

OPEN

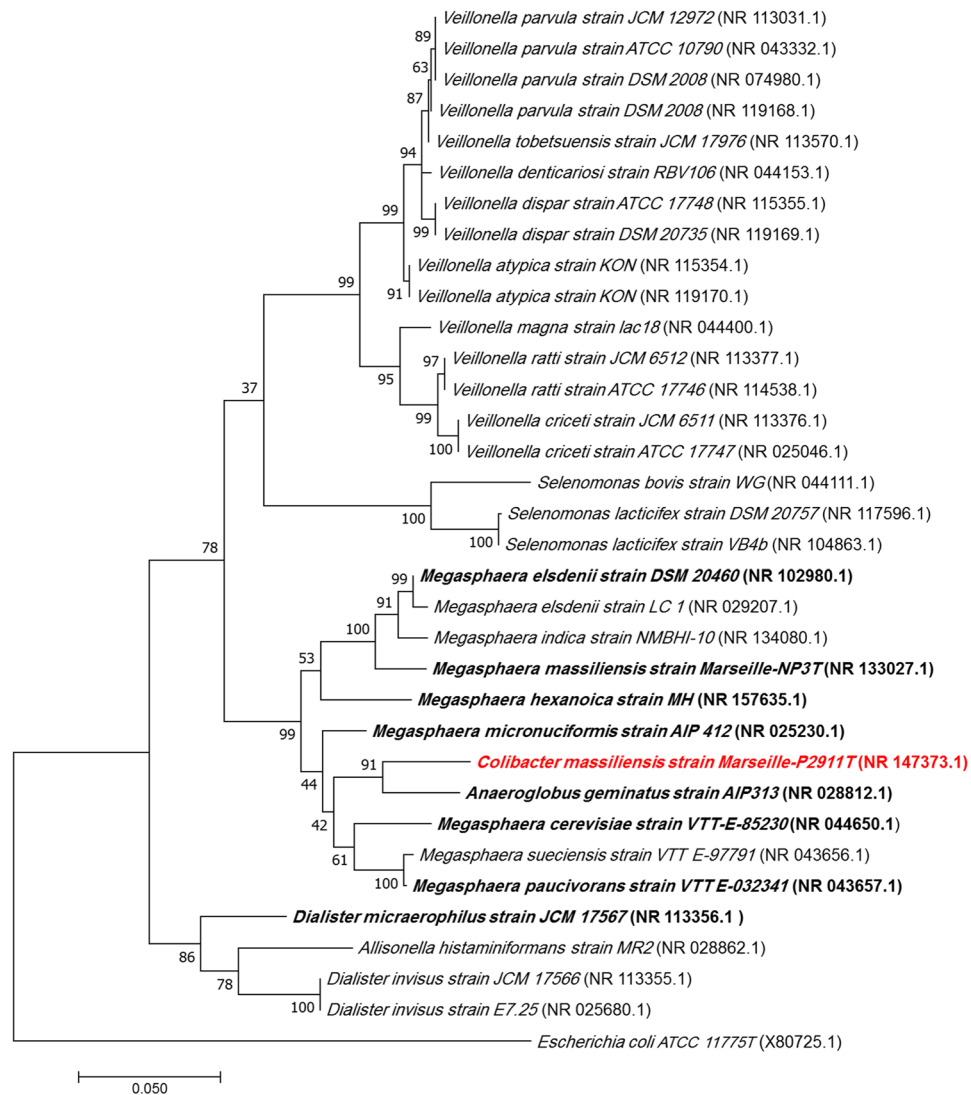
# *Colibacter massiliensis* gen. nov. sp. nov., a novel Gram-stain-positive anaerobic diplococcal bacterium, isolated from the human left colon

Hussein Anani<sup>1,2</sup>, Rita Abou Abdallah<sup>1,2</sup>, May Khoder<sup>2,3</sup>, Anthony Fontanini<sup>2,3</sup>, Morgane Mailhe<sup>2,3</sup>, Davide Ricaboni<sup>2,3</sup>, Didier Raoult<sup>2,3,4</sup> & Pierre-Edouard Fournier<sup>1,2\*</sup>

The gut microbiota is considered to play a key role in human health. As a consequence, deciphering its microbial diversity is mandatory. A polyphasic taxonogenomic strategy based on the combination of phenotypic and genomic analyses was used to characterize a new bacterium, strain Marseille-P2911. This strain was isolated from a left colon sample of a 60-year old man who underwent a colonoscopy for an etiological investigation of iron-deficiency anemia in Marseille, France. On the basis of 16S rRNA sequence comparison, the closest phylogenetic neighbor was *Anaeroglobus geminatus* (94.59% 16S rRNA gene sequence similarity) within the family Veillonellaceae. Cells were anaerobic, Gram-stain-positive, non-spore-forming, catalase/oxidase negative cocci grouped in pairs. The bacterium was able to grow at 37 °C after 2 days of incubation. Strain Marseille-P2911 exhibited a genome size of 1,715,864-bp with a 50.2% G + C content, and digital DNA-DNA hybridization (dDDH) and OrthoANI values with *A. geminatus* of only  $19.1 \pm 4.5\%$  and 74.42%, respectively. The latter value being lower than the threshold for genus delineation (80.5%), we propose the creation of the new genus *Colibacter* gen. nov., with strain Marseille-P2911<sup>T</sup> (=DSM 103304 = CSUR P2911) being the type strain of the new species *Colibacter massiliensis* gen. nov., sp. nov.

