HAL
archives-ouvertes.fr

Draft genome and description of Cohnella massiliensis sp. nov., a new bacterial species isolated from the blood culture of a hemodialysis patient

Rita Abou Abdallah, Jacques Bou Khalil, Claudia Andrieu, Enora Tomei, Nicholas Armstrong, Pierre-Edouard Fournier, Didier Raoult, Fadi Bittar

▶ To cite this version:

Rita Abou Abdallah, Jacques Bou Khalil, Claudia Andrieu, Enora Tomei, Nicholas Armstrong, et al.. Draft genome and description of Cohnella massiliensis sp. nov., a new bacterial species isolated from the blood culture of a hemodialysis patient. Archives of Microbiology, Springer Verlag, 2019, 201 (3), pp.305-312. 10.1007/s00203-018-1612-1 . hal-02202136

HAL Id: hal-02202136
https://hal-amu.archives-ouvertes.fr/hal-02202136

Submitted on 17 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Draft genome and description of *Cohnella massiliensis* sp. nov., a new bacterial species isolated from the blood culture of a hemodialysis patient

Rita Abou Abdallah[1] · Jacques Bou Khalil[2] · Claudia Andrieu[2] · Enora Tomeï[1] · Nicholas Armstrong[2] · Pierre-Edouard Fournier[1] · Didier Raoult[2] · Fadi Bittar[2]

Strain 6021052837^T was isolated from the blood culture of a hemodialysis patient on Chocolat PolyViteX medium at 37 °C after 2 days of incubation. Colonies could not be identified by our systematic MALDI-TOF Mass Spectrometry screening. The16S rRNA gene sequencing showed that the strain had 96% sequence identity with *Cohnella formosensis* (Genbank accession number JN806384), the phylogenetic closely related type strain of a species with standing in nomenclature, which putatively classifies it as a new species. The colonies cultivated on Columbia agar with 5% sheep blood medium at 37 °C after 24 h of incubation, are white pigmented, their size varied from 1.5 to 2 mm in diameter. Strain 6021052837^T is an aerobic, Gram-negative, motile, spore forming rod, which cannot grow microaerophilically or under anaerobic conditions. The major fatty acids are branched saturated fatty acids: 14-methyl-pentadecanoic acid (34%) and 12-methyl-tetradecanoic acid (31%). The 6.328 Mb long genome, composed of 25 contigs, has a G+C content of 57.24%. Out of the 5710 predicted genes, 5646 were protein-coding genes and 64 were RNAs. A total of 3239 genes (57.37%) were assigned as putative function (by COGs) and 288 genes were identified as ORFans (5.1%). Average genomic identity of orthologous gene sequences (AGIOS) of strain 6021052837^T against genomes of the type strains of related species ranged between 58.26 and 79.63%, respectively. According to our taxonogenomics results, we propose the creation of *Cohnella massiliensis* sp. nov. that contains the type strain 6021052837^T (= CSUR P2659, =DSM103435).

Keywords *Cohnella massiliensis* · Genome analysis · Blood · Taxono-genomics

✉ Fadi Bittar
fadi.bittar@univ-amu.fr

[1] Aix Marseille Univ, IRD, APHM, SSA, VITROME, IHU-Méditerranée Infection, Marseille, France

[2] Aix Marseille Univ, IRD, APHM, MEPHI, IHU-Méditerranée Infection, 19-21 Boulevard Jean Moulin, 13385 Marseille Cedex 05, France

Introduction

The genus *Cohnella* was proposed in 2006 by Kämpfer, after the isolation of *Cohnella thermotolerans* during a hygiene control checks in a starch-producing company in Sweden (Kämpfer et al. 2006). The new species was defined as belonging to new genus based on chemical properties and the 16S rRNA gene sequencing similarities which was < 94.5% with all other recognized species of the genus *Paenibacillus* except *Paenibacillus hongkongensis* that was reclassified as *Cohnella hongkongensis* (Kämpfer et al. 2006). At the time of writing, the genus contains 24 species with *Cohnella thermotolerans* as type species (http://www.bacterio.net). *Cohnella* species are found in different environments such as soil (Jiang et al. 2012), fresh water (Shiratori et al. 2010), root nodules of legumes (García-Fraile et al. 2008; Wang et al. 2012; Xu et al. 2014), and volcanic pond (Cho et al. 2007). Concerning human and animals, *Cohnella hongkongensis*

was isolated from the blood culture of a patient with neutropenic fever and *Cohnella cellulosilytica* was isolated from buffalo feaces (Khianngam et al. 2012). While the rest of the genus members are environmental species. Members of this genus within the family of *Paenibacillaceae* are in general Gram-positive, some of them are Gram-negative or Gram variable, endospore-forming, aerobic, rod-shaped organisms.

