Non contiguous-finished genome sequence and description of \textit{Bacillus massiliosenegalensis} sp. nov.

Dhamodharan Ramasamy\(^1\), Jean-Christophe Lagier\(^1\), Aurore Gorlas\(^1\), Didier Raoult\(^1\) and Pierre-Edouard Fournier\(^1\)*

\(^1\)Aix-Marseille Université, URMITE, Faculté de médecine, Marseille, France

Corresponding author: Pierre-Edouard Fournier (pierre-edouard.fournier@univmed.fr)

Keywords: \textit{Bacillus massiliosenegalensis}, genome, culturomics, taxono-genomics

\textit{Bacillus massiliosenegalensis} strain JC6\(^T\) sp. nov. is the type strain of \textit{Bacillus massiliosenegalensis} sp. nov., a new species within the genus \textit{Bacillus}. This strain was isolated from the fecal flora of a healthy Senegalese patient. \textit{B. massiliosenegalensis} is an aerobic Gram-positive rod-shaped bacterium. Here we describe the features of this organism, together with the complete genome sequence and annotation. The 4,981,278-bp long genome comprises a 4,957,301-bp chromosome and a 23,977-bp plasmid. The chromosome contains 4,925 protein-coding and 72 RNA genes, including 4 rRNA genes. The plasmid contains 29 protein-coding genes.

Introduction

\textit{Bacillus massiliosenegalensis} strain JC6\(^T\) (= CSUR P151 = DSM 25957) is the type strain of \textit{B. massiliosenegalensis} sp. nov., a new species within the genus \textit{Bacillus}. This bacterium is a Gram-positive, aerobic, catalase-positive and indole-negative bacillus that was isolated from the stool of a healthy Senegalese patient as part of a study aimed at individually cultivating all human enteric bacterial species [1,2].

Currently, bacterial taxonomy relies on a combination of various genetic and phenotypic criteria. However, the three main genetic criteria that are used, including 16S rRNA gene-based phylogeny and nucleotide similarity [3,4], DNA-DNA hybridization [5] and \textit{G+C} content suffer significant drawbacks and their cutoffs are not applicable to all genera and species. Over recent years, the introduction of high-throughput genome sequencing and proteomic analyses [6] provided a source of exhaustive information about characterized bacterial isolates. Such data may now be included among the criteria used for taxonomic identification. We recently proposed to use a polyphasic approach to describe new bacterial taxa that is based on their genome sequence, MALDI-TOF spectrum and main phenotypic characteristics [7-25].

Here we present a summary classification and a set of features for \textit{B. massiliosenegalensis} sp. nov. strain JC6\(^T\) together with the description of the complete genomic sequencing and annotation. These characteristics support the creation of the species \textit{B. massiliosenegalensis}.

The genus \textit{Bacillus} (Cohn 1872) was created in 1872 [26] and currently consists of mainly Gram-positive, motile, and spore-forming bacilli. Currently, 173 \textit{Bacillus} species and 4 subspecies are validly published [27]. Members of the genus \textit{Bacillus} are ubiquitous bacteria, mostly isolated from environmental sources. However, several species are associated with humans, either as pathogens or commensals [28].

Classification and features

A stool sample was collected from a healthy 16-year-old male Senegalese volunteer patient living in Dielmo (rural village in the Guinean-Sudanian zone in Senegal), who was included in a research protocol. Written assent was obtained from this individual. No written consent was needed from his guardians for this study because he was older than 15 years old (in accordance with the previous project approved by the Ministry of Health of Senegal and the assembled village population and as published elsewhere [29].)

Both this study and the assent procedure were approved by the National Ethics Committee of
Senegal (CNERS) and the Ethics Committee of the Institut Fédératif de Recherche IFR48, Faculty of Medicine, Marseille, France (agreement numbers 09-022 and 11-017). Several other new bacterial species were isolated from this specimen using various culture conditions, including the recently described *Alistipes senegalensis*, *Alistipes timonensis*, *Anaerococcus senegalensis*, *Bacillus timonensis*, *Clostridium senegalense*, *Peptoniphilus timonensis* and *Paenibacillus senegalensis*, *Herbaspirillum massilense*, *Kurthia massilenis*, *Brevibacterium senegalense*, *Aeromicrobium massilense*, *Cellulomonas massilense*, *Senegalemassilia anaerobia*, *Peptontophilus obesi*, *Peptontophilus senegalensis*, *Enterobacter massiliensis*, *Alistipes obesi*, *Peptontophilus grossensis*, *Brevibacillus massiliensis* [7-25].

