Genome sequence and description of *Paenibacillus ihuae* strain GD6 sp. nov., isolated from the stool of a 62-year-old Frenchman

C. Al-Bayssari, G. Dubourg, T. Cimmino, L. Okdah, N. Mathlouthi, T. T. Nguyen, C. Andrieu, J. Rathored, P.-E. Fournier, D. Raoult and J.-M. Rolain

Aix-Marseille Université IRD, APHM, MEPHI, IHU Méditerranée Infection, Marseille, France

Abstract

*Paenibacillus ihuae* strain GD6 (=CSUR P892 = DSMZ 45751T) is the new type strain collected from the stool of a 69-year-old Frenchman admitted to an intensive care unit and receiving a 10-day course of imipenem at the time of stool collection. This is a Gram-positive, facultative anaerobic, rod-shaped bacterium. We describe here the features of this organism, together with its complete genome sequence and annotation. The genome size is 6,719,043 bp with 49.6% G+C content and contains 6,211 protein-coding and 65 sRNA genes, including four 5S rRNA genes, one 16S rRNA gene and one 23S rRNA gene.

Keywords: Culturomics, *Paenibacillus ihuae*, taxonogenomics

Original Submission: 31 August 2017; Revised Submission: 14 February 2018; Accepted: 20 February 2018

Article published online: 27 February 2018

*Paenibacillus* is a genus of facultative anaerobic, endospore-forming Gram-positive bacteria, originally included within the genus *Bacillus* and then reclassified as a separate genus in 1993 [1]. Since this classification, additional transfer to the genus *Paenibacillus* and proposal for novel strains to be designated as *Paenibacillus* species have increased. Bacteria belonging to this genus are commonly found in the environment such as soil, water, rhizosphere, vegetable matter, forage and insect larvae, but few species have been linked to infections in humans [2,3], and it has been shown to produce a wide range of peptide antibiotics [4].

*Paenibacillus ihuae* strain GD6 was isolated from the stool of a 69-year-old man admitted to the intensive care unit and receiving a 10-day course of imipenem at the time of stool collection as part of a culturomics study aiming to isolate all bacterial species present in the human gut [5].

Here we present a summary of the classification and set of features for *Paenibacillus ihuae* sp. nov. strain GD6 (=CSUR P892 = DSMZ 45751T), together with the description of the complete genomic sequencing and annotation. These characteristics support the description of *Paenibacillus ihuae* sp. nov.

The stool sample was collected from a 69-year-old man living in France. The study was approved by the ethics committee of the Institut Fédératif de Recherche IFR48, Faculty of Medicine, Marseille, France, under agreement 09-002.

The faecal specimen was preserved at −80°C after collection. Strain GD6 was isolated on Columbia agar supplemented with 5% sheep’s blood (bioMérieux, Marcy l’Etoile, France) in aerobic condition at 37°C. Strain GD6 exhibited a 97.4% 16S rRNA sequence identity with *Paenibacillus typhae* strain xj7 5 (NR_109462.1), the phylogenetically closest bacterial species with standing in nomenclature (Fig. 1). Its 16S rRNA sequence was deposited in GenBank under accession number JX424768. As recommended by Stackebrandt and Ebers [6], this value was lower than the 98.7% 16S rRNA gene sequence threshold to delineate a new species without carrying out DNA-DNA hybridization, and a new species was thus identified. The spectrum of strain GD6 was added to our matrix-assisted desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) database.

