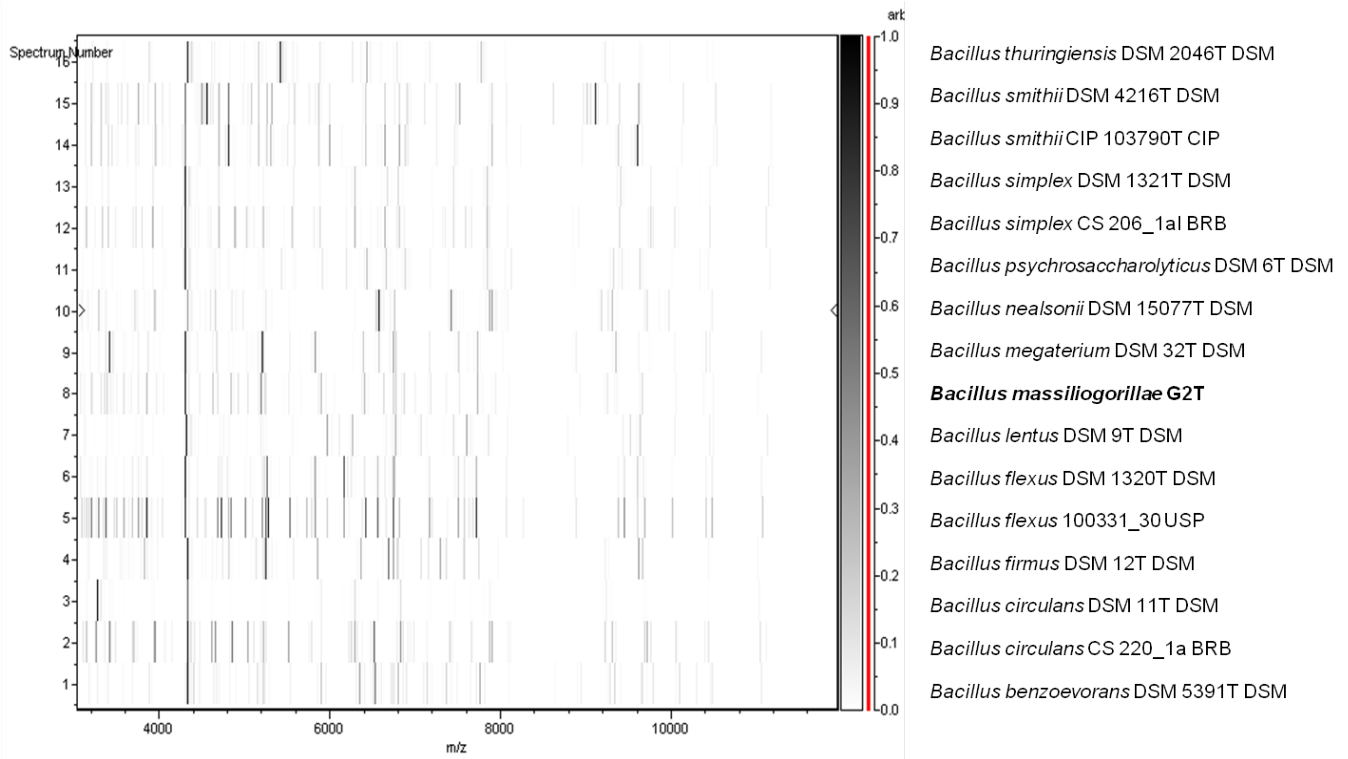


**Table 3.** Project information

| MIGS ID | Property | Term |
|---|---|---|
| MIGS-31 | Finishing quality | High-quality draft |
| MIGS-28 | Libraries used | 454 paired-end 3- kb libraries |
| MIGS-29 | Sequencing platform | 454 GS FLX Titanium |
| MIGS-31.2 | Sequencing coverage | 13× |
| MIGS-30 | Assemblers | Newbler version 2.5.3 |
| MIGS-32 | Gene calling method | Prodigal |
| | EMBL Date of Release | April 18, 2013 |
| | EMBL ID | CAVL000000000 |
| MIGS-13 | Project relevance | Study of the gorilla gut microbiome |

