Noncontiguous finished genome sequence and description of *Paenibacillus ihumii* sp. nov. strain AT5

A. H. Togo1, S. Khelafia1, J.-C. Lagier1, A. Caputo1, C. Robert1, P.-E. Fournier1, M. Maraninchi2, R. Valero2,3, D. Raoult1,4 and M. Million1

1) Aix Marseille Université, URMITE, Institut Hospitalier Universitaire Méditerranée-Infection, UM63, CNRS7278, IRD198, INSERM1095, 2) Aix Marseille Université, NORT “Nutrition, Obesity and Risk of Thrombosis”, INSERM1062, INRA1260, 3) APHM, CHU Hôpital de la Conception, Service Nutrition, Maladies Métaboliques et Endocrinologie, Marseille, France and 4) Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract

*Paenibacillus ihumii* sp. nov. strain AT5 (= CSUR P1981 = DSM 100664) is the type strain of *P. ihumii*. This bacterium was isolated from a stool sample from a morbidly obese French patient using the culturomics approach. The genome of this Gram-negative, facultative anaerobic, motile and spore-forming bacillus is 5,924,686 bp long. Genomic analysis identified 253 (5%) of 3,812 genes as ORFans and at least 2,599 (50.03%) of 5,194 orthologous proteins not shared with the closest phylogenetic species.

New Microbes and New Infections © 2016 The Authors. Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases.

Keywords: Culturomics, genome, *Paenibacillus ihumii* sp. nov., taxonogenomics

Original Submission: 24 November 2015; Revised Submission: 18 January 2016; Accepted: 21 January 2016

Article published online: 29 January 2016

Introduction

*Paenibacillus ihumii* strain AT5 (= CSUR P1981 = DSM 100664) is the type strain of *P. ihumii* sp. nov. This isolate is part of an exploratory study of the gut flora from obese patients before and after bariatric surgery using a microbial culturomics approach, the aim of which is to exhaustively explore the microbial ecosystem of gut flora by using different culture conditions [1]. This bacterium was isolated from a stool sample collected before bariatric surgery from a 33-year-old French woman living in Marseille with morbid obesity.

The conventional parameters used in the delineation of bacterial species include 16S rRNA sequence identity and phylogeny [2], genomic (G + C content) diversity and DNA-DNA hybridization (DDH) [3,4]. However, these methods present some shortfalls, mainly due to their cutoff values, which vary according to species or genera [5]. The advent of new technology tools such as high-throughput sequencing has enabled us to access descriptions of many bacterial species in the public nucleotide sequence library [6]. Recently we proposed including genomic data in a polyphasic approach to describe new bacterial taxa (taxonogenomics). This strategy considers phenotypic characteristics, genomic analysis and the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) spectrum and comparison [7,8]. These characteristics support the circumscription and description of the species *P. ihumii* as a novel bacterium. Here we provide a brief classification and set of characteristics for *P. ihumii* sp. nov. strain AT5 alongside the description of the complete genome sequencing and annotation.

Materials and Methods

Sample collection

A stool sample was collected from a 33-year-old obese Frenchwoman (body mass index 38.6 kg/m²; 100 kg, 1.61 m
isolation and identification of strain

Growth of *P. ihumii* strain AT5 was performed in May 2015. The sterile stool extract was preincubated in blood culture bottles enriched with rumen fluid and sheep’s blood as described elsewhere [1]. The culture was monitored for 30 days. On various days (days 1, 3, 7, 10, 15, 21 and 30), a seeding of the preincubated product was done on sheep’s blood–enriched Columbia agar (bioMérieux, Marcy l’Etoile, France) and incubated for 24 hours in an aerobic atmosphere at 37°C. The colonies that emerged were cultivated under the same conditions for isolation. They were then identified by MALDI-TOF as described elsewhere [9]. In short, one isolated bacterial colony was picked up with a pipette tip from a culture agar plate and spread as a thin smear on an MTP 384 MALDI-TOF target plate (Bruker Daltonics, Leipzig, Germany). Each smear was overlaid with 2 μL of matrix solution (saturated solution of α-cyano-4-hydroxycinnamic acid in 50% acetonitrile, 2.5% trifluoroacetic acid) and allowed to dry for 5 minutes. Measurements were performed with a Microflex spectrometer (Bruker). Spectra were recorded in the positive linear mode for the mass range from 2000 to 20 000 Da (parameter settings: ion source 1 (ISI), 20 kV; IS2, 18.5 kV; lens, 7 kV). A spectrum was obtained after 675 shots with variable laser power. The time of acquisition was between 30 seconds and 1 minute per spot. Identification was carried out as previously reported [10].

