SCIENTIFIC REP🞂RTS

OPEN

Draft genome and description of *Merdibacter massiliensis* gen.nov., sp. nov., a new bacterium genus isolated from the human ileum

Hussein Anani[1], Rita Abou Abdallah[1], Nisrine Chelkha[2], Anthony Fontanini[2], Davide Ricaboni[2], Morgane Mailhe[2], Didier Raoult[2,3] & Pierre-Edouard Fournier [ID][1]

We used phenotypic, genomic and phylogenetic information following the taxono-genomics approach to demonstrate that strain Marseille–P3254, isolated from an ileal sample of a 76-year old woman who underwent upper and lower digestive tract endoscopy for esophagitis and colonic polyp, is representative of a novel bacterial genus within the family *Erysipelotrichaceae* in the phylum *Firmicutes*. It is an anaerobic Gram-negative bacterium without catalase and oxidase activities. The genome of strain Marseille–P3254 is 2,468,496-bp long with a 40.1% G + C content. This new bacterium is most closely related to *Eubacterium dolichum*, with which it shares 90.7% 16S rRNA sequence similarity. In addition, genomic comparison using the digital DNA–DNA hybridization and OrthoANI analyses between the novel organism and the *E. dolichum* type strain revealed identities of 25.2 and 68.91%, respectively. The major fatty acids were $C_{16:0}$, $C_{18:1n9}$ and $C_{18:0}$. Based on these data, we propose the creation of the new genus *Merdibacter* gen. nov., with strain Marseille-P3254^T (=CSUR P3254 = DSM 103534) being the type strain of the new species *Merdibacter massiliensis* gen. nov., sp. nov.

A thorough knowledge of the gut microbiota composition appears essential to understand many aspects of health and diseases in humans. Culturomics, a new approach to study and decipher the human microbiota, based on the diversification of culture conditions and complementary to 16S rRNA metagenomics, has enabled the culture of more than 500 new bacterial species to date[1–3]. Most of these new species were characterized using the taxono-genomics strategy that combines phenotypic characteristics and whole genome sequencing analysis[4].

In 2016, we isolated the new bacterial strain Marseille-P3254 (=CSUR P3254 = DSM 103534), from an ileal sample of a 76-year-old patient who underwent upper and lower digestive tract endoscopy[5] for esophagitis and colonic polyp. Matrix-assisted desorption ionization–time of flight mass spectrometry (MALDI-TOF MS)[6] failed to identify this isolate at the species level. The strain, predicted to be affiliated with members of the family *Erysipelotrichaceae*, is most closely related to *Eubacterium dolichum* with which it exhibits a 16S rRNA sequence similarity of 90.7% (Fig. 1). The family *Erysipelotrichaceae* is composed of 13 genera with validly published names (*Allobaculum, Bulleidia, Catenibacterium, Catenisphaera, Coprobacillus Eggerthia, Erysipelothrix, Faecalicoccus, Faecalitalea, Holdemanella, Holdemania, Solobacterium* and *Turicibacter*)[7]. *Eubacterium dolichum*[8] was initially described as a member of the family *Eubacteriaceae*[9]. However, a recent study has questioned this classification and proposed that this species be classified within a new genus in the family *Erysipelotrichaceae*[10]. The aim of the current study was to describe and determine the exact taxonomic position of strain Marseille-P3254, for which the name *Merdibacter massiliensis* gen. nov., sp. nov. is proposed, on the basis of a polyphasic characterization of its phenotypic properties and genome comparison with closely related bacterial taxa.

[1]Aix Marseille Univ, Institut de Recherche pour le Développement (IRD), Service de Santé des Armées, AP-HM, UMR Vecteurs Infections Tropicales et Méditerranéennes (VITROME), Institut Hospitalo-Universitaire Méditerranée Infection, Marseille, France. [2]Aix-Marseille Université, Institut de Recherche pour le Développement (IRD), UMR Microbes Evolution Phylogeny and Infections (MEPHI), Institut Hospitalo-Universitaire Méditerranée-Infection, Marseille, France. [3]Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia. Correspondence and requests for materials should be addressed to P.-E.F. (email: pierre-edouard.fournier@univ-amu.fr)
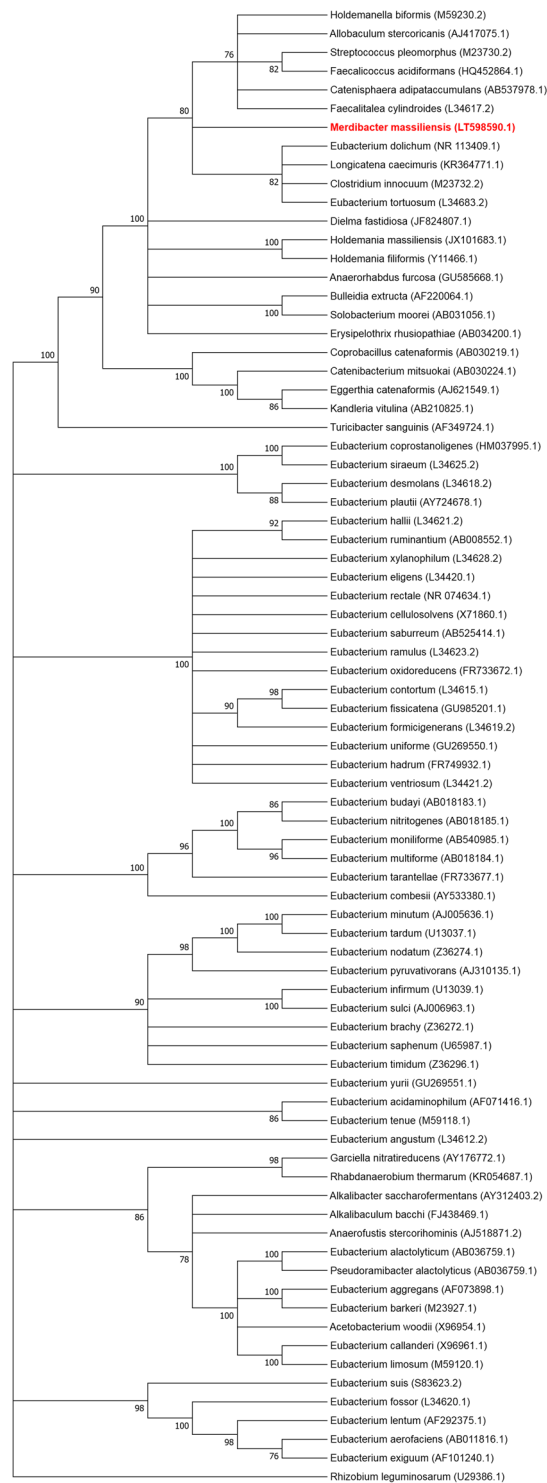
1