# Genome properties

The genome is 5,431,633 bp long (1 chromosome, but no plasmid) with a 34.95% G+C content (Figure 6 and Table 5). It is composed of 66 large contigs. Of the 5,276 predicted genes, 5,179 were protein-coding genes and 98 were RNAs (1 16S rRNA, 1 23S rRNA gene, 5 5S rRNA genes and 91 tRNA genes). A total of 3,801 genes (73.39%) were assigned a putative function (by COGS or by NR BLAST) and 368 genes were identified as ORFans (7.11%). The remaining genes were annotated as hypothetical proteins (666 genes, 12.86%). The distribution of genes into COGs functional categories is presented in Table 6. The properties and statistics of the genome are summarized in Tables 4 and 5.
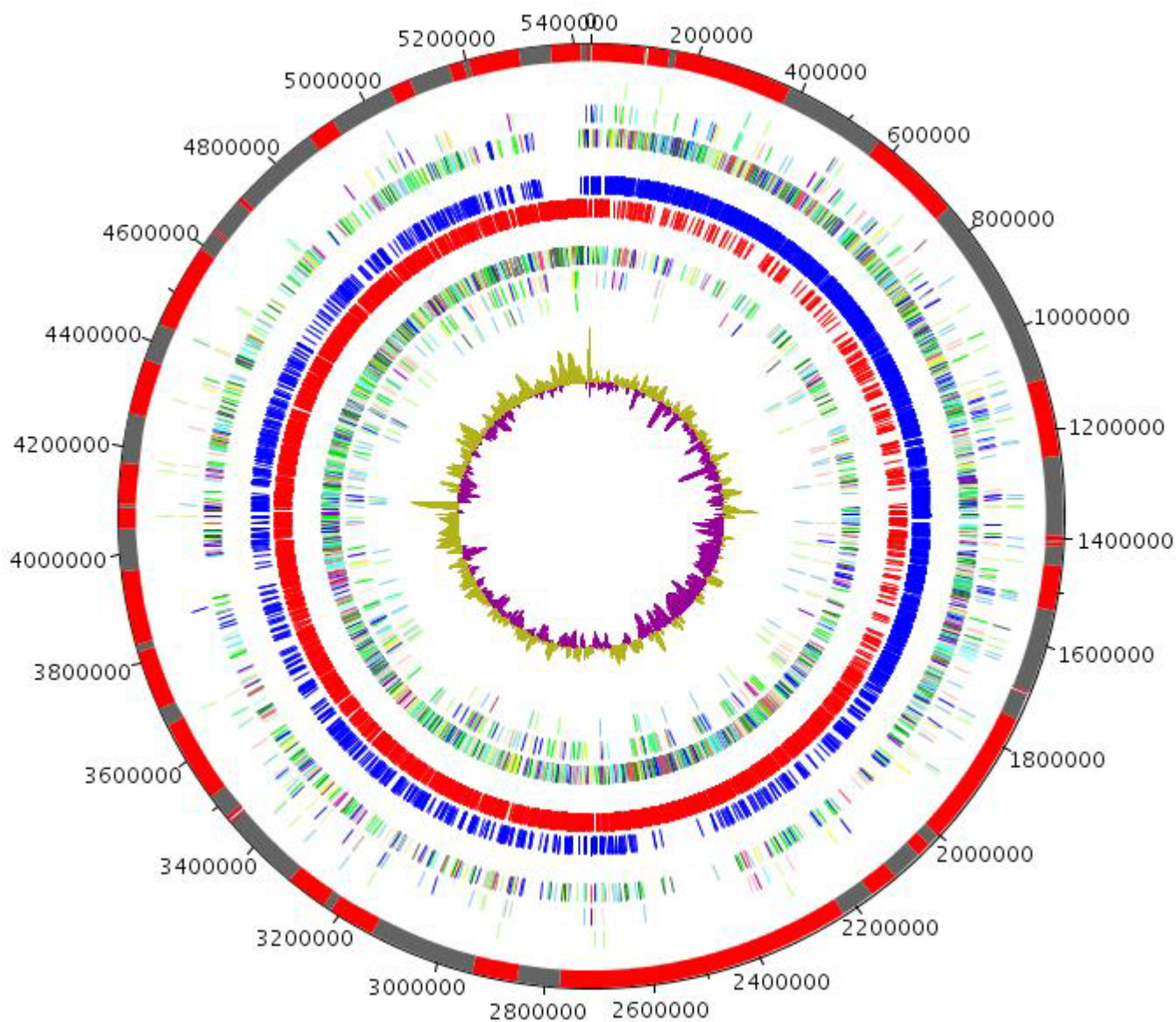


