Noncontiguous finished genome sequence and description of Nocardioides massiliensis sp. nov. GD13<sup>T</sup>

G. Dubourg<sup>1,2</sup>, S. A. Sankar<sup>1</sup>, J. Rathored<sup>1</sup>, J.-C. Lagier<sup>1,2</sup>, C. Robert<sup>1</sup>, C. Coudere<sup>2</sup>, L. Papazian<sup>3</sup>, D. Raoult<sup>1,2,4</sup> and P.-E. Fournier<sup>1,2</sup>

<sup>1</sup>Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, U9M 63, CNRS 7278, IRD 198, INSERM 1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de Médecine, Aix-Marseille Université, 2 Pôle des Maladies Infectieuses et Tropicales Clinique et Biologique, Fédération de Bactériologie–Hygéne–Virologie, Université, Hospital Centre Timane, Institut Hospitalo-Universitaire (IHU) Méditerranée Infection, Assistance Publique–Hôpitaux de Marseille, 3 Service de Réanimation Médicale-Dépistage Respiratoire et Infections Sévères, Hôpital Nord, Marseille, France and 4 Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdul Aziz University, Jeddah, Saudi Arabia

Abstract

Nocardioides massiliensis sp. nov strain GD13<sup>T</sup> is the type strain of N. massiliensis sp. nov., a new species within the genus Nocardioides. This strain was isolated from the faeces of a 62-year-old man admitted to intensive care for Guillain-Barré syndrome. Nocardioides massiliensis is a strictly aerobic Gram-positive rod. Herein we describe the features of this bacterium, together with the complete genome sequence and annotation. The 4 006 620 bp long genome contains 4132 protein-coding and 47 RNA genes.

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, taxonogenomics

Original Submission: 20 November 2015; Revised Submission: 31 December 2015; Accepted: 4 January 2016

Article published online: 9 January 2016

Introduction

Nocardioides massiliensis strain GD13<sup>T</sup> (= DSM 28216 = CSUR P894) is the type strain of N. massiliensis sp. nov. This bacterium was isolated from the human faeces of a 62-year-old man admitted to intensive care for a Guillain-Barré syndrome as part of an effort called culturomics to cultivate all bacterial species from the human gut [1,2]. The current taxonomic classification of prokaryotes is based on a combination of phenotypic and genotypic characteristics [3,4], which includes 16S rRNA gene phylogeny, G+C content and DNA-DNA hybridization. Although these tools are considered the reference standard, they have several drawbacks [5,6]. As a result of the declining cost of sequencing, the number of sequenced bacterial genomes has rapidly grown; to date, almost 40 000 bacterial genomes have been sequenced (GOLD Database, https://gold.jgi.doe.gov/). Hence, we recently proposed to use genomic information together with phenotypic criteria for the description of new bacterial species [7–20].

Here we present a summary classification and a set of features for Nocardioides massiliensis sp. nov. strain GD13<sup>T</sup> together with the description of the complete genomic sequencing and annotation. These characteristics support the circumscription of the species Nocardioides massiliensis. The genus Nocardioides was created in 1976 (Prauser, 1976). The first two species were Nocardioides albus and Nocardioides luteus [21,22]. These Gram-positive, aerobic and cardioform actinomycetes exhibit a high G+C content and are able to develop mycelium [21,22]. Currently the genus Nocardioides contains 72 species, including eight described in 2013 and species from the genera Pimelobacter and Arthrobacter that were reclassified in the genus Nocardioides [23–36] (LPSN, http://www.bacterio.net/). Members of the genus Nocardioides are mostly environmental bacteria found in soil or plants. To our knowledge, this
is the first report of isolation of a Nocardioides species from humans.

**Organism Information and Classification**

A stool sample was collected from a 62-year-old man admitted to the intensive care unit of Hôpital Nord, Marseille, France, in March 2012 [2]. The study was approved by the ethics committee of the Institut Fédératif de Recherche IFR48 (Faculty of Medicine, Marseille, France), under agreement 09-002. At the time of sampling, the patient had received a 10-day course of imipenem (3 g per day). The faecal specimen was preserved at −80°C after collection. Strain GD13T (Table 1) was isolated in March 2012 by cultivation on PVX agar (bioMérieux, Marcy l’Étoile, France) in an aerobic atmosphere at 28°C after 14 days of incubation. Strain GD13T exhibited a 98.07% 16S rRNA sequence identity with *N. massiliensis* strain GD13 T according to MIGS validly published name (Figure 1). Its 16S rRNA gene sequence was deposited in European Molecular Biology Laboratory—European Bioinformatics Institute (EMBL-EBI) under number HF952922. This value was lower than the 98.7% 16S rRNA gene sequence threshold recommended by Stackebrandt and Ebers [4] to delineate a new species without carrying out DNA-DNA hybridization.