Figure 1. 16S rRNA-based phylogenetic tree highlighting the position of *Merdibacter massiliensis* gen. nov. sp., nov., strain Marseille-P3254 (red) relative to other closely related bacterial taxa within the family *Erysipelotrichaceae* as well as members of the family *Eubacteriaceae*. Genbank accession numbers are presented in parentheses. Sequences were aligned using Muscle v3.8.31 with default parameters and phylogenetic relationship inferred using the Maximum Likelihood method, with 1,000 bootstrap replicates, within the MEGA software version 7.0. Only values above 70% were indicated. *Rhizobium leguminosarum* was used as outgroup.

Results

Strain identification and classification. Strain Marseille-P3254 was isolated from the ileal content of a 76-year-old woman who underwent upper and lower digestive tract endoscopy for esophagitis and colonic polyp. The patient gave an informed and written consent and the study was approved by the ethics committee

Figure 2. Scanning electron microscopy of *Merdibacter massiliensis* gen. nov., sp. nov., strain Marseille-P3254 using a Tabletop microscope TM 4000 plus (Hitachi, Tokyo, Japan). The scale bar represents 10 µm.

of the Institut Hospitalo-Universitaire Mediterranee Infection under number 2016-010. As no microorganism grew from the negative control, we are confident that *Merdibacter massiliensis* was not an experiment contamination. Strain Marseille-P3254 could not be identified by our systematic MALDI-TOF MS screening as the score was 1.763, suggesting that the corresponding species was not in the database (Figure S1). Moreover, strain Marseille-P3254 exhibited a 90.70% 16S rRNA sequence similarity with *Eubacterium dolichum* strain JCM 10413[T] (GenBank accession no. NR_113409), the phylogenetically closest bacterium with standing in nomenclature (Fig. 1). As this value is lower than the 95% threshold defined by Stackebrandt and Ebers for delineating a new genus, strain Marseille-P3254 was considered as representative of a putatively new genus within the family *Erysipelotrichaceae* in the phylum *Firmicutes*.

Phenotypic characteristics.   Growth was observed on 5% sheep blood-enriched Columbia agar (bioMérieux) at 37 °C and 45 °C after 5 days of incubation. Colonies from strain Marseille-P3254 showed neither pigmentation nor haemolysis. They were circular with a diameter of 0.5 to 1.5 mm, and transparent. Bacterial cells were Gram-negative, non-motile rods with a length of 1.50 to 2.78 µm and a width of 0.3 to 0.5 µm, as determined by electronic scanning microscopy (Fig. 2). Strain Marseille-P3254 grew only in anaerobic conditions. The sporulation test (20 minutes at 80 °C) was negative. In addition, this bacterium had no oxidase and catalase activities.

Using an API ZYM strip, a positive reaction was observed for alkaline and acid phosphatases but negative reactions were observed with esterase, esterase lipase, lipase, leucine arylamidase, valine arylamidase, cystine arylamidase, trypsin, α-chymotrypsin, naphtol-AS-BI-phosphohydrolase, α-galactosidase, ß-galactosidase, ß-glucuronidase, α-glucosidase, ß-glucosidase, N-acetyl-ß-glucosaminidase, α-mannosidase and α-fucosidase. Using an API 20NE strip, negative reactions were obtained for reduction of potassium nitrate, indole production from tryptophan, glucose fermentation, arginine hydrolysis, urea, aesculin, gelatin, p-nitrophenyl-ßD-galactopyranoside, and assimilation of glucose, arabinose, mannose, mannitol, N-acetyl-glucosamine, maltose, gluconate, caprate, adipate, malate, citrate and phenyl-acetate. Using an API 50 CH strip, strain Marseille-P3254 was able to metabolize glycerol, D-galactose, D-glucose, D-fructose, D-mannose, methyl-αD-glucopyranoside, N-acethylglucosamine, D-maltose, D-lactose, D-saccharose, D-trehalose, D-turanose, D-tagatose and potassium 5-Ketogluconate. However, negative reactions were obtained with erythritol, D-arabinose, L-arabinose, D-ribose, D-xylose, L-xylose, D-adonitol, methyl-ßD-xylopyranoside, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl-αD-mannopyranoside, amygdalin, arbutin, esculin, salicin, D-cellobiose, D-melibiose, inulin, D-melezitose, D-raffinose, starch, glycogen, xylitol, gentiobiose, D-lyxose, D-fucose, L-fucose, D-arabitol, L-arabitol, potassium gluconate and potassium 2-ketogluconate.

