



## OPEN Taxonogenomics of *Culturomica massiliensis* gen. nov., sp. nov., and *Emergencia timonensis* gen. nov., sp. nov. new bacteria isolated from human stool microbiota

Afaf Hamame<sup>1,2</sup>, Reham Magdy Wasfy<sup>1,2</sup>, Cheikh Ibrahima Lo<sup>2,3</sup>, Florence Fenollar<sup>2,3</sup>, Didier Raoult<sup>1,2</sup>, Pierre-Edouard Fournier<sup>2,3</sup> & Linda Houhamdi<sup>2,3</sup>✉

Two new bacterial strains, Marseille-P2698<sup>T</sup> (CSUR P2698 = DSM 103,121) and Marseille-P2260<sup>T</sup> (CSUR P2260 = DSM 101,844 = SN18), were isolated from human stools by the culturomic method. We used the taxonogenomic approach to fully describe these two new bacterial strains. The Marseille-P2698<sup>T</sup> strain was a Gram-negative, motile, non-spore-forming, rod-shaped bacterium. The Marseille-P2260<sup>T</sup> strain was a Gram-positive, motile, spore-forming rod-shaped bacterium. Major fatty acids found in Marseille-P2698<sup>T</sup> were C<sub>15:0 iso</sub> (63%), C<sub>15:0 anteiso</sub> (11%), and C<sub>17:0 3-OH iso</sub> (8%). Those found in Marseille-P2260<sup>T</sup> strain were C<sub>16:00</sub> (39%), C<sub>18:1n9</sub> (16%) and C<sub>18:1n7</sub> (14%). Strains Marseille-P2698<sup>T</sup> and Marseille-P2260<sup>T</sup> had 16S rRNA gene sequence similarities of 91.50% with *Odoribacter laneus*<sup>T</sup>, and of 90.98% and 95.07% with *Odoribacter splanchnicus*<sup>T</sup> and *Eubacterium sulci*<sup>T</sup>, respectively. The exhibited digital DNA-DNA Hybridization values lower than 20.7%, and Orthologous Average Nucleotide Identity values lower than 73% compared to their closest related bacterial species *O. splanchnicus*<sup>T</sup> and *E. sulci*<sup>T</sup> respectively. Phenotypic, biochemical, phylogenetic, and genomic results obtained by comparative analyses provided sufficient evidence that both of the two studied strains Marseille-P2698<sup>T</sup> and Marseille-P2260<sup>T</sup> are two new bacterial species and new bacterial genera for which the names *Culturomica massiliensis* gen. nov., sp. nov., and *Emergencia timonensis* gen. nov., sp. nov. were proposed, respectively.

The human gut microbiota is considered currently as one of the most active research fields in microbiology<sup>1</sup>. In fact, this microflora harbours a huge biodiversity of bacteria of which a large part is still unknown<sup>2</sup>. Researchers used a variety of strategies to speed up and simplify the description of new bacterial species by optimizing their in vitro growth conditions<sup>3-5</sup>. Culturomics is one among these strategies, which relies on a diversification of culture conditions that allowed the identification of several new bacterial species isolated from the human gastrointestinal tract<sup>3-5</sup>. Since 2012, it allowed the isolation of over 1000 different human-associated bacterial species, including several hundreds of new species<sup>3-6</sup>. This method highlighted the need to adopt taxonomic approaches to clinical microbiology by including the use of modern and reproducible tools, such as high throughput genomic and proteomic analyses.

In November 30, 2015, two putative new bacterial species, *Culturomica massiliensis* gen. nov., sp. nov., and *Emergencia timonensis* gen. nov., sp. nov., were isolated from patient's stools, and partially described<sup>7,8</sup>. The genomic sequencing of new described bacterial species constitutes currently a necessary step for performing their comparative taxogenomic descriptions with their closest related known species. In fact, several recent publications have used genomic descriptions to characterize the new species by comparison to their closest relatives strains<sup>9-11</sup>. The aim of our current study was to complete the phenotypic, taxonomic and genomic characterization proposal of new genera and new species of *Culturomica massiliensis* gen. nov., sp. nov., strain

<sup>1</sup>Aix Marseille Univ, IRD, AP-HM, MEPHI, Marseille, France. <sup>2</sup>IHU-Méditerranée Infection, Marseille, France. <sup>3</sup>Aix Marseille Univ, IRD, AP-HM, Service de Santé Des Armées (SSA), VITROME, 19-21 Boulevard Jean Moulin, 13385 Marseille Cedex 05, France. ✉email: lhouhamdi@yahoo.fr; linda.houhamdi@ap-hm.fr

Marseille-P2698<sup>T</sup>, and *Emergencia timonensis* gen. nov., sp. nov., strain Marseille-P2260<sup>T</sup> and formally expose the creation of both species.

## Materials and methods

**Sample collection and ethics approval.** In November 30, 2015, stool samples were collected from hospitalized patients in the Timone Hospital (Marseille, France) as a part of a study of human microbiota diversity. Patients provided signed informed consent<sup>7,8</sup>. The study protocol was approved by the ethics committee of the institut de recherche fédératif 48, under agreement number 09-022. In addition, all methods were performed in accordance with the relevant guidelines and regulations. Each sample was then cultured according to the culturomics method previously established in our laboratory<sup>3,5</sup>. Various types of bacterial colonies were isolated on 5% of sheep blood-enriched Columbia agar (bioMérieux®, Marcy l'Etoile, France). Bacterial colonies were then screened for identification by Matrix-Assisted Laser Desorption Ionization-Time Of Flight Mass Spectrometry (MALDI-TOF MS) instrument (Bruker Daltonics®, Bremen, Germany) as previously reported<sup>12</sup>. Both two strains studied herein had a MALDI-TOF score lower than 2.0, which did not allow their correct identification. Their spectra were then added to the local MALDI-TOF MS database (<https://www.mediterranee-infection.com/urms-data-base>).

**16S rRNA gene sequencing and identification.** The 16S rRNA sequences from both strains were directly extracted from their whole genomes sequences and then, compared by Basic Local Alignment Search Tool nucleotide (BLASTn) to the non-redundant (nr) databases<sup>13</sup>. The obtained sequence similarity percentages allowed identification of the closest species to each strain, and to predict if it was new species (<98.65% of similarity). Then, the phylogenetic tree was constructed based on these 16S rRNA gene sequences in comparison to the closest related species of each studied strain. Designated species sequences were downloaded from nr<sup>14</sup>, and aligned with ClustalW. Phylogenetic trees were constructed using MEGA 11 version 11.0.10 with the maximum likelihood method and 1000 bootstrap replications<sup>15</sup>.

**Phenotypic and biochemical characterizations.** Optimal culture conditions were determined by testing various incubation temperatures (25, 28, 37, 42, and 50 °C), atmospheres (aerobic, anaerobic and micro-aerophilic), NaCl concentrations (5, 5.5, 7.5, 10, 15, and 20% of NaCl) and pH levels (5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5). The morphology and motility were observed using a new-generation scanning electron microscope (Hitachi High-71 Technologies Corporation, Tokyo, Japan).

Furthermore, three semi-quantitative standardized micro-methods of Analytical Profile Index (API®, bioMérieux®) tests: API® 20A, API® 50 CH, and API® ZYM were used, according to the manufacturer's instructions<sup>16</sup>, in order to study carbohydrate metabolism and enzymatic activities.

Fatty acid methyl ester (FAME) analysis was explored by Gas Chromatography/Mass Spectrometry, as previously reported<sup>17,18</sup>. FAMES were separated using an Elite 5-MS column and monitored by mass spectrometry (Clarus 500—SQ 8 S, Perkin Elmer®, Courtaboeuf, France). Obtained spectra were compared with those contained in the repertory databases using MS Search 2.0 operated with the Standard Reference Database 1A (National Institute of Standards and Technology-NIST, Gaithersburg, USA), and FAMES mass spectral database (Wiley, Chichester, UK).