**Phenotypic and biochemical features**

Growth was tested at different temperatures (25, 30, 37, 45, 56°C). It occurred between 25 and 37°C after 24 to 48 hours of incubation. No growth was observed at 45 or 56°C. Colonies were 0.5 mm in diameter on 5% sheep’s blood—enriched Columbia agar (bioMérieux). Growth of the strain was tested under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems (bioMérieux) respectively, and under aerobic conditions, with or without 5% CO₂. It was achieved only aerobically. Gram staining showed Gram-positive, rod-shaped bacteria unable to form spores (Figure 2). A motility test was negative. Cells grown on agar were soft and yellowish in colour after 24 to 48 hours and could be grouped in small clumps (Figure 2). They had a mean width of 0.46 µm and a mean length of 1 µm (Figure 3).

Strain GD13T exhibits catalase activity, but was negative for oxidase. Using an API ZYM strip (bioMérieux), positive reactions were observed for esterase (C4), esterase lipase (C8), acid phosphatase, alkaline phosphatase, leucine arylamidase, naphthol-AS-BI-phosphohydrolase, α-glucosidase and β-glucosidase; negative reactions were observed for β-galactosidase, α-galactosidase, valine arylamidase, cystine arylamidase, β-glucuronidase, lipase (C14), trypsin, α-chymotrypsin, N-acetyl-β-glucosaminidase, α-mannosidase and α-fucosidase. Using rapid ID32A (bioMérieux), positive reactions were observed for arginine arylamidase, proline arylamidase and α-fucosidase and negative reactions were observed for arginine dihydrolase and N-acetyl-β-glucosaminidase. Using rapid ID32A (bioMérieux), positive reactions were observed for arginine arylamidase, proline arylamidase and α-fucosidase; negative reactions were observed for nitrate reduction, urease, production of indole, arginine dihydrolase, N-acetyl-β-glucosaminidase, serine arylamidase, histidin arylamidase, phenylalanine arylamidase and tyrosin arylamidase. Using an API 50 CH strip (bioMérieux), positive reactions were recorded for esculin hydrolysis and fermentation of d-lactose; negative reactions were obtained for fermentation of glycerol, erythritol, d-arabinose, L-arabinose, d-ribose, d-xylene, L-xylene, d-adonitol, methyl-β-d-xlyopyranoside, d-galactose, d-glucose, d-fructose, d-mannose, L-sorbose, L-rhamnose, dulcitol, inositol, d-mannitol, methyl-α-d-xlyopyranoside, methyl-α-d-glucopyranoside, N-acetylgulcosamine, amygdalin, arbutin, esculin ferric citrate, salicin, d-cellobiose, D-maltose, D-melibiose, D-}

### TABLE I. Classification and general features of *Nocardioides massiliensis* strain GD13T according to MIGS recommendations

| MIGS ID | Property | Term | Evidence code⁠* |
|---------|----------|------|----------------|
| MIGS-6.3 | Salinity | Unknown | IDA |
| MIGS-22 | Oxygen requirement | Aerobic | IDA |
| MIGS-6 | Energy source | Unknown | IDA |
| MIGS-13 | Biotic relationship | Free-living | IDA |
| MIGS-14 | Pathogenicity | Unknown | IDA |
| MIGS-4 | Geographic location | Marseille | IDA |
| MIGS-5 | Sample collection | March 2012 | IDA |
| MIGS-4.1 | Latitude | 43.296482 | IDA |
| MIGS-4.1 | Longitude | 5.36978 | IDA |
| MIGS-4.3 | Depth | Surface | IDA |
| MIGS-4.4 | Altitude | 0 m above sea level | IDA |

* MIGS, minimum information about a genome sequence.
* Evidence codes are as follows: IDA, inferred from direct assay; and TAS, traceable author statement (i.e. a direct report exists in the literature). These evidence codes are from the Gene Ontology project (http://www.geneontology.org/GO.evidence. html) [10]. If the evidence code is IDA, then the property should have been directly observed, for the purpose of this specific publication, for a live isolate by one of the authors, or by an expert or reputable institution mentioned in the acknowledgements.

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, 47–57. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).
trehalose, inulin, D-melezitose, D-raffinose, amidon, glycogen, gentiobiose, D-turanose, D-lyxose, D-tagatose, D-fucose, L-fucose, L-arabitol, D-sorbitol, D-saccharose, xylitol, D-arabitol, potassium-5-ketogluconate and potassium 2-ketogluconate. Cells were susceptible to penicillin G, amoxicillin, amoxicillin/clavulanic acid, ceftriaxone, imipenem, vancomycin, rifampicin, erythromycin, gentamicin, ciprofloxacin and trimethoprim/sulfamethoxazole and resistant to metronidazole. Compared to N. mesophilus, its phylogenetically closest neighbor, N. massiliensis differed in α-glucosidase, β-glucosidase, esterase–lipase production and utilization of D-glucose and D-mannose (Table 2). In addition, Nocardioides species clearly differed from Aeromicrobium massiliense, a representative species from another genus within the family Nocardioidaceae.