With  $10^{11}$  to  $10^{12}$  cells per gram<sup>1</sup>, the commensal microbiota that resides in the colon has become a focus of interest that is attracting the attention of scientists. Although this complex flora plays a role in homeostasis, it has been demonstrated to participate in triggering intestinal and extra-intestinal diseases in sensitive people<sup>2</sup>. The microbial diversity of the colon microbiota may result from a co-evolution with its host<sup>3</sup> but may be affected by environmental conditions<sup>4</sup>. In order to identify all bacteria (including uncultured and fastidious) present in the colon, we use the culturomics strategy based on diversified culture conditions (temperature, media, and atmosphere), and rapid bacterial identification using matrix-assisted desorption ionization-time of flight mass spectrometry (MALDI-TOF MS)<sup>5-7</sup>. For putative new taxa, we use the taxonogenomic method that combines phenotypic characteristics and whole genome sequencing analysis to describe new bacterial species<sup>8-10</sup>. In 2016, we isolated the new bacterial strain Marseille-P2911 (=CSUR P2911 = DSM 103304), from a left colon sample of a 60-year-old patient who underwent a colonoscopy for the etiological investigation of iron-deficiency anemia in Marseille, France<sup>11</sup>. This bacterium was identified by matrix-assisted desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) using a Microflex spectrometer<sup>12</sup> (Bruker Daltonics, Bremen, Germany). The strain was predicted to be affiliated with members of the family of Veillonellaceae but distinct from species with a validly published name. In the present study, we aimed at comparing strain Marseille-P2911 to its closely related phylogenetic neighbors, and at proposing the creation of the new genus *Colibacter massiliensis* gen. nov., sp. nov.

<sup>1</sup>Aix Marseille Univ, Institut de Recherche pour le Développement (IRD), Service de Santé des Armées, AP-HM, UMR Vecteurs Infections Tropicales et Méditerranéennes (VITROME), Marseille, France. <sup>2</sup>Institut Hospitalo-Universitaire Méditerranée-Infection, Marseille, France. <sup>3</sup>Aix-Marseille Université, Institut de Recherche pour le Développement (IRD), UMR Microbes Evolution Phylogeny and Infections (MEPHI), Marseille, France. <sup>4</sup>Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia. \*email: [pierre-edouard.fournier@univ-amu.fr](mailto:pierre-edouard.fournier@univ-amu.fr)

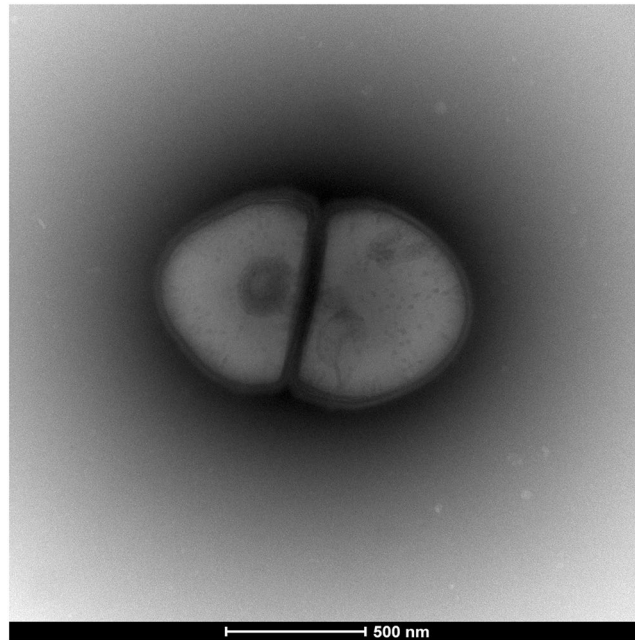


**Figure 1.** Phylogenetic tree highlighting the position of *Colibacter massiliensis* (in red) relative to other closely related bacterial taxa. Strains in bold are the eight closely related bacterial type strains with available genomes. The sequence of *Escherichia coli* ATCC 11775 served as outgroup. Genbank accession numbers for each 16S rRNA gene sequences are presented in parentheses. Sequences were here aligned by Muscle v3.8.31 with default parameters and finally 1247 nucleotides were considered to infer the phylogenetic relationship, using the Maximum Likelihood method with 1000 bootstrap replicates, within MEGA 7 software version 7.0.

## Results

**Strain identification and classification.** Strain Marseille-P2911 was isolated from the left colon liquid sample of a 60-year-old man who underwent a colonoscopy for the etiological investigation of an iron-deficiency anemia. The patient provided signed informed consent, and the study was approved by the ethics committee of the Institut Fédératif de Recherche IFR48 under number 2016-010. Strain Marseille-P2911 could not be identified by our systematic MALDI-TOF MS screening as the score was lower than 1.8, suggesting that the corresponding species was not in the database (Fig. S1). Moreover, strain Marseille-P2911 exhibited a 94.59% 16S rRNA sequence similarity with *Anaeroglobus geminatus* strain AIP 313.00 (GenBank accession no. AF338413), the phylogenetically-closest species with standing in nomenclature (Fig. 1). As this value is lower than the 95% threshold proposed by Stackebrandt and Ebers for defining a new genus, strain Marseille-P2911 was considered as a representative of a putatively new genus within the family Veillonellaceae in the phylum Firmicutes.

**Phenotypic characteristics.** Growth was observed on 5% sheep blood-enriched Columbia agar (BioMérieux) at 37 °C after 2 days of incubation. Colonies from strain Marseille-P2911 showed neither pigmentation nor haemolysis. They were circular with a diameter of 0.1 mm, and transparent. Bacterial cells were Gram-positive, non-motile diplococci with a diameter of 0.4 to 0.6 μm, as determined by transmission electron microscopy (Fig. 2). Strain Marseille-P2911 grew only in anaerobic conditions. The sporulation test (20 minutes at 80 °C) was negative. In addition, this bacterium had neither oxidase nor catalase activities. Biochemical



**Figure 2.** Transmission electron microscopy of *Colibacter massiliensis* using a Technai G2 electron microscope operating at 60 Kv (the scale bar represents 500 nm).

properties from strain Marseille-P2911, as determined by the use of API ZYM, API 20E and API 50 CH strips (bioMérieux), are detailed below in the description of *Colibacter massiliensis* gen. nov., sp. nov. paragraph.

According to the French Microbiology Society<sup>13</sup>, susceptibility tests showed that strain Marseille-P2911 was resistant to penicillin, piperacillin/tazobactam, ticarcillin-clavulanic acid, ceftioxin, ceftriaxone, ciprofloxacin, fosfomycin, trimethoprim-sulfamethoxazole, metronidazole, rifampin, amikacin and teicoplanin.

By comparison with closely related taxa, strain Marseille-P2911 differed in a combination of negative oxidase activity and positive sucrose/raffinose metabolism (Table 1). The most abundant fatty acids were saturated ones (83%), including dodecanoic acid that represented 25% of all fatty acids extracted from the bacterial cell wall. However, a specific 14:0 3-OH and several branched fatty acids were also described (Table S1).