The 16S rRNA PCR, coupled with sequencing, was performed using GeneAmp PCR System 2720 thermal cyclers (Applied Biosystems, Foster City, CA, USA) and ABI Prism 3130xl Genetic Analyser capillary sequencer (Applied Biosystems) respectively [11]. Chromas Pro 1.34 software (Technelysium, Tewantin, Australia) was used to correct sequences, and BLASTn searches were performed in National Center for Biotechnology Information (NCBI; http://blast.ncbi.nlm.nih.gov/gate1.inist.fr/Blast.cgi). Sequences were then aligned using Clustal W, and phylogenetic inferences were obtained using the neighbour-joining method with MEGA 6 (Molecular Evolutionary Genetics Analysis version 6) software. The numbers of nodes correspond to the percentages of bootstrap values obtained by repeating the analysis 1000 times in order to generate a consensus tree.

**Growth conditions**

Different growth temperatures (25, 28, 37, 45 and 55°C) were tested on sheep’s blood–enriched Columbia agar (bioMérieux). Growth of this strain was tested in an anaerobic atmosphere using the GENbag anaer system (bioMérieux), in a microaerophilic atmosphere using GENbag microaer system (bioMérieux) and in an aerobic atmosphere with or without 5% CO2. The saltiness of this species was tested using 5% NaCl on Schaedler agar with 5% sheep’s blood (bioMérieux) in an aerobic atmosphere.

**Biochemical, sporulation and motility assays**

Biochemical assays were performed using API Gallery systems: API ZYM (bioMérieux), API 20NE (bioMérieux) and API50 CH (bioMérieux). Detection of catalase (bioMérieux) and oxidase (Becton Dickinson, Franklin Lakes, NJ, USA) was also conducted. A thermal shock at 100°C for half an hour was carried out in order to test sporulation. A fresh colony was observed between blades and slats using a Leica DM 1000 photonic microscope (Leica Microsystems, Wetzlar, Germany) at 40× to assess the motility of the bacteria.

**Microscopy**

Transmission electron microscopy using a Tecnai G20 device (FEI Company, Limel-Brevannes, France) at an operating voltage of 60 kV was performed to observe the *P. ihumii* strain AT5 after negative colouration. Gram staining was performed and observed using a Leica DM 2500 photonic microscope with a 100× oil-immersion objective lens.

**Antibiotic susceptibility**

Antibiotic susceptibility of the strain was tested using 18 antibiotics, including amoxicillin 25 μg, amoxicillin 20 μg/clavulanic acid 10 μg, cefalexin 30 μg, ceftriaxone 30 μg, ciprofloxacin 5 μg, doxycycline 30 IU, erythromycin 15 IU, nitrofurantoin 300 μg, gentamicin 500 μg, gentamicin 15 μg, imipenem 10 μg, metronidazole 4 μg, oxacillin 5 μg, penicillin G 10 IU, rifampicin 30 μg, trimethoprim 1.25 μg/sulfamethoxazole 23.75 μg, tobramycin 10 μg and vancomycin 30 μg (i2a, Montpellier, France). The Scan 1200 was used to interpret the results (Interscience, Saint-Nom-La-Bretèche, France).

**Genome sequencing and assembly**

*P. ihumii* genomic DNA (gDNA) was sequenced on the MiSeq Technology (Illumina, San Diego, CA, USA) using the mate pair strategy. The gDNA was barcoded in order to be mixed with 11 other projects with the Nextera Mate Pair sample prep kit (Illumina). The gDNA was quantified by a Qubit assay using the high sensitivity kit (Thermo Fisher Scientific, Waltham, MA) to 145 ng/μL. The mate pair library was prepared with 1 μg of gDNA using the Nextera mate pair Illumina guide. The gDNA was then sequenced on the MiSeq and the sequencing reads were aligned to the *P. ihumii* AT5 genome assembly by using the paired-end reads as the query (145 ng/μL) and the genome assembly as the target. The paired-end reads were then mapped to the genome assembly using BWA and the resulting alignment was converted into SAM format using SAMtools.