The stool sample was diluted in phosphate-buffered saline (Life Technologies, Carlsbad, CA, USA). Obtained inoculum
(100 μL) was incubated for 24 to 48 hours on 5% sheep’s blood–enriched Columbia agar (bioMérieux) at 37°C. Growth was tested under aerobic and anaerobic conditions using AnaeroGen Compact (bioMérieux). Gram staining and electron microscopy were performed with a TechnaiG2 Cryo device (FEI Company, Limeil-Brévannes, France) at an operating voltage of 200 keV (Fig. 2). Cells were grown on 5% sheep’s blood–enriched agar for 24 hours. A bacterial suspension was prefixed in 5% (v/v) glutaraldehyde in phosphate-buffered saline (Thermo Fisher Scientific, Waltham, MA, USA) for at least 1 hour at room temperature, washed in the same buffer and then stained with 1% (w/v) ammonium molybdate 1%. Oxidase (Becton Dickinson, Le Pont-de-Claix, France) and catalase (bioMérieux) assays were performed separately. Biochemical tests were performed using an APIZYM strip (bioMérieux) and an API50CH strip (bioMérieux). In vitro susceptibility to antibiotics was determined using the disc diffusion method (i2a, Montpellier, France) on Muller-Hinton agar with 5% blood.

Colonies obtained were isolated on 5% sheep’s blood–enriched Columbia agar and were identified by MALDI-
Differential characteristics of Genomic comparison of generating a genome size of 6.71 Mb. It led to 11 scaffolds and 564 large contigs (>1500 bp), from Roche (Basel, Switzerland) with 90% identity and 40 bp as strategy. The assembly was performed using the gsAssembler from strain GD6 to our database. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Genome was annotated by RAST [8]. The predicted bacterial protein sequences were searched against the GenBank database and the Clusters of Orthologous Groups (COGs) databases using BLASTp (E value 1e-03, coverage 0.7, identity percentage 30%). The trRNAscanSE tool [9] was used to find tRNA genes, whereas ribosomal RNAs were found by RNAmmer [10]. The resistome was analysed with the ARG-ANNOT database [11]. The exhaustive bacteriocin database available in our laboratories (Bacteriocins of the Unité des Maladies Infectieuses et Tropicales Emergentes (URMITE); http://drissifatima.wixsite.com/bacteriocins) was performed by collecting all currently available sequences from the databases and from the National Center for Biotechnology Information. Protein sequences from this database allowed putative bacteriocins from human gut microbiota to be identified using BLASTp methodology [12].

The presence of polyketide synthases (PKS) and non-ribosomal peptide synthetase (NRPS) was analysed by MALDI-TOF MS. MALDI-TOF MS identification, measurement and analysis were performed as previously described [7] using a Microflex spectrometer (Bruker Daltonics, Bremen, Germany). No significant MALDI-TOF MS score was obtained for strain GD6 against the Bruker database, suggesting that our isolate was not a member of a known species. We added the spectrum to strain GD6 to our database.

Genomic DNA of strain GD6 was sequenced using MiSeq Technology (Illumina, San Diego, CA, USA) with the mate-pair strategy. The assembly was performed using the gsAssembler from Roche (Basel, Switzerland) with 90% identity and 40 bp as overlap. It led to 11 scaffolds and 564 large contigs (>1500 bp), generating a genome size of 6.71 Mb.

TABLE 1. Differential characteristics of Paenibacillus ihuae strain GD6 and phylogenetically close members of other Paenibacillus species

| Test | P. ihuae GD6 | P. graminis DSM 13188 | P. polymyxa DSM 365 | P. massiliensis DSM 16942 |
|------|--------------|------------------------|---------------------|---------------------------|
| Catalase | + | + | + | + |
| Spore presence | + | + | + | + |
| Anaerobic growth | + | + | + | + |
| Growth in presence of: | | | | |
| NaCl 5% | − | − | − | − |
| Glycerol | + | + | + | + |
| D-Arabino | + | − | − | − |
| L-Arabinose | + | + | + | + |
| D-Xylose | + | + | + | + |
| D-Ribose | − | − | + | + |
| D-Trehalose | + | + | + | + |
| D-Galactose | + | + | − | − |
| Starch | + | + | − | − |
| D-Glucose | + | + | + | + |
| D-Lactose | + | + | + | + |
| D-Mannose | + | + | + | + |
| D-Rhamnose | − | − | + | + |
| D-Mannitol | + | + | + | + |
| Inulin | − | − | + | + |
| D-Raffinose | + | + | − | − |
| D-Turanose | + | + | + | + |
| D-Melezitose | + | + | + | + |
| Methyl α-D-mannopyranoside | − | − | − | − |
| Methyl α-D-gluopyranoside | − | − | − | − |

+ , positive result; −, negative result; NA, data not available.