**Genome sequencing information and genome properties.** The genome size of strain Marseille-P2911 was 1,715,864 bp long with a 50.2% G + C content. It was assembled into 2 scaffolds. Of the 1,655 predicted genes, 1,567 were protein-coding genes and 62 were RNAs (one complete rRNA operon, three additional 5S rRNAs and 49 tRNA genes). A total of 1,350 genes (81.57%) were assigned a putative function (by COGs) and 305 genes (18.43%) were annotated as hypothetical proteins. The genome properties and distribution of genes into COGs functional categories are detailed in Table S2 and Fig. 3. Genes putatively gained by hypothetical lateral gene transfer (LGT) were classified according to the bacterial families of origin (Fig. 4). Although we cannot rule out the possibility that some of the transfers may be from as yet unidentified taxa, most hypothetical lateral gene transfer (LGT)-acquired genes were obtained from members of the Veillonellaceae (90.6%) and Selenomonadaceae (1.6%) families.

**Comparison with closely related bacterial strains.** The genome of strain Marseille-P2911 was compared to the available genomes of eight closely related bacterial type strains. This comparison revealed that the genome size of our strain (1.72 Mb) was larger than that of *Dialister microaerophilus* (1.28 Mb), but smaller than those of *Anaeroglobus geminatus*, *Megasphaera cerevisiae*, *M. elsdenii*, *M. hexanoica*, *M. massiliensis*, *M. micronuciformis* and *M. paucivorans* (1.8, 3.24, 2.5, 2.88, 2.74, 1.77 and 2.91 Mb, respectively) (Table S3). The G + C content of strain Marseille-P2911 (50.2 mol %) was equal to that of *M. massiliensis*, but greater than those of all compared species (Table S3) except *Megasphaera elsdenii* (52.8%). The gene content of strain Marseille-P2911 (1,655) was similar to that of *D. microaerophilus* but smaller than those of other compared genomes (Table S3). The distribution of genes into COG categories was similar in all nine compared genomes (Fig. 5). Strain Marseille-P2911 shared 1169, 1161, 1129, 1109, 1083, 1081, 1071 and 666 orthologous genes with *M. micronuciformis*, *M. cerevisiae*, *M. hexanoica*, *M. massiliensis*, *A. geminatus*, *M. elsdenii*, *M. paucivorans* and *D. microaerophilus*, respectively (Table 2). Moreover, the MAGi analysis showed that AGIOS (Average Genomic Identity of Orthologous gene Sequences) values ranged from 47.76% between *M. elsdenii* and *D. microaerophilus*, to 50.93% between *M. paucivorans* and *D. microaerophilus*, among studied bacterial strains. Regarding strain Marseille-P2911, AGIOS values ranged from 48.86% with *D. microaerophilus* to 50.78% with *A. geminatus* (Table 2). Using dDDH analysis, strain Marseille-P2911 exhibited values ranging from 18.1% with *M. massiliensis* to 27.1% with *D. microaerophilus* (Table S4). These values are lower than the 70% threshold used for delineating prokaryotic species, thus confirming that this strain represents a new species. Finally, strain Marseille-P2911 exhibited average nucleotide identity (ANI) values ranging from 63.37% with *D. microaerophilus* to 74.42% with

Characteristics	CM	AG	DM	MP	MMi	MC	ME	MMA	MH
Gram stain	+	–	–	–	–	–	–	–	–
Production of Catalase	–	–	–	–	–	–	–	–	–
Oxidase	–	–	–	–	–	–	–	+	–
Nitrate reductase	na	–	–	–	–	–	–	na	–
Gelatin hydrolysis	+	–	–	–	–	–	+	+	–
Utilisation of L-Arabinose	–	–	–	–	–	±	–	+	–
D-Galactose	+	+	–	–	–	–	–	+	w
D-Fructose	+	–	–	–	–	+	+	+	+
D-Mannose	+	+	–	–	–	–	–	+	±
D-Rhamnose	na	–	–	–	–	–	–	+	–
D-Mannitol	+	–	–	–	–	–	+	+	–
Salicin	+	–	–	–	–	–	–	+	–
D-Cellobiose	+	–	–	–	–	–	–	+	–
D-Maltose	+	–	–	–	–	–	+	+	–
D-Lactose	+	–	–	–	–	–	–	+	–
D-Trehalose	+	–	–	na	na	na	na	+	–
Inulin	–	+	na	–	+	–	–	na	–
Starch	–	–	–	na	na	na	na	na	–
Glycogen	–	+	na	–	+	–	–	na	–
D-Glucose	+	–	–	–	–	–	+	+	–
Raffinose	+	–	–	–	–	–	–	–	–
D-Xylose	–	–	–	–	–	–	–	+	–
Sucrose	+	–	–	na	–	na	–	–	–

**Table 1.** Compared phenotypic characteristics of studied species. Differential characteristics of *Colibacter massiliensis*<sup>11</sup> (CM), *Anaeroglobus geminatus*<sup>37</sup> (AG), *Dialister microaerophilus*<sup>38</sup> (DM), *Megasphaera paucivorans*<sup>39</sup> (MP), *Megasphaera micronuciformis*<sup>40</sup> (MMi), *Megasphaera cerevisiae*<sup>41,42</sup> (MC), *Megasphaera elsdenii*<sup>42,43</sup> (ME), *Megasphaera massiliensis*<sup>44</sup> (MMA) and *Megasphaera hexanoica*<sup>45</sup> (MH). +positive reaction, –negative reaction, ±variable, w week reaction, na data not available.

*A. geminatus*. An ANI value lower than 80.5% suggesting that two strains belong to distinct genera, we considered that strain Marseille-P2911 was representative of a new genus (Fig. 6). Consequently, based on the presented phenotypic and genomic data, we propose the creation of the new genus *Colibacter* gen. nov., with strain Marseille-P2911<sup>T</sup> being the type strain of the novel species *Colibacter massiliensis* gen. nov., sp. nov.