**Figure 6**. Graphical circular map of the genome. From outside in: contigs (red / grey), COG category of genes on the forward strand (three circles), genes on forward strand (blue circle), genes on the reverse strand (red circle), COG category on the reverse strand (three circles), GC content. The inner-most circle shows GC skew, purple and olive indicating negative and positive values, respectively.

**Table 4**. Nucleotide content and gene count levels of the genome

| Attribute | Value | % of total[a] |
|---|---|---|
| Genome size (bp) | 5,431,633 | 100 |
| Coding region (bp) | 4,561,287 | 83.98 |
| G+C content (bp) | 1,898,498 | 34.95 |
| Total genes | 5,276 | 100 |
| RNA genes | 98 | 1.84 |
| Protein-coding genes | 5,179 | 98.63 |
| Genes with function prediction | 3,801 | 73.39 |
| Genes assigned to COGs | 3,910 | 75.49 |
| Genes with peptide signals | 610 | 11.78 |
| Genes with transmembrane helices | 1,347 | 26.01 |

[a] The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome

**Table 5.** Number of genes associated with the 25 general COG functional categories

| Code | Value | % age[a] | Description |
|---|---|---|---|
| J | 180 | 3.48 | Translation, ribosomal structure and biogenesis |
| A | 0 | 0 | RNA processing and modification |
| K | 438 | 8.46 | Transcription |
| L | 191 | 3.69 | Replication, recombination and repair |
| B | 2 | 0.04 | Chromatin structure and dynamics |
| D | 42 | 0.81 | Cell cycle control, mitosis and meiosis |
| Y | 0 | 0 | Nuclear structure |
| V | 110 | 2.12 | Defense mechanisms |
| T | 275 | 5.31 | Signal transduction mechanisms |
| M | 182 | 3.51 | Cell wall/membrane biogenesis |
| N | 88 | 1.7 | Cell motility |
| Z | 0 | 0 | Cytoskeleton |
| W | 0 | 0 | Extracellular structures |
| U | 63 | 1.22 | Intracellular trafficking and secretion |
| O | 130 | 2.51 | Posttranslational modification, protein turnover, chaperones |
| C | 293 | 5.66 | Energy production and conversion |
| G | 247 | 4.77 | Carbohydrate transport and metabolism |
| E | 474 | 9.15 | Amino acid transport and metabolism |
| F | 110 | 2.12 | Nucleotide transport and metabolism |
| H | 177 | 3.42 | Coenzyme transport and metabolism |
| I | 188 | 3.63 | Lipid transport and metabolism |
| P | 300 | 5.79 | Inorganic ion transport and metabolism |
| Q | 133 | 2.57 | Secondary metabolites biosynthesis, transport and catabolism |
| R | 664 | 12.82 | General function prediction only |
| S | 344 | 6.64 | Function unknown |
| - | 1,269 | 24.50 | Not in COGs |

