NEW SPECIES

Genome sequence and description of *Traorella massiliensis* gen. nov., sp. nov., a new bacterial genus isolated from human left colon

M. Bonnet¹, M. Mailhe¹, D. Ricaboni¹,², N. Labas¹, M. Richez¹, V. Vitton³, A. Benezech³, P.-E. Fournier¹,²,³,⁴, M. Million¹, D. Raoult¹,² and S. Khelai²

¹ Aix-Marseille Université, IRD, APHM, MEPHI, ² Institut Hospitalo-Universitaire Méditerranée Infection, 3 Service de Gastroentérologie, Hôpital Nord, Assistance Publique-Hôpitaux de Marseille, Marseille, France and 4 Département des sciences cliniques et biomédicales, Luigi Sacco, Division des Maladies Infectieuses III, Université de Milan, Milan, Italy

Abstract

A strictly anaerobic, motile, non–spore-forming, Gram-negative, rod-shaped bacterium designated Marseille-P3110<sup>T</sup> was isolated from the left colon cleansing of a 76-year-old Frenchwoman. Its 16S ribosomal RNA (rRNA) gene showed a 93.2% similarity level with the 16S rRNA of *Dielma fastidiosa* strain JC13, the closest species with a validly published name. The genome of Marseille-P3110<sup>T</sup> is 2 607 061 bp long with 35.99% G+C content. Of the 2642 predicted genes, 2582 were protein-coding genes and 60 were RNAs, including five 16S rRNA genes.

© 2019 The Authors. Published by Elsevier Ltd.

Keywords: Culturomics, genome, gut microbiota, taxonogenomics, *Traorella massiliensis*

Original Submission: 17 January 2019; Revised Submission: 7 February 2019; Accepted: 13 February 2019

Article published online: 22 February 2019

Corresponding author: S. Khelai², Institut Hospitalo, Universitaire Méditerranée Infection, 19-21 Bd Jean Moulin, 13385 Marseille cedex 5, France.

E-mail: khelai²_saber@yahoo.fr

The first two authors contributed equally to this article, and both should be considered first author.

Introduction

Even though metagenomics has made it possible to explore the human microbiota, many bacteria remain to be cultivated. Culturomics is a culture-based approach that uses multiple culture conditions together with MALDI-TOF MS and 16S ribosomal RNA (rRNA) gene sequencing [1] for bacterial culture and identification. This approach was developed to improve the identification of uncultured bacteria by matching metagenomic unassigned sequences with identified and cultivated bacteria [1].

The project that allows the isolation of this new strain consists of the study the microbiota of different portions of the digestive tract. Its composition varies according to different factors such as pH, percentage of oxygen or health status [2].

During this project, we isolated a new bacterial genus, a member of the family *Erysipelotrichaceae*. This family was created in 2004 and actually regroups 11 genera [3]. Bacteria belonging to this family are all strictly anaerobic, except the genus *Erysipelothrix*, which contains facultative anaerobic or microaerophilic species. Most of these bacteria are Gram positive, non–spore forming and rod shaped. The bacteria belonging to this family seem to be highly immunogenic. The proportion of *Erysipelotrichaceae* is higher in patients with inflammatory bowel disease than in healthy individuals [4].

The new genus was described using the taxonogenomics approach. This approach combines next-generation sequencing, and phylogenetic and phenotypic techniques [5]. The MALDI-TOF MS protein profile plays also a role in the description of new bacterial species.

We describe *Traorella massiliensis* strain Marseille-P3110<sup>T</sup> (= CSUR P3110 = DSM 103514), a new genus isolated from the left colon cleansing of a 76-year-old Frenchwoman.
Materials and methods

Ethics and sample information
A sample from the left colon cleansing of a 76-year-old Frenchwoman who underwent a colonoscopy to check for colon polyps was collected at Hôpital Nord (Marseille, France) in May 2016. The study was authorized by the local ethics committee of the Institut Hospitalo-Universitaire Méditerranée-infection (Marseille, France) under agreement 2016-010, and the patient provided written informed consent. At the moment of the sample collection, she was treated with a proton pump inhibitor. The sample was transported in an antioxidant transport medium.

Strain identification by MALDI-TOF MS and 16S rRNA gene sequencing
The sample was seeded directly on Columbia medium supplemented with 5% sheep’s blood (COS) (bioMérieux, Marcy l’Etoile, France), and incubated under anaerobic conditions at...

FIG. 1. Phylogenetic tree highlighting position of Traorella massiliensis strain Marseille-P3110T relative to other close strains. GenBank accession numbers of each 16S rRNA are noted in parentheses. Sequences were aligned using Muscle version 3.8.31 with default parameters, and phylogenetic inferences were obtained using neighbour-joining method with 500 bootstrap replicates within MEGA6 software. Only bootstraps >95% are shown. Scale bar represents 0.02% nucleotide sequence divergence.