**Whole genomic sequencing and bioinformatic analyses.** First, bacterial DNA was extracted using the EZ1 DNeasy Blood Tissue Kit (Qiagen® GmbH, Hilden, Germany) in line with the manufacturer's protocol<sup>19</sup>. Whole-genome sequencing was performed using an Illumina® MiSeq sequencer (Illumina®, San Diego, CA, USA)<sup>20</sup>. Then, sequenced genomes were assembled using SPAdes 3.5.0 software<sup>21</sup>, which reduces short indels and the huge number of mismatches. Raw reads in contigs less than 700-bp-long were removed. Finally, the quality of the sequenced genome was checked using BLAST against the nr/nt database. This method allowed us to better explore the relationship between a submitted assembly of our new species to the International Nucleotide Sequence Database Collaboration (INSDC), i.e., DDBJ, ENA, or GenBank, and the assembly represented in the NCBI reference sequence (RefSeq) project. The global statistics section reported general statistics information including Gaps between scaffolds, number of scaffolds, number of contigs, total sequence length, and total ungapped length. Furthermore, taxonomic data were checked according to the best-matching-type strain with the declared new species repertory in NCBI<sup>22</sup>.

During annotation, genomic parameters were evaluated including transfer-messenger RNAs (tmRNAs) and transfer RNAs (tRNAs) using ARAGORN version 1.2 and ribosomal RNAs (rRNAs) using Barrnap version 0.9<sup>23,24</sup>. Generated file (.faa) was used for BLAST-P analyses against the Clusters of Orthologous Genes (COGs) database, and used for CRISPR-Cas identification<sup>25</sup>. Resistance genes were screened using ResFinder<sup>26</sup>. Other bioinformatic tools were also used such as AntiSMASH to search polyketide synthases (PKS) and non-ribosomal peptide synthetases (NRPS)<sup>27</sup>. Circular maps of the two genomes were generated using CGView (Circular Genome Viewer) software. This Java application converts XML or tab-delimited input into a Vector Graphics format<sup>28</sup>.

Besides, phylogenetic trees of interest were generated with the FastME 2.1.6.1 software to highlight the position of each new bacterial strain among its closest relatives<sup>29</sup>. Digital DNA-DNA Hybridization (dDDH) values were calculated to check the difference between the genomes using the following website (<https://ggdc.dsmz.de>). Critical limit was set at 70% below which a prokaryotic species may be considered as new<sup>30</sup>. Orthologous average nucleotide identity (OrthoANI) version 0.93.1 was also used to calculate genomic similarities between studied species and their related taxa.

**Ethics approval.** The study was approved by the ethics committee of the Institut de Recherche Fédératif 48 under Authorization number 09-022 with the consent of the patients.

## Results

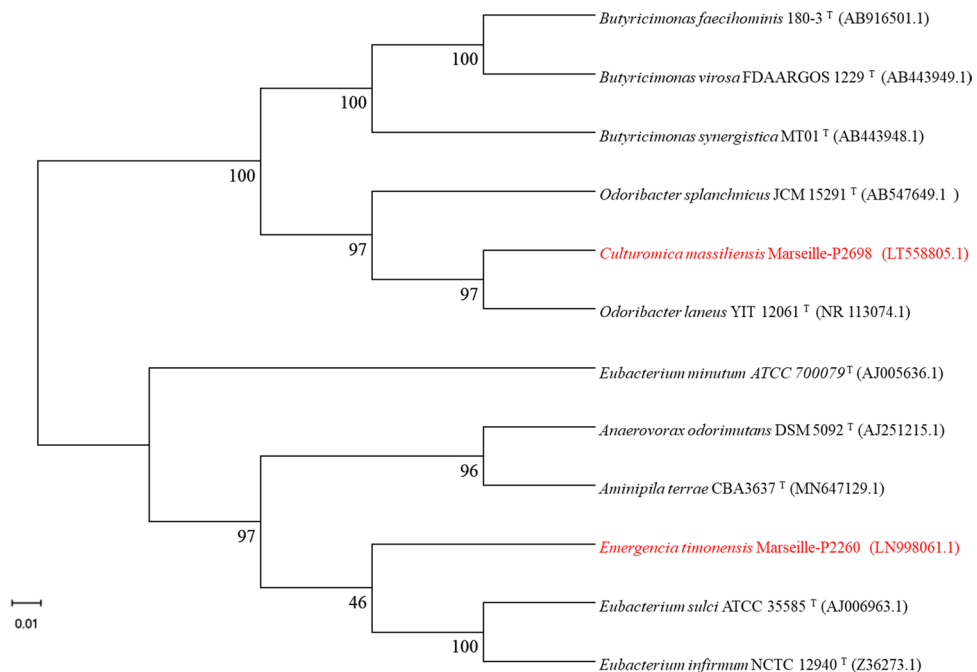
**Strain identification and phylogenetic analyses.** The species names *Culturomica massiliensis* gen. nov., sp. nov., and *Emergencia timonensis* gen. nov., sp. nov., had been previously proposed for the new species mainly as representative strains Marseille-P2698<sup>T</sup> and Marseille-P2260<sup>T</sup>, respectively<sup>7,8</sup>. As these previous descriptions were not exhaustive, we revisited the work by including phylogenetic, morphological, and genomic data.

Strain Marseille-P2698<sup>T</sup> had a 16S rRNA gene sequence similarity and a query coverage of 91.5% and 99% respectively with *Odoribacter laneus* strain YIT 12061<sup>T</sup> (Fig. 1). Strain Marseille-P2260<sup>T</sup> exhibited 16S rRNA gene similarity and a query coverage of 92.72%, and 100% respectively, with *Eubacterium sulci* strain ATCC 35585<sup>T</sup> (Fig. 1).

**Phenotypic and biochemical characterizations.** Growth of strains Marseille-P2698<sup>T</sup> and Marseille-P2260<sup>T</sup> occurred on 5% sheep blood-enriched Columbia agar (bioMérieux®), after 48 h of incubation at 37 °C in a strict anaerobic atmosphere. Optimal growth was obtained at pH 7. However, strain Marseille-P2260<sup>T</sup> did not tolerate NaCl, whereas strain Marseille-P2698<sup>T</sup> could grow with a NaCl concentration of 0.5%.

Strain Marseille-P2698<sup>T</sup> is a Gram-negative rod, strictly anaerobic, motile, and non-spore-forming with a size of 1.5–3 µm in length and 0.3 to 0.4 µm in diameter. It exhibits a positive catalase, but no oxidase activity. The colonies are circular, beige and from 0.7 to 1.2 mm in diameter.

Strain Marseille-P2260<sup>T</sup> is a Gram-positive, strictly anaerobic rod, ranging in length from 1 to 1.5 µm, and in diameter from 0.5 to 1 µm. It has no catalase or oxidase activity. Colonies of strain Marseille-P2260<sup>T</sup> are translucent with a diameter of 0.5 to 1 mm. The remaining cell characteristics of both strains compared to their closest relatives are summarized in Table 1. Using API® 50CH strips (bioMérieux®), positive reactions were obtained for both studied strains for glycerol, D-ribose, D-galactose, D-glucose, D-fructose, D-mannose, D-mannitol, D-sorbitol, N-acetylglucosamine, amygdalin, arbutin, esculin ferric citrate, salicin, D-cellobiose, D-maltose, D-lactose, D-saccharose, D-trehalose, D-melezitose, gentiobiose, D-tagatose. Using API® ZYM strips (bioMérieux®), positive activities were observed for esterase lipase (C8), leucine arylamidase, phosphatase acid, and naphthol-as-bi-phosphohydrolase for both studied strains. In contrast, phosphatase alkaline, esterase (C4), α-chymotrypsin, β-galactosidase, α-glucosidase, β-glucosidase, and N-acetyl-β-glucosaminidase were only positive for strain Marseille-P2260<sup>T</sup>. Using API® 20A strips, positive results for the two strains were obtained for D-glucose, D-mannitol, D-lactose, D-saccharose, D-maltose, salicin, esculin ferric citrate, glycerol, D-cellobiose,



**Figure 1.** Phylogenetic tree with the position of new species (strains Marseille-P2698<sup>T</sup> and Marseille-P2260<sup>T</sup>) among closely related species. The following phylogenetic tree was performed from the comparison of 16S rRNA sequences. The accession numbers of 16S rRNA gene are mentioned in parentheses. Bootstrap appears at the nodes. MUSCLE software was used to align sequences. The tree was designed with the MEGA-X software. The used methodology is the Maximum Likelihood method and Kimura 2-parameter model.