MALDI-TOF analysis

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) protein analysis was carried out as previously described [46]. Briefly, a pipette tip was used to pick one isolated bacterial colony from a culture agar plate and
spread it as a thin film on a MSP 96 MALDI-TOF target plate (Bruker Daltonics, Leipzig, Germany). Twenty distinct deposits from 20 isolated colonies were performed for strain GD13\(^T\). Each smear was overlaid with 2 \(\mu\)L of matrix solution (saturated solution of alpha-cyano-4-hydroxycinnamic acid in 50% acetonitrile, 2.5% trifluoracetic 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 of 2000 to 20 000 Da (parameter settings: ion source 1 (IS1), 20 kV; IS2, 18.5 kV; lens, 7 kV). A spectrum was obtained after 240 shots with variable laser power. The time of acquisition was between 30 seconds and 1 minute per spot. The 20 GD13\(^T\) spectra were imported into MALDI BioTyper 3.0 software (Bruker) and analysed by standard pattern matching (with default parameter settings) against the main spectra of 7379 bacteria, including 48 spectra from two Nocardioides species. It corresponded to the Bruker database combined with our own database constantly incremented with new species. The method of identification included the m/z from 3000 to 15 000 Da. For every spectrum, a maximum of 100 peaks were compared with spectra in the database. The resulting score enabled the identification or not of tested species: a score of \(\geq 2\) with a validly published species enabled identification at the species level, a score of \(\geq 1.7\) but <2 enabled identification at the genus level and a score of <1.7 did not enable any identification. No significant MALDI-TOF score was obtained for strain GD13\(^T\) against the Bruker database and against our own database, suggesting that our isolate was not a member of a known species. We added the spectrum from strain GD13\(^T\) to our database (Figure 4). Finally, the gel view showed the spectral differences with other members of the genus Nocardioides and Aeromicrobium massiliense, a member of the genus Aeromicrobium that is closely related to the genus Nocardioides (Figure 5).

**Genome Sequencing Information**

**Genome project history**

*Nocardioides massiliensis* GD13\(^T\) was selected for sequencing on the basis of its phenotypic differences, phylogenetic position and 16S rRNA sequence similarity to other members of the genus *Nocardioides*. It was part of a culturomics study aimed at isolating all bacterial species from the human digestive flora from patients treated with broad-spectrum antibiotics. It is the first sequenced genome of *N. massiliensis* sp. nov. The GenBank Bioproject number is PRJEB1962 and consists of 839 large contigs. Table 3 shows the project information and its association with minimum information about a genome sequence (MiGS) 2.0 compliance [47].

**Growth conditions and DNA isolation**

*Nocardioides massiliensis* strain GD13\(^T\) (= DSM 28216 = CSUR P894) was cultured aerobically on 5% sheep’s blood–enriched Columbia agar (bioMérieux). Bacteria grown on four petri dishes were resuspended in sterile water and centrifuged at 4°C at 2000 \(\times\) g for 20 minutes. Cell pellets were resuspended in 1 mL Tris/EDTA/NaCl (10 mM Tris/HCl (pH 7.0), 10 mM EDTA (pH 8.0), 300 mM NaCl) and recentrifuged under the same conditions. The pellets were then resuspended in 200 \(\mu\)L Tris-EDTA (TE) buffer and proteinase K and kept overnight at
37°C for cell lysis. DNA was purified with phenol/chloroform/isoamylalcohol (25:24:1), followed by an overnight precipitation with ethanol at −20°C. The DNA was resuspended in 205 μL TE buffer. DNA concentration was 401.63 ng/μL as measured by Qubit fluorometer using the high-sensitivity kit (Life Technologies, Carlsbad, CA, USA).

### Genome sequencing and assembly

Genomic DNA (gDNA) of *Nocardioides massiliensis* was sequenced on a MiSeq sequencer (Illumina, San Diego, CA, USA) with two applications: paired end and mate pair. The paired end and the mate pair strategies were barcoded in order to be mixed respectively with 14 other genomic projects and 11 other projects with the Nextera Mate pair kit (Illumina). The mate pair library was constructed from 1 μg of gDNA using the Nextera Mate Pair Illumina guide. The gDNA sample was simultaneously fragmented and tagged with a mate pair junction adapter. The profile of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) with a DNA7500 lab chip. The DNA fragments ranged in size from 1 to 8.4 kb with a mean size of 3.7 kb. No size selection was performed, and 240 ng tagmented fragments were circularized. The larger circularized DNA molecules were pooled into a single library for sequencing on the MiSeq. The pooled single strand library was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and paired-end sequencing with dual index reads was performed in a single 39-hour run at a 2 × 250 bp read length. Within this pooled run, the index representation was determined to be 3.93%. Total information of 3.27 Gb was obtained from a 374K/mm² density with 94.2% (6 719 000 clusters) of the clusters passing quality control (QC) filters. From the genome sequencing process, the 248 710 produced Illumina paired end reads for *Nocardioides massiliensis* were filtered according to the read qualities.