According to the French Microbiology Society, susceptibility tests showed that strain Marseille-P3254 was susceptible to cefoxitin, linezolid, vancomycin, trimethoprim-sulfamethoxazole, metronidazole and ciprofloxacin, but resistant to ticarcillin-clavulanic acid, cefepime, ceftriaxone, erythromycin, fosfomycin, rifampin, amikacin and teicoplanin.

By comparison with closely related taxa, strain Marseille-P3254 differed in a combination of negative catalase activity and positive galactose metabolism (Table 1). The major fatty acids were Hexadecanoic acid (34%) and 9-Octadecenoic acid (34%). Significant abundances of C16:0, C18:1n9 and C18:0 were also described. In addition, lighter aliphatic chains were detected such as C10:0, C12:0, C14:0 and C15:0 (Table S1).

Genome sequencing information and genome properties.   The genome size of strain Marseille-P3254 was 2,468,496-bp long with a 40.1% G + C content. It was assembled into 24 contigs. Of the 2,375 predicted genes, 2,315 were protein-coding genes and 60 were RNAs (one complete rRNA operon, three additional 5S rRNAs, 50 tRNAs and 4 non-coding RNAs). A total of 1,643 genes (69.1%) were assigned a putative function and 672 genes (28.2%) were annotated as hypothetical proteins. The genome properties and distribution of genes into COGs functional categories are detailed in Table S2. The *in silico* resistome of this multidrug resistant strain includes genes coding resistance to tetracycline (*tetW*), aminoglycoside (*ant6*), macrolide-lincosamide-streptogramin B (*lsa*) and oxazolidinone (*oxzln*) (Fig. 3).

Characteristics	MM	ED	FC	CI	DF	HB	HM	SP
Gram stain	−	+	+	+	−	+	+	+
Production of Catalase	−	−	−	−	−	−	+	−
Oxidase	−	−	−	Na	−	−	−	−
Nitrate reductase	−	−	−	−	−	−	−	−
Gelatin hydrolysis	−	−	−	+	na	+	na	−
Utilisation of L-Arabinose	−	−	−	na	−	+	−	−
D-Galactose	+	−	−	na	−	−	+	−
D-Fructose	+	+	−	na	−	−	+	−
D-Mannose	+	−	+	+	−	+	+	+
D-Rhamnose	−	−	−	na	−	−	−	−
D-Mannitol	−	−	−	+	−	+	+	−
D-Sorbitol	−	−	−	−	na	−	+	−
Amygdalin	−	−	−	na	na	−	+	−
Arbutin	−	na	na	na	na	−	+	−
Salicin	−	−	−	na	na	+	+	−
D-Cellobiose	−	−	−	+	−	−	+	−
D-Maltose	+	+	−	+	−	−	+	−
D-Lactose	+	−	−	−	−	−	+	−
D-Trehalose	+	+	−	+	−	−	−	−
Inulin	−	na	na	na	na	−	na	−
Starch	−	−	−	na	na	−	na	−
Glycogen	−	−	−	na	na	−	na	−
Xylitol	−	na	na	na	−	−	−	−
D-Arabitol	−	na	na	na	−	−	−	−
D-Glucose	−	−	+	+	−	+	+	+
Raffinose	−	−	+	+	−	+	−	−
D-Xylose	−	+	−	+	−	+	−	−

Table 1. Compared phenotypic characteristics of studied species. ED: *Eubacterium dolichum;* FC: *Faecalitalea cylindroides;* DF: *Dielma fastidiosa;* HB: *Holdemanella biformis;* HM: *Holdemania massiliensis:* SP: *Streptococcus pleomorphus;* MM: *Merdibacter massiliensis;* CI: *Clostridium innocuum.* Data for reference strains[5,8,10,39–41]. +: positive reaction, −: negative reaction, na: data not available.

In the genome from *Merdibacter massiliensis*, we identified genes coding enzymes involved in the pentose phosphate pathways (13 genes), including *aldoa* (fructosebiphosphate aldolase, 1 copy), *dera* (deoxyribose-phosphate aldolase, 1 copy), *rpel* (ribulose phosphate 3-epimerase, 2 copies), *pfkl* (6-phosphofructokinase, 1 copy), *pgd* (6-phosphogluconate dehydrogenase, 1 copy), *pgm1* (phosphoglucomutase, 2 copies), *pgm2* (phosphopento-mutase, 1 copy), *rpib* (ribose-5-phosphate isomerase B, 1 copy), *taldo1* (transaldolase, 1 copy) and *tkt* (transketolase, 2 copies). We also identified sugar phosphotransferase systems (52 predicted genes), including were *ulaB* (Ascorbate-specific PTS system, 5 copies), *bglF* (PTS system beta-glucoside-specific, 7 copies), *frwD* (PTS system fructose-like, 3 copies), *ptsg* (PTS system glucose-specific, 6 copies), *malX* (PTS system maltose-specific, 3 copies), *manX* (PTS system mannose-specific, 15 copies), *gmuB* (PTS system oligo-beta-mannoside-specific, 1 copy), *sorA* (PTS system sorbose-specific, 10 copies), *treP* (PTS system trehalose-specific, 1 copy) and *dhaM* (PTS-dependent dihydroxyacetone kinase, 1 copy). In addition, we identified several predicted proteins known to play a role in vitamin metabolism, including folate family ECF transporter S component, tetrahydrofolate ligase, 5-formyltetrahydrofolate cyclo-ligase, methylenetetrahydrofolate (folates), biotin transporter BioY, acetyl-CoA carboxylase biotin carboxylase subunit, acetyl-CoA carboxylase biotin carboxyl carrier protein, biotin–[acetyl-CoA-carboxylase] ligase (biotin), bifunctional riboflavin kinase/FAD synthetase (riboflavin) and cobalamin biosynthesis protein CobW (cobalamin).