© 2019 The Authors. Published by Elsevier Ltd. NMNI, 29, 100520
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
37°C during 3 days. Colonies were purified through subculture and identified by MALDI-TOF MS using a Microflex spectrometer and a MTP 96 MALDI-TOF target plate (Bruker Daltonics, Leipzig, Germany), as described previously [6,7]. The spectra obtained for each colony were matched against the MALDI Biotyper software version 3.0 (Bruker) and Unité des Maladies Infectieuses et Tropicales Emergentes (URMITE) databases using standard pattern matching (with default parameter settings). The identification scores used were as follows: a score over 1.9 allowed identification at the species level, while a score under 1.7 did not allow any identification. In the latter case, the colony was identified by sequencing its 16S rRNA gene as previously described [8]. According to Stackebrandt and Ebers [9], a similarity threshold of 98.65% was used to define a new species, whereas a threshold of 95% was used to define a new genus without performing DNA-DNA hybridization. Upon identification, a reference spectrum for strain Marseille-P3110T was incremented in the URMITE database.

16S rRNA phylogenetic tree
A custom Python script was used to automatically retrieve all species from the same order as the new genus and download 16S sequences from National Center for Biotechnology Information (NCBI) by parsing the NCBI eUtils results and NCBI taxonomy page. It only keeps sequences from type strains. In case of multiple sequences for one type strain, it selects the sequence obtaining the best identity rate from the BLASTn alignment with our sequence. The script then separates 16S sequences in two groups: one containing the sequences of strains from the same family (group a) and one containing the others (group b). It finally only keeps the 15 strains closest to the group and the closest to group b. If it is impossible to get 15 sequences from group a, the script selects more sequences from group b to get at least nine strains from both groups.

Growth conditions
Ideal growth conditions for strain Marseille-P3110T were determined by testing five growth temperatures (25, 28, 37, 45 and 56°C) in an aerobic atmosphere with or without 5% CO2, and under anaerobic and microaerophilic conditions using the GENbag anaer and GENbag microaer systems, respectively (bioMérieux). Different pH values (5, 5.5, 6.5, 7, 7.5 and 8) and NaCl concentrations (10, 15 and 20%) were also tested.

Morphologic, biochemical and antibiotic susceptibility tests
Phenotypic characteristics such as Gram staining, motility, sporulation, and catalase and oxidase activities were tested as previously described [3,10]. Biochemical analysis of strain Marseille-P3110T was carried out using API 50 CH, API 20A and API ZYM strips (bioMérieux) in an anaerobic atmosphere. Antibiotic susceptibility was tested using the disc diffusion method (i2a, Montpellier, France) [11] and according to European Committee on Antimicrobial Susceptibility Testing (EUCAST) 2015 recommendations.

In order to observe their morphology, the cells were fixed with 2.5% glutaraldehyde in 0.1 M cacodylate buffer for at least 1 hour at 4°C. A drop of cell suspension was deposited for approximately 5 minutes on glow-discharged formvar carbon film on 400 mesh nickel grids (FCF400-Ni, Electron Microscopy Sciences (EMS), Hatfield, PA, USA). The grids were dried on blotting paper, and cells were negatively stained for 10 seconds with 1% ammonium molybdate solution in filtered water at room temperature. Electron micrographs were acquired with a Morgagni 268D (Philips, Amsterdam, The Netherlands) transmission electron microscope operated at 80 keV.

Fatty acid methyl ester analysis by gas chromatography/mass spectrometry
Two samples were prepared with approximately 30 mg of bacterial biomass per tube collected from several culture plates. Fatty acid methyl esters (FAME) were prepared as described by Sasser [12]. Gas chromatography/mass spectrometry analyses were carried out as described previously [13]. Briefly, FAME were separated using an Elite 5-MS column and monitored by mass spectrometry (Clarus 500–SQ 8 S; PerkinElmer, Courtaboeuf, France). Spectral database search was performed using MS Search 2.0 operated with the Standard Reference Database IA (National Institute of Standards and Technology, Gaithersburg, MD, USA) and the FAMEs mass spectral database (Wiley, Chichester, UK).

Genomic DNA extraction and genome sequencing
After a lysozyme pretreatment and incubation at 37°C for 2 hours, DNA was extracted on the EZ1 biorobot (Qiagen, Waltham, MA, USA) and purified using the QIAquick PCR purification kit (Qiagen). A genome library was prepared using the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA). Libraries were sequenced using Illumina MiSeq (Illumina) in the Centre for Comparative Genomics and Evolution at the University of Victoria, British Columbia, Canada.

**TABLE 1. Classification and general features of *Traorella massiliensis* strain Marseille-P3110T**

| Property                  | Term                  |
|---------------------------|-----------------------|
| **Current classification**| **Domain: Bacteria**  |
|                           | **Phylum: Firmicutes**|
|                           | **Class: Erysipelotrichia** |
|                           | **Order: Erysipelotrichales** |
|                           | **Family: Erysipelotrichaceae** |
|                           | **Genus: Traorella**   |
| Species: *Traorella massiliensis* | **Type strain: Marseille-P3110T** |
| **Gram stain**            | Negative              |
| **Cell shape**            | Rod                   |
| **Motility**              | Motile                |
| **Sporulation**           | Nonsporulating        |
| **Temperature range**     | Mesophilic            |
| **Optimum temperature**   | 37°C                  |
The mate-pair library was prepared with 1.5 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 pattern of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) with a DNA 7500 labchip. The DNA fragments ranged in size from 1.5 to 11 kb, with an optimal size at 5.176 kb. No size selection was performed, and 378.6 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal at 1054 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 31.57 nmol/L.