TABLE 2. Genome features of Paenibacillus ihuae strain GD6

| Attribute | Value |
|-----------|-------|
| Size (bp) | 6,719,043 |
| G+C content (bp) | 49.6 |
| RNAs gene | 65 |
| SS rRNA | 4 |
| 16S rRNA | 1 |
| 23S rRNA | 1 |
| Protein-coding gene | 621 |
| Genes with unknown function | 445 |
| Genes assigned to COGS | 5284 |
| Genes associated to PKS or NRPS | 1 |
| Genes associated to toxin/antitoxin | 1 |
| Genes associated to resistome | 1 |

COGs, Clusters of Orthologous Groups database; G+C, guanine cytosine; NRPS, nonribosomal peptide synthetase; PKS, polyketide synthase.

TABLE 3. Genomic comparison of Paenibacillus ihuae strain GD6 with other Paenibacillus species

| Species | Strain | Size (Mb) | G+C% | Gene content |
|---------|--------|-----------|------|-------------|
| P. ihuae | GD6 | 6.71 | 49.6 | 621 |
| P. borealis | DSM 13188 | 8.15 | 51.4 | 7007 |
| P. graminis | RSA19 | 6.98 | 50.30 | 6379 |
| P. polymyxa | DSM 365 | 5.78 | 45.5 | 5031 |
| P. massiliensis | DSM 16942 | 6.39 | 48.5 | 5461 |

TABLE 4. Distribution of genes into COGs functional categories

| Code | Value | % of total | Description |
|------|-------|------------|-------------|
| A    | 2     | 0.023      | RNA processing and modification |
| B    | 1     | 0.016      | Chromatin structure and dynamics |
| C    | 204   | 3.28       | Energy production and conversion |
| D    | 50    | 0.80       | Cell cycle control, cell division, chromosome partitioning |
| E    | 364   | 5.90       | Amino acid transport and metabolism |
| F    | 129   | 2.07       | Nucleotide transport and metabolism |
| G    | 591   | 9.51       | Carbohydrate transport and metabolism |
| H    | 193   | 3.10       | Coenzyme transport and metabolism |
| I    | 108   | 1.73       | Lipid transport and metabolism |
| J    | 222   | 3.60       | Translation, ribosomal structure and biogenesis |
| K    | 578   | 9.30       |Ribosomal protein synthesis, ribosome biogenesis |
| L    | 195   | 3.13       | Replication, recombination and repair |
| M    | 263   | 4.23       |Cell wall/membrane/envelope biogenesis |
| N    | 97    | 1.56       |Cell motility |
| O    | 136   | 2.23       | Posttranslational modification, protein turnover, chaperones |
| P    | 290   | 4.67       | Inorganic ion transport and metabolism |
| Q    | 91    | 1.46       |Secondary metabolites biosynthesis, transport and catabolism |
| R    | 641   | 10.32      |General function prediction only |
| S    | 445   | 7.16       |Function unknown |
| T    | 484   | 7.89       |Signal transduction mechanisms |
| U    | 65    | 1.04       | Intracellular trafficking, secretion and vesicular transport |
| V    | 134   | 2.15       |Defense mechanisms |
| W    | 0     | 0.00       |Extracellular structures |
| X    | 0     | 0.00       |Nuclear structure |
| Z    | 927   | 14.92      |Not in COGs |

COGs, Clusters of Orthologous Groups database.
discriminating genes with large size using a database realized in our laboratory; predicted proteins were compared against nonredundant GenBank database using BLASTp and finally examined using antiSMASH [15]. PHAST was used to identify phage sequences [13].