The fecal specimen was conserved at -80°C after collection. Strain JC6T was (Table 1) was isolated in January 2011 by cultivation on 5% sheep blood-enriched Brain Heart infusion (BHI) agar (Becton Dickinson, Heidelberg, Germany). The strain exhibited a 97.3% nucleotide sequence similarity with *B. siralis* (Pettersson et al. 2000), the phylogenetically closest *Bacillus* species (Figure 1).

Table 1. Classification and general features of *Bacillus massiliosenegalensis* strain JC6T according to the MIGS recommendations [30]

| MIGS ID | Property | Term | Evidence code |
|---------|----------|------|---------------|
| MIGS-6.3 | Salinity | Growth in BHI medium + 5% NaCl | IDA |
| MIGS-22 | Oxygen requirement | Aerobic | IDA |
| MIGS-6 | Habitat | Human gut | IDA |
| MIGS-15 | Biotic relationship | Free living | IDA |
| MIGS-14 | Isolation | Human feces | NAS |
| MIGS-4 | Geographic location | Senegal | IDA |
| MIGS-5 | Sample collection time | September 2010 | IDA |
| MIGS-4.1 | Latitude | 13.7167 | IDA |
| MIGS-4.1 | Longitude | -16.4167 | IDA |
| MIGS-4.3 | Depth | Surface | IDA |
| MIGS-4.4 | Altitude | 51 m above sea level | IDA |

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 [42]. 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.
Different growth temperatures (25, 30, 37, 45°C) were tested. Growth occurred between 25°C and 45°C, and the optimal growth was observed at 30°C. Colonies were translucent and 2 mm in diameter on blood-enriched Columbia agar. Growth of the strain was tested under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems respectively (BioMérieux) in the presence of air with or without 5% CO₂. Growth was achieved in aerobic condition (with or without CO₂), and weak growth was observed in microaerophilic and anaerobic conditions. Gram staining showed a rod-shaped Gram-positive bacterium (Figure 2). The motility test was positive by means of peritrichous flagella. Cells have a mean diameter of 0.65 µm and a mean length of 3.076 µm in electron microscopy (Figure 3).

Strain JC6° exhibited catalase activity but not oxidase activity. Using the API 50CH system, we observed positive reactions for aesculin, D-cellobiose, D-glucose, D-maltose, N-acetyl-glucosamine and D-trehalose. Using the API ZYM system, a positive reaction was observed for α-glucosidase and weak reactions were observed for alkaline phosphatase, esterase lipase, valine arylamidase and trypsin. Using the API 20E system, a positive reaction was observed for nitrate reduction and negative reactions were observed for indole production and urease. B. massiliosenegalensis is susceptible to amoxicillin, ceftriaxone, imipenem, trimethoprim/sulfamethoxazole, gentamicin, ciprofloxacin, rifampicin and vancomycin, but resistant to metronidazole and erythromycin. The differential phenotypic characteristics with other Bacillus species are summarized in Table 2.
Figure 2. Gram staining of *B. massiliosenegalensis* strain JC6\(^T\).

Figure 3. Transmission electron microscopy of *B. massiliosenegalensis* strain JC6\(^T\), using a Morgani 268D (Philips) at an operating voltage of 60kV. The scale bar represents 500 nm.
**Bacillus massiliosenegalensis** sp. nov.

Table 2. Differential characteristics of *B. massiliosenegalensis* strain JC6<sup>T</sup>, *B. timonensis* strain MM10403188<sup>T</sup>, *B. amyloliquefaciens* strain FZB42, *B. siralis* strain 171544<sup>T</sup> and *B. thuringiensis* strain BMB171

| Properties                      | *B. massiliosenegalensis* | *B. timonensis* | *B. amyloliquefaciens* | *B. siralis* | *B. thuringiensis* |
|---------------------------------|---------------------------|-----------------|-------------------------|--------------|-------------------|
| **Cell diameter (µm)**          | 0.65                      | 0.66            | 0.8                     | 0.5 to 0.8   | 1.0               |
| **Oxygen requirement**          | aerobic                   | aerobic         | aerobic                 | aerobic      | facultative anaerobic |
| Pigment production              | –                         | –               | –                       | +            | –                 |
| Gram stain                      | +                         | –               | +                       | +            | +                 |
| Salt requirement                | +                         | +               | +                       | +            | +                 |
| Motility                        | +                         | +               | +                       | –            | –                 |
| Endospore formation             | +                         | +               | +                       | +            | +                 |