The mate pair library was constructed from 1 μg of gDNA using the Nextera Mate Pair Illumina guide. The gDNA sample was simultaneously fragmented and tagged with a mate pair junction adapter. The profile of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) with a DNA7500 lab chip. The DNA fragments ranged in size from 1 to 8.4 kb with a mean size of 3.7 kb. No size selection was performed, and 240 ng tagmented fragments were circularized. The larger circularized DNA molecules were pooled into a single library for sequencing on the MiSeq. The pooled single strand library was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and paired-end sequencing with dual index reads was performed in a single 39-hour run at a 2 × 250 bp read length. Within this pooled run, the index representation was determined to be 3.93%. Total information of 3.27 Gb was obtained from a 374K/mm² density with 94.2% (6 719 000 clusters) of the clusters passing quality control (QC) filters. From the genome sequencing process, the 248 710 produced Illumina paired end reads for *Nocardioides massiliensis* were filtered according to the read qualities.

The mate pair library was constructed from 1 μg of gDNA using the Nextera Mate Pair Illumina guide. The gDNA sample was simultaneously fragmented and tagged with a mate pair junction adapter. The profile of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) with a DNA7500 lab chip. The DNA fragments ranged in size from 1 to 8.4 kb with a mean size of 3.7 kb. No size selection was performed, and 240 ng tagmented fragments were circularized. The larger circularized DNA molecules were pooled into a single library for sequencing on the MiSeq. The pooled single strand library was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and paired-end sequencing with dual index reads was performed in a single 39-hour run at a 2 × 250 bp read length. Within this pooled run, the index representation was determined to be 3.93%. Total information of 3.27 Gb was obtained from a 374K/mm² density with 94.2% (6 719 000 clusters) of the clusters passing quality control (QC) filters. From the genome sequencing process, the 248 710 produced Illumina paired end reads for *Nocardioides massiliensis* were filtered according to the read qualities.

The mate pair library was constructed from 1 μg of gDNA using the Nextera Mate Pair Illumina guide. The gDNA sample was simultaneously fragmented and tagged with a mate pair junction adapter. The profile of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) with a DNA7500 lab chip. The DNA fragments ranged in size from 1 to 8.4 kb with a mean size of 3.7 kb. No size selection was performed, and 240 ng tagmented fragments were circularized. The larger circularized DNA molecules were pooled into a single library for sequencing on the MiSeq. The pooled single strand library was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and paired-end sequencing with dual index reads was performed in a single 39-hour run at a 2 × 250 bp read length. Within this pooled run, the index representation was determined to be 3.93%. Total information of 3.27 Gb was obtained from a 374K/mm² density with 94.2% (6 719 000 clusters) of the clusters passing quality control (QC) filters. From the genome sequencing process, the 248 710 produced Illumina paired end reads for *Nocardioides massiliensis* were filtered according to the read qualities.
FIG. 4. Reference mass spectrum from Nocardioides massiliensis strain GD13\textsuperscript{T}. Spectra from 17 individual colonies were compared and reference spectrum was generated.

FIG. 5. Gel view comparing Nocardioides massiliensis strain GD13\textsuperscript{T}. Gel view displays raw spectra of loaded spectrum files arranged in pseudo-gel-like look. x-axis records m/z value. Left y-axis displays running spectrum number originating from subsequent spectra loading. Peak intensity is expressed in arbitrary units by greyscale scheme code, as indicated on right y-axis. Displayed species are indicated at left.
physically sheared to smaller sized fragments with a mean size of 839 bp on the Covaris device S2 in microtubes (Woburn, MA, USA). The library’s profile and the quantitation were visualized on a High Sensitivity Bioanalyzer LabChip. The libraries were normalized to 2 nM and pooled. After a denaturation step and dilution at 10 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 sequencing run was performed in a single 39-hour run at a 2 × 250 bp read length. Total mate paired run information of 3.2 Gb was obtained from a genome sequencing process, the 1 027 586 predicted genes, 4132 were protein-coding genes and 47 were RNA genes. Three rRNA genes (one 16S rRNA, one 23S rRNA and one 5S rRNA) and 44 predicted tRNA genes 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.