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

## Comparison with other *Bacillus* species genomes

Here, we compared the genome of *B. massiliogorillae* strain G2$^T$ with those of *B. psychrosaccharolyticus* strain ATCC 23296, *B. megaterium* strain DSM 319 and *B. thuringiensis* strain ATCC 10792 (Table 6). The draft genome of *B. massiliogorillae* is larger in size than those of *B. psychrosaccharolyticus* and *B. megaterium* (5.43 vs 4.59 and 5.1 Mb, respectively) and smaller in size than that of *B. thuringiensis* (5.43 vs 6.26 Mb). *B. massiliogorillae* has a lower G+C content than *B. psychrosaccharolyticus* (34.95% vs 38.8%) and *B. megaterium* (34.95% vs 38.1%) but slightly higher than that *B. thuringiensis* (34.95% vs 34.8%). The protein content of *B. massiliogorillae* is higher than those of *B. psychrosaccharolyticus* and *B.*

*megaterium* (5,179 vs 4,832 and 5,100 respectively) and fewer than that of *B. thuringiensis* (5,179 vs 6,243) (Table 6). In addition, *B. massiliogorillae* shares 1,936, 1,966 and 1,877 orthologous genes with *B. psychrosaccharolyticus*, *B. megaterium* and *B. thuringiensis* respectively (Table 6). The nucleotide sequence identity of orthologous genes ranges from 68.46 to 70.15% among *Bacillus* species, and from 69.28 to 70.15% between *B. massiliogorillae* and other *Bacillus* species (Table 6), thus confirming its new species status. Table 6 summarizes the number of orthologous genes and the average percentage of nucleotide sequence identity between the different genomes studied.

**Table 6.** The number of orthologous proteins shared between genomes$^†$

| | *B. massiliogorillae* | *B. psychrosaccharolyticus* | *B. megaterium* | *B. thuringiensis* |
|---|---|---|---|---|
| *B. massiliogorillae* | **5,179** | 70.15 | 69.28 | 69.66 |
| *B. psychrosaccharolyticus* | 1,936 | **4,832** | 68.74 | 68.46 |
| *B. megaterium* | 1,966 | 1,962 | **5,100** | 69.86 |
| *B. thuringiensis* | 1,877 | 1,873 | 1,903 | **6,243** |

$^†$Lower left triangle- shared orthologous, upper right triangle- average percentage similarity of nucleotides corresponding to orthologous proteins shared between genomes, bold- number of proteins per genome

## Conclusion

On the basis of phenotypic (Table 2), phylogenetic and genomic analyses (taxonogenomics) (Table 6), we formally propose the creation of *Bacillus massiliogorillae* sp. nov. that contains the strain G2$^T$. This strain has been found in a stool sample collected from gorilla in Cameroon.

### Description of *Bacillus massiliogorillae* sp. nov.

*Bacillus massiliogorillae* (ma.sil.io.go.ril'ae. L. gen. masc. n. *massiliogorillae*, combination of Massilia, the Latin name of Marseille, where strain G2$^T$ was isolated, and of Gorilla, the Latin name of the gorilla, from which the stool sample was obtained).

*B. massiliogorillae* is an aerobic Gram-variable bacterium. Optimal growth is achieved aerobically. No growth is observed in microaerophilic or anaerobic conditions. Growth occurs on axenic media between 25 and 45°C, with optimal growth observed at 37°C. Cells stain Gram-positive or negative, are rod-shaped, endospore-forming, motile and have a mean diameter of 1 μm (range 0.8 to 1.2 μm) and a mean length of 5.4 μm (range 3.2 to 7.5 μm). Colonies are grey opaque and 2-5 mm in diameter on blood-enriched BHI agar.