## Discussion

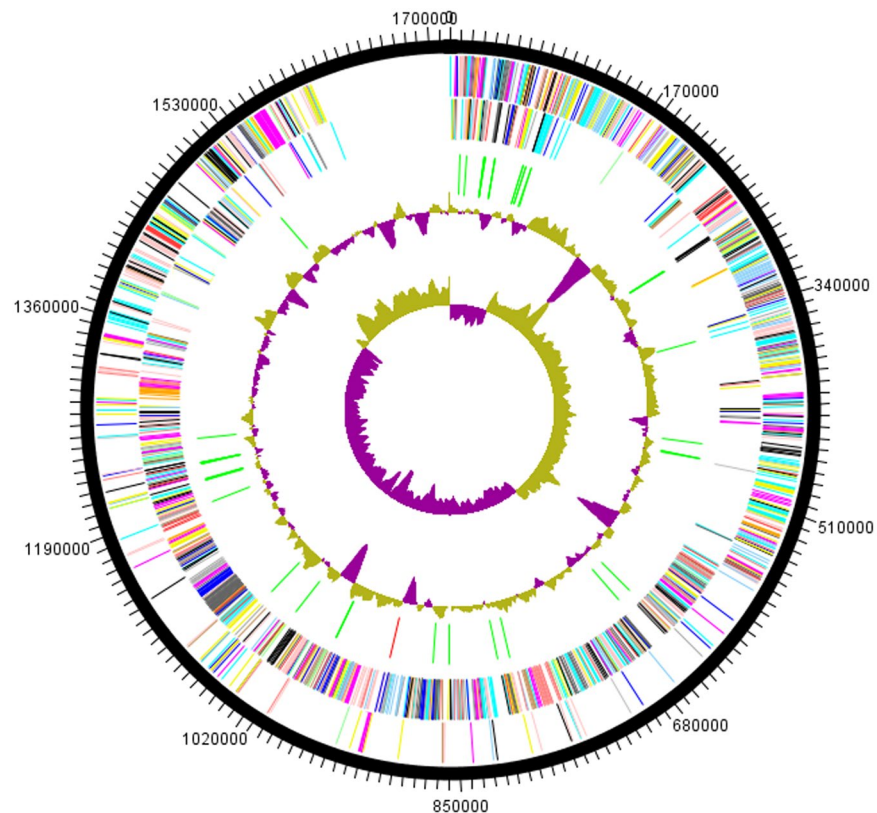
The gastrointestinal tract harbors a complex microbial microflora whose dynamic composition is important for health<sup>14</sup>. Here we aimed at describing a new bacterial species to enrich the knowledge on the human microbiome, using the culturomics and taxonogenomic strategies<sup>7,10,15</sup>.

The phylogenetic and phenotypic analysis of the new strain Marseille-P2911 revealed several distinct traits when compared to other members of the family Veillonellaceae<sup>16</sup>, suggesting that it could be classified in a new species of a new genus. The family Veillonellaceae is currently made of six genera of Gram-negative bacteria, including *Veillonella* (twelve species), *Megasphaera* and *Dialister* (five species each), *Allisonella*, *Anaeroglobus* and *Negativicoccus* (one species each)<sup>17</sup>.

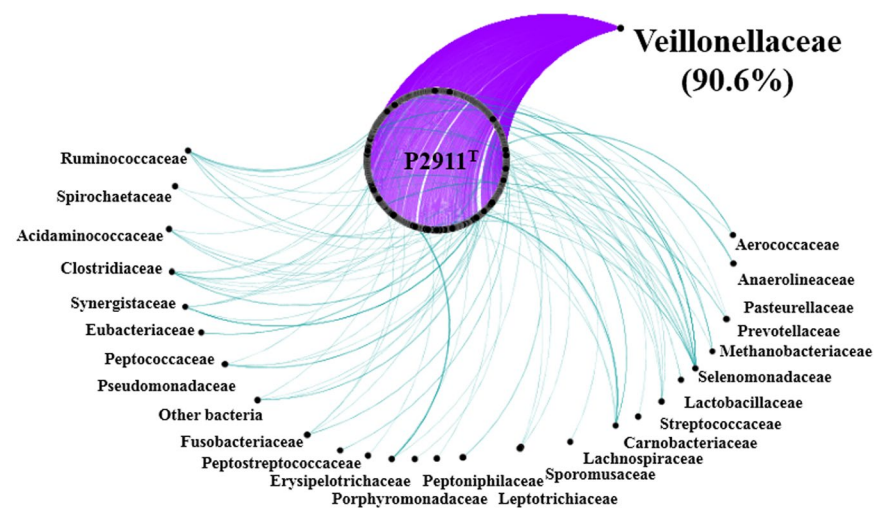
As strain Marseille-P2911, many members of the family Veillonellaceae were detected in humans. For example, *Allisonella histaminiformans* was previously isolated from the vagina and *A. geminatus* was isolated from the gastrointestinal tract, stool and skin<sup>17</sup>. Moreover, these two species were demonstrated to be pathogenic in several diseases such as community-acquired pneumonia and advanced caries<sup>17</sup>.

The genomic content of strain Marseille-P2911 (dDDH, orthoANI, and AGIOS values) comforted its new species status. We observed a significant similarity to genes from the family Veillonellaceae (90.6%) (Fig. 4). A small rate of hypothetical lateral gene transfer (9.4%) from other bacterial families was observed, notably several toxin/antitoxin system-related genes putatively acquired from the families Eubacteriaceae, Lachnospiraceae, Lactobacillaceae, Selenomonadaceae and Streptococcaceae). So, we formally propose the creation of the new genus and species *Colibacter massiliensis* gen. nov., sp. nov., within the family Veillonellaceae. The type strain, Marseille-P2911<sup>T</sup>, was deposited in the DSMZ and CSUR collections under accession numbers DSM 103304 and CSUR P2911, respectively. The 16S rRNA and genome sequences are available in GenBank under accession numbers LT576403 and FMIY00000000, respectively. The Digital Protologue TaxoNumbers (<http://imedea.uibcsic.es/dprotologue/index.php>) of *C. massiliensis* gen. nov., sp. nov. is GA00103.

**Description of *Colibacter* gen. nov.** *Colibacter* (Co.li.bac'ter, N.L. masc. n. *colibacter*, composed of colon, the organ from which the strain was first isolated, and bacter, a rod. *Colibacter*, a bacterium from the colon). Cells are diplococci, anaerobic, Gram-positive, nonmotile, asporogenous, catalase and oxidase negative.