For the genomic comparison, we compared the genome sequence of N. massiliensis strain GD13T (GenBank accession no. CCX000000000) with those of Nocardioides alkalitoleraens strain DSM 16699 (GenBank accession no. AUFN000000000), Nocardioides halotolerans strain DSM 19273 (AUGT000000000), Nocardioides insulare strain DSM 17944 (AUGU000000000) and Aeromicrobium massiliense strain JC14T (CAHG000000000). To estimate the nucleotide sequence similarity at the genome level between N. massiliensis and other members of the family Nocardioidaceae, we determined the average genomic identity of orthologous gene sequences (AGIOS) parameter using the MAGi (Marseille Average genomic identity) software as follows: orthologous proteins were detected using the Proteinortho software (version 1.4) [54] using the following parameters: E value 1e-5, 30% percentage of identity, 50% coverage and algebraic connectivity of 50%; after fetching the corresponding nucleotide sequences of orthologous proteins for each pair of genomes, we determined the mean percentage of nucleotide sequence identity using the Needleman-Wunsch global alignment algorithm; Artemis [55] was used for data management, and DNAPlotter [56] was used for visualization of genomic features. The Mauve alignment tool was used for multiple genomic sequence alignment and visualization [57].

**Genome properties**

The genome of N. massiliensis strain GD13T is 4 006 620 bp long with a 71.0% G+C content (Table 4, Figure 6). Of the 4179 predicted genes, 4132 were protein-coding genes and 47 were RNAs. Three rRNA genes (one 16S rRNA, one 23S rRNA and one 5S rRNA) and 44 predicted tRNA genes were identified in the genome. A total of 2783 genes (66.59%) were assigned a putative function, and 382 genes (9.24%) were identified as ORFans. The remaining genes were annotated as hypothetical proteins. The properties and the statistics of the genome are summarized in Table 4. The distribution of genes into COGs functional categories is presented in Table 5.

**Extended insights from the genome sequence**

The draft genome of N. massiliensis has a larger size than that of A. massiliense (4.00 and 3.32 Mb respectively) but smaller than

---

**Table 3. Project information**

| MIGS ID | Property | Term |
|---------|----------|------|
| MIGS-31 | Finishing quality | High-quality draft |
| MIGS-28 | Libraries used | Mite pair |
| MIGS-29 | Sequencing platform | MiSeq |
| MIGS-31.2 | Fold coverage | Bfx |
| MIGS-30 | Assemblers | CLCGENOMICSWB4 |
| MIGS-32 | Gene calling method | Prodigal |
| EMBL-EBI Date of Release | 18 September 2014 |
| NCBI project ID | PRJEB2092 |
| EMBL-EBI accession | CCX000000000 |
| MIGS-13 | Project relevance | Study of human gut microbiota |

---

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

| Attribute       | Value     | % of total
|-----------------|-----------|------------|
| Size (bp)       | 4 006 620 | 100        |
| G+C content (bp)| 2 844 700 | 71.0       |
| Coding region (bp) | 3 668 187 | 91.55     |
| Total genes     | 4179      | 100        |
| RNA genes       | 47        | 1.12       |
| Protein-coding genes | 4132   | 98.87     |
| Genes with function prediction | 2408 | 57.62 |
| Genes assigned to COGs | 2783 | 66.59 |
| Genes with peptide signals | 262 | 6.36 |
| Genes with transmembrane helices | 899 | 21.51 |

COGs, Clusters of Orthologous Groups database.

*Total is based on either size of genome in base pairs or total number of protein-coding genes in annotated genome.*
N. alkalitolerans, N. halotolerans and N. insulae (4.85, 4.86 and 4.06 Mb respectively). The G+C content of N. massiliensis is higher than those of N. insulae (71.0 and 70.8% respectively) but less than that of A. massiliense, N. alkalitolerans and N. halotolerans (72.6, 73.1 and 72% respectively). The gene content of N. massiliensis is larger than those of A. massiliense and N. insulae (4179, 3328 and 3702 respectively) and smaller than those of N. alkalitolerans and N. halotolerans (4634 and 4633 respectively). However, the distribution of genes into COGs categories was similar in all compared genomes (Figure 7). In addition, N. massiliensis shared 1447, 1745, 1693 and 1598 orthologous genes with A. massiliense, N. alkalitolerans, N. halotolerans and N. insulae. Among Nocardioides species, AGIOS values ranged from 75.67% between N. alkalitolerans and N. insulae to 77.56% between N. alkalitolerans and N. halotolerans. When considering N. massiliensis, AGIOS values were 75.06, 76.67 and 76.82 with N. insulae, N. alkalitolerans and N. halotolerans respectively, in the same range as among other Nocardioides species (Table 6).