Properties	<i>Culturomica massiliensis</i> gen. nov., sp. nov. Marseille-P2698 <sup>T</sup>	<i>Odoribacter laneus</i> YIT 12061 <sup>T</sup>	<i>Odoribacter splanchnicus</i> JCM 15291 <sup>T</sup>	<i>Butyricimonas synergistica</i> MT01 <sup>T</sup>	<i>Emergencia timonensis</i> gen. nov., sp. nov. Marseille-P2260 <sup>T</sup>	<i>Eubacterium sulci</i> ATCC 35585 <sup>T</sup>	<i>Eubacterium infirmum</i> NCTC 12940 <sup>T</sup>
Gram stain	–	–	–	–	+	–	+
Cell shape	Rod	Rod	Fusiform	Rod	Rod Bacilli	Rod	Rod
Motility	+	NA	–	–	+	–	–
Cell diameter(μm)	0.7–1.2	NA	NA	1.0	0.5–1.0	1.0	1.0
Length(μm)	0.3–0.4 × 1.5–3	9.2 × 1.15	0.7 × 1.0–5.0	1.5	1 to 1.5	0.5 × 1–2	0.5 × 1–2
Endospore	–	–	–	–	–	–	–
Optimum growth temperature	37 °C	37 °C	37 °C	37 °C	37 °C	37 °C	37 °C
Oxygen tolerance	Strict anaerobic	Anaerobic	Anaerobic	Anaerobic	Strict anaerobic	Anaerobic	Strict anaerobic
Incubation	4 days	4 days	2–3 days	2 days	3 days	3 days	7 days
Salt tolerance	0.5%	NA	NA	NA	–	NA	NA
Optimum pH	7	NA	NA	6–7.5	7	NA	NA
Acid phosphatase	+	+	NA	NA	+	NA	NA
Catalase	+	NA	–	–	–	NA	–
Oxidase	–	NA	NA	–	–	NA	NA
Indole	–	+	+	+	–	–	–
Urease	–	–	–	–	–	–	–
βGalactosidase	–	+	–	–	+	–	NA
Ribose	+	NA	NA	NA	+	–	–
Mannose	+	–	+	–	+	–	NA
Mannitol	+	–	NA	–	+	–	–
Sucrose	+	–	–	–	+	–	–
Glucose	+	–	+	+	+	NA	–
Fructose	+	NA	+	NA	+	NA	NA
Maltose	+	–	NA	–	+	–	–
Sorbitol	+	–	NA	–	+	–	NA
Starch	+	NA	NA	NA	–	NA	NA
Lactose	+	–	–	–	+	–	–
Source	Human faeces in diabetic patient	Human faeces	Human Abdominal abscess	Mouse feces	Human faeces	Human gingival sulcus	human periodontal pockets

**Table 1.** Phenotypic characteristics of strains Marseille-P2698<sup>T</sup> and Marseille-P2260<sup>T</sup> compared with closely related species. + positive result. – negative result. NA not available data.

D-mannose, D-melezitose, D-sorbitol, L-rhamnose, and D-trehalose. However, strain Marseille-P2698<sup>T</sup> was positive to hydrolysis of gelatin, unlike strain Marseille-P2260<sup>T</sup> which was negative.

The most abundant fatty acids of strain Marseille-P2698<sup>T</sup> were 13-methyl-tetradecanoic acid (63%), 12-methyl-tetradecanoic acid (11%), and 3-hydroxy-15-methyl-hexadecanoic acid (8%). Several other branched structures, mainly iso, were also detected. Specific 3-hydroxy structures were detected, mainly branched as well. For strain Marseille-P2260<sup>T</sup>, the major fatty acid was hexadecanoic acid (39%) (Table 2).

**Genomic properties and analyses.** The whole genome of strain Marseille-P2698<sup>T</sup> was composed of 14 contigs, for a total size of 4,410,591 bp, with a G+C content of 43 mol% (Fig. 2). This genome contained 3679 genes, of which 3487 were protein-coding genes. In addition, 59 RNA sequences were also identified and distributed as follows: 8 rRNAs (three 16S, two 23S, and three 5S), 51 tRNAs, and 1 tmRNA.

Genome from strain Marseille-P2260<sup>T</sup> was composed of 9 contigs, with a size of 4,661,482 bp, and a 45.8 mol% G + C content (Fig. 2). Genome annotation identified 4380 genes, of which 4288 were protein-coding genes. There were 56 RNA sequences including 5 rRNAs (one 16S, one 23S, and three 5S), 51 tRNAs, and 1 tmRNA.

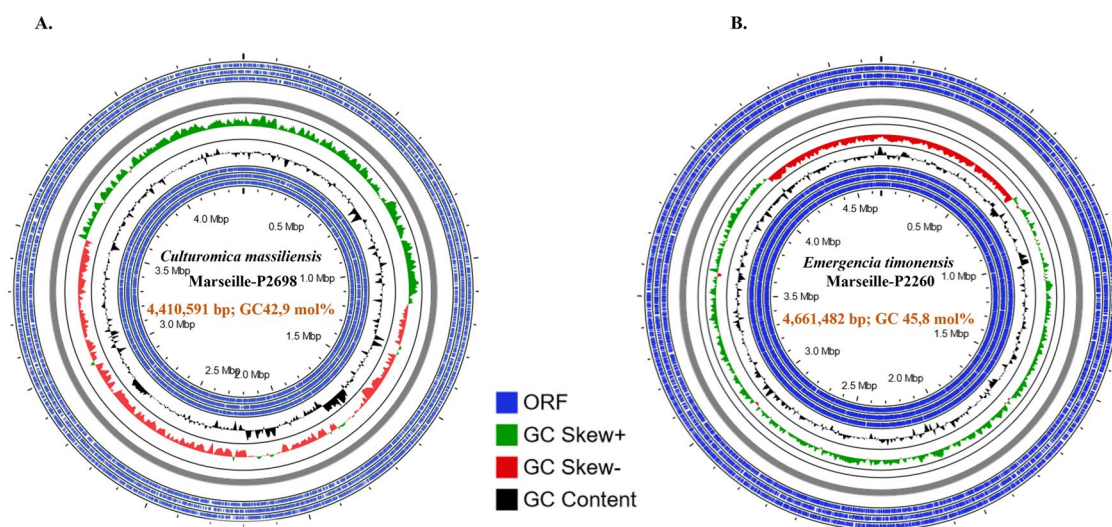
*Odoribacter laneus*<sup>T</sup>, *O. splanchnicus*<sup>T</sup>, *Butyricimonas synergistica*<sup>T</sup>, *B. faecihominis*<sup>T</sup> and *B. virosa*<sup>T</sup> exhibited genome sizes ranging from 3.77 to 4.81 Mbp. The closest bacteria of strains Marseille-P2698<sup>T</sup> and Marseille-P2260<sup>T</sup> are presented in Table 3. *Eubacterium sulci*<sup>T</sup>, *E. infirmum*<sup>T</sup>, *A. terrae*<sup>T</sup>, *E. minutum*<sup>T</sup>, and *Aminipila odorimitans*<sup>T</sup> exhibited genome sizes ranging from 1.73 to 4.66 Mbp.