In silico search for virulence factors showed the presence of six proteins conferring a potential pathogenicity with high identity percentage. One of these proteins exhibited 97% identity (100% sequence coverage) with its ortholog (Transcription regulator Immr) in the pathogenic bacterium *Clostridium difficile* strain R2029. In addition, a predicted protein exhibited 100% similarity with the virulence-associated protein E from *Streptococcus suis*. The other four predicted proteins were conjugal transfer protein Tcpe (86% identity), transcription regulator yobd (85%), antirestriction protein ArdA (84%) and excisionate HTH (81%).

A total of 714 protein-coding genes were likely to be acquired vertically or from closely related species, as their closest orthologs belonged to members of the family *Erysipelotrichaceae* to which strain Marseille-P3254 also belonged (Fig. 4). In addition, 1601 genes were likely to be gained by lateral gene transfer (LGT) from members of other bacterial families (Fig. 4). Most lateral gene transfer (LGT)-acquired genes were obtained from *Erysipelotrichaceae* (30.84%) followed by other bacteria (17.14%), *Clostridiaceae* (17.10%) and *Eubacteriaceae* (15.42%).
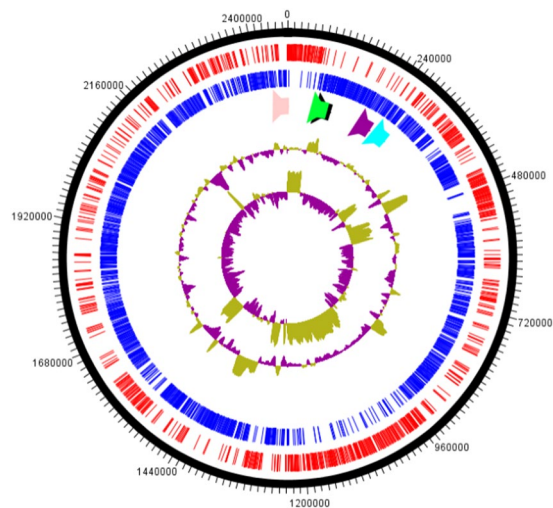
Figure 3. Graphical circular map of the genome from strain Marseille-P3254. From outside to the center: genes on the forward strand colored in dark blue, genes on the reverse strand colored in red, the five resistance genes identified by comparison with the ARG-ANNOT database colored in pink, green, black, purple and light blue, G + C content and G + C skew.
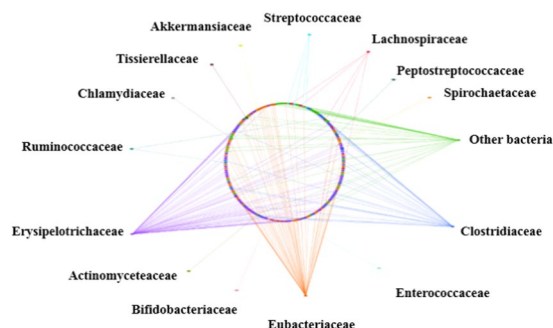

Figure 4. Network showing the origin of predicted protein-coding genes in strain Marseille-P3254 according to bacterial families. A total of 714 protein-coding genes were likely to be acquired vertically or from closely related species within the family *Erysipelotrichaceae*, and 1601 genes were likely to be gained by lateral gene transfer from members of other bacterial families.