The strain 6021052837^T was isolated from the blood culture of a hemodialysis patient in February 2016 at La Timone University Hospital in Marseille, France. A polyphasic strategy combining phenotypic and genomic characteristics was used to classify this new species (Ramasamy et al. 2014).

Materials and methods

Strain isolation and identification

Strain 6021052837^T was isolated in February 2016 on Chocolat PolyViteX medium (BioMérieux, Marcy l'Etoile, France) at 37 °C after 2 days of incubation of blood sample of a hemodialysis patient. Matrix assisted laser-desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) protein analysis was carried out on the colonies obtained using a Microflex LT spectrometer (Bruker Daltonics, Leipzig, Germany) as previously described (Seng et al. 2009; Dubourg et al. 2015). Strain 6021052837^T spectra were imported into the MALDI BioTyper software (version 2.0, Bruker) and analyzed by standard pattern matching (with default parameter settings) against the main spectra included in the database [Bruker database constantly updated with MEPHI database (http://www.mediterranee-infection.com/article.php?larub=280&titre=urms-database)]. In addition, 16S rRNA gene sequencing was performed on this strain, as previously described (Bittar et al. 2014). DNA extraction was realized using an EZ1 DNA Tissue Kit and BioRobot EZ1 Advanced XL (Qiagen, Courtaboeuf, France). The DNA extract was amplified using PCR technology and universal primers fD1 and rP2 (Eurogentec, Angers, France) (Weisburg et al. 1991). Then, the sequencing of the amplified products was performed using the Big Dye® Terminator v1.1 Cycle Sequencing Kit and ABI Prism 3130xl Genetic Analyzer capillary sequencer (Applied Biosystems), as previously described (Morel et al. 2015). The 16S rRNA gene sequence was compared with those available in GenBank (http://www.ncbi.nlm.nih.gov/genbank/).

Phenotypic, morphological, and biochemical characterization

Different pH using a Columbia agar with 5% sheep blood COS medium (BioMérieux, Marcy l'Etoile, France) with NaCl, $MgCl_2$, $MgSO_4$, KCl, $CaCl_2$, and glucose and different growth temperatures (4, 25, 28, 37 and 45 °C) were tested. The pH was modified by adding HCl to the medium and measured with a pH meter. Moreover, growth of strain 6021052837^T was tested under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems, respectively (BioMérieux), and under aerobic conditions, with or without 5% CO_2. To observe cells morphology, they were fixed with 2.5% glutaraldehyde in 0.1 M cacodylate buffer for at least 1 h at 4 °C. A drop of cell suspension was deposited for approximately 5 min on glow-discharged formvar carbon film on 400 mesh nickel grids (FCF400-Ni, EMS). The grids were dried on blotting paper and cells were negatively stained for 10 s with 1% ammonium molybdate solution in filtered water at RT. Electron micrographs were acquired with a Tecnai G20 Cryo (FEI) transmission electron microscope operated at 200 keV. API 50CH system (BioMérieux) was used for carbohydrate metabolism, as recommended by the manufacturer and performed in duplicate. The standard disc method was applied for antimicrobial susceptibility testing according to the Société Française de Microbiologie (SFM). The cellular fatty acid methyl ester (FAME) analysis was performed by Gas Chromatography/Mass Spectrometry (GC/MS). Two samples were prepared with approximately 55 mg of bacterial biomass per tube harvested from several culture plates. Fatty acid methyl esters were prepared as described (Sasser 2006). GC/MS analyses were carried out as previously described (Dione et al. 2016). Briefly, fatty acid methyl esters were separated using an Elite 5-MS column and monitored by mass spectrometry (Clarus 500—SQ 8 S, Perkin Elmer, Courtaboeuf, France). A spectral database search was performed using MS Search 2.0 operated with the Standard Reference Database 1A (NIST, Gaithersburg, USA) and the FAMEs mass spectral database (Wiley, Chichester, UK).

Genomic DNA preparation and genome sequencing

Genomic DNA (gDNA) from strain 6021052837^T was extracted using the EZ1 biorobot (Qiagen) with EZ1 DNA tissues kit as described previously (Abdallah et al. 2017). Genomic DNA was sequenced on the MiSeq Technology (Illumina Inc, San Diego, CA, USA) with the two applications: paired end and mate pair as described previously (Abdallah et al. 2017).

Genome assembly

The genome's assembly was performed as described previously using a pipeline which made it possible to create an assembly with different kinds of software [Velvet (Zerbino and Birney 2008), Spades (Bankevich et al. 2012) and Soap Denovo (Luo et al. 2012)], on trimmed [MiSeq and

Trimmomatic software (Bolger et al. 2014)] or untrimmed data (only MiSeq software) (Abdallah et al. 2017).