Strain Marseille-P2698<sup>T</sup> shared dDDH values of 21.6% with *Odoribacter laneus*<sup>T</sup>, 20.6% with *O. splanchnicus*<sup>T</sup>, 30.6% with *Butyricimonas faecihominis*<sup>T</sup>, 21.8% with *B. virosa*<sup>T</sup>, and 18.9% with *B. synergistica*<sup>T</sup>. Strain Marseille-P2260<sup>T</sup> exhibited dDDH values of 20.1% with *Eubacterium sulci*<sup>T</sup>, 21.1% with *E. infirmum*<sup>T</sup>, 23.1% with *A. terrae*<sup>T</sup>, 24.4% with *E. minutum*, and 21.1% with *Aminipila odorimitans* (Table 4).

The phylogenetic relationships of strains Marseille-P2698<sup>T</sup> and Marseille-P2260<sup>T</sup> with relative strains, based on whole-genome sequencing, is represented in Fig. 3.

	Fatty acids	Name	Mean relative % (a)
	15:0 iso	13-methyl-tetradecanoic acid	63.4 ± 1.8
	15:0 anteiso	12-methyl-tetradecanoic acid	10.9 ± 0.2
	17:0 3-OH iso	3-hydroxy-15-methyl-Hexadecanoic acid	8.0 ± 0.7
	16:00	Hexadecanoic acid	7.8 ± 0.3
	16:0 3-OH	3-hydroxy-Hexadecanoic acid	3.1 ± 0.3
	5:0 iso	3-methyl-Butanoic acid	1.2 ± 0.0
<i>Culturomica</i>	18:2n6	9,12-Octadecadienoic acid	1.1 ± 0.1
<i>massiliensis</i>	18:1n9	9-Octadecenoic acid	1.0 ± 0.1
Strain	15:00	Pentadecanoic acid	TR
Marseille-	14:00	Tetradecanoic acid	TR
P2698 <sup>T</sup>	18:00	Octadecanoic acid	TR
	15:0 3-OH iso	3-hydroxy-13-methyl-Tetradecanoic acid	TR
	16:0 iso	14-methyl-Pentadecanoic acid	TR
	18:1n7	11-Octadecenoic acid	TR
	17:0 iso	15-methyl-Hexadecanoic acid	TR
	14:0 iso	12-methyl-Tridecanoic acid	TR
	16:0 3-OH iso	3-hydroxy-14-methyl-Pentadecanoic acid	TR
	17:00	Heptadecanoic acid	TR
	15:00	3-OH anteiso 3-hydroxy-12-methyl-Tetradecanoic acid	TR
	16:00	Hexadecanoic acid	39.2 ± 2.4
	18:1n9	9-Octadecenoic acid	16.5 ± 1.3
	18:1n7	11-Octadecenoic acid	14.5 ± 4.8
<i>Emergencia</i>	14:00	Tetradecanoic acid	11.9 ± 0.6
<i>timonensis</i>	18:00	Octadecanoic acid	9.5 ± 1.4
Strain	18:1n3	15-Octadecenoic acid	5.6 ± 0.9
Marseille-	18:2n6	9,12-Octadecadienoic acid	1.1 ± 0.3
P2260 <sup>T</sup>	15:00	Pentadecanoic acid	TR
	15:0 anteiso	12-methyl-tetradecanoic acid	TR
	12:00	Dodecanoic acid	TR

**Table 2.** Cellular fatty acids composition of strains Marseille-P2698<sup>T</sup> and Marseille-P2260<sup>T</sup>. a: Mean peak area percentage. TR trace amount.



**Figure 2.** Circular genome map of strains Marseille-P2698<sup>T</sup> (left) and Marseille-P2260<sup>T</sup> (right) generated by the CGView software. From outside to the center: blue rings demonstrate the ORFs (Open Reading Frames) on both forward and reverse strands, green and red rings represent both positive and negative GC skew respectively, and black ring represents the GC content plot.

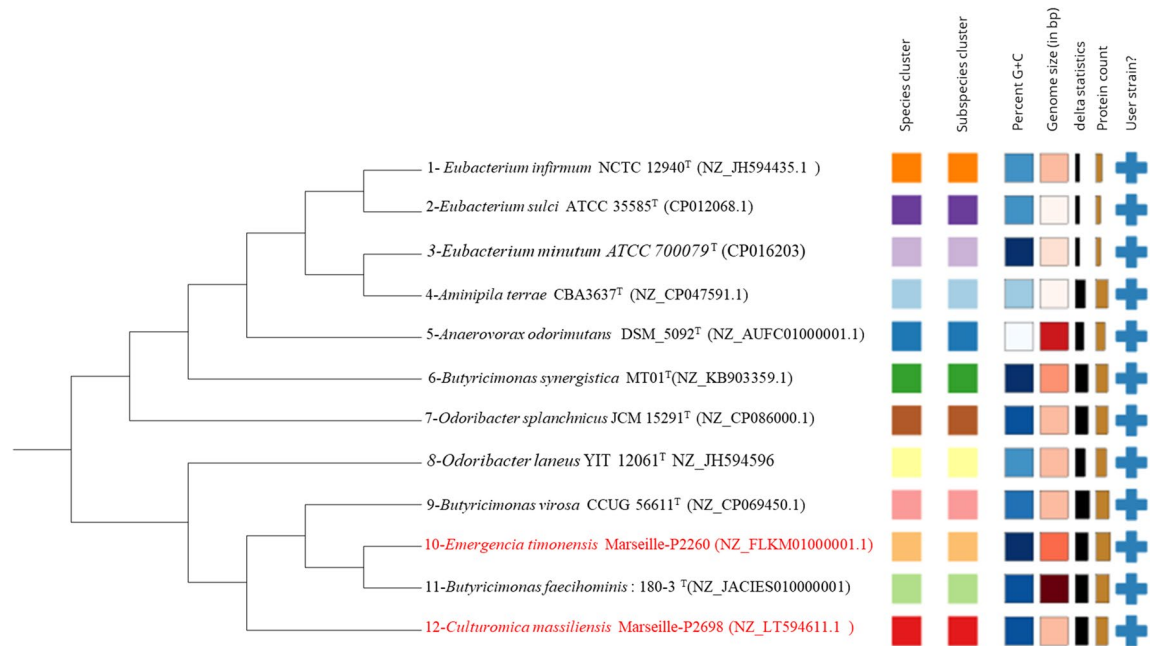
	Organisms	INSDC/RefSeq	Size (M bp)	G + C %	Protein-coding genes	rRNA	tRNA	tmRNA	genes	Genes assigned to COGs	Contigs	Bases	CDS
	<b>Strain Marseille-P2698<sup>T</sup></b>	FLSN01000000	4.41	43	3487	8	50	1	3679	2113	11	4410591	3620
<i>Culturomica</i>	<i>Odoribacter laneus</i> YIT 12,061 <sup>T</sup>	NZ_JH594596	3.77	40.6	2993	10	53	3	3100	2995	42	3,794,909	3095
<i>massiliensis</i>	<i>Odoribacter splanchnicus</i> JCM 15291 <sup>T</sup>	CP086000.1	4.39	43.4	3545	12	59	1	3757	2106	1	4,393,364	3685
strain	<i>Butyricimonas synergistica</i> MT01 <sup>T</sup>	NR_041690.1	4.77	43.8	3741	8	54	1	3922	2113	16	4,770,838	3859
Marseille-P2698 <sup>T</sup>	<i>Butyricimonas faecihominis</i> 180-3 <sup>T</sup>	NR_126194.1	4.79	42.9	3832	4	57	1	4017	2156	29	4,793,705	3955
	<i>Butyricimonas virosa</i> CCUG 56611 <sup>T</sup>	CP069450.1	4.81	42.4	3837	15	57	1	4032	2169	1	4,813,143	3959
	<b>Strain Marseille-P2260<sup>T</sup></b>	FLKM01000000	4.66	45.8	4288	5	50	1	4380	2895	9	4,661,482	4324
<i>Emergencia</i>	<i>Eubacterium sulci</i> ATCC 35585 <sup>T</sup>	CP012068.1	1.73	39.9	3062	6	40	1	1646	1235	1	1,739,380	1599
<i>timonensis</i>	<i>Eubacterium infirmum</i> NCTC 12940 <sup>T</sup>	NZ_JH594435.1	1.91	39.9	3510	4	43	1	1781	1305	13	1,910,927	1733
strain	<i>Aminipila terrae</i> CBA3637 <sup>T</sup>	CP047591.1	3.51	37	3102	18	65	1	3581	2346	1	3,512,404	3497
Marseille-P2260 <sup>T</sup>	<i>Eubacterium minutum</i> ATCC 700,079 <sup>T</sup>	CP016203	1.9	45.79	1454	7	43	1	1620	1035	124	1,903,428	1246
	<i>Anaerovorax odorimutans</i> DSM_5092 <sup>T</sup>	NZ_AUFC01000001.1	3.26	31.5	2894	4	42	1	2944	1976	54	3,263,520	2897

**Table 3.** Summary of comparative genomes and characteristics for strains Marseille-P2698<sup>T</sup> and Marseille-P2260<sup>T</sup>.