**Production of**

| Acid phosphatase                | w                          | na              | +                       | na           | +                 |
| Catalase                        | +                         | –               | +                       | –            | +                 |
| Oxidase                         | –                          | +               | +                       | +            | +                 |
| Nitrate reductase               | +                          | na              | +                       | +            | +                 |
| Urease                          | -                          | na              | –                       | na           | +                 |
| β-galactosidase                 | na                         | +               | v                       | na           | –                 |
| N-acetyl-glucosamine            | +                          | +               | +                       | na           | +                 |

**Acid from**

| L-Arabinose                     | –                          | +               | +                       | –            | na                |
| Ribose                          | –                          | –               | +                       | –            | +                 |
| Mannose                         | –                          | –               | +                       | –            | +                 |
| Mannitol                        | –                          | –               | +                       | –            | +                 |
| Sucrose                         | –                          | –               | +                       | –            | v                 |
| D-glucose                       | +                          | –               | +                       | –            | +                 |
| D-fructose                      | –                          | –               | +                       | –            | +                 |
| D-maltose                       | +                          | –               | +                       | –            | +                 |
| D-lactose                       | –                          | +               | +                       | –            | +                 |

**Hydrolysis of**

| Gelatin                         | –                          | –               | +                       | +            | +                 |

**G+C content (mol%)**

| 37.6                           | 37.3                       | 46.48           | na                      | 35.18        |

**Habitat**

| human gut                      | human gut                  | Soil            | silage                  | soil         |

na = data not available; w = weak, v = variable reaction
Matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) MS protein analysis was carried out as previously described [8,43] using a Microflex spectrometer (Bruker Daltonics, Germany). Spectra were compared with the Bruker database that contained the main spectra from 3,769 bacteria including 129 spectra from 98 validly named Bacillus species. No significant score was obtained, thus suggesting that our isolate was not a member of a known species. We incremented our database with spectrum from strain JC6T (Figure 4). Finally, the gel view allows us to highlight the spectra differences with other Bacillus genera members 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 study of the human digestive flora aiming at isolating all bacterial species within human feces [1,2]. It was the 268th genome of a Bacillus species, and the first genome of *B. massiliosenegalensis* sp. nov. The Genbank accession number is CAHJ00000000 and consists of 102 contigs. Table 3 shows the project information and its association with MIGS version 2.0 compliance [44].

**Growth conditions and DNA isolation**

*B. massiliosenegalensis* sp. nov. strain JC6T, CSUR P151, DSM 25957, was grown aerobically on 5% sheep blood-enriched BHI agar at 37°C. Five petri dishes were spread and cultivated bacteria were resuspended in 3×100µl of G2 buffer (EZ1 DNA Tissue kit, Qiagen). A first mechanical lysis was performed by glass powder on the Fastprep-24 device (MP Biomedicals, USA) using 2×20 seconds cycles. DNA was then treated with 2.5 µg/µL lysozyme for 30 minutes at 37°C, and extracted using a BioRobot EZ1 Advanced XL (Qiagen). The DNA was then concentrated and purified using a Qiamp kit (Qiagen). The yield and the concentration was measured by the Quant-it Picogreen kit (Invitrogen) on the Genios_Tecan fluorometer at 36.8 ng/µl.

![Figure 4](http://standardsingenomics.org)  
**Figure 4.** Reference mass spectrum from *B. massiliosenegalensis* strain JC6T. Spectra from 12 individual colonies were compared and a reference spectrum was generated.
Bacillus massiliosenegalensis sp. nov.