Comparison with closely related bacterial strains. The genome of strain Marseille P3254 was compared to the available genomes of eight closely related bacterial type strains (Table S3). Core-genome-based phylogenetic relationships of strain Marseille-P3254 and the closest species with standing in nomenclature are presented in Fig. S2. The distribution of genes into COG categories was similar in all nine compared genomes (Fig. 5). Strain Marseille-P3254 shared 975, 584, 966, 835, 905, 860, 1,137 and 754 orthologous genes with *Eubacterium dolichum*, *Faecalitalea cylindroides*, *Dielma fastidiosa*, *Holdemanella biformis*, *Streptococcus pleomorphus*, *Holdemania massiliensis*, *Clostridium innocuum* and *Anaerorhabdus furcosa*, respectively (Table 2). Moreover, MAGi (Marseille Average Genomic identity)[11] analysis showed that AGIOS (Average Genomic Identity of Orthologous gene Sequences)[4] values ranged from 49.2% between *Anaerorhabdus furcosa* and *Holdemania massiliensis*, to 52.39% between *Anaerorhabdus furcosa* and *Holdemanella biformis*, among studied genera with standing in nomenclature. Regarding strain Marseille-P3254, the range of AGIOS value varied from 50.25% with *Holdemania massiliensis* to 52.14% with *Dielma fastidiosa* (Table 2). Strain Marseille-P3254 was closer to *Dielma fastidiosa*, with 52.14% genomic identity, but shared more orthologous genes (1,137) with *Clostridium innocuum*. Finally, dDDH (digital DNA-DNA hybridization)[12,13] estimation of strain Marseille-P3254 against the available genomes for type strains of species, with standing in nomenclature ranged from 18.7% between *Dielma fastidiosa* and *Streptococcus pleomorphus*, to 41.1% between *Eubacterium dolichum* and *Faecalitalea cylindroides* (Table S4). These values are lower than 70% the cutoff used for delineating prokaryotic species, thus confirming that this strain represents a genus distinct from these other bacterial strains. Ortho Average nucleotide identity[14] (OrthoANI) values ranged between 63.52 and 73.48% (orthoANI value lower than 80.5%) confirming again that this strain is distinct from these other tested bacterial taxa (Fig. 6). Cumulatively, the above data suggest that strain Marseille-P3254 represents a novel species within a new genus in the family *Erysipelotrichaceae*, for which the name *Merdibacter massiliensis* gen. nov., sp. nov., is proposed.
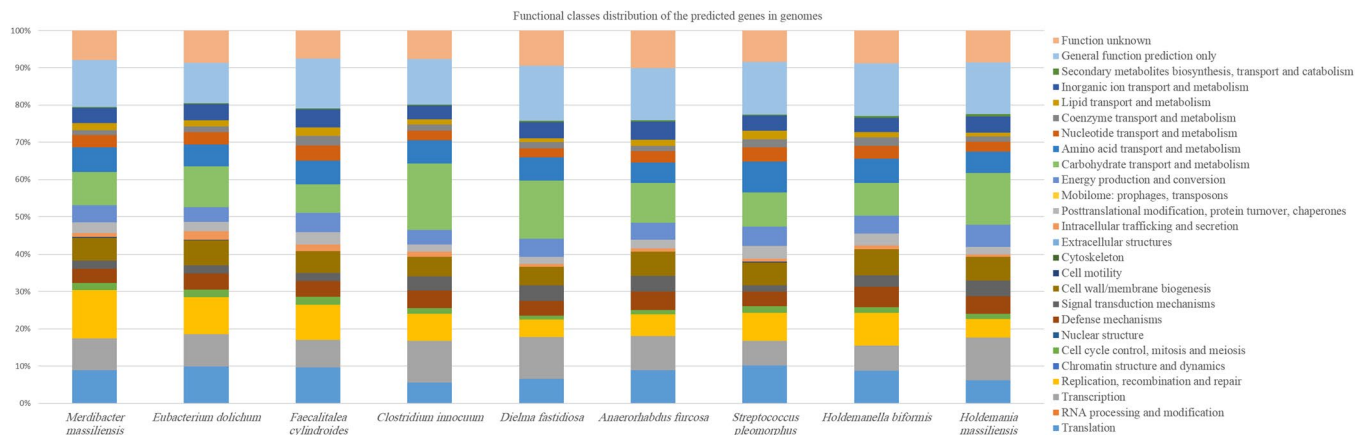
Figure 5. Distribution of functional classes of predicted genes according to the clusters of orthologous groups of proteins of *Merdibacter massiliensis* gen. nov., sp. nov., strain Marseille-P3254 and other compared bacterial taxa.
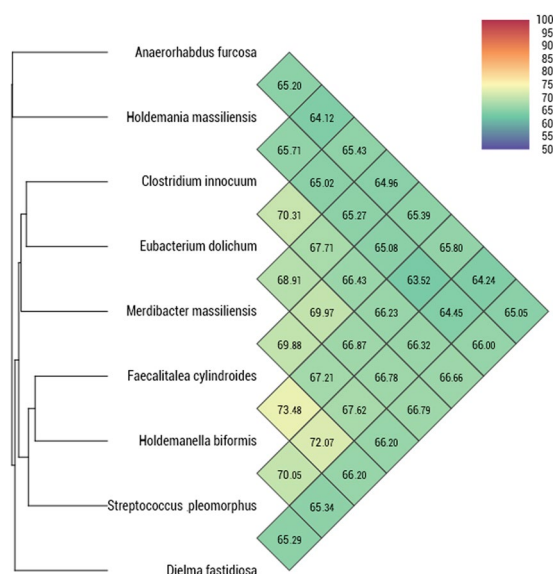

Figure 6. Heatmap generated with OrthoANI values calculated using the OAT software between *Merdibacter massiliensis* gen. nov., sp. nov., strain Marseille-P3254 and other closely related taxa with standing in nomenclature.

	ED	FC	DF	HB	MM	SP	HM	CI	AF
ED	**2,190**	507	955	769	975	811	839	1,098	744
FC	51.55%	**1,971**	490	574	584	631	459	612	424
DF	51.31%	51.065%	**3,497**	779	966	820	1,083	1,209	916
HB	51.61%	51.735%	51.20%	**2,392**	835	895	727	934	663
MM	51.77%	51.20%	52.14%	51.47%	**2,411**	905	860	1,137	754
SP	51.64%	51.34%	51.13%	51.38%	51.29%	**2,005**	742	990	722
HM	50.12%	49.87%	50.60%	49.61%	50.25%	50.49%	**3,486**	1,097	917
CI	50.71%	50.48%	50.79%	50.46%	50.81%	50.65%	50.64%	**4,702**	932
AF	51.65%	51.57%	51.07%	52.39%	51.42%	51.47%	49.20%	50.21%	**2,404**

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. ED: *Eubacterium dolichum;* FC: *Faecalitalea cylindroides;* DF: *Dielma fastidiosa;* HB: *Holdemanella biformis;* HM: *Holdemania massiliensis:* SP: *Streptococcus pleomorphus;* MM: *Merdibacter massiliensis;* CI: *Clostridium innocuum;* AF: *Anaerorhabdus furcosa.*