The libraries were normalized at 2 nM and pooled. After 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 sequencing run were performed in a single 39-hour run at a 2 × 251 bp read length.

Total information of 5.1 Gb was obtained from a 544K/mm² cluster density with a cluster passing quality control filters of 96.8% (10 139 000 passing filter paired reads). Within this run, the index representation for strain Marseille-P3110T was determined to be of 8.93%. The 905 502 paired reads were trimmed then assembled in nine scaffolds.

Genome assembly
The genome assembly was performed with a pipeline that enabled creation of an assembly with different software (Velvet [14], Spades [15] and Soap Denovo [16]) on trimmed (MiSeq and Trimmomatic [17] software) or untrimmed data (only MiSeq software). For each of the six assemblies performed, GapCloser [16] was used to reduce gaps. Then contamination with Phage Phix was identified (BLASTn against Phage Phix174 DNA sequence) and eliminated. Finally, scaffolds under 800 bp in size were removed, and scaffolds with a depth value lower than 25% of the mean depth were removed (identified as possible contaminants). The best assembly was selected by using different criteria (number of scaffolds, N50, number of N). For the studied strain, Spades gave the best assembly, with a depth coverage of 128×.
Genome annotation and comparison

Open reading frames (ORFs) were predicted using Prodigal [18] with default parameters, but the predicted ORFs were excluded if they spanned a sequencing gap region (containing N). The predicted bacterial protein sequences were searched against the Clusters of Orthologous Groups (COGs) database using BLASTP (E value 1e-03, coverage 0.7 and identity percentage 30%). If no hit was found, then the predicted bacterial protein sequences were searched against the NR database using BLASTP with an E value of 1e-03, coverage of 0.7 and identity percentage of 30%, and if the sequence length was smaller than 80 aa, we used an E value of 1e-05. The tRNAscanSE tool [19] was used to find transfer RNA genes, while ribosomal RNAs were found using RNAmer [20]. Lipoprotein signal peptides and the number of transmembrane helices were predicted using Phobius [21]. ORFans were identified if all the BLASTP performed did not give positive results (E value smaller than 1e-03 for ORFs with sequence size larger than 80 aa or E value smaller than 1e-05 for ORFs with sequence length smaller 80 aa). Such parameter thresholds have already been used in previous studies to define ORFans. Pfam-conserved domains (Pfam-A and Pfam-B domains) were searched on each protein with the HMMscan of the HMMER3 suite [6]. PKS and NRPS were searched against the ClusterMine360 [17] database. Resistome was analysed by using the ARG-ANNOT database [22].

Species used for genomic comparison were identified in the 16S rRNA tree using PhyloPattern software [23]. For each selected species, the complete genome sequence, proteome sequence and ORFeome sequence were retrieved from the NCBI. If one specific strain did not have a complete and available genome, a complete genome of the same species used was used. If ORFeome and proteome were not predicted, Prodigal was used with default parameters to predict them. All proteome were analysed with proteinOrtho [24]. Then, for each couple of genomes, a similarity score was computed. This score is the mean value of nucleotide similarity between all couples of orthologous genes between the two genomes studied (average genomic identity of orthologous gene sequences (AGIOS) tool). An annotation of the entire proteome was performed in the multiagent software system DAGOBAH [25], which includes Figenix [26] libraries that provide pipeline analysis, and by using PhyloPattern [23] for tree manipulation. To evaluate the genomic similarity between studied genomes,
we determined two parameters, digital DNA-DNA hybridization (dDDH), which exhibits a high correlation with DNA-DNA hybridization (DDH) [27,28], and AGIOS [29], which was designed to be independent from DDH [31]. Genome-to-Genome Distance Calculator (GGDC) analysis was performed using the GGDC web server as previously reported [28].

**TABLE 4. Nucleotide content and gene count levels of genome**

| Attribute                        | Genome (total) | % of total |
|----------------------------------|----------------|------------|
| Size (bp)                        | 2 607 061      | 100        |
| G+C content (%)                  | 937 722        | 35.99      |
| Coding region (bp)               | 2 334 961      | 90.02      |
| Total genes                      | 2642           | 100        |
| RNA genes                        | 60             | 2.27       |
| Protein-coding genes             | 2582           | 100        |
| Genes with function prediction   | 1757           | 68.05      |
| Genes assigned to COGs           | 1468           | 56.86      |
| Genes with peptide signals       | 296            | 11.46      |
| ORFans genes                     | 279            | 10.81      |
| Genes associated with PKS or NRPS| 1              | 0.039      |
| No. of antibiotic resistance genes| 11            | 0.039      |
| No. of genes associated with Pfam-A domains | 2308 | 87 |

COGs, Clusters of Orthologous Groups database; PKS, polyketide synthase; NRPS, nonribosomal peptide synthase. *Total is based on either size of genome in base pairs or total number of protein-coding genes in annotated genome.