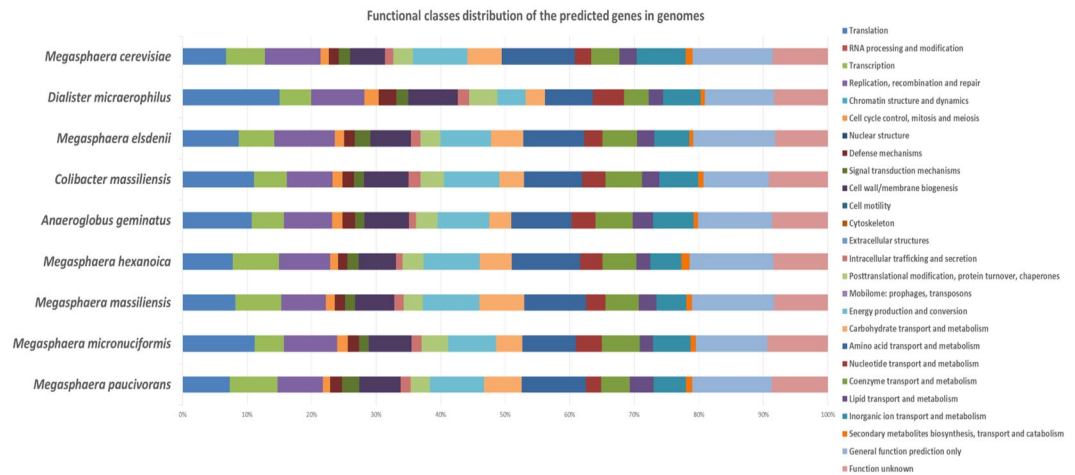
**Figure 3.** Graphical circular map of the genome. From outside to the center: genes on the forward strand colored by Clusters of Orthologous Groups of proteins (COG) categories (only genes assigned to COG). Genes on the reverse strand colored by COG categories (only genes assigned to COG). RNA genes (tRNAs green, rRNA red). G + C content and G + C skew.



**Figure 4.** Network showing the origin of genes in *C. massiliensis* according to the bacterial families.

**Description of *Colibacter massiliensis* gen. nov., sp. nov.** *Colibacter massiliensis* (mas.si.li.en'sis; L. fem. adj. *massiliensis*, for Massilia, the Roman name of Marseille, where the type strain was first isolated).

In addition to the description features of the genus, cells have a diameter varying from 0.4 to 0.6  $\mu\text{m}$ . Colonies grown on 5% sheep blood-enriched Columbia agar (bioMérieux) are circular and transparent after 2 days of incubation in anaerobic atmosphere. Growth occurs at 37 °C. Cells grow anaerobically only. Using an API ZYM strip, a positive reaction was observed for alkaline phosphatase, esterase (C4), lipase esterase (C8), leucine arylamidase, valine arylamidase, cystine arylamidase,  $\alpha$ -chymotrypsine, phosphatase acid, naphthol-AS-BI-phosphohydrolase,  $\beta$ -galactosidase,  $\alpha$ -glucosidase,  $\beta$ -glucosidase and N-acetyl- $\beta$ -glucosaminidase, but negative reactions were



**Figure 5.** Distribution of functional classes of predicted genes according to the clusters of orthologous groups of proteins of *Colibacter massiliensis* among other bacterial taxa type strain.

	MP	MMi	MMa	MH	AG	CM	ME	DM	MC
MP	<b>2717</b>	1089	1309	1340	972	1071	1300	672	1418
MMi	50.45	<b>1745</b>	1154	1162	1090	1169	1124	665	1201
MMa	49.55	50.37	<b>2536</b>	1405	1029	1109	1441	675	1444
MH	50.26	50.18	50.6	<b>2785</b>	1019	1129	1367	692	1442
AG	49.85	50.59	50.39	50.58	<b>1969</b>	1083	1003	594	1075
CM	49.81	50.42	50.46	50.31	50.78	<b>1661</b>	1081	666	1161
ME	48.97	49.83	50.78	50.42	50.4	50.38	<b>2333</b>	683	1434
DM	50.93	50.1	48.48	49.22	49.07	48.86	47.76	<b>1245</b>	683
MC	50.61	50.47	50.08	50.3	50.41	50.08	49.79	50.18	<b>3086</b>

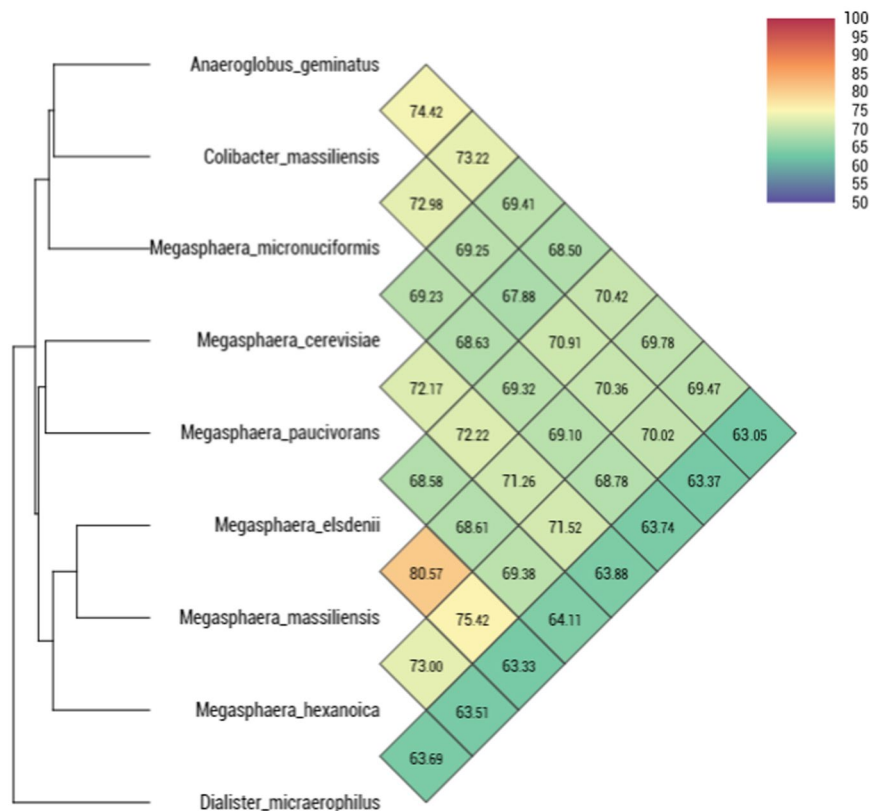
**Table 2.** Numbers of orthologous proteins shared between genomes (upper right) and AGIOS values (lower left). The numbers of proteins per genome are indicated in bold. *Colibacter massiliensis* (CM), *Anaeroglobus geminatus* (AG), *Dialister microaerophilus* (DM), *Megasphaera paucivorans* (MP), *Megasphaera micronuciformis* (MMi), *Megasphaera cerevisiae* (MC), *Megasphaera elsdenii* (ME), *Megasphaera massiliensis* (MMa) and *Megasphaera hexanoica* (MH).