Discussion

Based on the diversification of culture conditions, the new microbial approach "culturomics" aiming at deciphering the complexe diversity of the human microbiota, allowed isolation of more than 1,500 different human bacterial species over the past 6 years, including more than 500 new species[1,15]. Using the taxono-genomics concept combining the genomic and phenotypic properties of a putative new taxa[15], we have characterized a new bacterial species representing a new genus within the family *Erysipelotrichaceae* in a human ileal specimen. One of the human gut microbiota characteristics is the richness of the enzymes related to central metabolism (like the pentose phosphate pathway) and sugar phosphotransferase systems (PTS)[16,17]. In addition, the metabolic phenotype observed for strain Marseille-P3254 was similar to that of *Eubacterium dolichum*, its closest phylogenetic neighbor and also a gut microorganism. As a consequence, as *Merdibacter massiliensis* exhibits metabolic characteristics that enable harvesting nutrients and energy from the diet, and are consistent with a role in metabolic homeostasis in the human gut, we believe that it may survive in the human gut[18–20]. Thus, we propose the creation of the new genus and species *Merdibacter massiliensis* gen. nov., sp. nov., a member of the *Erysipelotrichaceae* family, closely related to *E. dolichum* that was recently proposed to belong to a new genus within this family[10]. Bioinformatic analysis predicted that our strain was a putative pathogen. This hypothesis was supported by the presence of several proteins associated to pathogenesis (hemolysin III, type II toxin-antitoxin system Hica), all of which were acquired from to the family *Clostridiaceae* (Fig. 4). In addition, strain Marseille-P3254 shared several protein-coding genes with *Streptococcus suis*[21], a pathogen causing gastro-intestinal tract infections in human, including a virulence-associated protein E. However, the ileal lavage being part of the exploration of a colonic polyp, no clear association of *M. massiliensis* with pathogenesis can be inferred in this patient. The network of *M. massiliensis* exhibited a remarkable inheritance of genes from bacterial families distinct from its parent family (Fig. 4). The Digital Protologue TaxoNumber (http://imedea.uibcsic.es/dprotologue/index.php) of *M. massiliensis* gen. nov., sp. nov. is TA00782.

Description of *Merdibacter* gen. nov.
Merdibacter (Mer.di.ba'c.ter, L. masc. n. *Merdibacter*, composed of merdi from *merda*, 'excrement' and *bacter*, 'a rod'. *Merdibacter*, 'a rod from faeces'). Cells are anaerobic, Gram-negative, nonmotile and asporogenous rods. Catalase and oxidase activities are negative.

Description of *Merdibacter massiliensis* gen. nov., sp. nov.
Merdibacter massiliensis (mas.si.li.en'sis. L. masc. adj. massiliensis, of Massilia, the Latin name of Marseille where strain Marseille-P3254 was first isolated).

In addition, the description features of the genus, cells have a length of 1.5 μm to 2.78 μm and a width of 0.3 to 0.5 μm. Colonies grown on 5% sheep blood-enriched Columbia agar (bioMérieux) are circular and transparent after 5 days of incubation in anaerobic atmosphere, varying in size from 0.5 to 1.5 mm in diameter. Growth occurs at 37 °C (optimum) and 45 °C. Cells grow anaerobically only. Using an API ZYM strip, a positive reaction is observed for alkaline and acid phosphatases but negative reactions are observed with esterase, esterase lipase, lipase, leucine arylamidase, valine arylamidase, cystine arylamidase, trypsin, α-chymotrypsin, naphtol-AS-BI-phosphohydrolase, α-galactosidase, ß-galactosidase, ß-glucuronidase, α-glucosidase, ß-glucosidase, N-acetyl-ß-glucosaminidase, α-mannosidase and α-fucosidase. Using an API 20NE strip, negative reactions are obtained for reduction of potassium nitrate, indole production from tryptophan, glucose fermentation, arginine hydrolysis, urea, aesculin, gelatin, p-nitrophenyl-ßD-galactopyranoside, and assimilation of glucose, arabinose, mannose, mannitol, N-acetyl-glucosamine, maltose, gluconate, caprate, adipate, malate, citrate and phenyl-acetate. Using an API 50 CH strip, strain Marseille-P3254 was able to metabolize glycerol, D-galactose, D-glucose, D-fructose, D-mannose, methyl-αD-glucopyranoside, N-acethylglucosamine, D-maltose, D-lactose, D-saccharose, D-trehalose, D-turanose, D-tagatose and potassium 5-Ketogluconate. However, negative reactions are obtained with erythritol, D-arabinose, L-arabinose, D-ribose, D-xylose, L-xylose, D-adonitol, methyl-ßD-xylopyranoside, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl-αD-mannopyranoside, amygdalin, arbutin, esculin, salicin, D-cellobiose, D-melibiose, inulin, D-melezitose, D-raffinose, starch, glycogen, xylitol, gentiobiose, D-lyxose, D-fucose, L-fucose, D-arabitol, L-arabitol, potassium gluconate and potassium 2-ketogluconate. The most abundant fatty acids are hexadecanoic acid ($C_{16:0}$), 9-Octadecenoic acid ($C_{18:1n9}$) and Octadecanoic acid ($C_{18:0}$). The genome is 2,468,496-bp long and its $G + C$ content is 40.1%.

The type strain, Marseille-P3254^T, isolated from the ileum of a patient, was deposited in the CSUR and DSMZ collections under accession numbers CSUR P3254 and DSM 103534, respectively. The 16S rRNA and genome sequences are available in GenBank under accession numbers LT598590 and FTLC00000000, respectively.