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

| Code | Value | % of total |
|------|-------|------------|
| [J]  | 191   | 7.40       |
| [A]  | 0     | 0          |
| [K]  | 125   | 4.84       |
| [L]  | 88    | 3.41       |
| [B]  | 0     | 0          |
| [D]  | 32    | 1.24       |
| [T]  | 0     | 0          |
| [V]  | 73    | 2.83       |
| [Y]  | 0     | 0          |
| [V]  | 73    | 2.83       |
| [T]  | 64    | 2.48       |
| [M]  | 96    | 3.72       |
| [N]  | 7     | 0.27       |
| [Z]  | 0     | 0          |
| [W]  | 1     | 0.04       |
| [U]  | 18    | 0.70       |
| [O]  | 53    | 2.05       |
| [X]  | 58    | 2.25       |
| [C]  | 73    | 2.83       |
| [G]  | 155   | 6.00       |
| [E]  | 116   | 4.49       |
| [F]  | 69    | 2.67       |
| [H]  | 47    | 1.82       |
| [I]  | 42    | 1.63       |
| [P]  | 68    | 2.63       |
| [Q]  | 21    | 0.81       |
| [R]  | 152   | 5.89       |
| [S]  | 74    | 2.87       |
| [S]  | 1114  | 43.14      |

COGs, Clusters of Orthologous Groups database. *Total is based on total number of protein-coding genes in annotated genome.

FIG. 5. Graphical circular map of chromosome. From outside to centre, genes on forward strain coloured by COGs categories (only gene assigned to COGs), RNA genes (tRNAs green, rRNAs red), G+C content and G+C skew. COGs, Clusters of Orthologous Groups database; rRNA, ribosomal RNA; tRNA, transfer RNA.
Results

Strain identification and phylogenetic analyses
Strain Marseille-P3110T was first cultivated on COS under anaerobic atmosphere at 37°C. The sample containing this bacterium was transported in an antioxidant transport medium and seeded directly onto COS. Strain Marseille-P3110T could not be identified using MALDI-TOF MS, and therefore the 16S rRNA was sequenced. The resulting sequence (accession no. LT615365) showed a 93.2% similarity level with the 16S rRNA gene of Dielma fastidiosa strain JC13, the closest species with a validly published name [30] (Fig. 1). Because this 16S rRNA nucleotide sequence similarity was lower than 95%, strain Marseille-P3110T is considered to be a new genus according to the threshold described by Stackebrandt and Ebers [9]. This new genus belongs to the family Erysipelotrichaceae for which we suggest the name Traorella, with Traorella massiliensis as type species and Marseille-P3110T (= CSUR P3110 = DSM 103514) as type strain (Table 1). Consequently, the reference protein spectra for Traorella massiliensis (Fig. 2(A)) were incremented in the URMITE database (http://www.mediterranee-infection.com/article.php?laref=256&tite=urms-database). A gel view was also performed to observe the spectra’s comparison with the closest bacteria (Fig. 2(B)).

Phenotypic description
The bacterium was cultivated at 37°C in anaerobic conditions. No growth was noted in aerobic and microaerophilic conditions. The optimal growth condition was observed after 72 hours in anaerobic conditions. Growth was observed at pH 7, 7.5 and 8 but not at pH 5, 5.5 and 6.5. No growth was observed for the salinities tested (10, 15 and 20%). Colonies of the strain Marseille-P3110T were motile, non-spore forming, Gram negative and rod shaped (Fig. 3). The strain Marseille-P3110T forms regular white colonies with a mean diameter of 1 mm. Individual cells have a width ranging from 0.28 to 0.37 μm and a length ranging from 2.5 to 3.1 μm (Fig. 4). No catalase and oxidase activities were observed. Using an API ZYM strip, an API 20A

| Organism                     | INSDC          | Size (Mb) | G+C (%) | Protein-coding genes | Total genes |
|------------------------------|----------------|-----------|---------|----------------------|-------------|
| Traorella massiliensis strain Marseille-P3110T | FNLJ00000000 | 2.61      | 35.99   | 2582                 | 2582        |
| Faecalitalea cylindroides strain ATCC 27803 | AMV100000000 | 1.95      | 34.68   | 1841                 | 2057        |
| Erysipelothrix rhusiopathiae strain ATCC 19414 | ACLK00000000 | 1.75      | 36.47   | 1613                 | 1645        |
| Solibacterium moorei strain F0204 | AUKY00000000 | 2.01      | 36.78   | 2181                 | 2035        |
| Eubacterium dolchum strain CAG373 | ABAV00000000 | 2.03      | 37.40   | 2076                 | 1927        |
| Erysipelothrix tantilla strain DSM 14972 | AREO00000000 | 1.93      | 36.83   | 1750                 | 1792        |
| Erysipelotrichus ramosum strain DSM 1402 | ABFX00000000 | 3.24      | 31.39   | 2941                 | 3169        |
| Holdemanella bifurca strain DSM 3989 | ABY700000000 | 2.52      | 33.79   | 2248                 | 2529        |
| Dielma fastidiosa strain DSM 26099 | CAEN00000000 | 3.62      | 39.97   | 3321                 | 3496        |

INSDC, International Nucleotide Sequence Database Collaboration.