The Illumina reads were then aligned against the genome assembly using BWA and the resulting alignment was converted into SAM format using SAMtools. The genome was assembled using Allpaths-LG with the Illumina reads as the query and the genome assembly as the target. The genome was then annotated using Prokka with the Illumina reads as the query and the genome assembly as the target. The genome was then subjected to a quality control check using Quast and then submitted to the European Nucleotide Archive (ENA) with the accession number MG980125.
sample was simultaneously fragmented and tagged with a mate pair junction adapter. The pattern of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) using a DNA 7500 lab chip. DNA fragments were ranged in size from 1.5 to 11 kb, with an optimal size of 3.987 kb. No size selection was performed, and only 334 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared into small fragments with an optimal size of 1051 bp on the Covaris device S2 in T6 tubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity BioAnalyzer LabChip (Agilent), and the final concentration library was measured at 2.90 nmol/L. The libraries were normalized at 2 nM and pooled. Following a denaturation step and dilution at 15 pM, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and a sequencing run were performed in a single 2 × 301 bp run. In total, 7.3 Gb of information was obtained from a 51 K/mm² cluster density, with a cluster passing quality control filters of 97.0% (12 079 000 passing filter paired reads). Within this run, the index representation for P. ihumii was determined to 10.02%. The 1 210 259 paired reads were trimmed and assembled to 12 scaffolds using the SPAdes software [12].

Genome annotation and comparison

Open reading frames (ORFs) were predicted using Prodigal [13] with default parameters. Nevertheless, the predicted ORFs were excluded if they spanned a sequencing gap region (containing N). The predicted bacterial protein sequences were searched against the GenBank database and the Clusters of Orthologous Groups (COGs) database using BLASTP (E value 1e-03, coverage 0.7 and 30% identity). If no hit was found, it searched against the NR database using BLASTP with an E value of 1e-03, coverage 0.7 and 30% identity. If the sequence length was smaller than 80 amino acids, we used an E value of 1e-05. The tRNAs and rRNAs were predicted using the tRNA Scan-SE and RNAmmer tools respectively [14,15]. SignalP and TMHMM were used to foresee the signal peptides and the number of transmembrane helices respectively [16,17]. Mobile genetic elements were predicted using PHAST and RAST [18,19]. ORFans were identified if their BLASTP E value was lower than 1e-03 for an alignment length greater than 80 amino acids. If alignment lengths were smaller than 80 amino acids, we used an E value of 1e-05. Artemis and DNA Plotter were used for data management and visualization of genomic features respectively [20,21]. Genomes were automatically retrieved from the 16S rRNA tree using Xegen software (PhyloPattern) [22]. For each selected genome, complete genome sequence, proteome genome sequence and Orfeome genome sequence were retrieved from the FTP site of National Center for Biotechnology Information (NCBI). All proteomes were analyzed using proteinOrtho [23]. A similarity score was then computed for each pair of genomes. This score is the mean value of nucleotide similarity between all orthologous pairs in the two genomes studied (average genomic identity of orthologous gene sequences, AGIOS) [6]. For the genomic comparison, we used P. ihumii strain AT5 (CYXK00000000), P. fonticola strain DSM21315 (ARM700000000), P. peariae strain KCTC 3763 (CP011512), P. stellifer strain DSM 14472 (CP009286), P. terrae strain HPL-003 (CP003107) and P. borealis strain DSM 13188 (CP009285). An annotation of the entire proteome was performed to define the distribution of functional classes of predicted genes according to the clusters of orthologous groups of proteins (using the same method as for the genome annotation). P. ihumii genome was locally aligned two by two by using the BLAT algorithm [24,25] against each of the selected genomes previously cited, and DDH values were estimated from a generalized model. The DDH threshold is less than 70%.

FIG. 1. Gram staining of Paenibacillus ihumii strain AT5.

FIG. 2. Transmission electron microscopy of Paenibacillus ihumii strain AT5 using Tecnai G20 (FEI Company) at operating voltage of 60 kV. Scale bar = 500 nm.
for a species to be considered as new species [26]. Annotation and comparison processes were performed using the Multi-Agent software system DAGOBAH [27], including Figenix [28] libraries that provide pipeline analysis.