Figure 5. Gel view comparing *B. massiliosenegalensis* sp. nov strain JC6 and other *Bacillus* species. The gel view displays the raw spectra of 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 with which a peak is displayed and the peak intensity in arbitrary units. Displayed species are indicated on the left.

| Table 3. Project information |
|-----------------------------|
| **MIGS ID** | **Property** | **Term** |
| MIGS-31 | Finishing quality | High-quality draft |
| MIGS-28 | Libraries used | 454 GS shotgun and paired-end 3-kb libraries |
| MIGS-29 | Sequencing platform | 454 GS FLX Titanium |
| MIGS-31.2 | Sequencing coverage | 54.25× |
| MIGS-30 | Assemblers | Newbler version 2.5.3 |
| MIGS-32 | Gene calling method | PRODIGAL |
| | Genbank Date of Release | June 1, 2012 |
| | Gold ID | Gi18838 |
| | NCBI project ID | CAHJ00000000 |
| MIGS-13 | Project relevance | Study of the human gut microbiome |
Genome sequencing and assembly

Both a shotgun and a 3kb paired-end libraries were constructed. The shotgun library was constructed with 500 ng of DNA as described by the manufacturer with the GS Rapid library Prep kit (Roche). For the paired-end library, 5µg of DNA was mechanically fragmented on the Hydroshear device (Digilab, Holliston, MA, USA) with an enrichment size at 3-4kb. The DNA fragmentation was visualized using an Agilent 2100 BioAnalyzer on a DNA labchip 7500 with an optimal size of 3.2kb. 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 optimal size at 555 bp. After PCR amplification through 17 cycles followed by double size selection, the single stranded paired-end library was then quantified on the Quant-it Ribogreen kit (Invitrogen) on the Agilent 2100 BioAnalyzer at 21pg/µL. The library concentration equivalence was calculated at 6.94E+07 molecules/µL. Libraries were stored at -20°C until further use.

The shotgun library was clonally amplified with 2 cpb in 4 emPCR reactions, whereas the 3kb paired-end library was amplified with 1cpb in 9 emPCR reactions and 0.5 cpb in 2 emPCR with the GS Titanium SV emPCR Kit (Lib-L) v2 (Roche). The yield of the shotgun emPCR reaction was 8.12%, and those of the 2 kinds of paired-end emPCR reactions were 7.8% and 11.2%, respectively. Such results were in the 5-20% quality range expected from the Roche procedure. For sequencing, the shotgun and paired-end libraries were loaded onto the 1/2 region and 4 1/4 regions of a PTP Picotiterplate 70×75 (Roche), respectively. The sequencing reactions were performed using a 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 969,014 passed filter wells were obtained and generated 274 Mb of sequences with a length average of 286 bp. These sequences were assembled using Newbler with 90% identity and 40 bp as overlap. The final assembly identified 31 scaffolds and 129 large contigs (>1,500 bp), generating a genome size of 5.05 Mb, which corresponds to a 54.25-fold coverage.

Genome annotation

Open Reading Frames (ORFs) were predicted using Prodigal [45] with default parameters. However, the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched against the GenBank [46] and Clusters of Orthologous Groups (COG) databases using BLASTP. tRNAs and ribosomal RNAs were predicted using the tRNAscanSE [47] and RNAmmer [48] tools, respectively. Lipoprotein signal peptides and numbers of transmembrane helices were predicted using SignalP [49] and TMHMM [50], respectively. ORFans were identified if their BLASTP E-value was lower than 1e-03 for alignment length greater than 80 amino acids. If alignment lengths were smaller than 80 amino acids, we used an E-value of 1e-05. Such parameter thresholds have already been used in previous works to define ORFans. Artemis [51] was used for data management and DNA Plotter [52] was used for visualization of genomic features. The alignment tool, Mauve, was used for multiple genomic sequence alignment [53]. To estimate the mean level of nucleotide sequence similarity at the genome level between _B. massiliosenegalensis_ and 5 other _Bacillus_ genomes (Table 4), orthologous proteins were detected using Proteinortho [54] and we compared genomes two by two and determined the mean percentage of nucleotide sequence identity among orthologous ORFs using BLASTn.

| Species                  | Strain | Genome accession number | Genome Size (Mb) | G+C content % |
|-------------------------|--------|-------------------------|------------------|---------------|
| _B. massiliosenegalensis_ | JC6    | CAHJ00000000            | 4,981,278        | 37.60         |
| _B. timonensis_          | MM10403188 | CAET00000000              | 4,632,049        | 37.30         |
| _B. cereus_              | 03BB102 | NC_012472                | 5,269,628        | 35.40         |
| _B. licheniformis_       | ATCC 14580 | NC_006270                | 4,222,597        | 46.20         |
| _B. subtilis_            | 168    | NC_000964                | 4,215,606        | 43.50         |
| _B. thuringiensis_       | BMB171 | NC_014171                | 5,330,088        | 35.30         |