Genome annotation

Open Reading Frames (ORFs) were predicted using Prodigal (Hyatt et al. 2010) with default parameters. However, the predicted ORFs were excluded if spanning a sequencing gap region (contained N). The predicted bacterial protein sequences were searched in the Clusters of Orthologous Groups (COG) database using BLASTP (E value of $1e^{-03}$, coverage 0.7 and identity percent 30%). If no hit was found, it searches against the NR database using BLASTP, with the same parameters. If the sequence's length was smaller than 80 amino acids, the E value of $1e^{-05}$ was used. The tRNAScanSE (Lowe and Eddy 1997) tool was used to find transfer RNA genes, whereas ribosomal RNA genes were found using RNAmmer (Lagesen et al. 2007). ORFans were identified if all the BLASTP performed did not give positive results (E value smaller than $1e^{-03}$ for ORFs with sequence size superior to 80 aa or E-value smaller than $1e^{-05}$ for ORFs with sequence length smaller than 80 aa). Such parameter thresholds have already been used in previous works to define ORFans. A typical genes were identified by two criteria: genes with atypical G+C contents (genes with G+C% higher or lower than average G+C% plus or minus twice the standard deviation, respectively) and genes with codon usage bias score (calculated with CodonW) higher or lower than average score plus or minus twice the standard deviation. Argannot database was used to detect resistance genes (Gupta et al. 2014). ACLAME database (Leplae et al. 2010), Cluster MINE 360 (Conway and Boddy 2013) and MvirDB (Zhou et al. 2007) were used to determine respectively the mobilome, PKS or NRPS and virulence factors. These sequences were also submitted to Rapid Annotation using Subsystem Technology (RAST) (Aziz et al. 2008).

16S RNA phylogenetic tree

Phylogenetic trees highlighting the position of the new bacterium relative to other species was constructed. First, all 16S rRNA sequences of *Cohnella* species with validly published names were downloaded. Sequences were aligned using clustalW and phylogenetic inferences obtained using the approximately maximum-likelihood and neighbor joining analyses method. Kimura 2-parameter model was used within the MEGA 7 software (Tamura et al. 2011).

Comparative genomic analysis

Comparative genomic analysis was realized against available genomes of three *Cohnella* species and two *Paenibacillus* species. To determine the mean percentage of nucleotide sequences between orthologous genes, we used Average Genomic Identity of Orthologous gene Sequences (AGIOS) homemade software. OrthoAni was also used to compare genome similarity (Ouk Kim et al. 2016).

Strain and sequences deposition

Strain 6021052837^{T} has been deposited in two microbial culture collections: the German collection of microorganisms (Deutsche Sammlung von Mikroorganismen und Zellkulturen, DSMZ), under the accession number DSM 103,435, and the French culture collection (Collection de Souches de l'Unité des Rickettsies, CSUR), under the accession number CSUR 2659. The 16S rRNA gene and genome sequences are available in GenBank database under accession numbers LT223697 and FWCJ00000000, respectively. The Digital Protologue database (http://imedea.uib-csic.es/dprotologue) taxon number for strain 6021052837^{T} is TA00569.

Results and discussion

Phylogenetic affiliation

Strain 6021052837^{T} was isolated from the blood sample of a hemodialysis patient in February 2016 on Chocolat PolyViteX medium at 37 °C after 2 days of incubation. It could not be identified by our systematic MALDI-TOF screening as the score was < 1.7, thus leading to the 16S rRNA gene sequencing. The result showed that the novel isolate had 96% sequence identity with *Cohnella formosensis* (Genbank accession number JN806384), the phylogenetic most closely related species with standing in nomenclature. The percentage of sequence identity with other related species presented in the phylogenetic tree are respectively 95.8%, 95.1%, 95.1%, 95%, 94.5% for *C. lubricants, C. panacarvi, C. xylanilytica, C. thermotolerans, C. fontinalis*. This classified it as a putative new species of the genus *Cohnella* (Stackebrandt and Ebers 2006). Figure 1a, b illustrates the phylogenetic trees of strain 6021052837^{T} relative to other closest type species whose names have been validly published using maximum-likelihood and neighbor-joining methods, respectively.