	CM	OL	OS	BS	BF	BV	ET	EM	AT	AO	ES	EI
CM	100%	21.6% (± 4.7%)	20.6 (± 4.7%)	18.9 (± 4.5%)	30.6 (± 4.9%)	21.8 (± 4.7%)	17.8 (± 4.4%)	16 (± 4.4%)	17.1 (± 4.4%)	21.5 (± 4.7%)	18.3 (± 4.5%)	18.2 (± 4.5%)
OL		100%	13.3 (± 6.5%)	18.6 (± 4.66%)	18.5 (± 4.5%)	19.9 (± 4.6%)	18.1 (± 4.5%)	17.7 (± 4.5%)	18.1 (± 4.5%)	21 (± 4.7%)	18.2 (± 4.6%)	18.1 (± 4%)
OS			100%	23.7 (± 4.8%)	19.2 (± 4.6%)	23 (± 4.8%)	19.3 (± 4.5%)	18.5 (± 4.5%)	19.1 (± 4.6%)	21.5 (± 4.7%)	19.1 (± 4.6%)	17.3 (± 4.5%)
BS				100%	21 (± 4.7%)	20.9 (± 4.7%)	19.6 (± 4.6%)	20.9 (± 4.7%)	20.6 (± 4.6%)	22.9 (± 4.8%)	20.9 (± 4.7%)	18.1 (± 4.5%)
BF					100%	42.2 (± 5.1%)	43.4 (± 5.1%)	3.7 (± 2%)	3.7 (± 2%)	30.4 (± 4.9%)	3.7 (± 2%)	3.7 (± 2%)
BV						100%	18 (± 4.5%)	17.7 (± 4.5%)	17.9 (± 4.5%)	21 (± 4.7%)	18.3 (± 4.5%)	18.3 (± 4.6%)
ET							100%	24.4 (± 4.8%)	23.1 (± 4.8%)	21.1 (± 4.7%)	20.1 (± 4.6%)	21.1 (± 4.6%)
EM								100%	29.5 (± 4.9%)	31.2 (± 4.9%)	25.4 (± 4.8%)	24.9 (± 4.8%)
AT									100%	23 (± 4.8%)	28.8 (± 4.9%)	24.4 (± 4.8%)
AO										100%	19.8 (± 4.6%)	23.4 (± 4.8%)
ES											100%	28.9 (± 4.9%)
EI												100%

**Table 4.** Comparative digital DNA-DNA Hybridization (dDDH) values (%) between studied bacterial genomes.

CM. *Culturomica massiliensis* gen. nov., sp. nov., Marseille-P2698<sup>T</sup>. OL. *Odoribacter laneus*<sup>T</sup>. OS. *Odoribacter splanchnicus*<sup>T</sup>. BS. *Butyricimonas synergistica*<sup>T</sup>. BF. *Butyricimonas faecihominis*<sup>T</sup>. BV. *Butyricimonas virosa*<sup>T</sup>. ET. *Emergencia timonensis* gen. nov., sp. nov., Marseille-P2260<sup>T</sup>. EM. *Eubacterium minutum*<sup>T</sup>. AT. *Aminipila terrae*<sup>T</sup>. AO. *Anaerovorax odorimutans*<sup>T</sup>. ES. *Eubacterium sulci*<sup>T</sup>. EI. *Eubacterium infirmum*<sup>T</sup>.

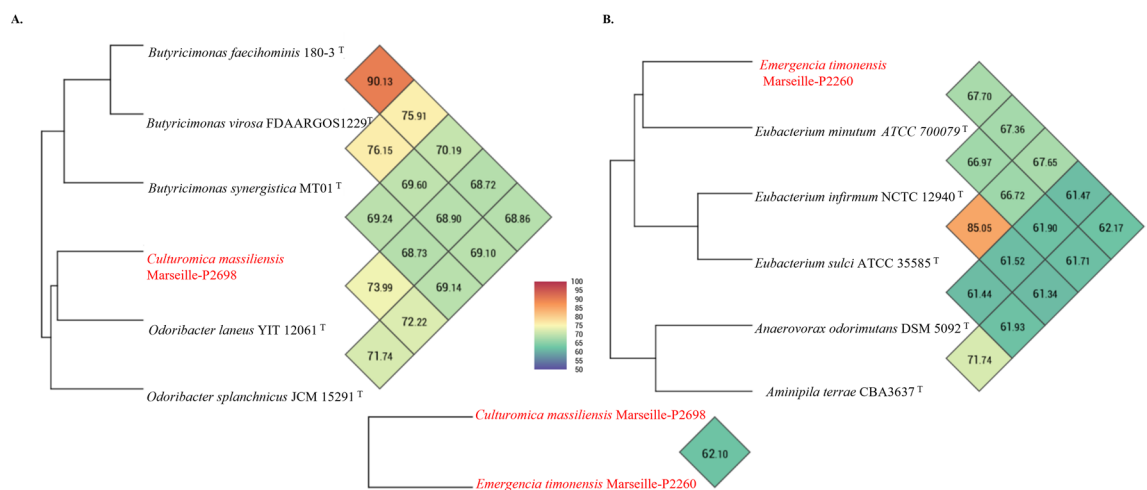


**Figure 3.** Whole-Genome sequence-based phylogenetic tree of strains Marseille-P2698<sup>T</sup> and Marseille-P2260<sup>T</sup> using TYGS design tree with FastME 2.1.4 software based on Genome BLAST Distance Phylogeny (GBDP) parameters. Distances were calculated from the genome sequences, and branch lengths were calculated by GBDP distance formula d5.