Table 4. List of _Bacillus_ species genomes used for genomic comparison
Genome properties

The genome is 4,981,278 bp long (chromosome: 4,957,301 bp, plasmid: 23,977 bp) with a GC content of 37.60% (Figure 6 and Table 5). Of the 4,997 predicted chromosomal genes, 4,925 were protein-coding genes and 72 were RNAs. A total of 3,554 genes (72.16%) were assigned a putative function. ORFans accounted for 338 (6.86%) of the genes identified. The remaining genes were annotated as hypothetical proteins. The properties and statistics of the genome are summarized in Tables 5 and 6. The distribution of genes into COGs functional categories is presented in Table 6. The 23,977 bp-long plasmid contains 29 protein-coding genes.

Figure 6. Graphical circular map of the chromosome. From outside to the center: Genes on the forward strand (colored by COG categories), genes on the reverse strand (colored by COG categories), RNA genes (tRNAs green, rRNAs red), GC content, and GC skew.
Table 5. Nucleotide content and gene count levels of the chromosome

| Attribute                        | Value    | % of totala |
|----------------------------------|----------|-------------|
| Genome size (bp)                 | 4,957,301|             |
| DNA coding region (bp)           | 4,222,509| 85.17       |
| DNA G+C content (bp)             | 1,863,945| 37.6        |
| Total genes                      | 4,997    | 100         |
| RNA genes                        | 72       | 1.44        |
| Protein-coding genes             | 4,925    | 98.55       |
| Genes with function prediction   | 3,554    | 72.16       |
| Genes assigned to COGs           | 3,142    | 63.79       |
| Genes with peptide signals       | 223      | 4.52        |
| Genes with transmembrane helices | 1,329    | 26.59       |

aThe 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 6. Number of genes associated with the 25 general COG functional categories

| Code | Value | % of totala | Description                                      |
|------|-------|-------------|--------------------------------------------------|
| J    | 167   | 3.39        | Translation                                      |
| A    | 0     | 0           | RNA processing and modification                  |
| K    | 252   | 5.11        | Transcription                                    |
| L    | 171   | 3.47        | Replication, recombination and repair            |
| B    | 0     | 0           | Chromatin structure and dynamics                 |
| D    | 37    | 0.75        | Cell cycle control, mitosis and meiosis          |
| Y    | 0     | 0           | Nuclear structure                                |
| V    | 82    | 1.66        | Defense mechanisms                               |
| T    | 146   | 2.96        | Signal transduction mechanisms                   |
| M    | 162   | 3.28        | Cell wall/membrane biogenesis                    |
| N    | 23    | 0.46        | Cell motility                                    |
| Z    | 0     | 0           | Cytoskeleton                                     |
| W    | 0     | 0           | Extracellular structures                         |
| U    | 32    | 0.64        | Intracellular trafficking and secretion          |
| O    | 105   | 2.13        | Posttranslational modification, protein turnover, chaperones |
| C    | 238   | 4.83        | Energy production and conversion                 |
| G    | 186   | 3.77        | Carbohydrate transport and metabolism            |
| E    | 262   | 5.31        | Amino acid transport and metabolism              |
| F    | 72    | 1.46        | Nucleotide transport and metabolism              |
| H    | 112   | 2.27        | Coenzyme transport and metabolism                |
| I    | 85    | 1.72        | Lipid transport and metabolism                   |
| P    | 227   | 4.61        | Inorganic ion transport and metabolism           |
| Q    | 29    | 0.58        | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 407   | 8.26        | General function prediction only                 |
| S    | 347   | 7.04        | Function unknown                                 |
| -    | 412   | 8.36        | Not in COGs                                      |

aThe total is based on the total number of protein coding genes in the annotated genome.
Comparison with the genomes from other Bacillus species