FIG. 6. Distribution of functional classes of predicted genes according to clusters of orthologous groups of proteins.
strip and an API 50 CH strip, positive enzymatic activities included esterase C4, leucine arylamidase, N-acetyl-β-glucosaminidase, α-fucosidase, α-glucosidase, acid phosphatase and naphthol-AS-Bi-phosphohydrolase. No activity was found for the following enzymes: valine arylamidase, β-galactosidase, esterase lipase C8, protease, urease, alkaline phosphatase, lipase C14, cystine arylamidase, trypsin, β-glucuronidase, α-chymotrypsin, α-galactosidase, β-glucosidase and α-mannosidase. No acid production was observed from D-glucose, D-lactose, D-sucrose, D-maltose, salicin, D-cellobiose, D-mannose, D-raf (D-melibiose, glycerol, glycogen), D-arabinose, L-arabinose, erythromycin, esculin ferric citrate, amygdalin, D-cellobiose, arbutin, D-ribose, D-tagatose and potassium 5-ketogluconate as presented in Table 2. Strain Marseille-P3110T differed from those of closely related species with an available genome: Dielma fastidiosa (DSM 26099; CAEN000000000), Erysipelatoclostridium ramosum (DSM 1402; ABBFX000000000), Faecalitalea cylindroides (ATCC 27803; AWI000000000), Erysipelothrix rhusiopathiae (ATCC 19414; ACLK000000000), Solobacterium moorei (FOU004; AUKY000000000), Eubacterium dolichum (CAG:375; ABAAW000000000), Erysipelothrix tonsillarum (DSM 14972; AREOO000000000) and Holdemanae bifors (DSM 3989; ABYT000000000) (Table 6). The draft genome sequence of strain Marseille-P3110T is smaller than that of Dielma fastidiosa and E. ramosum (3615, 2607 and 3235 MB respectively), but larger than that of F. cylindroides, E. rhusiopathiae, S. moorei, E. dolichum, E. tonsillarum and H. bifors (1946, 1746, 1927, 2035, 1931 and 2518 MB respectively). The G+C content of strain Marseille-P3110T is 33.79% respectively) (Table 6). The genome of Marseille-P3110T is 2 607 061 bp long with 35.99% G+C content (Fig. 5 and Table 4). It is composed of nine scaffolds (composed of 14 contigs). Of the 2642 predicted genes, 2582 were protein-coding genes and 60 were RNAs (five genes are 5S rRNA, five genes are 16S rRNA, five genes are 23S rRNA, 45 genes are transfer RNA genes). A total of 1757 genes (68.05%) were assigned as putative function (by COGs or by NR BLAST). A total of 279 genes were identified as ORFans (10.81%). The remaining genes were annotated as hypothetical proteins (472 genes, 18.28%). The distribution of genes into COGs functional categories is presented in Table 5.

**Genome properties**

The genome of Marseille-P3110T is 2 607 061 bp long with 35.99% G+C content (Fig. 5 and Table 4). It is composed of nine scaffolds (composed of 14 contigs). Of the 2642 predicted genes, 2582 were protein-coding genes and 60 were RNAs (five genes are 5S rRNA, five genes are 16S rRNA, five genes are 23S rRNA, 45 genes are transfer RNA genes). A total of 1757 genes (68.05%) were assigned as putative function (by COGs or by NR BLAST). A total of 279 genes were identified as ORFans (10.81%). The remaining genes were annotated as hypothetical proteins (472 genes, 18.28%). The distribution of genes into COGs functional categories is presented in Table 5.

**Genome comparison**

Genomic characteristics of strain Marseille-P3110T were compared to those of closely related species with an available genome: Dielma fastidiosa (DSM 26099; CAEN000000000), Erysipelatoclostridium ramosum (DSM 1402; ABBFX000000000), Faecalitalea cylindroides (ATCC 27803; AWI000000000), Erysipelothrix rhusiopathiae (ATCC 19414; ACLK000000000), Solobacterium moorei (FOU004; AUKY000000000), Eubacterium dolichum (CAG:375; ABAAW000000000), Erysipelothrix tonsillarum (DSM 14972; AREOO000000000) and Holdemanae bifors (DSM 3989; ABYT000000000) (Table 6). The draft genome sequence of strain Marseille-P3110T is smaller than that of Dielma fastidiosa and E. ramosum (3615, 2607 and 3235 MB respectively), but larger than that of F. cylindroides, E. rhusiopathiae, S. moorei, E. dolichum, E. tonsillarum and H. bifors (1946, 1746, 1927, 2035, 1931 and 2518 MB respectively). The G+C content of strain Marseille-P3110T is smaller than that of Solobacterium moorei, Eubacterium dolichum, Dielma fastidiosa, Erysipelothrix rhusiopathiae and Erysipelothrix tonsillarum (35.99, 36.78, 37.4, 39.97, 36.47 and 36.83% respectively), but larger than that of F. cylindroides, E. ramosum and H. bifors (34.68, 31.39 and 33.79% respectively) (Table 6).