Results

Phenotypic and biochemical characterization

The P. ihumii strain AT5 is a Gram-negative motile rod which is catalase and oxidase negative (Fig. 1). The growth of the strain occurred between 28 to 55°C, but optimal growth was observed at 37°C after 24 hours of incubation in an aerobic atmosphere. The colonies were approximately 1 to 2 mm in diameter and grey on 5% sheep’s blood–enriched Columbia agar. Cells had a diameter ranging from 0.50 to 1.75 μm, with a mean diameter of 1 μm measured using electron microscopy (Fig. 2). No growth of this bacterium was observed using 5% NaCl on Schaedler agar with 5% sheep’s blood. This bacterium is a facultative anaerobe bacillus but can also grow in a microaerophilic atmosphere. The strain was able to form spores. Table 1 summarizes the classification and main features of P. ihumii.

Of all the 18 antibiotics tested, the P. ihumii strain AT5 was susceptible to all of them except metronidazole and tobramycin. Using an API ZYM strip, we observed that P. ihumii possesses alkaline phosphatase, esterase (C4), esterase lipase (C8), naphthol-AS-BI-phosphohydrolase, α-galactosidase (melibiose), β-galactosidase (hydrodase), α-glucosidase (maltase) and β-glucosidase (cellulose) activities. However, there are no activities of lipase (C14), leucin arylamidase, valine arylamidase, cystine arylamidase, trypsin, α-chymotrypsin, phosphatase acid, β-glucuronidase, N-acetyl-β-glucosaminidase, α-mannosidase and α-fucosidase. Using an API 20 NE strip, positive reactions were obtained for nitrate reduction, β-galactosidase, arabinose, β-glucosidase, mannose, mannotol, N-acetyl-glucosamine and maltose. Negative reactions were obtained for L-tryptophan, D-glucose, L-arginine, urea, gelatine, potassium gluconate, capric acid, adipic acid, malic acid, trisodium citrate and phenyl acetic acid. Using an

TABLE 2. Differential characteristics of Paenibacillus ihumii strain AT5, Paenibacillus borealis, Paenibacillus fonticola, Paenibacillus peoriae, Paenibacillus stellifer and Paenibacillus terrae

| Property           | P. ihumii | P. borealis | P. fonticola | P. peoriae | P. stellifer | P. terrae |
|--------------------|-----------|-------------|--------------|------------|--------------|-----------|
| Cell diameter width/length (μm) | 0.5/1.75  | 0.7–1/3–5  | 0.8–1/2–12.4 | 0.5–1/3–6  | 0.6–0.8/2.5–5 | 1.3–1.8/4–7 |
| Oxygen requirement  | +/−       | +/−         | +/−          | +/−        | +/−          | +/−       |
| Gram stain         | −         | −           | −            | −          | −            | −         |
| Sporulation        | +         | +           | +            | +          | +            | +         |
| Temperature range  | −−        | +           | +            | +          | +            | +         |
| Optimum temperature | 37°C      | 37°C        | 37°C         | 37°C       | 37°C         | 37°C      |
| Spore formation    | +         | +           | +            | +          | +            | +         |
| Production of:     | −         | +           | −            | −          | −            | −         |
| Carbohydrates      | −         | −           | −            | −          | −            | −         |
| Nitrate reductase  | −         | −           | −            | −          | −            | −         |
| Urease             | −         | −           | −            | −          | −            | −         |
| N-acetyl-glucosamine| +         | +           | −            | +          | +            | +         |
| Utilization of:    | +         | +           | −            | −          | −            | −         |
| N-Arabinose        | +         | +           | −            | −          | −            | −         |
| N-Ribose           | +         | +           | −            | −          | −            | −         |
| N-Mannose          | +         | +           | −            | −          | −            | −         |
| N-Manitol          | +         | +           | −            | −          | −            | −         |
| N-Glucose          | +         | +           | −            | −          | −            | −         |
| N-Fructose         | +         | +           | −            | −          | −            | −         |
| N-Maltose          | +         | +           | −            | −          | −            | −         |
| N-Lactose          | +         | +           | −            | −          | −            | −         |
| N-Xylose           | +         | +           | −            | −          | −            | −         |
| Habitat            | Human gut | Spruce forest humus | Warm springs | Soil/rotting vegetation | Soil | Food-packing board |
| Genome size         | 5.92      | 8.16        | 6.30         | 5.77       | 5.66         | 6.08      |
| DNA G + C content  | 50.2      | 51.39       | 47.68        | 46.44      | 53.54        | 46.77      |

*+, positive result; −, negative result; +/−, facultative anaerobic; v, variable result; NA, data not available.