obtained for lipase, trypsin,  $\alpha$ -galactosidase,  $\beta$ -glucuronidase,  $\alpha$ -mannosidase and  $\alpha$ -fucosidase. Using an API 20E strip, a positive reaction were obtained for  $\beta$ -galactosidase, arginine dihydrolase, urea and gelatin hydrolysis, glucose fermentation, mannitol, inositol, sorbitol, rhamnose, sucrose, melibiose, amygdalin and arabinose. Using an API 50 CH strip, a positive result was shown for D-ribose, D-galactose, D-glucose, D-fructose, D-mannose, D-mannitol, D-sorbitol, N-acetylglucosamine, amygdalin, arbutin, esculin iron citrate, salicin, D-cellobiose, D-maltose, D-lactose, D-melibiose, D-saccharose, D-trehalose, D-melezitose, D-raffinose and D-tagatose. In contrast, negative reactions were observed for glycerol, methyl- $\alpha$ D-glucopyranoside, D-turanose, potassium 5-Ketogluconate, erythritol, D-arabinose, L-arabinose, D-xylose, L-xylose, D-adonitol, methyl- $\beta$ D-xylopyranoside, L-sorbose, L-rhamnose, dulcitol, inositol, methyl- $\alpha$ D-mannopyranoside, inulin, starch, glycogen, xylitol, gentiobiose, D-lyxose, D-fucose, L-fucose, D-arabitol, L-arabitol, potassium gluconate and potassium 2-ketogluconate. The major fatty acid found for this strain was dodecanoic acid. The genome is 1,715,864 bp long with 50.2% G + C content.

The type strain, Marseille-P2911<sup>T</sup>, isolated from the left colon of a patient, was deposited in the DSMZ and CSUR collections under accession numbers DSM 103304 and CSUR P2911, respectively. The 16S rRNA and genome sequences are available in GenBank under accession numbers LT576403 and FM1Y00000000, respectively.

## Materials and Methods

**Strain isolation and phenotypic tests.** The left colon liquid sample of a 60-year-old-man, who underwent a colonoscopy for an aetiological investigation of iron-deficiency anemia, was initially collected in La Timone Hospital in Marseille, France. From this sample, strain Marseille-P2911 was isolated after 3 days of pre-incubation in an anaerobic blood culture bottle (VersaTREK REDOX 2, Thermo Scientific, Villebon sur Yvette, France) supplemented with 5 mL of 0.2  $\mu$ m-filtered rumen. This enriched liquid medium was then inoculated on 5% sheep blood-enriched Columbia agar (BioMérieux, Marcy l'Etoile, France) followed by an incubation at 37 °C in anaerobic atmosphere (AnaeroGEN Compact, Oxoid, Thermo Scientific, Dardilly, France). MALDI-TOF mass spectrometry (MS) protein analysis was carried out using a Microflex spectrometer<sup>18</sup> (Bruker Daltonics,



**Figure 6.** Heatmap generated with OrthoANI values calculated using the OAT software between *Colibacter massiliensis* and other closely related taxa with standing in nomenclature.

Bremen, Germany). Strain Marseille-P2911 spectra were imported into the MALDI BioTyper software (version 2.0, Bruker) and analysed by standard pattern matching (with default parameter settings). Interpretation of the scores was performed as previously described<sup>19</sup>.

Moreover, the 16S rRNA gene was sequenced using the fD1-rP2 primer pairs as previously described<sup>20</sup>, using a 3130-XL sequencer (Applied Biosciences, Saint Aubin, France). A phylogenetic tree was obtained using the Maximum Likelihood method and Kimura 2-parameter within the MEGA 7 software<sup>21</sup>. Several growth temperatures (20, 28, 37, 45 and 55 °C) on 5% sheep blood-enriched Columbia agar medium (BioMérieux, Marcy l'Etoile, France) were tested. Growth of strain Marseille-P2911 was tested under different atmospheres (anaerobic, aerobic and microaerophilic (CampyGEN, Oxoid). API ZYM and API 50CH strips (BioMérieux) were used to evaluate the biochemical properties of the strain test according to the manufacturer's instructions. All API experiments were performed under anaerobic conditions. Using API 50CH, API 20E and API ZYM strips, strain Marseille-P2911 was incubated for 48, 24 and 4 hours, respectively. The standard disc method was applied for antimicrobial susceptibility testing according to the French Microbiology Society<sup>13</sup>. Finally, cellular fatty acid methyl ester (FAME) analysis was performed by GC/MS. Two samples were prepared with approximately 18 mg of bacterial biomass per tube harvested from several culture plates. Briefly, fatty acid methyl esters were separated using an Elite 5-MS column and monitored by mass spectrometry (Clarus 500 - SQ 8S, Perkin Elmer, Courtabouef, France)<sup>22</sup>. GC/MS analyses were carried out as described before<sup>23</sup>. Spectral database search was performed using MS Search 2.0 operated with the Standard Reference Database 1 A (NIST, Gaithersburg, USA) and the FAMES mass spectral database (Wiley, Chichester, UK). For transmission electronic microscopy, detection formvar-coated grids were dropped onto a 40 µL bacterial suspension before incubation at 37 °C for 30 minutes. Then, the grids were incubated on 1% ammonium molybdate for 10 seconds, dried on blotting paper and finally observed using a Tecnai G20 transmission electron microscope (FEI, Limeil-Brevannes, France) at an operating voltage of 60 Kv. All methods were performed in accordance with the relevant guidelines and regulations.