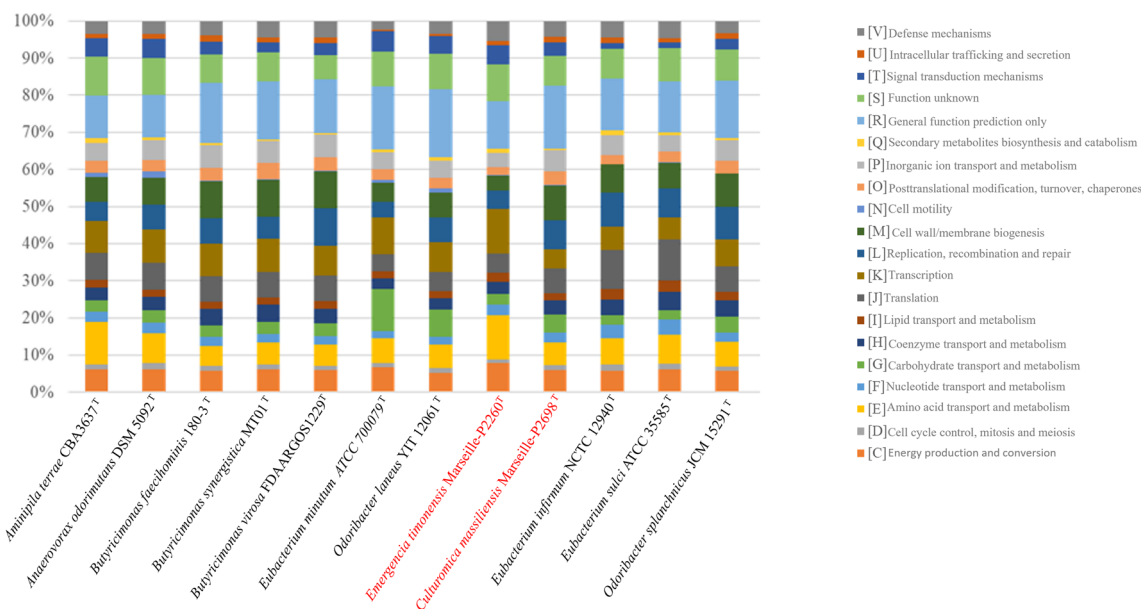
The obtained dDDH values were lower than the 70% threshold used for delineating prokaryotic species<sup>4</sup>. In addition, strain Marseille-P2698<sup>T</sup> exhibited OrthoANI values of 73.99% with *O. laneus*<sup>T</sup>, 72.22% with *O. splanchnicus*<sup>T</sup>, and 69.24% with *B. synergistica*<sup>T</sup>. Strain Marseille-P2260<sup>T</sup> had an OrthoANI value of 67.65% with *E. sulci*<sup>T</sup> and 67.36% with *E. infirmum*<sup>T</sup>. These values were lower than 95%, also suggesting that strains Marseille-P2698<sup>T</sup> and Marseille-P2260<sup>T</sup> belonged to distinct species (Fig. 4).

Genes encoding proteins are divided into several categories according to their functions and others have unknown functions. COGs of strains Marseille-P2698<sup>T</sup> and Marseille-P2260<sup>T</sup> have different functional distributions from their closest species (Fig. 5).

**Antibiotic resistance genes and defense mechanisms.** Using the ResFinder software, antibiotic resistance genes (ARG) *erm* (F) and *tet* (Q) were detected within the genome of strain Marseille-P2698<sup>T</sup>, with identity percentages of 100% and 99.8%, respectively. For strain Marseille-P2260<sup>T</sup>, ARG included *erm* (B), *tet* (M), and *tet* (O), with identity percentage ranging from 99.7% to 100% (Table 5).



**Figure 4.** Heatmap and phylogenetic trees showing the average nucleotide identity based on calculated orthology (OrthoANI) of Marseille-P2698<sup>T</sup> (A) and Marseille-P2260<sup>T</sup> (B) compared to their closest related bacterial species.



**Figure 5.** Distribution of functional classes of predicted genes expressing group of proteins that clusters according to functions.

	Resistance gene	Identity	Alignment Length/Gene Length	Position in contig	Phenotype
	<b>Macrolide</b>				
	erm(F)	100	801/801	1,826,565.0.1827365	erythromycin,lincomycin,clindamycin,quinupristin,pristinamycin ia,virginiamycin s
	<b>Lincosamide</b>				
Culturomica	erm(F)	100	801/801	1,826,565.0.1827365	erythromycin,lincomycin,clindamycin,quinupristin,pristinamycin ia,virginiamycin s
massiliensis	<b>Streptogramin b</b>				
Marseille-P2698 <sup>T</sup>	erm(F)	100	801/801	1,826,565.0.1827365	erythromycin,lincomycin,clindamycin,quinupristin,pristinamycin ia,virginiamycin s
	<b>Tetracycline</b>				
	tet(Q)	99.8	1926/1926	1,055,038.0.1056963	doxycycline,tetracycline,minocycline
	<b>Macrolide</b>				
	erm(B)	100	738/738	1,359,368.0.1360105	erythromycin,lincomycin,clindamycin,quinupristin,pristinamycin ia,virginiamycin s
Emergencia	<b>Lincosamide</b>				
timonensis	erm(B)	99.86	738/738	1,359,368.0.1360105	erythromycin,lincomycin,clindamycin,quinupristin,pristinamycin ia,virginiamycin s
Marseille-P2260 <sup>T</sup>	<b>Streptogramin b</b>				
	erm(B)	99.9	738/738	1,359,368.0.1360105	erythromycin,lincomycin,clindamycin,quinupristin,pristinamycin ia,virginiamycin s
	<b>Tetracycline</b>				
	tet(M)	99.84	1920/1920	1,222,665.0.1224584	doxycycline,tetracycline,minocycline
	tet(O)	99.79	1920/1920	2,008,154.0.2010073	doxycycline,tetracycline,minocycline

**Table 5.** Antibiotic resistance genes detected in the genomes strains Marseille-P2698<sup>T</sup> and Marseille-P2260<sup>T</sup> using ResFinder software.

The genomes of strains Marseille-P2698<sup>T</sup> and Marseille-P2260<sup>T</sup> contained CRISPR-Cas modules (Table 6). Strain Marseille-P2698<sup>T</sup> contains one defense mechanism composed of polyketide synthases (PKS) and non-ribosomal peptide synthetases (NRPS) enzymes. NRPS-PKS was previously demonstrated to have a role in the biosynthesis of pharmaceutically-important natural products<sup>31</sup>. Using ANTI SMASH software, NRPS-PKS-like genes cluster had been detected in genome of strain Marseille-P2698<sup>T</sup>, and β-lactone; containing potent medicinal properties<sup>32, 33</sup>, had been detected in strain Marseille-P2698<sup>T</sup> genome. In addition, according to our results in ToxFinder-1.0 software detecting toxins<sup>34</sup>, our two strains did not possess toxin genes.



Bacteria	Element	CRISPR Id /gene name	Start	End	Spacer /Gene	Repeat consensus /cas genes	Orientation
	CRISPR		112	199	1	ATGGGACTCTTTTGTGAAATA	
	Cas cluster	CAS	71,544	73,337	1	Cas3_0_I	+
	Cas cluster	CAS-TypeIB	613,607	622,650	7	Cas6_0_I-III, Cas7_3_IB, Cas5_1_IB, Cas3_0_I,	+
		Cas6_0_I-III				Cas4_0_I-II, Cas1_0_I-II-III, Cas2_0_I-II-III-V	
<i>Culturomica</i>	CRISPR	Cas7_3_IB	622,870	625,330	37	CTTTAATTGAACTAAGGTAGAATTGAAAC	+
<i>massiliensis</i>	CRISPR	Cas5_1_IB	992,376	993,764	21	GTTTCAATACTACTTAGTCTATTAAAAG	+
Marseille-P2698 <sup>T</sup>	CRISPR	Cas3_0_I	2,048,339	2,049,405	15	GTCTCAATGCCGTGATTACAGCTATTCTCTAAATC	+
	Cas cluster	Cas4_0_I-II	2,049,852	2,057,279	6	Cas2_0_I-II-III-V, Cas1_0_I-II-III, Cmr6_0_IIIb,	+
						Cmr4_0_IIIb, Cmr3_0_IIIb, Cas10_1_IIIb	
	CRISPR	Cas1_0_I-II-III	3,461,867	3,462,226	5	ATTTCAATTCTACTCCAGTTCTATTAAAAT	+
	CRISPR	Cas2_0_I-II-III-V	3,750,732	3,750,812	1	AATTCTCTTTAGCGTTGGTTTGTG	+
	CRISPR		882	1086	2	GGCATATAATTATCAATACAGTATTTCCCTATTTT	
	CRISPR		300,401	300,525	1	TTGCGGCAACGCCGTGAGATTTTGAATCATTATA	+
	CRISPR		522,369	522,503	1	TTGCAGCAACGCTGTCAATATCGTTGTTCATTAT	+
						ACCATGA	
	CRISPR		1,380,564	1,381,183	9	GTTCGCATCCTCATGGATGCGTGGATTGAAAT	-
	CRISPR	CAS-TypeIC	1,429,920	1,430,223	4	GTCTTGCTCCGCATGGAGCAAGTGGATTGAAAT	+
	Cas cluster		1,453,194	1,461,082	7	Cas3_0_I, Cas5_0_IC, Cas8c_0_IC, Cas7_0_IC, Cas4_0_I-	
						II, Cas1_0_IC, Cas2_0_I-II-III-V	
<i>Emergencia</i>	CRISPR		1,461,248	1,463,328	30	GTCCCACTCCGCATGGAGTGCCTGGATTGAAAT	+
<i>timonensis</i>	CRISPR		1,608,002	1,608,146	1	AAAGCATAAGGACTTTCATCAATTTGTTTCAAAAT	-
						TGTGACAAGTCCATTA	
Marseille-P2260 <sup>T</sup>	CRISPR		1,944,682	1,944,814	1	GGTATAATGAATCCATAACCAACGACGTTGTCGCA	+
						AGGTAG	
	CRISPR		225,209	225,286	1	ATCCAAGCCAATTCTTATCGAGGCT	-
	CRISPR		530,585	530,712	1	TATTTTCAGCGGAATGCACATGGTATAATAACGCA	-
	CRISPR		853,819	853,957	1	TGGTATAATGGTTCCACGATCTTACGGCGTTGCCGA	+
						AGTGTAGGAACCA	
	CRISPR		1,673,013	1,673,146	1	CATTATACGACAAAACGACCCAAATCAGAGATTTG	+
						GGCGTT	
	CRISPR		1,798,696	1,798,841	1	TCCTTATGCTTAAACGCCTAAATCTGTGATTAGC	-
						TGCGTTTTGTCGTATAA	
	CRISPR		1,910,685	1,910,831	1	AATGAACGCTTCCTTGTGACGAATCTGATTAAGAT	-
						TGTGCGTTTATTATACCA	
	CRISPR		1,927,335	1,927,459	1	CCGTTTCATCTTTCGGCGTTGCCGCAAGGTAGGAA	+
	CRISPR		2,025,969	2,026,111	1	TGTTCTGCTAACTGACTCCATATCCAGCCCTCTGT	+
						GTTACGCAGATTT	