The gene content of strain Marseille-P3110T is smaller than that of Dielma fastidiosa and E. ramosum (3319, 2582 and

---

**TABLE 7. Numbers of orthologous proteins shared between genomes (upper right) and AGIOS values obtained (lower left)**

|       | Tm | Fc | Ed | Er | Et | Hb | Sm | Era | Df |
|-------|----|----|----|----|----|----|----|-----|----|
| Tm    | 2582 | 769 | 696 | 605 | 614 | 757 | 663 | 768 | 940 |
| Fc    | 57.72 | 2057 | 696 | 534 | 541 | 851 | 606 | 723 | 797 |
| Ed    | 59.24 | 59.86 | 1927 | 484 | 501 | 691 | 529 | 632 | 785 |
| Er    | 58.60 | 57.06 | 57.12 | 1645 | 769 | 527 | 567 | 579 | 647 |
| Et    | 62.91 | 56.24 | 56.25 | 70.60 | 1792 | 544 | 580 | 590 | 654 |
| Hb    | 55.70 | 60.78 | 57.01 | 56.11 | 54.85 | 2529 | 614 | 722 | 800 |
| Sm    | 58.87 | 57.70 | 58.21 | 57.81 | 58.48 | 56.99 | 2035 | 595 | 708 |
| Era   | 59.80 | 56.40 | 56.80 | 57.56 | 58.29 | 57.36 | 56.95 | 3169 | 836 |
| Df    | 59.29 | 56.88 | 58.13 | 56.47 | 56.39 | 55.88 | 56.07 | 56.88 | 3319 |

Number of proteins per genome is indicated in bold.

Df, Dielma fastidiosa DSM 26099; Ed, Eubacterium dolichum CAG, 375; Er, Erysipelothrix rhusiopathiae ATCC 19414; Era, Erysipelatoclostridium ramosum DSM 1402; Et, Erysipelothrix tonsillarum DSM 14972; Fc, Faecalitalea cylindroides ATCC 27803; Hb, Holdemanae bifors DSM 3989; Sm, Solobacterium moorei FOU004; Tm, Traorella massiliensis Marseille-P3110T.

© 2019 The Authors. Published by Elsevier Ltd.
but larger than that of *F. cylindroides*, *E. rhusiopathiae*, *S. moorei*, *E. dolichum*, *E. tonsillarum* and *H. biformis* (2057, 1645, 2035, 1927, 1792 and 2529 respectively). Distribution of genes into COGs categories was similar among all compared genomes (Fig. 6).

Among species with standing in nomenclature, AGIOS values ranged from 54.85 between *H. biformis* and *E. tonsillarum* to 70.60 between *E. tonsillarum* and *E. rhusiopathiae*. When compared to strain Marseille-P3110T, AGIOS values ranged from 55.70 with *H. biformis* to 62.91 with *E. tonsillarum* (Table 7). Among species with standing in nomenclature, dDDH values ranged from 15.10% between *F. cylindroides* and *S. moorei* to 44.10% between *F. cylindroides* and *E. tonsillarum*. dDDH values between strain Marseille-P3110T and compared species ranged from 20.20% with *E. tonsillarum* to 27.80% with *E. rhusiopathiae* (Table 8).

**Conclusion**

Considering the specific phenotypic properties of strain Marseille-P3110T, including its low matching MALDI-TOF MS score, the 93.2% 16S rRNA similarity level with *Dielma fastidiosa*, and its genomic analysis, we propose the creation of a new genus within the family *Erysipelotrichaceae*, named *Traorella*, with *Traorella massiliensis* as type species and strain Marseille-P3110T as type strain.

**Description of *Traorella* gen. nov.**

*Traorella* (tra.o.rel ‘la, N.L. fem. gen. n., *Traore*, ‘of Traore,’ the family name of Sory Ibrahima Traore, a Malian microbiologist, for his contribution to the description of the human gut microbiota). Bacteria belonging to this genus are strictly anaerobic, non–spore forming, motile, Gram negative and rod shaped. The type species is *Traorella massiliensis*.