New Microbes and New Infections © 2016 The Authors. Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases, NMNI, 10, 142–150

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
API 50 CH strip, we demonstrated that *P. ihumii* is able to ferment L-arabinose, D-ribose, D-xylene, methyl-β-D-xylopyranoside, D-galactose, D-glucose, D-fructose, D-mannose, D-mannitol, methyl-D-D-glucopyranoside, N-acetylgluicosamine, amygdalin, arbutin, esculin ferric citrate, salicin, D-cellobioso, D-maltose, D-lactose, D-melibiose, D-saccharose, D-trehalose, D-rafifinose, starch, glycerin, gentiobiase and D-lyxose. No fermentation were recorded for glycerol, erythritol, D-arabinose, L-xylene, D-adonitol, L-sorbose, L-rhamnose, dulcitol, inositol, D-sorbitol, methyl-α-D-mannopyranoside, inulin, D-melezitose, xylitol, D-turanose, D-tagatose, D-fucose, L-fucose, D-arabitol, L-arabitol, potassium gluconate, potassium 2-ketogluconate and potassium 5-ketogluconate. Table 2 presents a comparison of the different characteristics with other representatives of the *Paenibacillus* genus.

**Phylogenetic analysis**

No match could be found from the spectrum generated from clean *P. ihumii* strain AT5 spots and those in the Bruker database (Fig. 3). The phylogenetic analysis, performed using 16S rRNA sequences, showed that *P. ihumii* sp. nov. strain AT5 exhibited 98.2% identity with *Paenibacillus lentus* [29], classified in the *Paenibacillaceae* family created by Ash in 1993 [30]. However, this percentage remains lower than the 98.7% 16S rRNA gene sequence threshold recommended by Stackebrandt and Ebers [2] to delineate a new species. This enables us to say that the *P. ihumii* strain AT5 is a new species within the *Paenibacillaceae* family (Table 1). A neighbour-joining phylogenetic tree (Fig. 4) based on 16S rRNA gene sequences shows the relationships between *P. ihumii* and some related taxa. The

**FIG. 3.** MALDI-TOF reference mass spectrum from *Paenibacillus ihumii* strain AT5. MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.

**FIG. 4.** Phylogenetic tree highlighting position of *Paenibacillus ihumii* strain AT5 relative to other close species. Sequences are recovered using nucleotide blast against 16S rRNA Database of Silva ‘All-Species Living Tree’ project (LTIPs119). Sequences were aligned using muscle and phylogenetic inferences obtained using approximately maximum likelihood method within Fast Tree software. Numbers at nodes are support local values computed using Shimodaira-Hasegawa test. Corresponding GenBank accession numbers for 16S rRNA genes are indicated at right of strains in tree.

*P. ihumii* 16S rRNA sequence was deposited in European Molecular Biology Laboratory—European Bioinformatics Institute (EMBL-EBI) under accession number LN881615. A gel view was performed in order to observe spectra differences between *P. ihumii* and other close bacteria (Fig. 5).
Genome properties

The genome of *P. ihumii* strain AT5 (Fig. 6) is 5,924,686 bp long with 50.20% G + C content (Table 3). It consists of 12 scaffolds (composed of 16 contigs). Of the 5274 predicted genes, 5194 were protein-coding genes and 80 were RNAs (six 5S rRNA, five 16S rRNA, two 23S rRNA, 67 tRNA). A total of 3812 genes (73.39%) were assigned as having a putative function (by cogs or by NR blast), and 253 genes (4.87%) were identified as ORFans. The remaining genes (896 genes, 17.25%) were annotated as hypothetical proteins. Using ARG-ANNOT [31], no resistance genes were found. Nevertheless, 15 genes associated to polyketide synthase or nonribosomal peptide synthetase [32] were discovered through genome analysis and implicated in the production of secondary metabolites. The distribution of genes into COGs functional categories is presented in Table 4.