Phenotypic results

Bacterial growth, morphology and characteristics for strain 6021052837^{T} are presented in the description of the species, Table 1 and Supplementary Fig. 1. API 50CH system results, after 24 h of incubation at 37 °C, are detailed in Supplementary Table 1. When compared to the phenotypic characteristics of other *Cohnella* species, strain 6021052837^{T} exhibited the differences detailed in Table 1. The major fatty acids

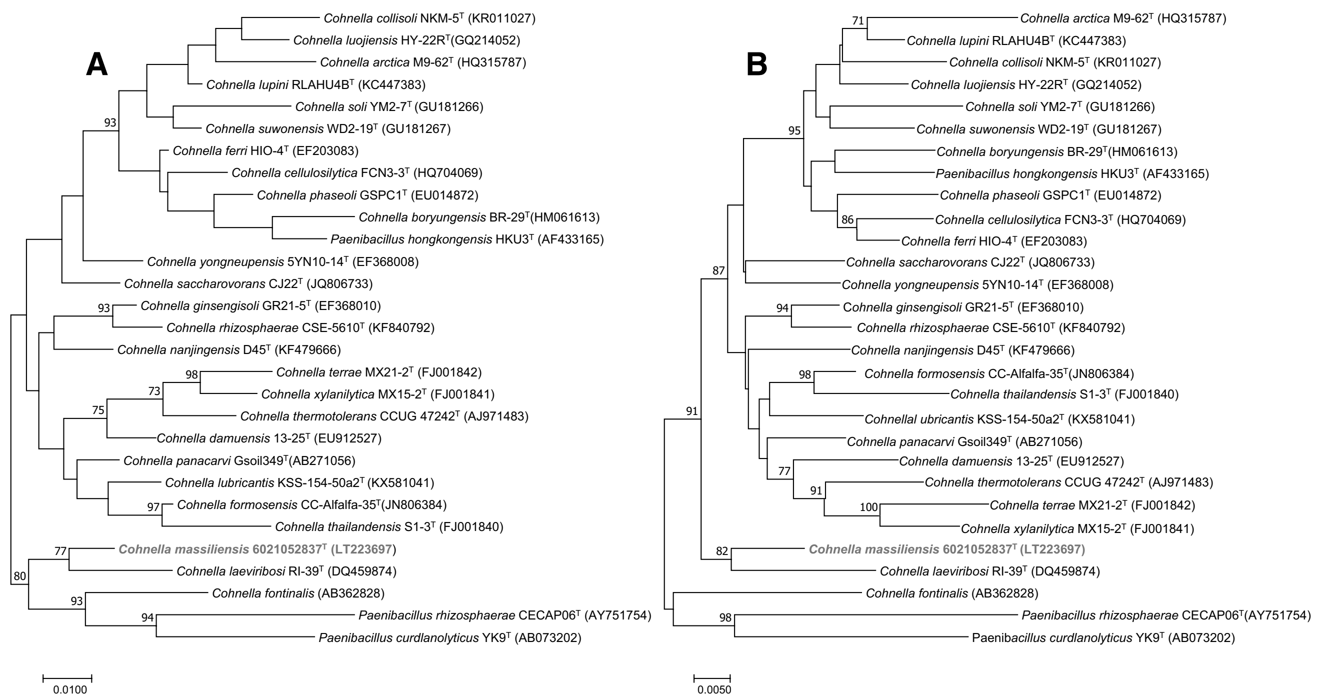

Fig. 1 Phylogenetic trees highlighting the position of *C. massiliensis* strain 6021052837^{T} relative to type strains within the genus *Cohnella*. Sequences were aligned by clustalW with default parameters and phylogenetic inferences were performed using Maximum Likelihood method (**a**) or neighbor-joining method (**b**) with 1000 bootstrap replicates, within MEGA7 software

found for this strain were branched saturated fatty acids: 14-methyl-pentadecanoic acid (34%) and 12-methyl-tetradecanoic acid (31%). Other unsaturated, branched or saturated fatty acids are described in Table 2.

Genome properties

The genome of strain 6021052837^{T} is 6.328 Mb long (Supplementary Fig. 2) with 57.24% GC content. It is assembled into 25 contigs. Out of the 5,710 predicted genes, 5,646 were protein-coding genes and 64 were RNAs (3 genes are 5S rRNA, 2 genes are 16S rRNA, 2 genes are 23S rRNA, 57 genes are tRNA genes). A total of 3881 genes (68.74%) were assigned as putative function and 288 genes were identified as ORFans (5.1%). The remaining genes (1312 genes) were annotated as hypothetical proteins (23.24%). The genome properties as well as the distribution of genes into COGs functional categories are detailed in Tables 3 and 4. Among carbohydrates analyzed via API 50CH system, genes coding for mannose, L-rhamnose, L-arabinose, D-ribose, maltose and saccharose were found in the draft genome.