Currently, more than 50 complete genome sequences are available for Bacillus species. Here, we compared the genome sequence of *B. massiliosenegalensis* strain JC6T with those of another 5 Bacillus species (Table 3). The genome sequence of *B. massiliosenegalensis* (4.98 Mb) is larger than those from *B. cereus*, *B. licheniformis*, *B. subtilis*, and *B. timonensis* (4.89, 4.22, 4.21, and 4.63 Mb, respectively) but smaller than *B. thuringiensis* (5.33 Mb). The G+C content of *B. massiliosenegalensis* is higher than *B. cereus*, *B. thuringiensis* and *B. timonensis* (37.6, 35.4, 35.3 and 37.3%, respectively) but lower than *B. licheniformis* and *B. subtilis* (46.2 and 43.5%, respectively). *B. massiliosenegalensis* has more predicted genes than *B. licheniformis*, *B. subtilis*, and *B. timonensis* (4,895, 4,295, 4,226, 4,610, respectively) but fewer genes than *B. thuringiensis* and *B. cereus* (5,341 and 5,331, respectively). However, the distribution of genes into COG categories was similar in all 6 compared genomes (Figure 7). In addition, *B. massiliosenegalensis* shared 2,022, 2,052, 2,009, 1,968 and 2,036 orthologous genes with *B. timonensis*, *B. cereus*, *B. licheniformis*, *B. subtilis* and *B. thuringiensis* respectively. The average nucleotide sequence identity ranged from 66.27 to 92.71% among the 6 Bacillus species, and from 67.05 to 70.42% between *B. massiliosenegalensis* and other Bacillus compared, thus confirming its new species status (Table 7). Finally, the plasmid genome of *B. massiliosenegalensis* is closely matching the *Bacillus cereus* Q1 plasmid (GenBank accession number CP000228).

![Figure 7. Distribution of functional classes of predicted genes on *B. massiliosenegalensis* (dark blue), *B. cereus* (green), *B. licheniformis* (red), *B. thuringiensis* (yellow) *B. timonensis* (brown) and *B. subtilis* (light blue) chromosomes according to the clusters of orthologous groups of proteins.](image-url)
Table 7. Genomic comparison of *B. massiliosenegalensis* with five other *Bacillus* species†.

| Species               | *B. massiliosenegalensis* | *B. timonensis* | *B. cereus* | *B. licheniformis* | *B. subtilis* | *B. thuringiensis* |
|-----------------------|----------------------------|-----------------|-------------|--------------------|--------------|-------------------|
| *B. massiliosenegalensis* | 4,895                      | 2,022           | 2,052       | 2,009              | 1,968        | 2,036             |
| *B. timonensis*        | 70.42                      | 4,610           | 2,038       | 1,976              | 1,962        | 2,028             |
| *B. cereus*            | 68.94                      | 69.35           | 5,331       | 2,116              | 2,111        | 2,779             |
| *B. licheniformis*     | 67.05                      | 67.19           | 66.29       | 4,295              | 2,350        | 2,095             |
| *B. subtilis*          | 67.70                      | 67.93           | 67.28       | 74.01              | 4,226        | 2,087             |
| *B. thuringiensis*     | 68.91                      | 69.32           | 92.71       | 66.27              | 67.30        | 5,341             |

†The numbers of orthologous protein shared between genomes (above diagonal), average percentage similarity of nucleotides corresponding to orthologous protein shared between genomes (below diagonal) and the numbers of proteins per genome (bold).

### Conclusion

On the basis of phenotypic, phylogenetic and genome analysis, we formally propose the creation of *Bacillus massiliosenegalensis* sp. nov. which contains the strain JC6\textsuperscript{t}. This strain has been found in Senegal.

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

*Bacillus massiliosenegalensis* (mas.si.li.o.se.ne.gal.en'sis. L. gen. masc. n. *massiliosenegalensis*, contraction of the Latin names of Marseille and Senegal, where strain JC6\textsuperscript{t} was cultivated and collected, respectively.) *B. massiliosenegalensis* is an aerobic Gram-positive bacterium. Optimal growth is achieved aerobically and weak growth is observed under 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, are rod-shaped, endospore-forming, motile and have a mean diameter of 0.65 µm and a mean length of 3.076 µm. Peritrichous flagellae were observed. Colonies are translucent and 2 mm in diameter on blood-enriched BHI agar.