Genome comparison

*P. borealis, P. fonticola, P. peoriae, P. stellifer* and *P. terrae* are species closely related to *P. ihumii* with available genomes (Table 1) and were consequently chosen for this comparative analysis. The G + C content of *P. ihumii* is smaller than that of...
P. borealis and P. stellifer (50.20, 51.39 and 53.54% respectively) but larger than that of P. fonticola, P. peoriae and P. terrae (47.68, 46.44 and 46.77% respectively). The gene content of P. ihumii is smaller than that of P. fonticola, P. borealis and P. terrae (5194, 5645, 6213 and 5525 respectively) but larger than that of P. peoriae and P. stellifer (5122 and 4464 respectively). Fig. 7 shows that the distribution of genes into COGs categories was similar across all compared genomes. In addition, P. ihumii shared 2595, 2141, 1998, 2217 and 2414 orthologous genes with P. fonticola, P. peoriae, P. stellifer, P. terrae and P. borealis respectively (Table 5). The AGIOS values ranged from 57.59 to 87.31% among compared Paenibacillus species with the exception of P. ihumii. When P. ihumii was compared to other Paenibacillus species, the AGIOS value ranged from 57.88% for P. fonticola to 67.99% for P. peoriae (Table 5). DDH was 28% ± 2.43 for P. fonticola, 17.5% ± 2.23 for P. peoriae, 18.2% ± 2.26 for P. stellifer, 17.6% ± 2.23 for P. terrae and 17.6% ± 2.24 for P. borealis (Table 6). These data confirm P. ihumii as a unique species.

Conclusion
On the basis of phenotypic, genomic and phylogenetic analyses, we formally propose the creation of P. ihumii sp. nov., which contains the strain ATS. This bacterium was isolated from a stool sample of a 33-year-old morbidly obese Frenchwoman living in Marseille.

Taxonomic and Nomenclatural Proposals

Description of P. ihumii strain ATS sp. nov.
Paenibacillus ihumii (i.hum.i). N.L. gen. n. ihumii, based on the acronym IHUMI, the Institut Hospitalo-Universitaire Médi terranée-Infection in Marseille, France, where the type strain was isolated.

Cells are Gram-negative, spore-forming, motile, rod-shaped bacilli with a size of 0.5–1.75 μm. Colonies are grey, with a diameter of 1–2 mm on 5% sheep’s blood–enriched Columbia agar. The strain is catalase and oxidase negative. It has an optimum growth temperature of 37°C and is a facultative anaerobe, able to grow in a microaerophilic atmosphere.

Using API Gallery systems, positive reactions were observed for alkaline phosphatase, esterase (C4), esterase lipase (C8), naphthol-AS-BI-phosphohydrolase, α-galactosidase, β-galactosidase, α-glucosidase, β-glucosidase, nitrate reduction, β-galactosidase, mannose, mannitol, N-acetylglucosamine, arabinose, maltose, d-arabinose, d-ribose, d-xylene, methyl-β-D-xylopyranoside, d-galactose, d-glucose, d-fructose, d-mannose, d-mannitol, methyl-α-D-glucopyranoside, N-acetylglucosamine, amygdalin, arbutin, esculin ferric citrate, salicin, D-cellobiose, D-maltose, D-lactose, D-melibiose, D-saccharose, D-trehalose, D-raffinose, starch, glycogen, gentiobiose and D-lyxose.

Cells are susceptible to amoxicillin, amoxicillin/clavulanic acid, cefalexin, ceftriaxone, ciprofloxacin, doxycycline, erythromycin, nitrofurantoin, gentamicin, imipenem, oxacillin, penicillin G, rifampicin, trimethoprim/sulfamethoxazole and vancomycin, and are resistant to metronidazole and tobramycin.

The length of the genome is 5 924 686 bp with 50% G + C content. The 16S rRNA gene sequence and whole-genome shotgun sequence of the P. ihumii strain ATS were deposited

| Attribute          | Genome (total) | Value     | % of total |
|--------------------|----------------|-----------|------------|
| size (bp)          | 5 924 686      | 50%       | 100        |
| G+C content (bp)   | 2 973 782      | 50.20     |            |
| coding region (bp) | 5 065 111      | 85.49     |            |
| extrachromosomal elements | 0 | 0 | |
| total genes        | 5274           | 100       |            |
| RNA genes          | 80             | 1.51      |            |
| protein-coding genes | 5194       | 98.48     |            |
| genes with function prediction | 3812 | 73.39 | |
| genes assigned to COGs | 3881 | 74.72 | |
| genes with peptide signals | 743  | 14.30 | |
| genes with transmembrane helices | 1327 | 25.54 | |

COGs, Clusters of Orthologous Groups database.