**FIG. 6.** Graphical circular map of chromosome of Nocardioides massiliensis strain GD13<sup>T</sup>. From outside in, outer two circles show ORFs oriented in forward (coloured by COGs categories) and reverse (coloured by COGs categories) direction respectively. Third circle marks rRNA gene operon (red) and tRNA genes (green). Fourth circle shows G+C% content plot. Innermost circle shows GC skew, with purple indicating negative values and olive positive values.

**TABLE 5.** Number of genes associated with 25 general COGs functional categories

| Code | Value | %  | Description                                      |
|------|-------|----|--------------------------------------------------|
| J    | 166   | 4.01| Translation                                      |
| A    | 1     | 0.02| RNA processing and modification                  |
| K    | 240   | 5.80| Transcription                                    |
| L    | 214   | 5.17| Cell cycle control, mitosis and meiosis          |
| B    | 1     | 0.02| Nuclear structure                                |
| D    | 28    | 0.67| Defense mechanisms                               |
| T    | 112   | 2.71| Signal transduction mechanisms                   |
| Y    | 0     | 0   | Cell wall/membrane biogenesis                    |
| V    | 49    | 1.18| Call motility                                    |
| Z    | 0     | 0   | Cytoskeleton                                     |
| W    | 0     | 0   | Extracellular structures                         |
| U    | 43    | 1.04| Intracellular trafficking and secretion          |
| C    | 229   | 5.54| Energy production and conversion                 |
| G    | 159   | 3.84| Carbohydrate transport and metabolism            |
| E    | 315   | 7.62| Amino acid transport and metabolism              |
| F    | 73    | 1.76| Nucleotide transport and metabolism              |
| H    | 126   | 3.04| Coenzyme transport and metabolism                |
| I    | 268   | 6.48| Lipid transport and metabolism                   |
| P    | 213   | 5.15| Inorganic ion transport and metabolism           |
| Q    | 178   | 4.30| Secondary metabolites biosynthesis, transport and catabolism |
| R    | 460   | 11.13| General function prediction only                 |
| S    | 190   | 4.59| Function unknown                                 |
|     | 1396  | 33.4| Not in COGs                                     |

COGs, Clusters of Orthologous Groups database.
Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Nocardioides massiliensis* sp. nov. The strain was isolated from a stool sample from a 62-year-old man admitted to intensive care in Marseille, France, for Guillain-Barré syndrome and who had received 10 days of broad-spectrum antibiotics.

Description of *Nocardioides massiliensis* strain GD13<sup>T</sup> sp. nov

*Nocardioides massiliensis* (mas.si.li.en’sis, L. masc. adj., from *massiliensis* of Massilia, the ancient Roman name for Marseille, France, where the type strain was isolated).

Cells are Gram-positive, nonsporulating, nonmotile and rod-shaped bacilli with a mean diameter and length of 0.46 and 1.0 μm respectively. Colonies are 0.5 mm in diameter and are smooth and pale yellow on 5% sheep’s blood–enriched Columbia agar (bioMérieux). Growth occurred aerobi
caly between 25 and 37°C.

Cells exhibit catalase activity but are oxidase negative. Positive reactions are observed for esterase (C4), esterase lipase (C8), acid phosphatase, alkaline phosphatase, leucine arylamidase, naphthol-AS-BI-phosphohydrolase, α-glucosidase, β-glucosidase, arginine arylamidase, proline arylamidase, esculin hydrolysis and fermentation of D-lactose. Negative reactions are observed for β-galactosidase, valine arylamidase, cystine arylamidase, phenylalanine arylamidase, tyrosin arylamidase, fermentation of glycerol, erythritol, d-arabinose, l-arabinose, d-ribose, d-xylene, l-xylene, d-adonitol, methyl-β-D-xylanopyranoside, d-galactose, d-glucose, d-fructose, d-mannose, l-sorbitose, l-rhamnose, dulcitol, inositol, d-mannitol, methyl-β-D-d glucopyranoside, N-acetylglucosaminidase, amylglucosidase, arbutin, esculin ferric citrate, salicin, d-cellobiose, β-maltose, d-melibiose, d- trehalose, inulin, d-melezitose, d-raffinose, amidon, glycogen, gentiobiose, d-turanose, d-lyxose, d-tagatose, d-fucose, l-fucose, l-arabinol, d-sorbitol, d-saccharose, xylitol, d-arabitol, potassium-5-ketogluconate and potassium 2-ketogluconate. Cells are susceptible to penicillin G, amoxicillin, amoxicillin-clavulanic acid, ceftriaxone, imipenem, vancomycin, rifampicin, erythromycin, gentamicin, ciprofloxacin and trimethoprim/sulfamethoxazole and resistant to metronidazole.

The G+C content of the genome is 71.0%. The 16S rRNA and whole-genome sequence of *N. massiliensis* strain GD13<sup>T</sup> are deposited in EMBL-EBI under the accession numbers HF952922 and CCXJ00000000 respectively. The type strain GD13<sup>T</sup> (= CSUR P894 = DSM 28216) was isolated from the faecal flora of a 62-year-old man admitted to intensive care in Marseille, France, for Guillain-Barré syndrome who had received 10 days of broad-spectrum antibiotics.