Comparison with other *Cohnella* species genomes

The genome of *Cohnella massiliensis* was compared to those of three other *Cohnella* species, for which genome sequences are available, and two *Paenibacillus* species. The draft genome sequence of *Cohnella massiliensis* (6.328 Mb), is smaller than these of *Paenibacillus rhizosphaerae* (7.773 Mb), but larger than those of *Cohnella panacarvi, Cohnella thermotolerans, Paenibacillus curdlanolyticus and Cohnella laeviribosi* (4.773, 5.051, 5.453 and 4.476 Mb, respectively). The G+C content of *C. massiliensis* (57.2%) is lower than those of *C. thermotolerans* and *C. laeviribosi* (58.3 and 57.3%, respectively), but higher than those of *C. panacarvi, P. curdlanolyticus* and *P. rhizosphaerae* (54.5, 51.9 and 52.7%, respectively).

The distribution of genes into the different COG categories was almost similar between all compared species, although *P. rhizosphaereae* showed a different distribution as shown in Supplementary Fig. 3. The strain 6021052837^{T} shares 2206, 2086, 2086, 2375, 2360 orthologous genes with *P. rhizosphaerae, P. curdlanolyticus, C. panacarvi, C. laeviribosi*, and *C. thermotolerans*, respectively (Supplementary Table 2). Moreover, MAGi analysis showed that the Average Genomic Identity of Orthologus Gene Sequence (AGIOS) ranges from 58.26% with *P. curdlanolyticus* to 79.63% with *C. laeviribosi* (Supplementary Table 2). *C. laevribosi*, which is the phylogenetic neighbor of the new isolate (Fig. 1), is as well the closest species to strain 6021052837^{T} in term of genomic identity and number of shared orthologous genes. The low16S sequence similarity (< 97%) of strain 6021052837^{T} with all species within the genus *Cohnella*, enabled us to omit the

Table 1 Differential phenotypic characteristics between *C. massiliensis* strain 6021052837^T and other type strains of *Cohnella* species

Characteristic	1	2	3	4	5	6	7	8
Gram reaction	−	+	+	+	+	−	var	+
Catalase	+	+	na	+	+	+	+	+
Oxidase	+	−	+	+	+	+	+	+
Temperature range (°C)	37–45	37–52	20–55	37–45	18–45	25–55	< 55	< 55
Glycerol	−	+	na	−	na	−	−	−
D-arabinose	−	+	na	+	na	na	−	−
L-Arabinose	−	+	w	+	+	na	−	−
D-Ribose	−	+	+	na	−	−	−	−
D-Xylose	−	+	−	+	+	−	na	+
L-Xylose	−	−	na	−	−	na	−	−
Adonitol	−	+	−	na	−	−	−	−
D-Glucose	+	+	+	+	+	+	+	+
Fructose	+	+	+	na	+	−	w	−
D-Mannose	+	+	+	+	+	+	+	+
Sorbose	−	−	na	na	na	−	−	−
Rhamnose	+	+	+	na	+	−	−	+
Mannitol	−	+	−	+	−	+	−	+
Sorbitol	−	+	−	+	−	−	−	+
Amygdalin	+	+	na	na	na	−	−	+
Arbutin	+	−	+	+	na	−	w	−
D-Cellobiose	+	+	+	+	+	−	+	+
Maltose	+	+	+	+	+	+	+	+
Lactose	+	+	−	+	na	−	+	+
Melibiose	+	+	+	+	+	−	+	+
Trehalose	+	+	−	+	+	−	−	+
Inulin	+	−	na	na	na	−	−	+
Melezitose	+	−	na	−	na	−	−	−
Raffinose	+	+	−	+	+	−	−	−
Starch	+	+	na	na	−	na	+	w
Glycogen	+	+	na	na	−	−	−	+
Xylitol	+	+	na	+	−	−	−	−
Gentiobiose	var	−	na	+	na	−	+	−
Turanose	−	+	na	+	na	−	−	+
D-Lyxose	−	+	na	+	na	−	−	−
D-Tagatose	−	−	na	−	na	−	−	−
L-Fucose	+	+	na	+	na	−	−	−
D-Arabitol	−	+	-	+	na	−	−	+
L-Arabitol	−	+	na	−	na	−	−	+

var variable, + positive result, − negative result, *na* data not available, *w* weakly positive

Species: 1—*C. massiliensis* 6021052837^T; 2—*C. laeviribosi* RI-39^T (Cho et al. 2007); 3—*C. thermotolerans* CCUG 47242^T (Kämpfer 2006); 4—*C. xylanilytica* MX15-2^T (Khianngam et al. 2010); 5—*C. panacarvi* Gsoil 349^T (Yoon et al. 2007); 6—*C. fontinalis* YT-1101^T (Shiratori et al. 2010); 7—*C. lupini* RLAHU4B^T (David Flores-Felix et al. 2014); *C. formosensis* CC-Alfalfa-35^T (Hameed 2013)

DNA–DNA hybridization (DDH) study (Stackebrandt and Ebers 2006). Finally, the average nucleotide identity values (ANI) ranged between 69 and 79.2% observed with *P. curdlanolyticus* and *C. laeviribosi*, respectively, confirming that this strain is distinct from these other *Cohnella* species (Richter and Rosselló-Móra 2009) (Supplementary Fig. 4).