TABLE 3. Nucleotide content and gene count levels of the chromosome

| Code | Value | % of total | Description               |
|------|-------|------------|---------------------------|
| J    | 271   | 5.21       | Translation               |
| A    | 0     | 0.0        | RNA processing and modification |
| K    | 462   | 8.89       | Transcription             |
| L    | 133   | 2.56       | Replication, recombination and repair |
| B    | 1    | 0.01       | Chlamydia structure and dynamics |
| D    | 64    | 1.23       | Cell cycle control, mitosis and meiosis |
| Y    | 0    | 0          | Nuclear structure         |
| V    | 165   | 3.17       | Defence mechanisms        |
| T    | 281   | 5.41       | Signal transduction mechanisms |
| M    | 214   | 4.12       | Cell wall/membrane biogenesis |
| N    | 81    | 1.55       | Cell motility             |
| Z    | 9     | 0.17       | Cytoskeleton              |
| W    | 21    | 0.40       | Extracellular structures   |
| U    | 49    | 0.94       | Intracellular and trafficking secretion |
| O    | 159   | 3.06       | Post-translational modification, protein turnover, chaperones |
| C    | 150   | 2.88       | Energy production and conversion |
| G    | 623   | 11.99      | Carbohydrate transport and metabolism |
| E    | 269   | 5.17       | Amino acid transport and metabolism |
| F    | 110   | 2.11       | Nucleotide transport and metabolism |
| H    | 211   | 4.06       | Coenzyme transport and metabolism |
| I    | 125   | 2.40       | Lipid transport and metabolism |
| P    | 212   | 4.08       | Inorganic ion transport and metabolism |
| Q    | 95    | 1.82       | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 399   | 7.68       | General function prediction only |
|      | 1313  | 25.27      | Not in COGs               |

COGs, Clusters of Orthologous Groups database.

TABLE 4. Number of genes associated with 26 general COGs functional categories

New Microbes and New Infections © 2016 The Authors. Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases, NMNI, 10, 142–150
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
The AT5 type strain (= CSUR P1981 = DSM 100664) was isolated from a stool sample from an obese Frenchwoman. The habitat of this microorganism is the human digestive tract.

TABLE 5. Numbers of orthologous proteins shared between genomes (upper right), average percentage similarity of nucleotides corresponding to orthologous protein shared between genomes (lower left) and numbers of proteins per genome (bold)

|                | Paenibacillus ihumii | Paenibacillus fonticola | Paenibacillus peoriae | Paenibacillus stellifer | Paenibacillus terrae | Paenibacillus borealis |
|----------------|----------------------|-------------------------|----------------------|------------------------|----------------------|------------------------|
| **P. ihumii**  | 5194                 | 5995                    | 2141                 | 2328                   | 2217                 | 2414                   |
| **P. fonticola** | 57.88                | 5645                    | 2249                 | 2067                   | 2328                 | 2317                   |
| **P. peoriae** | 67.99                | 67.76                   | 5122                 | 2084                   | 2792                 | 2448                   |
| **P. stellifer** | 58.11                | 57.59                   | 4464                 | 2141                   | 2345                 |                       |
| **P. terrae**  | 67.96                | 67.73                   | 87.31                | 57.69                   | **5525**             | 2549                   |
| **P. borealis** | 58.25                | 57.88                   | 58.20                | 72.26                   | 58.19                | **6213**               |

TABLE 6. Pairwise comparison of Paenibacillus ihumii with other species using GGDC, formula 2 (DDH estimates based on identities/HSP length)

|                | Paenibacillus ihumii | Paenibacillus fonticola | Paenibacillus peoriae | Paenibacillus stellifer | Paenibacillus terrae | Paenibacillus borealis |
|----------------|----------------------|-------------------------|----------------------|------------------------|----------------------|------------------------|
| **P. ihumii**  | 100%                 | 28% ± 2.43              | 17.5% ± 2.23         | 18.2% ± 2.26           | 17.6% ± 2.23         | 17.6% ± 2.24           |
| **P. fonticola** | 100%                 | 16.8% ± 2.21            | 17.1% ± 2.22         | 18.2% ± 2.26           | 17.2% ± 2.22         | 17% ± 2.22             |
| **P. peoriae** | 100%                 | 18.2% ± 2.26            | 33.7% ± 2.47         | 18.3% ± 2.26           | 19.7% ± 2.3          | 18.1% ± 2.25           |
| **P. stellifer** | 100%                 |                         |                      | 100%                   |                      |                        |
| **P. terrae**  | 100%                 |                         |                      | 100%                   |                      |                        |
| **P. borealis** | 100%                 |                         |                      | 100%                   |                      |                        |

*Confidence intervals indicate inherent uncertainty in estimating DDH values from intergenomic distances based on models derived from empirical test data sets (which are always limited in size). These results are in accordance with the 16S rRNA and phylogenomic analyses as well as the GGDC results. DDH, DNA-DNA hybridization; GGDC, Genome-to-Genome Distance Calculator; HSP, high-scoring segment pairs.