### Table 6. Pairwise comparison of *Nocardioides massiliensis* strain GD13<sup>T</sup> with other species showing (upper right) numbers of orthologous proteins shared between compared genomes and (lower left) AGIOS values

|              | N. massiliensis | Aeromicrobium massiliense | N. alkalitolerans | N. halotolerans | N. insulae |
|--------------|----------------|--------------------------|-------------------|----------------|-----------|
| *N. massiliensis* | 4132<sup>a</sup> | 1447                     | 1745              | 1693           | 1598      |
| A. massiliense | 75.17          | 2297<sup>a</sup>         | 1605              | 1501           | 1517      |
| N. alkalitolerans | 76.67      | 75.61                    | 4585<sup>a</sup>  | 1822           | 1822      |
| N. halotolerans | 76.82        | 75.14                    | 77.56             | 4677<sup>a</sup> | 1753      |
| N. insulae    | 75.06         | 74.12                    | 75.67             | 76.73          | 36.48<sup>b</sup> |

AGIOS, average genomic identity of orthologous gene sequences.

<sup>a</sup>Numbers of proteins per genome.
Acknowledgements

The authors thank the Xegen Company (http://www.xegen.fr) for automating the genomic annotation process. They also thank K. Griffiths for English-language review and C. Andrieu for administrative assistance. This study was funded by the Fondation Méditerranée Infection.

Conflict of Interest

None declared.

References

[1] Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturomes: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 2012;18:1185–93.
[2] Dubourg G, Lagier JC, Armougom F, Robert C, Hamad I, Brouqui P, et al. Culturomics and pyrosequencing evidence of the reduction in gut microbiota diversity in patients with broad-spectrum antibiotics. Int J Antimicrob Agents 2014;4:117–24.
[3] Tindall BJ, Rosselló-Mora R, Busse HJ, Ludwig W, Kämpfer P. Notes on the characterization of prokaryote strains for taxonomic purposes. Int J Syst Evol Microbiol 2010;60:249–66.
[4] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152–5.
[5] Stackebrandt E, Frederiksen W, 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(Pt 3):1043–7.
[6] Rosselló-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; 2006. p. 23–50.
[7] 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 2012;64(Pt 3):960–1.
[8] Fournier PE, Lagier JC, Dubourg G, Raoult D. From culturomics to taxonomogenomics: a need to change the taxonomy of prokaryotes in clinical microbiology. Anaerobe 2015;36:73–8.
[9] 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;6:116–25.
[10] Lagier JC, Armougom F, Mishra AK, Nguyen TT, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of Alastipes timanensis sp. nov. Stand Genomic Sci 2012;6:315–24.
[11] Mishra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of Costidiurn sene- galense sp. nov. Stand Genomic Sci 2012;6:386–95.
[12] Mishra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of Peptoniphilus timonensis sp. nov. Stand Genomic Sci 2012;7:1–11.
[13] Mishra AK, Lagier JC, Rivet R, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of Pectinibacillus senegalensis sp. nov. Stand Genomic Sci 2012;7:80–81.
[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;7:200–9.
[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;7:221–32.
[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;7:233–45.
[17] Ramasamy D, Kokcha S, Lagier JC, Nguyen TT, Raoult D, Fournier PE. Genome sequence and description of Aeromicrobium massiliense sp. nov. Stand Genomic Sci 2012;7:246–57.
[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;7:258–70.
[19] Lagier JC, El Karkouri K, Rivet R, Couderc C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of Cellulomonas massiliensis sp. nov. Stand Genomic Sci 2013;7:343–56.
[20] Lagier JC, El Karkouri K, Mishra AK, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of Enterobacteriaceae sp. nov. Stand Genomic Sci 2013;7:399–412.
[21] Prauser H. Nocardioceutica, a new genus of the order Actinomycetales. Int J Syst Bacteriol 1976;26:58–65.
[22] Prauser H. Nocardioceuticella luteus spec. nov. Zeitschr Allgem Mikrobiol 1984;24:647–8.
[23] Alas-Villagas C, Jurado V, Laiz L, Miller AZ, Saiz-Jimenez C. Nocardioceuticella albatrossiae sp. nov., isolated from Roman catacombs. Int J Syst Bacteriol 2013;63:1280–4.
[24] Cho Y, Jang GI, Hwang CY, Kim EH, Cho BC. Nocardioceuticella sabulosatricola sp. nov., isolated from biofilm formed in coastal seawater. Int J Syst Bacteriol 2013;63:3800–6.
[25] Cho Y, Jang GI, Cho BC. Nocardioceuticella maringulina sp. nov., isolated from coastal seawater. Int J Syst Bacteriol 2013;63:2594–9.
[26] Du HJ, Wei YZ, Su J, Liu HY, Ma BP, Guo BLY. Nocardioceuticella peritale sp. nov., isolated from surface-sterilized roots of Perilla frutescens. Int J Syst Bacteriol 2013;63:1068–72.
[27] Han JH, Kim TS, Joungh Y, Kim MN, Shin KS, Bae T, et al. Nocardioceuticella endophytica sp. nov. and Nocardioceuticella conyzicola sp. nov., isolated from herbaceous plant roots. Int J Syst Bacteriol 2013;63:6730–9.
[28] Kim JK, Liu QM, Park HY, Kang MS, Kim SC, Im WT, et al. Nocardioceuticella panacotellate sp. nov., isolated from soil of a ginseng field, with ginsenoside converting activity. Antonie van Leeuwenhoek 2013;103:1385–93.
[29] Liu Q, Xin YH, Liu HC, Zhou YG, Wen Y. Nocardioceuticella szechwanensis sp. nov. and Nocardioceuticella psychrotolerans sp. nov., isolated from a glacier. Int J Syst Bacteriol 2013;63:129–33.
[30] Cui Y, Woo SG, Lee J, Sinha S, Kang MS, Jin L, et al. Nocardioceuticella daeogensis sp. nov., a nitrate-reducing bacterium isolated from acti- vated sludge of an industrial wastewater treatment plant. Int J Syst Bacteriol 2013;63:3727–32.
[31] Collins MD, Dorsch M, Stackebrandt E. Transfer of Pimelobacter to Terrabacter gen. nov. as Terrabacter terrarum comb. nov. and of Pimelobacter jensenii to Nocardioceuticella as Nocardioceuticella jenseni comb. nov. Int J Syst Bacteriol 1989;39:1–6.
[32] Fan X, Qiao Y, Gao X, Zhang XH. Nocardioceuticella pacificus sp. nov., isolated from deep sub-seafloor sediment of South Pacific Gyre. Int J Syst Bacteriol 2014;64:2217–22.
[33] Glaeser SP, McInroy JA, Busse HJ, Kämpfer P. Nocardioceuticella zeae sp. nov., isolated from the stem of Zea mays. Int J Syst Bacteriol 2014;64:2491–6.
[34] Lee SD, Seong CN. Nocardioceuticella opuntiae sp. nov., isolated from a soil of cactus. Int J Syst Bacteriol 2014(4):2094–9.
[35] Zhang DF, Zhong JM, Zhang XM, Jiang Z, Zhou EM, Tian XP. Nocardioides nanhaiensis sp. nov., an actinobacterium isolated from a marine sediment sample. Int J Syst Bacteriol 2014;64(8):2718–22.
Lee SD, Lee DW. Nocardioides rubroscoriae sp. nov., isolated from volcanic ash. Antonie van Leeuwenhoek 2014;105:1017–23.

Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 1990;87:4576–9.

Garrity GM, Holt JG. The road map to the manual. In: Garrity GM, Boone DR, Castenholz RW, editors. Bergey’s manual of systematic bacteriology. 2nd ed. vol. I. New York: Springer; 2001. p. 119–69.

Stackebrandt E, Ebers J, Rainey P, Ward-Rainey N. Proposal for a new hierarchic classification system, Actinobacteria classis nov. Int J Syst Bacteriol 1997;47:479–91.

Skerman V, McGowan V, Sneath P. Approved lists of bacterial names. Int J Syst Bacteriol 1980;30:225–420.

Buchanan RE. Studies in the nomenclature and classification of bacteria. II. The primary subdivisions of the Schizomycetes. J Bacteriol 1917;2:155–64.

Zhi XY, Li WJ, Stackebrandt E. An update of the structure and 16S rRNA gene sequence–based definition of higher ranks of the class Actinobacteria, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. Int J Syst Evol Microbiol 2009;59:589–608.

Nesterenko OA, Kvasnikov EI, Nogina TM. Nocardioidaceae fam. nov., a new family of the order Actinomycetales Buchanan 1917. Mikrobiol Zh 1985;47:3–12.

Validation list no. 34. Int J Syst Bacteriol 1990;40:320–1.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–9.

Seng P, Drancourt M, Gouriet F, La Scala B, Fourrier 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.

Field D, Garrity G, Gray T, Morrison N, Selingut J, Sterk P, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol 2008;26:541–7.

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.

Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2012;40:D48–53.

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.

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

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides. SignalP 3.0. J Mol Biol 2004;340:783–95.

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.

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.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. Bioinformatics 2000;16:944–5.

Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 2009;25:119–20.

Darling AC, Mau B, Blatov FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004;14:1394–403.