Conclusion

Consequently, all the above results lead to the conclusion that strain 6021052837^T is a novel species of the genus *Cohnella* for which the name *Cohnella massiliensis* is proposed.

Table 2 Cellular fatty acid composition (%) of strain 6021052837^T, *C. formosensis* strain CC-Alfalfa-35^T and *C. laevribosi* RI-39^T

Fatty acids	Name	6021052837^T	CC-Alfalfa-35^T	RI-39^T
16:0 iso	14-Methyl-pentadecanoic acid	34.3	32.8	40.5
15:0 anteiso	12-Methyl-tetradecanoic acid	31.4	31.4	22
16:00	Hexadecanoic acid	8.3		9.2
15:00	Pentadecanoic acid	5.5		1.3
15:0 iso	13-Methyl-tetradecanoic acid	5.3	8	11.7
18:00	Octadecanoic acid	1.2		
14:0 iso	12-Methyl-tridecanoic acid	3.5	7.1	3.9
17:0 anteiso	14-Methyl-hexadecanoic acid	2.9	2	5.8
18:1n9	9-Octadecenoic acid	1.3		
16:1 iso	14-Methylpentadecenoic acid	1.2		

Table 3 Nucleotide content and gene count levels of the genome of *Cohnella massiliensis* strain 6021052837^T

Characteristics	Number	Percent[a]
Size (bp)	6,328,479	100
Number of G+C	3,620,803	57.24
Total genes	5710	100
Protein coding genes	5646	98.88
Total RNA genes	64	1.12
tRNA genes	57	1
rRNA (5S, 16S, 23S) genes	7	0.12
Coding sequence size	5,494,158	86.82
Protein coding sequence size	5,480,178	86.6
tRNA coding sequence size	4466	0.07
rRNA coding sequence size	9514	0.15
Proteins associated to COGs	3239	57.37
Proteins not associated to COGs	2407	42.63
Proteins associated to ORFan	288	5.1
Proteins associated to hypothetical protein	1312	23.24
Proteins associated to predicted function by NR	807	14.29
Proteins associated to unknown function of COGs ([S])	165	2.92
Proteins associated to function prediction (nr + COGs not [S])	3881	68.74
Atypique G+C percent genes	320	5.67
Genes associated to resistance genes	0	0
Gene associated to PKS or NRPS	15	0.27
Genes associated to mobilome	2396	42.44
Genes associated to virulence	1042	18.46

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

Description of *Cohnella massiliensis* sp. nov

Cohnella massiliensis (mas.si.li.en'sis, L. fem. adj. *massiliensis*, of *Massilia*, the Roman name of Marseille, France, where type strain was isolated).

Cells are aerobic, Gram-negative, motile, spore-forming, catalase and oxidase-positive rods. The rod is approximately 2.4 µm long and 0.5 µm in diameter. Colonies on Columbia agar with 5% sheep blood COS medium are white, vary in size from 1.5 to 2 mm in diameter. Colonies are mucoid and cannot be identified as single entities after prolonged incubation. The optimal growth temperature is 37 °C. It grows at a salt concentration from 0 to 5 g/l after 24 h of incubation and survives from pH 5–8 with an optimal pH at 7. Acid production from carbohydrates is observed for D-glucose, D-mannose, aesculin, D-maltose, D-trehalose, glycerol, D-fructose, D-galactose, L-rhamnose, methyl-αD-mannopyranoside, methyl-αD-glucopyranoside, *N*-acetylglucosamine, amygdalin, arbutin, salicin,

Table 4 Number of genes associated with the 25 general COG functional categories (Strain 6021052837^T)