Catalase positive but oxidase negative. Using the API 50CH system (BioMerieux), a positive reaction is obtained for D-glucose, D-fructose, D-ribose, N-acetylglucosamin, amygdalin, arbutin, aesculin, salicin, cellobiose, maltose, D-lactose, D-trehalose, D-saccharose, and hydrolysis of starch. Using the API ZYM system, positive reactions are obtained for esterase (C4), esterase lipase (C8), phosphatase acid, α- glucosidase and N-acetyl-β-glucosaminidase. Using API 20NE, there are neither nitrate reduction nor indole production but urease reaction was positive. Susceptible to amoxicillin, nitrofurantoin, erythromycin, doxycycline, rifampin, vancomycin, gentamycin and imipenem but resistant to trimethoprim-sulfamethoxazole, ciprofloxacin, ceftriaxon and amoxicillin-clavulanic acid.

The G+C content of the genome is 34.95%. The 16S rRNA and genome sequences are deposited in GenBank under accession numbers JX650055 and CAVL00000000, respectively. The type strain G2$^T$ (= CSUR P206 = DSM 26159) was isolated from the fecal flora of a *Gorilla gorilla gorilla* from Cameroon.

## Acknowledgements

## References

1.  Cohn F. Untersuchungen über Bakterien. *Beitrage zur Biologie der Pflanzen Heft* 1872; **1**:127-224.

2.  Abstract for the genus *Bacillus*. NamesforLife, LLC. Retrieved September 22, 2013. http://doi.namesforlife.com/10.1601/nm.4857

3.  Jernigan JA, Stephens DS, Ashford DA, Omenaca C, Topiel MS, Galbraith M, Tapper M, Fisk TL, Zaki S, Popovic T, *et al*. Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States. *Emerg Infect Dis* 2001; **7**:933-944. PubMed http://dx.doi.org/10.3201/eid0706.010604

4.  Bottone EJ. *Bacillus cereus*, a volatile human pathogen. *Clin Microbiol Rev* 2010; **23**:382-398. PubMed http://dx.doi.org/10.1128/CMR.00073-09

5.  Mandell GL, Bennett JE, Dolin R. (2010) Principles and Practice of Infectious Diseases. Elsevier. 4320 p.

6.  Klee SR, Ozel M, Appel B, Boesch C, Ellerbrok H, Jacob D, Holland G, Leendertz FH, Pauli G, Grunow R, Nattermann H. Characterization of *Bacillus anthracis*-like bacteria isolated from wild great apes from Cote d'Ivoire and Cameroon. *J Bacteriol* 2006; **188**:5333-5344. PubMed http://dx.doi.org/10.1128/JB.00303-06

7.  Leendertz FH, Yumlu S, Pauli G, Boesch C, Couacy-Hymann E, Vigilant L, Junglen S, Schenk S, Ellerbrok H. A new *Bacillus anthracis* found in wild chimpanzees and a gorilla from West and Central Africa. *PLoS Pathog* 2006; **2**:e8. PubMed http://dx.doi.org/10.1371/journal.ppat.0020008

8.  Leendertz FH, Lankester F, Guislain P, Néel C, Drori O, Dupain J, Speede S, Reed P, Wolfe N, Loul S, *et al*. Anthrax in Western and Central African great apes. *Am J Primatol* 2006; **68**:928-933. PubMed http://dx.doi.org/10.1002/ajp.20298

9.  Sentausa E, Fournier PE. Advantages and limitations of genomics in prokaryotic taxonomy. *Clin Microbiol Infect* 2013. PubMed http://dx.doi.org/10.1111/1469-0691.12181

10. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archae, Bacteria*, and *Eukarya. Proc Natl Acad Sci USA* 1990; **87**:4576-4579. PubMed http://dx.doi.org/10.1073/pnas.87.12.4576

11. Gibbons NE, Murray RGE. Proposals concerning the higher taxa of *Bacteria. Int J Syst Bacteriol* 1978; **28**:1-6. http://dx.doi.org/10.1099/00207713-28-1-1

12. Garrity GM, Holt JG. The Road Map to the Manual. In: Garrity GM, Boone DR, Castenholz RW (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 1, Springer, New York, 2001, p. 119-169.