Materials and Methods

Strain isolation and phenotypic tests. As per our culturomics procedure[2], the fresh ileal wash sample was collected in sterile vial and then inoculated in an anaerobic blood culture vial (Becton Dickinson, Pont de Claix, France) enriched with 5 mL of sterile sheep blood (BioMérieux) and 5 mL of filter-sterilized (Thermo Fisher Scientific, Vilbon-sur-Yvette, France) rumen fluid (3 successive filtrations using filters with 0.8 μm, 0.45 μm, and 0.2 μm pore sizes). In parallel, an anaerobic blood culture vial (Becton Dickinson) enriched with 5 mL of sterile sheep blood (bioMérieux) was inoculated with 5 mL of filter-sterilized (Thermo Fisher Scientific) rumen fluid as a negative control to verify the sterility of the nutrient. After 7 days of incubation at 37 °C, the suspension was inoculated on 5% sheep blood-enriched Columbia agar (BioMérieux, Marcy l'Etoile, France) in anaerobic atmosphere (anaeroGEN, Oxoid, Dardilly, France). Isolated colonies were identified using MALDI-TOF mass spectrometry (MS) protein analysis and a Microflex spectrometer (Bruker Daltonics, Bremen, Germany)[22]. Spectra from strain Marseille-P3254 were imported into the MALDI BioTyper software (version 2.0, Bruker) and analyzed by standard pattern matching (with default parameter settings) against 11805 bacterial spectra in the Bruker database

(7854 spectra) enriched with spectra from bacteria isolated in our laboratory as part of the culturomics study (3951 spectra). Interpretation of the scores was performed as previously described[23].

Moreover, the 16S rRNA gene was sequenced using the fD1-rP2 primer pair as previously described[24], 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[25]. Different growth temperatures (20, 28, 37, 45 and 55 °C) were tested. Growth of strain Marseille-P3254 was tested under different atmospheres (anaerobic, aerobic and microaerophilic) (CampyGEN, Oxoid). API ZYM, API NE and API 50CH strips (BioMérieux) were used to evaluate the biochemical properties of the strain according to the manufacturer's instructions. For scanning electronic microscopy, a colony was collected from agar and immersed into a 2.5% glutaraldehyde fixative solution. The slide was gently washed in water; air dried and examined with approximately 60 centimeters in height and 33 cm in width to evaluate bacterial structure on a TM4000 microscope. The standard disc method was applied for antimicrobial susceptibility testing according to the French Microbiology Society. Finally, cellular fatty acid methyl ester (FAME) analysis was performed by GC/MS. Two samples were prepared with approximately 25 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, Courtaboeuf, France)[26]. GC/MS analyses were carried out as previously described[27,28]. Spectral database search was performed using MS Search 2.0 operated with the Standard Reference Database 1A (NIST, Gaithersburg, USA) and the FAME mass spectral database (Wiley, Chichester, UK).

DNA extraction and genome sequencing. Genomic DNA (gDNA) of strain Marseille-P3254 was extracted in two steps: a mechanical treatment was first performed by acid-washed glass beads (G4649-500g Sigma) using a FastPrep BIO 101 instrument (Qbiogene, Strasbourg, France) at maximum speed (6.5 m/sec) for 90 sec. Then, after a 2.5 hour lysozyme incubation at 37 °C, DNA was extracted using an EZ1 biorobot (Qiagen) with EZ1 DNA Tissue kit. The elution volume was 50 μL. gDNA was quantified by a Qubit assay with the high sensitivity kit (Life technologies, Carlsbad, CA, USA) at 313 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 other 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 profile 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 kb up to 11 kb with an optimal size at 2.38 kb. No size selection was performed and 114.4 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal at 993 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 10.51 nmol/l.

The library was normalized at 2 nM and, 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 hour run in a 2 × 151 bp format.

Total information of 2.8 Gbp was obtained with 274 K/mm^2 cluster density with a cluster passing quality control filters of 97.6% (5,537,000 passing filter paired reads). Within this run, the index representation for strain Marseille-P3254 was determined to 9.48%. The 524,686 paired reads were assembled using the SPAdes version 3.10.1 software[29]. 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[30] and GapFiller[31] were used to combine contigs with default parameters. Finally, manual finishing was performed by using similarity searches based on BLAST searches, and synteny blocks were detected by progressive alignment using the Mauve software[32].

Genome annotation and genome comparison. The genome was annotated as previously described[23]. In addition, we compared the genome from strain Marseille-P3254 to those of the closely related species *Eubacterium dolichum* strain JCM 10413^T (accession number ABAW00000000), *Faecalitalea cylindroides* strain JCM 10261^T (AWVI00000000), *Dielma fastidiosa* strain DSM 26099^T (CAEN00000000), *Holdemanella biformis* strain DSM 3989^T (ABYT00000000), *Streptococcus pleomorphus* strain DSM 20574^T (ATUT00000000), *Holdemania massiliensis* strain DSM 26143^T (CALK00000000), *Clostridium innocuum* strain DSM 1286^T (AGYV00000000) and *Anaerorhabdus furcosa* strain ATCC 25662^T (FUWY00000000). For this, we used the Genome-to Genome Distance Calculator (GGDC) web server available at http://ggdc.dsmz.de to estimate the overall similarity among compared genomes and to replace the wet-lab DNA–DNA hybridization (DDH) by a digital DDH (dDDH)[12,13]. The average nucleotide identity at the genomic level was also estimated using the orthoANI[14] and AGIOS[4,33] software. Antibiotic resistance genes (ARG) were searched using the ARG-ANNOT database and Bio-Edit interface[34]. Assembled sequences were searched against the ARG-ANNOT database under moderately stringent conditions (e-value of 10^{-5}) 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 PathogenFinder 1.1[35]. Predicted protein sequences of strain Marseille-P3254 were used as queries to search the NCBI GenBank non-redundant protein sequence database. These results were formatted to generate a network of protein sequences using the Cytoscape tool[36]. Finally, predicted protein sequences were searched against KEGG PATHWAY[37] and PATRIC 3.5.31[38] databases for the screening of genes involved in metabolic pathways.