Phylogenetic relationships with closely related species were determined by MEGA6. The evolutionary history was concluded by using the maximum likelihood method based on the JTT matrix-based model. We compared the genome sequence of Paenibacillus ihuae strain GD6 with those of Paenibacillus graminis strain RSA19T (NZ_ASSG00000000.1), Paenibacillus polymyxa strain DSM 365T (NZ_JMIQ00000000.1), Paenibacillus massiliensis strain DSM 16942T (NZ_ARIL00000000.1), Paenibacillus typhae strain CGMCC 1.11012T (NZ_FNDX00000000.1) and Paenibacillus borealis strain DSM 13188T (NZ_CP009285.1).

*Paenibacillus ihuae* growth was obtained either on aerobic and anaerobic conditions on 5% sheep’s blood—enriched Columbia agar at 37°C. Gram staining showed elongated-shaped Gram positive bacilli. The motility test was positive. Cells grown in trypticase soy broth medium have flagellum, as observed by electron microscopy (Fig. 2). Strain GD6 exhibits positive catalase and negative oxidase activity. Acid production was also observed using an API 50 CH strip (bioMérieux).

Differential phenotype characteristics between *P. ihuae* and other species are shown in Table 1. *Paenibacillus ihuae* strain GD6 was resistant to oxacillin and metronidazole but was susceptible to cephalosporins, carbapenems, vancomycin, telcoplanin, lincomycin, gentamycin, amikacin, trimethoprim/sulfamethoxazole, rifampicin and fosfomycin.

The genome of *Paenibacillus ihuae* strain GD6 is 6,719,043 bp long with 49.6% G+C content. It is composed of 13 scaffolds.

**FIG. 3.** Phylogenetic of cluster representative of nonribosomal peptide synthase *Paenibacillus ihuae* strain GD6. Comparison of nonribosomal peptide synthetase (NRPS) of *Paenibacillus* sp. GD6 with closer cluster in other species. Percentages of identity are indicated for homologs found in cluster of WP 054942807.1 nonribosomal peptide synthetase of *Paenibacillus* sp. GD6 and closer cluster in other species.
(CTED00000000.1) comprising 600 contigs (LN831198 to LN831210). The phylogenetic tree highlights the position based on 16S rDNA of *Paenibacillus ihuae* strain GD6 (Fig. 1). A total of 6211 protein-coding genes are annotated; 65 were RNAs (including four 5S, one 16S and one 23S). The properties of the genome and the comparison with other genomes are summarized in Tables 2 and 3, respectively. The distribution of genes into COGs functional categories is presented in Table 4.

The analysis of the resistome shows the absence of resistance genes. *In silico* analysis for PKS and NRPS revealed the presence of a NRPS organized as a highly modular mode in a massive multidomain enzyme organized with upstream enzyme clustering of condensation (C), adenylation (A), thiolation (T) or peptidyl carrier. The nonribosomal polyketide synthase (NRPKs) had a size of 3369 bp and a G+C content of 48%. This cluster showed 88% similarity with the NRPKs of *Paenibacillus* sp. FSL R5-0912 (Fig. 3).

Here we compared the genome of *Paenibacillus ihuae* strain GD6 with those of *Paenibacillus graminis* RSA19\(^T\), *Paenibacillus polymyxa* DSM 365\(^T\), *Paenibacillus massiliensis* DSM 16942\(^T\), *Paenibacillus typhae* CGMCC 1.11012\(^T\) and *Paenibacillus borealis* DSM 13188\(^T\) (Table 5).