No oxidase activity detected. Negative for indole. Presence of catalase activity. Using the API 50CH system, positive reactions are observed for aesculin, D-cellobiose, D-glucose, D-maltose, N-acetyl-glucosamine and D-trehalose. Using the API ZYM system, a positive reaction is observed for α-glucosidase and weak reactions are observed for alkaline phosphatase, esterase lipase, valine arylamidase and trypsin. Using the API 20E system, a positive reaction is observed for nitrate reduction and a negative reaction is observed for urease. *B. massiliosenegalensis* is susceptible to amoxicillin, ceftriaxone, imipenem, trimethoprim/sulfamethoxazole, gentamicin, ciprofloxacin, rifampicin and vancomycin, but resistant to metronidazole and erythromycin.

The G+C content of the genome is 37.60%. The 16S rRNA and genome sequences are deposited in Genbank and EMBL under accession numbers JF824800 and CAHJ00000000, respectively. The type strain is JC6\textsuperscript{t} (= CSUR P151 = DSM 25957) was isolated from the fecal flora of a healthy Senegalese patient.

### Acknowledgements

The authors thank the Xegen Company (www.xegen.fr) for automating the genomic annotation process. This study was funded by the Méditerranée Infection Foundation.

### References

1. Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, Bittar F, Fournous G, Gimenez G, Maraninchi M, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect* 2012; 18:1185-1193. [PubMed](https://www.ncbi.nlm.nih.gov/pubmed/22169938)

2. Dubourg G, Lagier JC, Armougom F, Robert C, Hamad I, Brouqui P. The gut microbiota of a pa-
Bacillus massiliosenegalensis sp. nov.

tient with resistant tuberculosis is more comprehensively studied by culturotics than by
metagenomics. *Eur J Clin Microbiol Infect Dis* 2013; 32:637-645. PubMed 
http://dx.doi.org/10.1007/s10096-012-1787-3

3. Tindall BJ, Rossello-Móra R, Busse HJ, Ludwig W, Kämpfer P. Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol* 2010; 60:249-266. PubMed http://dx.doi.org/10.1099/ijs.0.016949-0

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

5. Rossello-Mora R. DNA-DNA Reassociation Methods Applied to Microbial Taxonomy and Their Critical Evaluation. In: Stackebrandt E (ed), Molecular Identification, Systematics, and population Structure of Prokaryotes. Springer, Berlin, 2006, p. 23-50.

6. Welker M, Moore ER. Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. *Syst Appl Microbiol* 2011; 34:2-11. PubMed http://dx.doi.org/10.1016/j.syapm.2010.11.013

7. Kokcha S, Mishra AK, Lagier JC, Million M, Leroy Q, Raoult D, Fournier PE. Non-contiguous-finished genome sequence and description of *Bacillus timonensis* sp. nov. *Stand Genomic Sci* 2012; 6:346-355. PubMed http://dx.doi.org/10.4056/sigs.2776064

8. 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; 6:116-125. PubMed http://dx.doi.org/10.4056/sigs.2415480

9. Mishra AK, Gimenez G, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Alistipes senegalensis* sp. nov. *Stand Genomic Sci* 2012; 6:304-314. http://dx.doi.org/10.4056/sigs.2625821

10. 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; 6:315-324. PubMed

11. 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; 6:386-395. PubMed

12. Mishra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous-finished genome sequence and description of *Peptoniphilus timonensis* sp. nov. [epub.]. *Stand Genomic Sci* 2012; 7:1-11. PubMed http://dx.doi.org/10.4056/sigs.2956294

13. Mishra AK, Lagier JC, Rivet R, Raoult D, Fournier PE. Non-contiguous-finished genome sequence and description of *Paenibacillus senegalensis* sp. nov. *Stand Genomic Sci* 2012; 7:200-209. PubMed

14. Lagier JC, Gimenez G, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Herbaspirillum massiliense* sp. nov. *Stand Genomic Sci* 2012; 7:233-245. PubMed http://dx.doi.org/10.4056/sigs.3206554

15. 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.3256677

16. Kokcha S, Ramasamy D, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Brevibacterium senegalense* sp. nov. *Stand Genomic Sci* 2012; 7:258-270. PubMed http://dx.doi.org/10.4056/sigs.3316719

17. Ramasamy D, Kokcha S, Lagier JC, N’Guyen TT, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Aeromicrobium massiliense* sp. nov. *Stand Genomic Sci* 2012; 7:343-356. PubMed http://dx.doi.org/10.4056/sigs.3246665