Acknowledgements

The authors thank the Xegen Company (http://www.xegen.fr/) for automating the genomic annotation process. This study was...
funded by the Fondation Méditerranée Infection. We thank C. Andrieu for administrative assistance.

Conflict of Interest

None declared.

References

[1] Lagier JC, Aragouf M, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturotics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 2012;18:1185–93.
[2] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;32:132–5.
[3] Stackebrandt E, Frederiksen W, Garrity GM, Grimont PA, Kämpfer P, Maiden MC, et al. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Microbiol 2002;52:1043–7.
[4] Rossello-Mora R. DNA-DNA reassocation methods applied to microbial taxonomy and their critical evaluation. In: Stackebrandt E, editor. Molecular identification, systematics, and population structure of prokaryotes. Berlin: Springer Verlag; 2006. p. 23–50.
[5] Welker M, Moore ERB. Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. Syst Appl Microbiol 2011;34:2–11.
[6] Ramasamy D, Mishra AK, Lagier JC, Padmanabhan R, Rossi M, Sentausa E, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol 2014;64(pt 2):384–91.
[7] Kokcha S, Mishra AK, Lagier JC, Million M, Leroy Q, Raoult D, et al. Non contiguous-finished genome sequence and description of Bacillus timonensis sp. nov. Stand Genomic Sci 2012;6:346–55.
[8] Mishra AK, Lagier JC, Nguyen TT, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of P. senegalensis sp. nov. Stand Genomic Sci 2013;7:370–81.
[9] Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 2009;49:543–51.
[10] Hugon P, Ramasamy D, Lagier JC, Rivet R, Couderc C, Raoult D, et al. Non contiguous-finished genome sequence and description of Alloiococcus obstes sp. nov. Stand Genomic Sci 2013;7:427–39.
[11] Nikamg VE, Hyunh HTT, Abdoubharam G, Ruym R, Drancourt M. Diversity of human-associated Methionine-requiring bacterium isolates revealed by multiplex DNA sequence typing. Curr Microbiol 2015;70:810–5.
[12] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 2012;19:455–77.
[13] Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:119.
[14] 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.
[15] 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.
[16] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340:783–95.
[17] Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001;305:567–80.
[18] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phase search tool. Nucleic Acids Res 2011;39(Web Server issue): W347–52.
[19] Overbeek R, Olson R, Rusch GD, Olsen GJ, Davis JJ, Dusz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 2014;42(Database issue):D206–14.
[20] Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 2012;28:646–9.
[21] Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 2009;25:119–20.
[22] Gouret P, Thompson JD, Pontarotti P. PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinformatics 2009:10:298.
[23] Lehner M, Findeis S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics 2011;12:124.
[24] Kent WJ. BLAT—the BLAST-like alignment tool. Genome Res 2002;12:656–64.
[25] Auch AF, van Jm M, Klenk HP, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci 2010:2:117–34.
[26] Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 2013:14:60.
[27] Gouret P, Paganini J, Dainat J, Louati D, Darbo E, Pontarotti P, et al. Integration of evolutionary biology concepts for functional annotation and automation of complex research in evolution: the multi-agent software system DAGOBIAH. In: Pontarotti P, editor. Evolutionary biology—concepts, biodiversity, macroevolution and genome evolution. Berlin: Springer Verlag; 2011. p. 71–87.
[28] Gouret P, Viselli V, Balandraud N, Gilles A, Pontarotti P, Danchin EG. FIGENIX: intelligent automation of genomic annotation: expertise integration in a new software platform. BMC Bioinformatics 2005:6:198.
[29] Li YF, Calley JN, Ebert PJ, Helmes EB. Paenibacillus lentus sp. nov., a β-mannanolytic bacterium isolated from mixed soil samples in a selective enrichment using guar gum as the sole carbon source. Int J Syst Evol Microbiol 2014;64(pt 4):1166–72.
[30] Ash C, Priest RF, 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 Van Leeuwenhoek 1993–1994;64:253–60.
[31] Gupta SK, Padmanabhan BR, Dine 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.
[32] Conway KR, Boddy CN. ClusterMine360: a database of microbial PKS/NRPS biosynthesis. Nucleic Acids Res 2013;41(Database issue):D402–7.