**Table 6.** Identification of CRISPR-Cas type and subtype genes in user-submitted sequence for strains Marseille-P2698<sup>T</sup> and Marseille-P2260<sup>T</sup>.

## Conclusion

Culturomics combined with Taxogenomics allowed the isolation and the full characterization of two new bacterial species isolated from human intestine. Primitive phylogenetic comparisons were previously performed<sup>7,8</sup>, while a thorough phylogenetic and genomic analyses were performed in our study. The results of phenotypic, biochemical, phylogenetic, and genomic analyses obtained for both studied strains proved that they belong to new species. Thus, we propose the creation of two new bacterial genera and new species, *Culturomica massiliensis* gen. nov., sp. nov., and *Emergencia timonensis* gen. nov., sp. nov.

**Description of *Culturomica* gen. nov.** *Culturomica* (*Cul.tu.ro.mi'ca*. N.L. fem, *culturomica*, referring to a new method of diversified bacterial culture). Cells are anaerobic, Gram-negative, motile and rod-shaped. The type species is *Culturomica massiliensis* gen. nov., sp. nov.. It was isolated from human feces.

**Description of *Culturomica massiliensis* gen. nov., sp. nov.** *Culturomica massiliensis* (*mas.si.li.en'sis*. L. fem. adj. *massiliensis*, form “*Massilia*”, the Latin name of Marseille, France, where the type strain was isolated). Bacterial cells are Gram-negative and rod-shaped. They are motile and non-spore-forming, with lengths ranging from 1.5 to 3.0 µm, and diameters from 0.3 to 0.4 µm. The optimal growth conditions are 37 °C, a strict anaerobic atmosphere, and neutral pH. On 5% sheep blood-enriched Columbia agar, colonies appear small, circular, beige to white, and measure between 0.7 and 1.2 mm in diameter. They ferment D-galactose, D-glucose, D-fructose, D-mannose, inositol, D-mannitol, D-sorbitol, N-acetylglucosamine, amygdalin, arbutin, esculin, salicin, D-cellobiose, D-maltose, D-glucose, glycerol, D-lactose, D-saccharose, D-trehalose, D-melezitose,

Amidon, gentiobiose, D-tagatose, potassium gluconate and potassium 5-ketogluconate. Enzymatic activities for phosphatase al-kaline, esterase, esterase lipase, leucine arylamidase,  $\alpha$ -chymotrypsin, phosphatase acid, naphthol-as-bi-phosphohydrolase,  $\beta$ -galactosidase,  $\alpha$ -glucosidase,  $\beta$ -glucosidase, and N-acetyl- $\beta$ -glucosaminidase are present. The major cell wall fatty acids are C<sub>15:0 iso</sub> (63%), C<sub>15:0 anteiso</sub> (11%), and C<sub>17:0 3-OH iso</sub> (8%). The genome size from strain Marseille-P2698<sup>T</sup> is 4.41 Mbp-long with a G+C content of 43 mol%.

The type strain Marseille-P2698<sup>T</sup> (CSUR P2698 = DSM 103,121) was isolated from a stool specimen of a 66-year-old man with diabetes mellitus.

The accession numbers for the genomic and 16S rRNA gene sequences of strain Marseille-P2698<sup>T</sup> are deposited in the GenBank database under references FLSN00000000 and LT558805, respectively.

**Description of *Emergencia* gen. nov.** *Emergencia timonensis* (*e.mer.gencia* N.L. fem. n., *Emergencia*, for emergence, in reference to the discovery of emerging human bacteria).

**Description of *Emergencia timonensis* gen. nov., sp. nov.** *Emergencia timonensis* (*ti.mo.nen'sis*. L. fem. adj., *timonensis*, from Timone, the name of a university hospital in Marseille, France, where the type strain was isolated).

Bacterial cells are Gram-positive rod-shaped and bacilli. They are motile, non-spore-forming, with lengths ranging from 1.0 to 1.5  $\mu$ m, and a diameter of 0.5 to 1.0  $\mu$ m. The optimal growth conditions are 37 °C, a strict anaerobic atmosphere, and pH.7. On 5% sheep, blood-enriched Columbia agar, colonies of strain Marseille-P2260<sup>T</sup> appear small, translucent, and measure from 0.7 to 1.2 mm in diameter. Cells ferment D-glucose, D-mannitol, D-lactose, D-saccharose, D-maltose, salicin, esculin ferric citrate, glycerol, D-cellobiose, D-mannose, D-melezitose, D-sorbitol, L-rhamnose, D-trehalose, glycerol, D-arabinose, L-arabinose, D-ribose, D-xylose, D-galactose, D-glucose, D-fructose, D-mannose, L-rhamnose, dulcitol, D-mannitol, D-sorbitol, Methyl- $\alpha$ -glucopyranoside, N-acetylglucosamine, amygdalin, arbutin, esculin, salicin, D-cellobiose, D-maltose, D-lactose, D-melibiose, D-saccharose, D-trehalose, D-melezitose, D-raffinose, gentiobiose, D-turanose, D-tagalose, L-fucose, and potassium gluconate. In addition, enzymatic activities such as esterase lipase, leucine arylamidase, acid phosphatase, and naphthol phosphohydrolase are present. The major cell wall fatty acids are C<sub>16:00</sub> (39%), C<sub>18:1n9</sub> (16%), and C<sub>18:1n7</sub> (14%).

The genome size from strain Marseille-P2260<sup>T</sup> is 4.66 Mbp with a G+C content of 45.8 mol%.

The type strain Marseille-P2260<sup>T</sup> (CSUR P2260 = DSM 101,844 = SN18) was isolated from a stool sample the feces of a healthy patient with an unremarkable medical history.

## Data availability

The accession numbers for the genomic and 16S rRNA gene sequences are deposited in the GenBank database under references: FLSN00000000 and LT558805 respectively for Marseille-P2698<sup>T</sup> strain, and FLKM00000000 and LN998061 respectively for Marseille-P2260<sup>T</sup> strain.