**Extraction and genome sequencing.** After a pretreatment by lysozyme incubation at 37 °C for 2 hours, DNA of strain Marseille-P2911 was extracted using an EZ1 biorobot (Qiagen) with the EZ1 DNA Tissue kit. The elution volume was 50 µL. Genomic DNA (gDNA) was quantified by a Qubit assay with the high sensitivity kit (Life technologies, Carlsbad, CA, USA) at 92.6 ng/µL. Genomic DNA was sequenced on a MiSeq sequencer (Illumina Inc, San Diego, CA, USA) with the Mate Pair strategy. The gDNA was barcoded in order to be mixed with 11 others projects with the Nextera Mate Pair sample prep kit (Illumina). The Mate Pair library was prepared with 1.5 µg of genomic DNA 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 Inc, Santa Clara, CA, USA) with a DNA 7500 labchip. DNA

fragments ranged in size from 1.5 kb up to 11 kb with an optimal size at 8.4 kb. No size selection was performed and 600 ng of tagged fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal at 706 bp on the Covaris device S2 in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies Inc, Santa Clara, CA, USA) and the final concentration library was measured at 13.256 nmol/l. The libraries were normalized at 2 nM after a denaturation step and dilution at 15 pM, 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-hours run in a 2 × 251-bp format. Total information of 8.3 gb was obtained from a 910 K/mm<sup>2</sup> cluster density with a cluster passing quality control filters of 92.8% (16,316,000 passing filter paired reads). Within this run, the index representation for strain Marseille-P2911 was determined to 8.4%. The 688,244 paired reads were quality-checked using FastQC, trimmed using Trimmomatic version 0.36.6<sup>24</sup> and assembled in two scaffolds using the SPAdes version 3.5.0 software<sup>25</sup>. The option “careful” was used in order to reduce the number of mismatches and short indels. Default parameters were applied for k values, i.e., k-mer values of 127, 99, 77, 55, 33, and 21. SSPACE<sup>26</sup> and GapFiller<sup>27</sup> were used to combine contigs, using default parameters<sup>28,29</sup>.

**Genome annotation and genome comparison.** The genome was annotated as previously described<sup>19</sup>. In addition, we used the Genome-to Genome Distance Calculator (GGDC) web server available at (<http://ggdc.dsmz.de>) to estimate the overall similarity among the compared genomes and to replace the wet-lab DNA–DNA hybridization (DDH) by a digital DDH (dDDH)<sup>30,31</sup>. Average nucleotide identity analysis was also estimated using the orthoANI<sup>32</sup> and MAGI<sup>33</sup> softwares. Antibiotic resistance genes (ARG) were searched using the ARG-ANNOT database and Bio-Edit interface<sup>34</sup>. Assembled sequences were searched against the ARG-ANNOT database under moderately stringent conditions (e-value of 10<sup>-5</sup>) for the in silico ARG prediction. These putative ARGs were further confirmed through a BLAST search against non-redundant (nr) database in GenBank.

The presence of pathogenesis-related proteins was investigated using PathogeneFinder 1.1<sup>35</sup>. Finally, predicted protein sequences of strain Marseille-P2911 were used as queries to search the NCBI GenBank non-redundant protein sequence database. These results were formatted to generate network of protein sequences using the Cytoscape tool<sup>36</sup>.

Received: 6 February 2019; Accepted: 6 November 2019;

Published online: 20 November 2019

## References

- Berg, R. The indigenous gastrointestinal microflora. *Trends in Microbiology* **4**, 430–435 (1996).
- Shanahan, F. The colonic microbiota in health and disease. *Current Opinion in Gastroenterology* **29**, 49–54 (2013).
- Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine. *Cell* **124**, 837–848 (2006).
- Chow, J., Lee, S. M., Shen, Y., Khosravi, A. & Mazmanian, S. K. Host–Bacterial Symbiosis in Health and Disease. In *Advances in Immunology* **107**, 243–274 (Elsevier, 2010).
- Lagier, J.-C. *et al.* Microbial culturomics: paradigm shift in the human gut microbiome study. *Clinical Microbiology and Infection* **18**, 1185–1193 (2012).
- Lagier, J.-C. *et al.* The Rebirth of Culture in Microbiology through the Example of Culturomics To Study Human Gut Microbiota. *Clinical Microbiology Reviews* **28**, 237–264 (2015).
- Lagier, J.-C. *et al.* Culture of previously uncultured members of the human gut microbiota by culturomics. *Nature Microbiology* **1** (2016).
- Ramasamy, D. *et al.* A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. *International Journal of Systematic and Evolutionary Microbiology* **64**, 384–391 (2014).
- Fournier, P.-E. & Drancourt, M. New Microbes New Infections promotes modern prokaryotic taxonomy: a new section “TaxonoGenomics: new genomes of microorganisms in humans”. *New Microbes and New Infections* **7**, 48–49 (2015).
- Abdallah, R. A. *et al.* The impact of culturomics on taxonomy in clinical microbiology. *Antonie van Leeuwenhoek* **110**, 1327–1337 (2017).
- Mailhe, M. *et al.* ‘Colibacter massiliensis’ gen. nov., sp. nov., a new bacterial species isolated from human left colon. *New Microbes and New Infections* **17**, 23–24 (2017).
- Seng, P. *et al.* Ongoing Revolution in Bacteriology: Routine Identification of Bacteria by Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry. *Clinical Infectious Diseases* **49**, 543–551 (2009).
- Leclercq, R. *et al.* EUCAST expert rules in antimicrobial susceptibility testing. *Clinical Microbiology and Infection* **19**, 141–160 (2013).
- Cremer, J., Arnoldini, M. & Hwa, T. Effect of water flow and chemical environment on microbiota growth and composition in the human colon. *Proceedings of the National Academy of Sciences* **114**, 6438–6443 (2017).
- Fournier, P.-E., Lagier, J.-C., Dubourg, G. & Raoult, D. From culturomics to taxonomogenomics: A need to change the taxonomy of prokaryotes in clinical microbiology. *Anaerobe* **36**, 73–78 (2015).
- Rogosa, M. Transfer of *Peptostreptococcus elsdenii* Gutierrez *et al.* to a New Genus, *Megasphaera* CM. *elsdenii* (Gutierrez *et al.*) comb. nov. **3** (2018).
- Marchandin, H. & Jumas-Bilak, Estelle. The Family Veillonellaceae. The Prokaryotes: Firmicutes and Tenericutes. 433–453 (2013).
- Seng, P. *et al.* Identification of Rare Pathogenic Bacteria in a Clinical Microbiology Laboratory: Impact of Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry. *Journal of Clinical Microbiology* **51**, 2182–2194 (2013).
- Hadjadj, L. *et al.* Non contiguous-finished genome sequence and description of *Microbacterium gorillae* sp. nov. *Standards in Genomic Sciences* **11** (2016).
- Drancourt, M. *et al.* 16S Ribosomal DNA Sequence Analysis of a Large Collection of Environmental and Clinical Unidentifiable Bacterial Isolates. *J. CLIN. MICROBIOL.* **38**, 8 (2000).
- Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* **33**, 1870–1874 (2016).
- Sasser, M. Bacterial Identification by Gas Chromatographic Analysis of Fatty Acids Methyl Esters (GC-FAME). **6** (2006).
- Dione, N. *et al.* Genome sequence and description of *Anaerobaculum massiliensis* sp. nov. *New Microbes and New Infections* **10**, 66–76 (2016).
- Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).



25. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* **19**, 455–477 (2012).
26. Boetzer, M. *et al.* Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011).
27. Boetzer, M. & Pirovano, W. Toward almost closed genomes with GapFiller. *Genome B* (2012).
28. Anani, H. *et al.* Draft genome and description of *Merdibacter massiliensis* gen.nov., sp. nov., a new bacterium genus isolated from the human ileum. *Scientific Reports* **9** (2019).
29. Anani, H., Khodor, M., Raoult, D. & Fournier, P.-E. Whole-Genome Sequence of French Clinical *Olivibacter jilunii* Strain P8502. *Microbiology Resource Announcements* **8** (2019).
30. Auch, A. F. Digital DNA-DNA hybridization. *Standards in Genomic Sciences* **18** (2010).
31. Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P. & Göker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**, 60 (2013).
32. Lee, I., Ouk Kim, Y., Park, S.-C. & Chun, J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *International Journal of Systematic and Evolutionary Microbiology* **66**, 1100–1103 (2016).
33. Weber, T. *et al.* antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research* **43**, W237–W243 (2015).
34. Gupta, S. K. *et al.* ARG-ANNOT, a New Bioinformatic Tool To Discover Antibiotic Resistance Genes in Bacterial Genomes. *Antimicrobial Agents and Chemotherapy* **58**, 212–220 (2014).
35. Cosentino, S., Voldby Larsen, M., Møller Aarestrup, F. & Lund, O. PathogenFinder - Distinguishing Friend from Foe Using Bacterial Whole Genome Sequence Data. *PLoS ONE* **8**, e77302 (2013).
36. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431–432 (2011).
37. Carlier, J. P. *Anaeroglobus geminatus* gen. nov., sp. nov., a novel member of the family Veillonellaceae. *International Journal of Systematic and Evolutionary Microbiology* **52**, 983–986 (2002).
38. Jumas-Bilak, E. *Dialister micraerophilus* sp. nov. and *Dialister propionificiens* sp. nov., isolated from human clinical samples. *International Journal of Systematic and Evolutionary Microbiology* **55**, 2471–2478 (2005).
39. Juvonen, R. *Megasphaera paucivorans* sp. nov., *Megasphaera sueciensis* sp. nov. and *Pectinatus haikarae* sp. nov., isolated from brewery samples, and emended description of the genus *Pectinatus*. *International Journal of Systematic and Evolutionary Microbiology* **56**, 695–702 (2006).
40. Marchandin, H. Phylogenetic analysis of some *Sporomusa* sub-branch members isolated from human clinical specimens: description of *Megasphaera micronuciformis* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* **53**, 547–553 (2003).
41. Engelmann, U. & Weiss, N. *Megasphaera cerevisiae* sp. nov.: A New Gram-negative Obligately Anaerobic Coccus Isolated from Spoiled Beer. *Systematic and Applied Microbiology* **6**, 287–290 (1985).
42. Shetty, S. A., Marathe, N. P., Lanjekar, V., Ranade, D. & Shouche, Y. S. Comparative Genome Analysis of *Megasphaera* sp. Reveals Niche Specialization and Its Potential Role in the Human Gut. *PLoS ONE* **8**, e79353 (2013).
43. Gutierrez, J., Davis, R. E., Lindahl, I. L. & Warwick, E. J. Bacterial changes in the rumen during the onset of feed-lot bloat of cattle and characteristics of *Peptostreptococcus elsdenii* n. sp. *Applied Microbiology* **7**, 16–22 (1959).
44. Padmanabhan, R. *et al.* Non-contiguous finished genome sequence and description of *Megasphaera massiliensis* sp. nov. *Standards in Genomic Sciences* **8**, 525–538 (2013).
45. Jeon, B. S., Kim, S. & Sang, B.-I. *Megasphaera hexanoica* sp. nov., a medium-chain carboxylic acid-producing bacterium isolated from a cow rumen. *International Journal of Systematic and Evolutionary Microbiology* **67**, 2114–2120 (2017).

## Acknowledgements

The study was supported by the Méditerranée Infection foundation, the National Research Agency under the program “Investissements d’avenir”, reference ANR-10-IAHU-03 and by Région Provence Alpes Côte d’Azur and European funding FEDER PRIMI.

## Author contributions

H.A. performed the genomic analysis and phenotypic tests, R.A. and M.K. participated in the genomic analysis, A.F. did the electronic microscopy, D.Ri. and M.M. isolated the bacterium, D.Ra., P.E.F. and H.A. wrote the paper and designed the study. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-019-53791-1>.

**Correspondence** and requests for materials should be addressed to P.-E.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019