18. Lagier JC, Ramasamy D, Rivet R, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Cellulomonas massiliensis* sp. nov. *Stand Genomic Sci* 2012; 7:357-369. PubMed http://dx.doi.org/10.4056/sigs.32766871

276 Standards in Genomic Sciences
21. Mishra AK, Lagier JC, Nguyen TT, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of Peptoniphilus senegalensis sp. nov. Stand Genomic Sci 2013; 7:370-381. http://dx.doi.org/10.4056/sigs.3366764

22. Lagier JC, El Karkouri K, Mishra AK, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of Enterobacter massiliensis sp. nov. Stand Genomic Sci 2013; 7:399-412. http://dx.doi.org/10.4056/sigs.3396830

23. Hugon P, Ramasamy D, Lagier JC, Rivet R, Couderc C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of Alstipes obesi sp. nov. Stand Genomic Sci 2013; 7:427-439. http://dx.doi.org/10.4056/sigs.3336746

24. Mishra AK, Hugon P, Robert C, Couderc C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of Peptoniphilus grossensis sp. nov. Stand Genomic Sci 2012; 7:320-330. PubMed

25. Hugon P, Mishra AK, Lagier JC, Nguyen TT, Couderc C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of Brevibacillus massiliensis sp. nov. Stand Genomic Sci 2013; 8:1-14. http://dx.doi.org/10.4056/sigs.3466975

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

27. Taxon Abstract for the genus Bacillus. NamesforLife, LLC. Retrieved March 10, 2013. doi:10.1601/nm.4857

28. Didelot X, Barker M, Falush D, Priest FG. Evolution of pathogenicity in the Bacillus cereus group. Syst Appl Microbiol 2009; 32:81-90. PubMed http://dx.doi.org/10.1016/j.syapm.2009.01.001

29. Trape JF, Tall A, Diagne N, Ndiall O, Ly AB, Faye J, Dieye-Ba F, Roucher C, Bouganali C, Badiane A, et al. Malaria morbidity and pyrethroid resistance after the introduction of insecticide-treated bednets and artemisinin-based combination therapies: a longitudinal study. Lancet Infect Dis 2011; 11:925-932. PubMed http://dx.doi.org/10.1016/S1473-3099(11)70194-3

30. 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

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

32. Skerman VBD, Sneath PHA. Approved list of bacterial names. Int J Syst Bacteriol 1980; 30:225-420. http://dx.doi.org/10.1099/00207713-30-1-225

33. Murray RGE. The higher taxa, or, a place for everything…? In: Krieg NR, Holt JG (eds), Bergey’s Manual of Systematic Bacteriology, First edition, volume 1, The Williams & Wilkins Co., Baltimore, 1984, p. 31-34.

34. 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

35. Garrity GM, Holt J. 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.

36. 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/ijsem.0.022855-0

37. 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.

38. Prevot AR. Dictionnaire des bactéries pathogens. In: Hauduroy P, Ehringer G, Guillot G, Magrou J, Prevot AR, Rosset, Urbain A (eds). Paris, Masson, 1953, p.1-692.

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

40. 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.

41. Mathews WC, Caperna J, Toerner JG, Barber RE, Morgenstern H. Neutropenia is a risk factor for

http://standardsingenomics.org
Bacillus massiliosenegalensis sp. nov.

Gram-negative Bacillus bacteremia in human immunodeficiency virus-infected patients: results of a nested case-control study. *Am J Epidemiol* 1998; 148:1175-1183. PubMed http://dx.doi.org/10.1093/oxfordjournals.aje.a009606

42. 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

43. 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

44. 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

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

46. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res* 2012; 40:D48-D53. PubMed http://dx.doi.org/10.1093/nar/gkr1202

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

48. 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

49. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004; 340:783-795. PubMed http://dx.doi.org/10.1016/j.jmb.2004.05.028

50. 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; 305:567-580. PubMed http://dx.doi.org/10.1006/jmbi.2000.4315

51. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. Artemis: sequence visualization and annotation. *Bioinformatics* 2000; 16:944-945. PubMed http://dx.doi.org/10.1093/bioinformatics/16.10.944

52. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 2009; 25:119-120. PubMed http://dx.doi.org/10.1093/bioinformatics/btn578

53. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004; 14:1394-1403. PubMed http://dx.doi.org/10.1101/gr.2289704

54. 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