13. Murray RGE. The Higher Taxa, or, a Place for Everything...? In: Holt JG (ed), Bergey's Manual of Systematic Bacteriology, First Edition, Volume 1, The Williams and Wilkins Co., Baltimore, 1984, p. 31-34.

14. Ludwig W, Schleifer KH, Whitman WB. Class I. *Bacilli* class nov. *In:* De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 3, Springer-Verlag, New York, 2009, p. 19-20.

15. List of new names and new combinations previously effectively, but not validly, published. List no. 132. *Int J Syst Evol Microbiol* 2010; **60**:469-472. http://dx.doi.org/10.1099/ijs.0.022855-0

16. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; **30**:225-420. http://dx.doi.org/10.1099/00207713-30-1-225

17. Prévot AR. In: Hauderoy P, Ehringer G, Guillot G, Magrou. J., Prévot AR, Rosset D, Urbain A (eds), Dictionnaire des Bactéries Pathogènes, Second Edition, Masson et Cie, Paris, 1953, p. 1-692.

18. Fischer A. Untersuchungen über bakterien. *Jahrbücher für Wissenschaftliche Botanik* 1895; **27**:1-163.

19. Gibson T, Gordon RE. Genus I. *Bacillus* Cohn 1872, 174; Nom. gen. cons. Nomencl. Comm. Intern. Soc. Microbiol. 1937, 28; Opin. A. Jud. Comm. 1955, 39. In: Buchanan RE, Gibbons NE (eds), *Bergey's Manual of Determinative Bacteriology*, Eighth Edition, The Williams and Wilkins Co., Baltimore, 1974, p. 529-550.

20. Cohn F. Untersuchungen über Bakterien. *Beitr Biol Pflanz* 1872; **1**:127-224.

21. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29. PubMed http://dx.doi.org/10.1038/75556

22. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 2011; **28**:2731-2739. PubMed http://dx.doi.org/10.1093/molbev/msr121

23. Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* 2006; **33**:152-155.

24. Heyrman J, Logan NA, Rodríguez-Díaz M, Scheldeman P, Lebbe L, Swings J, Heyndrickx M, De Vos P. Study of mural painting isolates, leading to the transfer of '*Bacillus maroccanus*' and '*Bacillus carotarum*' to *Bacillus simplex*, emended description of *Bacillus simplex*, re-examination of the strains previously attributed to '*Bacillus macroides*' and description of *Bacillus muralis* sp. nov. *Int J Syst Evol Microbiol* 2005; **55**:119-131. PubMed http://dx.doi.org/10.1099/ijs.0.63221-0

25. Larkin JM, Stokes JL. Taxonomy of psychrophilic strains of *Bacillus*. *J Bacteriol* 1967; **94**:889-895. PubMed

26. De Paolis MR, Lippi D. Use of metabolic and molecular methods for the identification of a *Bacillus* strain isolated from paper affected by foxing. *Microbiol Res* 2008; **163**:121-131. PubMed http://dx.doi.org/10.1016/j.micres.2007.06.002

27. Roux V, El Karkouri K, Lagier JC, Robert C, Raoult D. Non-contiguous finished genome sequence and description of *Kurthia massiliensis* sp. nov. *Stand Genomic Sci* 2012; **7**:221-232. PubMed http://dx.doi.org/10.4056/sigs.3206554

28. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, Raoult D. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin Infect Dis* 2009; **49**:543-551. PubMed http://dx.doi.org/10.1086/600885

29. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, *et al*. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. PubMed http://dx.doi.org/10.1038/nbt1360

30. Prodigal. http://prodigal.ornl.gov

31. GenBank database.

32. Lowe TM, Eddy SR. t-RNAscan-SE: a program for improved detection of transfer RNA gene in genomic sequence. *Nucleic Acids Res* 1997; **25**:955-964. PubMed

33. 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**:3100-3108. PubMed http://dx.doi.org/10.1093/nar/gkm160

34. 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. PubMed http://dx.doi.org/10.1186/1471-2105-12-124