Received: 13 October 2022; Accepted: 18 May 2023

Published online: 25 May 2023

## References

- Sankar, S. A., Lagier, J. C., Pontarotti, P., Raoult, D. & Fournier, P. E. The human gut microbiome, a taxonomic conundrum. *Syst. Appl. Microbiol.* **38**(4), 276–286. <https://doi.org/10.1016/J.SYAPM.2015.03.004> (2015).
- Fontaine, B., Perrard, A. & Bouchet, P. 21 years of shelf life between discovery and description of new species. *Curr. Biol.* **22**(22), R943–R944. <https://doi.org/10.1016/J.CUB.2012.10.029> (2012).
- Lagier, J. C. *et al.* Current and past strategies for bacterial culture in clinical microbiology. *Clin. Microbiol. Rev.* **28**(1), 208–236. <https://doi.org/10.1128/CMR.00110-14> (2015).
- Lagier, J. C. *et al.* Microbial culturomics: Paradigm shift in the human gut microbiome study. *Clin. Microbiol. Infect.* **18**(12), 1185–1193. <https://doi.org/10.1111/1469-0691.12023> (2012).
- Lagier, J. C. *et al.* The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. *Clin. Microbiol. Rev.* **28**(1), 237–264. <https://doi.org/10.1128/CMR.00014-14/ASSET/EC9AA70E-2915-4655-BCD0-6A04B11458FC/ASSETS/GRAPHIC/ZCM0011524990005.JPEG> (2015).
- Abdallah, R. A. *et al.* The impact of culturomics on taxonomy in clinical microbiology. *Antonie Van Leeuwenhoek* **110**(10), 1327–1337. <https://doi.org/10.1007/S10482-017-0871-1> (2017).
- Bessis, S., Ndongo, S., Lagier, J. C., Fournier, P. E. & Raoult, D. “Emergencia timonensis”, a new bacterium isolated from the stool of a healthy patient. *New Microbes New Infect.* **12**, 73–75. <https://doi.org/10.1016/J.NMNI.2016.05.001> (2016).
- Ndongo, S., Lagier, J.-C., Fournier, P.-E., Raoult, D. & Khelafia, S. NEW SPECIES “*Culturomica massiliensis*”, a new bacterium isolated from the human gut. *New Microbes New Infect.* **13**, 60–61. <https://doi.org/10.1016/j.nmni.2016.06.010> (2016).
- Ndongo, S. *et al.* Genome analysis and description of *Xanthomonas massiliensis* sp. nov., a new species isolated from human faeces. *New Microbes New Infect.* **26**, 63. <https://doi.org/10.1016/J.NMNI.2018.06.005> (2018).
- Ndongo, S. *et al.* *Vitreoscilla massiliensis* sp. nov., isolated from the stool of an Amazonian patient. *Curr. Microbiol.* **78**(8), 3313–3320. <https://doi.org/10.1007/S00284-021-02577-8> (2021).
- Zgheib, R. *et al.* *Buttiauxella massiliensis* sp. nov., isolated from a human bone infection. *Curr. Microbiol.* <https://doi.org/10.1007/S00284-021-02714-3> (2022).
- Seng, P. *et al.* Ongoing revolution in bacteriology: Routine identification of bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin. Infect. Dis.* **49**(4), 543–551. <https://doi.org/10.1086/600885> (2009).
- BLAST: Basic Local Alignment Search Tool. Accessed August 29, 2022. <https://blast.ncbi.nlm.nih.gov/Blast.cgi>.
- Parte, A. C., Carbasse, J. S., Meier-Kolthoff, J. P., Reimer, L. C. & Göker, M. List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* **70**(11), 5607–5612. <https://doi.org/10.1099/IJSEM.0.004332> (2020).
- Kumar, S., Stecher, G., Li, M., Niyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**(6), 1547. <https://doi.org/10.1093/MOLBEV/MSY096> (2018).
- BioMérieux France: *Diagnostic Microbiologique Clinique ou Industriel*. Accessed February 3, 2022. <https://www.biomerieux.fr/>

17. Dione, N. *et al.* Genome sequence and description of *Anaerobaculum massiliensis* sp. nov. *New Microbes New Infect.* **10**, 66–76. <https://doi.org/10.1016/j.nmni.2016.01.002> (2016).
18. Sasser M. *Culturing of the Bacteria* (Microb ID, 2006).
19. Anani, H. *et al.* *Colibacter massiliensis* gen. nov. sp. nov., a novel Gram-stain-positive anaerobic diplococcal bacterium, isolated from the human left colon. *Sci. Rep.* <https://doi.org/10.1038/S41598-019-53791-1> (2019).
20. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**(8), 1621–1624. <https://doi.org/10.1038/ismej.2012.8> (2012).
21. Pribelski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A. Using SPAdes De Novo Assembler. *Curr. Protoc. Bioinform.* <https://doi.org/10.1002/CPBI.102> (2020).
22. Home - Assembly - NCBI. Accessed February 4, 2022. <https://www.ncbi.nlm.nih.gov/assembly/>
23. Hyatt, D. *et al.* Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**(1), 1–11. <https://doi.org/10.1186/1471-2105-11-119/TABLES/5> (2010).
24. Laslett, D. & Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **32**(1), 11–16. <https://doi.org/10.1093/NAR/GKH152> (2004).
25. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**(1), 33–36. <https://doi.org/10.1093/NAR/28.1.33> (2000).
26. Bortolaia, V. *et al.* ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.* **75**(12), 3491–3500. <https://doi.org/10.1093/jac/dkaa345> (2020).
27. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**(14), 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153> (2014).
28. Stothard, P., Grant, J. R. & Van Domselaar, G. Visualizing and comparing circular genomes using the CGView family of tools. *Brief Bioinform.* **20**(4), 1576–1582. <https://doi.org/10.1093/bib/bbx081> (2018).
29. Meier-Kolthoff, J. P. & Göker, M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* <https://doi.org/10.1038/S41467-019-10210-3> (2019).
30. 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 Bioinform.* **14**(1), 1–14. <https://doi.org/10.1186/1471-2105-14-60/TABLES/2> (2013).
31. Akey, D. L., Gehret, J. J., Khare, D. & Smith, J. L. Insights from the Sea: Structural biology of marine polyketide synthases. *Nat. Prod. Rep.* **29**(10), 1038. <https://doi.org/10.1039/C2NP20016C> (2012).
32. Robinson, S. L. & Wackett, L. P. Rings of power: Enzymatic routes to  $\beta$ -Lactones. *Compr. Nat. Prod. III* **2020**, 323–345. <https://doi.org/10.1016/B978-0-12-409547-2.14634-7> (2020).
33. Wang, J., Shi, Y. & Jiang, D.  $\beta$ -lactone derivatives and their anticancer activities: A short review. *Curr. Top. Med. Chem.* **21**(18), 1645–1656. <https://doi.org/10.2174/1568026621666210402142150> (2021).
34. ResFinder 4.1. Accessed May 31, 2022. <https://cge.food.dtu.dk/services/ResFinder/>

## Author contributions

Conceptualization and methodology: D.R., P.E.F., F.F., L.H.; Software and analysis: A.H.; Validation: L.H., P.E.F., C.I.L.; Investigation: A.H., R.M.W.; Writing-original draft preparation: A.H., R.M.W., P.E.F., L.H.; Supervision, L.H., P.E.F.; Project administration, P.E.F., D.R.; Funding acquisition, D.R. All authors have read and agreed to the published version of the manuscript.

## Funding

This work was supported by the French Government under the « Investissements d'avenir » (Investments for the Future) program managed by the Agence Nationale de la Recherche (ANR, fr: National Agency for Research), (reference: Méditerranée Infection 10-IAHU-03). This work was also supported by Région Provence Alpes Côte d'Azur and European funding FEDER PRIMI.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to L.H.

**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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023