Code	Value	% of total[a]	Description
[J]	215	3.81	Translation
[A]	0	0	RNA processing and modification
[K]	331	5.86	Transcription
[L]	125	2.21	Replication, recombination and repair
[B]	1	0.02	Chromatin structure and dynamics
[D]	55	0.97	Cell cycle control, mitosis and meiosis
[Y]	0	0	Nuclear structure
[V]	96	1.7	Defense mechanisms
[T]	194	3.44	Signal transduction mechanisms
[M]	179	3.17	Cell wall/membrane biogenesis
[N]	51	0.9	Cell motility
[Z]	3	0.05	Cytoskeleton
[W]	5	0.09	Extracellular structures
[U]	50	0.89	Intracellular trafficking and secretion
[O]	104	1.84	Post-translational modification, protein turnover, chaperones
[X]	90	1.59	Mobilome: prophages, transposons
[C]	160	2.83	Energy production and conversion
[G]	518	9.17	Carbohydrate transport and metabolism
[E]	282	4.99	Amino acid transport and metabolism
[F]	109	1.93	Nucleotide transport and metabolism
[H]	175	3.1	Coenzyme transport and metabolism
[I]	123	2.18	Lipid transport and metabolism
[P]	248	4.39	Inorganic ion transport and metabolism
[Q]	86	1.52	Secondary metabolites biosynthesis, transport and catabolism
[R]	324	5.74	General function prediction only
[S]	165	2.92	Function unknown
_	2407	42.63	Not in COGs

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

D-cellobiose, D-lactose, D-melibiose, D-saccharose, inulin, D-melezitose, D-raffinose, starch, glycogen, xylitol, L-fucose and potassium gluconate. The most abundant fatty acids are branched saturated structures: 14-methyl-pentadecanoic acid (34%) and 12-methyl-tetradecanoic acid (31%). The cells were susceptible to all the antibiotics tested: amoxicillin, imipenem, amoxicillin/clavulanic acid, cefotaxime, penicillin, clindamycin, oxacillin, tobramycin, ceftriaxone, doxycycline, erythromycin and trimethoprim/sulfamethoxazole. The G+C content of the genome is 57.24%. The type strain is *Cohnella massiliensis* strain 6021052837^T (= CSUR P2659, =DSM 103435). The strain was isolated from the blood culture of a hemodialysis patient at La Timone University Hospital in Marseille, France.

Acknowledgements This work was supported by the French Government under the "Investissements d'avenir" (Investments for the Future) program managed by the Agence Nationale de la Recherche (ANR, fr: National Agency for Research) (reference: Méditerranée Infection 10-IAHU-03), and also by the "Fondation Méditerranée Infection".

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

Abdallah RA, Cimmino T, Baron S et al (2017) Description of *Chryseobacterium timonianum* sp. nov., isolated from a patient with pneumonia. Antonie Van Leeuwenhoek 1–12. https://doi.org/10.1007/s10482-017-0885-8

Aziz RK, Bartels D, Best AA et al (2008) The RAST server: rapid annotations using subsystems technology. BMC Genom 9:75. https://doi.org/10.1186/1471-2164-9-75

Bankevich A, Nurk S, Antipov D et al (2012) SPAdes—a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021

Bittar F, Keita MB, Lagier JC et al (2014) *Gorilla gorilla gorilla* gut: a potential reservoir of pathogenic bacteria as revealed using culturomics and molecular tools. Sci Rep 4:7174. https://doi.org/10.1038/srep07174

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Cho E-A, Lee J-S, Lee KC et al (2007) *Cohnella laeviribosi* sp. nov., isolated from a volcanic pond. Int J Syst Evol Microbiol 57:2902–2907. https://doi.org/10.1099/ijs.0.64844-0

Conway KR, Boddy CN (2013) ClusterMine360: a database of microbial PKS/NRPS biosynthesis. Nucleic Acids Res 41:D402–D407. https://doi.org/10.1093/nar/gks993

Dubourg G, Cimmino T, Senkar SA et al (2015) Noncontiguous finished genome sequence and description of *Paenibacillus antibioticophila* sp. nov. GD11(T), the type strain of *Paenibacillus antibioticophila*. New Microbes New Infect 8:137–147. https://doi.org/10.1016/j.nmni.2015.10.006

Flores-Felix D, Carro J, Ramirez-Bahena L M-H, et al (2014) *Cohnella lupini* sp nov., an endophytic bacterium isolated from root nodules of *Lupinus albus*. Int J Syst Evol Microbiol 64:83–87. https://doi.org/10.1099/ijs.0.050849-0

García-Fraile P, Velázquez E, Mateos PF et al (2008) *Cohnella phaseoli* sp. nov., isolated from root nodules of *Phaseolus coccineus* in Spain, and emended description of the genus *Cohnella*. Int J Syst Evol Microbiol 58:1855–1859. https://doi.org/10.1099/ijs.0.65468-0

Gupta SK, Padmanabhan BR, Diene SM et al (2014) ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 58:212–220. https://doi.org/10.1128/AAC.01310-13