|                          | *P. ihuae* | *P. graminis* | *P. polymyxa* | *P. massiliensis* | *P. typhae* | *P. borealis* |
|--------------------------|------------|---------------|---------------|-------------------|-------------|--------------|
| GD6                      | 100 ± 00%  | 23.2 ± 2.5%   | 21.1 ± 2.5%   | 19.6 ± 2.4%       | 20.9 ± 3.5% | 24.9 ± 2.3   |
| *P. graminis* RSA19\(^T\)| 100 ± 00%  | 23.8 ± 2.5%   | 21 ± 3.4%     | 19.2 ± 2.4%       | 20.5 ± 2.2% | 21.2 ± 2.3   |
| *P. polymyxa* DSM 365\(^T\)| 100 ± 00%  | 20.2 ± 2.4%   | 100 ± 00%     | 100 ± 00%         | 21.7 ± 2.3  | 100 ± 00%    |
| *P. massiliensis* DSM 16942\(^T\)| 100 ± 00%  | 20.7 ± 2.3    | 21.1 ± 2.5%   | 19.6 ± 2.4%       | 24.9 ± 2.3  |
| *P. typhae* CGMCC 1.11012\(^T\)| 100 ± 00%  | 23.7 ± 2.3    | 21.1 ± 2.5%   | 19.6 ± 2.4%       | 24.9 ± 2.3  |
| *P. borealis* DSM 13188\(^T\)| 100 ± 00%  | 23.7 ± 2.3    | 21.1 ± 2.5%   | 19.6 ± 2.4%       | 24.9 ± 2.3  |

Pairwise comparison performed using GGDC, formula 2 (DDH estimates based on identities/HSP length). dDDH values are 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).

**Conflict of interest**

None declared.

**Acknowledgements**

We thank L. Hadjadj for technical assistance, Aix-Marseille Université IRD, APHM, MEPHI, IHU Méditerranée Infection, Marseille, France. This project was funded by the French
Centre National de la Recherche Scientifique and Infectiopole Sud Foundation. This work was also supported by the French government under the ‘Investissements d’avenir’ (Investments for the Future) programme managed by the Agence Nationale de la Recherche (Méditerranée Infection 10-IAHU-03).

References

[1] Ash C, Priest FG, Collins MD. Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Proposal for the creation of a new genus Paenibacillus. Antonie Leeuwenhoek 1993;1994:64:253–60.

[2] Pinho-Gomes AC, Nasir A, Mosca R, Mirza S, Kadir I. Intraoperative diagnosis of mitral valve endocarditis secondary to Paenibacillus provencensis. Ann R Coll Surg Engl 2017;99:e54–5.

[3] Ouyang J, Pei Z, Lutwick L, Dalal S, Yang L, Cassai N, et al. Case report: Paenibacillus thiaminolyticus: a new cause of human infection, inducing bacteremia in a patient on hemodialysis. Ann Clin Lab Sci 2008;38:393–400.

[4] Li J, Beatty PK, Shah S, Jensen SE. Use of PCR-targeted mutagenesis to disrupt production of fusaricidin-type antifungal antibiotics in Paenibacillus polymyxa. Appl Environ Microbiol 2007;73:3480–9.

[5] 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.

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

[7] Seng P, Rollin JM, Fournier PE, La SB, Drancourt M, Raoult D. MALDI-TOF—mass spectrometry applications in clinical microbiology. Future Microbiol 2010;5:1733–54.

[8] Aziz RK, Bartels D, Best AA, Dejongh M, Dusz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 2008;9:75.

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

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

[11] Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 2014;58:212–20.

[12] Drissi F, Buffet S, Raoult D, Merhej V. Common occurrence of antibiotic agents in human intestinal microbiota. Front Microbiol 2015;6:441.

[13] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res 2011;39(Web Server issue):W347–52.

[14] Auch AF, Klenk HP, Goker M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. Stand Genomic Sci 2010;2:142–8.

[15] Medema MH, Blin K, Cimermancic P, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Research 2011;39(Web Server issue):W339–46. https://doi.org/10.1093/nar/gkr466.