**Description of *Traorella massiliensis* sp. nov.**

*Traorella massiliensis* (mas.si.li.en ‘sis, L. masc. adj., *massiliensis*, ‘of Massilia,’ the Latin name of Marseille, where strain Marseille-P3110T was isolated) is an anaerobic Gram-negative and motile bacilli with a mean length of 2.8 μm and a mean diameter of 0.33 μm. Strain Marseille-P3110T forms regular white colonies with a mean diameter of 1 mm. This bacterium is catalase and oxidase negative, and non–spore forming. Optimal growth was observed after 72 hours of incubation in anaerobic conditions at 37°C. Strain Marseille-P3110T was susceptible to amoxicillin/
clavulanic acid, clindamycin, metronidazole, imipenem, tobra-
mycin, fosfomycin, erythromycin and amoxicillin but was
resistant to rifampicin, gentamicin, vancomycin, colistin,
trimethoprim/sulfamethoxazole, oxacillin and doxycycline.

The major cellular fatty acid was 9-octadecenoic acid. The
genome of strain Marseille-P3110 T (= CSUR P3110 = DSM
accession numbers LT615365 and FNLJ00000000, respectively.
are available in the European Bioinformatics Institute/European
35.99% G+C content. The 16S rRNA and genome sequences
old Frenchwoman.

References

None declared.

Acknowledgements

The authors thank the Xegen Company (www.xegen.fr) for
automating the genomic annotation process. This study was
funded by the Fondation Méditerranée Infection. We thank M.
Lardière for English-language review.

Conflict of interest

None declared.

References

[1] Lagier JC, Khelafia S, Alou MT, Ndongo S, Dione N, Hugon P, et al.
Culture of previously uncultured members of the gut micro-
biota by culturomics. Nat Microbiol 2016;1(7):16203.
[2] Claesson MJ, Csakus S, O’Sullivan O, Greene-Diniz R, de Weerd H,
Flannery E, et al. Composition, variability, and temporal stability of the
intestinal microbiota of the elderly. Proc Natl Acad Sci U S A
2011;108:10351–4.
[3] Ramasamy D, Lagier JC, Nguyen TT, Raoult D, Fournier PE. Non
contiguous—finished genome sequence and description of Deinobac-
tidiospora gen. nov., sp. nov., a new member of the family Erysipelo-
trichaceae. Stand Genomic Sci 2013;8:336–51.
[4] Kaakoush NO. Insights into the role of Erysipelotrichaceae in the human
host. Front Cell Infect Microbiol 2015;5:84.
[5] Fournier PE, Drancourt M. New microbes in new infections promotes
modern prokaryotic taxonomy: a new section \"Taxogenomics: new
genomes of microorganisms in humans. New Microb New Infect
2015;7:48–9.
[6] Seng P, Abat C, Rolain JM, Colson P, Lagier JC, Gouriet F, et al.
Identification of rare pathogenic bacteria in a clinical microbiology
lab obtained from matrix-assisted laser desorption ionization—time
of-flight mass spectrometry. J Clin Microbiol 2013;51:2182–94.
[7] 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.
[8] Drancourt M, Bollet C, Carlizoz A, Martelini R, Gayral JP, Raoult D. 16S
ribosomal DNA sequence analysis of a large collection of environ-
mental and clinical unidentified bacterial isolates. J Clin Microbiol
2000;38:3623–30.
[9] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished
gold standards. Microbiol Today 2006;33:152–5.
[10] Lagier JC, Hugon P, Khelafia S, Fournier PE, La Scola B, Raoult D. The
rebirth of culture in microbiology through the example of culturomics
to study human gut microbiota. Clin Microbiol Rev 2015;28:237–64.
[11] Matyschek E, Brown DF, Kahlmeter G. Development of the EUCAST
disk diffusion antimicrobial susceptibility testing method and its
implementation in routine microbiology laboratories. Clin Microbiol
Infect 2014;20:235–66.
[12] Sasser M. Bacterial identification by gas chromatographic analysis of
fatty acids methyl esters (GC-FAME). Newark, NY: Microbial ID; 2006.
[13] Dione N, Sankar SA, Lagier JC, Khelafia S, Michele C, Armstrong N,
et al. Genome sequence and description of Anaerosalibacter massiliensis
sp. nov. New Microb New Infect 2016;10:66–76.
[14] Zerbino DR, Birney E. Velvet: algorithms for de novo short read as-
sembly using de Bruijn graphs. Genome Res 2008;18:821–9.
[15] 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 2015;19:455–77.
[16] Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an
empirically improved memory-efficient short-read de novo assembler.
GigaScience 2012;1:18.
[17] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimming for
Illumina sequence data. Bioinformatics 2014;30(11):1414–20.
[18] Huytatt D, Chen GL, LoCastio PF, Land ML, Larimer FW, Hauser LJ.
Prodigal: prokaryotic gene recognition and translation initiation site
detection. BMC Bioinform 2010;11:119.
[19] 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.
[20] 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;35:3080–1.
[21] Kall L, Krogh A, Sonnhammer ELL. A combined transmembrane to-
pology and signal peptide prediction method. J Mol Biol 2004;338:
1027–36.
[22] Darling ACE, Mau B, Blatner FR, Perna NT. Mauve: multiple alignment
of conserved genomic sequence with rearrangements. Genome Res
2004;14:394–403.
[23] Gouret P, Thompson JD, Pontarotti P. PhyloPattern: regular expres-
sions to identify complex patterns in phylogenetic trees. BMC Bio-
inform 2009;10:129.
[24] Lechner M, Findeib S, Steiner L, Marz M, Stadler PF, Prohaska SJ.
Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC
Bioinform 2011;12:124.
[25] Gouret P, Pagnanini J, Dainat J, Loutai D, Barbo 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 DAGOBAH. In: Pontarotti P, editor. Evolutionary
biology: concepts, biodiversity, macroevolution and genome evolution.
Berlin: Springer; 2011. p. 71.
[26] 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 2015;19:455–77.
[27] Auch AF, von Jan M, Klenk HP, Göker M. Digital DNA–DNA hy-
dridization in a clinical microbiology laboratory. J Clin Microbiol
2015;53:1027–36.
[28] Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome
definition by gas chromatographic analysis of fatty acids methyl esters (GC-FAME). Newark, NY: Microbial ID; 2006.
[29] Darling ACE, Mau B, Blatner FR, Perna NT. Mauve: multiple alignment
of conserved genomic sequence with rearrangements. Genome Res
2004;14:394–403.
[30] Gouret P, Thompson JD, Pontarotti P. PhyloPattern: regular expres-
sions to identify complex patterns in phylogenetic trees. BMC Bio-
inform 2009;10:129.
[31] Lechner M, Findeib S, Steiner L, Marz M, Stadler PF, Prohaska SJ.
Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC
Bioinform 2011;12:124.
[32] Gouret P, Pagnanini J, Dainat J, Loutai D, Barbo 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 DAGOBAH. In: Pontarotti P, editor. Evolutionary
biology: concepts, biodiversity, macroevolution and genome evolution.
Berlin: Springer; 2011. p. 71–87.
[33] Gouret P, Pagnanini J, Dainat J, Loutai D, Barbo 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 DAGOBAH. In: Pontarotti P, editor. Evolutionary
biology: concepts, biodiversity, macroevolution and genome evolution.
Berlin: Springer; 2011. p. 71–87.
[34] Gouret P, Pagnanini J, Dainat J, Loutai D, Barbo 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 DAGOBAH. In: Pontarotti P, editor. Evolutionary
biology: concepts, biodiversity, macroevolution and genome evolution.
Berlin: Springer; 2011. p. 71–87.
[35] Gouret P, Pagnanini J, Dainat J, Loutai D, Barbo 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 DAGOBAH. In: Pontarotti P, editor. Evolutionary
biology: concepts, biodiversity, macroevolution and genome evolution.
Berlin: Springer; 2011. p. 71–87.
[29] Ramasamy D, Mishra AK, Lagier JC, Padmanabhan R, Rossi M, Sentauxa E, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol 2014;64:84–91.