Hameed A, Hung MH, Lin SY et al (2013) *Cohnella formosensis* sp. nov., a xylanolytic bacterium isolated from the rhizosphere of *Medicago sativa L.* Int J Syst Evol Microbiol 63:2806–2812. https://doi.org/10.1099/ijs.0.045831-0

Hyatt D, Chen G-L, LoCascio PF et al (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform 11:119. https://doi.org/10.1186/1471-2105-11-119

Jiang F, Dai J, Wang Y et al (2012) *Cohnella arctica* sp. nov., isolated from Arctic tundra soil. Int J Syst Evol Microbiol 62:817–821. https://doi.org/10.1099/ijs.0.030247-0

Kämpfer P, Rosselló-Mora R, Falsen E et al (2006) *Cohnella thermotolerans* gen. nov., sp. nov., and classification of "*Paenibacillus hongkongensis*" as *Cohnella hongkongensis* sp. nov. Int J Syst Evol Microbiol 56:781–786. https://doi.org/10.1099/ijs.0.63985-0

Khianngam S, Tanasupawat S, Akaracharanya A et al (2010) *Cohnella xylanilytica* sp nov and *Cohnella terrae* sp nov., xylanolytic bacteria from soil. Int J Syst Evol Microbiol 60:2913–2917. https://doi.org/10.1099/ijs.0.017855-0

Khianngam S, Tanasupawat S, Akaracharanya A et al (2012) *Cohnella cellulosilytica* sp. nov., isolated from buffalo faeces. Int J Syst Evol Microbiol 62:1921–1925. https://doi.org/10.1099/ijs.0.032607-0

Lagesen K, Hallin P, Rødland EA et al (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108. https://doi.org/10.1093/nar/gkm160

Leplae R, Lima-Mendez G, Toussaint A (2010) ACLAME: a classification of mobile genetic elements, update 2010. Nucleic Acids Res 38:D57–D61. https://doi.org/10.1093/nar/gkp938

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

Luo R, Liu B, Xie Y et al (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 2012 1:1 1:18. https://doi.org/10.1186/2047-217X-1-18

Morel A-S, Dubourg G, Prudent E et al (2015) Complementarity between targeted real-time specific PCR and conventional broad-range 16S rDNA PCR in the syndrome-driven diagnosis of infectious diseases. Eur J Clin Microbiol Infect Dis 34:561–570. https://doi.org/10.1007/s10096-014-2263-z

Ouk Kim Y, Chun J, Lee I, Park S-C (2016) OrthoANI: An improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol 66:1100–1103. https://doi.org/10.1099/ijsem.0.000760

Ramasamy D, Mishra AK, Lagier J-C et al (2014) A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol 64:384–391. https://doi.org/10.1099/ijs.0.057091-0

Richter M, Rosselló-Móra R (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci USA 106:19126–19131. https://doi.org/10.1073/pnas.0906412106

Sasser M (2006) Bacterial identification by gas chromatographic analysis of fatty acids methyl esters (GC-FAME). Technical Note 101. MIDI Inc., Newark, DE

Seng P, Drancourt M, Gouriet F et al (2009) Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 49:543–551. https://doi.org/10.1086/600885

Shiratori H, Tagami Y, Beppu T et al (2010) *Cohnella fontinalis* sp. nov., a xylanolytic bacterium isolated from fresh water. Int J Syst Evol Microbiol 60:1344–1348. https://doi.org/10.1099/ijs.0.014605-0

Stackebrandt E, Ebers J (2006) Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 33:152–

Tamura K, Peterson D, Peterson N et al (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739. https://doi.org/10.1093/molbev/msr121

Wang L-Y, Chen S-F, Wang L et al (2012) *Cohnella plantaginis* sp. nov., a novel nitrogen-fixing species isolated from plantain rhizosphere soil. Antonie Van Leeuwenhoek 102:83–89. https://doi.org/10.1007/s10482-012-9715-1

Weisburg WG, Barns SM, Pelletier DA et al (1991) 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol 173:697–703. https://doi.org/10.1128/jb.173.2.697-703.1991

Xu L, Zhang Y, Wang L et al (2014) Diversity of endophytic bacteria associated with nodules of two indigenous legumes at different altitudes of the Qilian Mountains in China. Syst Appl Microbiol 37:457–465. https://doi.org/10.1016/j.syapm.2014.05.009

Yoon M-H, Ten LN, Im W-T (2007) *Cohnella panacarvi* sp. nov., a xylanolytic bacterium isolated from ginseng cultivating soil. J Microbiol Biotechnol 17:913–918

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829. https://doi.org/10.1101/gr.074492.107

Zhou CE, Smith J, Lam M et al (2007) MvirDB–a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. Nucleic Acids Res 35:D391–D394. https://doi.org/10.1093/nar/gkl791