[30] Oren A, Garrity GM. List of new names and new combinations previously effectively, but not validly, published. Int J Syst Evol Microbiol 2016;66:4299–305.

[31] Stackebrandt E, Goebel BM. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Evol Microbiol 1994;44:846–9.

[32] Moore WEC, Johnson JL, Holdeman LV. Emendation of Bacteroidaceae and Butyrivibrio and descriptions of Desulfomonas gen. nov. and ten new species in the genera Desulfomonas, Butyrivibrio, Eubacterium, Clostridium, and Ruminococcus. Int J Syst Evol Microbiol 1976;26:238–52.

[33] De Maesschalck C, Van Immerseel F, Eeckhaut V, De Baere S, Cnockaert M, Craubels S, et al. Faecalitalea acidiformans gen. nov., sp. nov., isolated from the chicken caecum, and reclassification of Streptococcus pleomorphus, Eubacterium biforme and Eubacterium cylindroides as Faecalitalea acidiformans comb. nov., Holdemania biformis gen. nov., comb. nov. and Faecalitalea cylindroides gen. nov., comb. nov., respectively, within the family Erysipelotrichaceae. Int J Syst Evol Microbiol 2014;64(Pt 11):3877–84.

[34] Greetham HL, Gibson GR, Giffard C, Hippe H, Merkhoffer B, Steiner U, et al. Allobaculum stercoricanis gen. nov., sp. nov., isolated from canine feces. Anaerobe 2004;10:301–7.

[35] Mishra AK, Lagier JC, Pfleiderer A, Nguyen TT, Caputo A, Raoult D, et al. Non-contiguous finished genome sequence and description of Holdemania massiliensis sp. nov. Stand Genomic Sci 2013;9:395–409.

[36] Kageyama A, Benno Y. Phylogenetic and phenotypic characterization of some Eubacterium-like isolates from human feces: description of Solobacterium moorei gen. nov., sp. nov. Microbiol Immunol 2000;44:223–7.

[37] Downes J, Olsvik B, Hiom SJ, Spratt DA, Cheeseman SL, Olsen I, et al. Bulleidia extructa gen. nov., sp. nov., isolated from the oral cavity. Int J Syst Evol Microbiol 2000;50(Pt 3):979–83.

© 2019 The Authors. Published by Elsevier Ltd. NMNI, 29, 100520
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).