References

1. Lagier, J.-C. *et al.* Microbial culturomics: paradigm shift in the human gut microbiome study. *Clinical Microbiology and Infection* **18**, 1185–1193 (2012).
2. 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).
3. Lagier, J.-C. *et al.* Culture of previously uncultured members of the human gut microbiota by culturomics. *Nature Microbiology* **1** (2016).
4. 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).
5. Ricaboni, D. *et al.* 'Merdibacter massiliensis' gen. nov., sp. nov., isolated from human ileum. *New Microbes and New Infections* **15**, 89–91 (2017).
6. 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).
7. Kanno, M. *et al.* *Catenisphaera adipataccumulans* gen. nov., sp. nov., a member of the family *Erysipelotrichaceae* isolated from an anaerobic digester. *International Journal of Systematic and Evolutionary Microbiology* **65**, 805–810 (2015).
8. Moore, W. E. C. *et al.* Emendation of *Bacteroidaceae* and *Butyrivibrio* and Descriptions of *Desulfornonas* gen. nov. and ten New Species in the Genera *Desulfomonas, Butyrivibrio, Eubacterium, Clostridium*, and *Ruminococcus. International Journal of Systematic Bacteriology* p.238–252 (1976).
9. Ludwig, W. *et al.* Family II. *Eubacteriaceae* fam. nov. In: De Vos, P. *et al.* (editors): Bergey's Manual of Systematic Bacteriology, second edition, vol. 3 (*The Firmicutes*), Springer, Dordrecht, Heidelberg, London, New York, p. 865 (2009).
10. Paek, J. *et al.* Description of *Absiella argi* gen. nov., sp. nov., and transfer of *Eubacterium dolichum* and *Eubacterium tortuosum* to the genus *Absiella* as *Absiella dolichum* comb. nov. and *Absiella tortuosum* comb. nov. *Anaerobe* **48**, 70–75 (2017).
11. Lo, C. I. *et al.* Genome sequence and description of *Pantoea septica* strain FF5. *Standards in Genomic. Sciences* **10**, 103 (2015).
12. Auch, A. F. Digital DNA-DNA hybridization. *Standards in Genomic Sciences* 18 (2010).
13. 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).
14. 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).
15. Abdallah, R. A. *et al.* The impact of culturomics on taxonomy in clinical microbiology. *Antonie van Leeuwenhoek* **110**, 1327–1337 (2017).
16. Zoetendal, E. G. *et al.* The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *The ISME Journal* **6**, 1415–1426 (2012).
17. Heinken, A. & Thiele, I. Systems biology of host-microbe metabolomics: Host-microbe metabolomics. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **7**, 195–219 (2015).
18. Goh, Y. J. & Klaenhammer, T. R. Genetic Mechanisms of Prebiotic Oligosaccharide Metabolism in Probiotic Microbes. *Annual Review of Food Science and Technology* **6**, 137–156 (2015).
19. Flint, H. J., Scott, K. P., Louis, P. & Duncan, S. H. The role of the gut microbiota in nutrition and health. *Nature Reviews Gastroenterology & Hepatology* **9**, 577–589 (2012).
20. Wang, B., Yao, M., Lv, L., Ling, Z. & Li, L. The Human Microbiota in Health and Disease. *Engineering* **3**, 71–82 (2017).
21. Ferrando, M. L. & Schultsz, C. A hypothetical model of host-pathogen interaction of *Streptococcus suis* in the gastro-intestinal tract. *Gut Microbes.* **7**(2), 154–162 (2016).
22. 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).
23. Hadjadj, L. *et al.* Non contiguous-finished genome sequence and description of *Microbacterium gorillae* sp. nov. *Standards in Genomic Sciences* **11** (2016).
24. 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).
25. 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).
26. Sasser, M. Bacterial Identification by Gas Chromatographic Analysis of Fatty Acids Methyl Esters (GC-FAME) TechNote_101. MIDI Inc, Newark.
27. Dione, N. *et al.* Genome sequence and description of *Anaerosalibacter massiliensis* sp. nov. *New Microbes and New Infections* **10**, 66–76 (2016).
28. Sasser, M. Bacterial Identification by Gas Chromatographic Analysis of Fatty Acids Methyl Esters (GC-FAME). 6 (2006).
29. 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).
30. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011).
31. Boetzer, M. & Pirovano, W. Toward almost closed genomes with GapFiller. *Genome Biol.* **13**, R56 (2012).
32. Darling, A. E., Tritt, A., Eisen, J. A. & Facciotti, M. T. Mauve Assembly Metrics. *Bioinformatics* **27**, 2756–2757 (2011).
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. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research* **45**, D353–D361 (2017).
38. Wattam, A. R. *et al.* PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Research* **42**(D1), D581–D591 (2014).
39. Ramasamy, D. *et al.* Non contiguous-finished genome sequence and description of *Dielma fastidiosa* gen. nov., sp. nov., a new member of the Family *Erysipelotrichaceae. Standards in Genomic Sciences* **8**(2), 336–351 (2013).
40. De Maesschalck, C. *et al.* *Faecalicoccus acidiformans* gen. nov., sp. nov., isolated from the chicken caecum, and reclassification of *Streptococcus pleomorphus* (Barnes *et al.* 1977), *Eubacterium biforme* (Eggerth 1935) and *Eubacterium cylindroides* (Cato *et al.* 1974) as *Faecalicoccus pleomorphus* comb. nov., *Holdemanella biformis* gen. nov., comb. nov. and *Faecalitalea cylindroides* gen. nov., comb. nov., respectively, within the family *Erysipelotrichaceae. International Journal of Systematic and Evolutionary Microbiology* **64**, 3877–3884 (2014).
41. Mishra, A. K. *et al.* Non-contiguous finished genome sequence and description of *Holdemania massiliensis* sp. nov. *Standards in Genomic Sciences i* **9**, 395–409 (2013).

Acknowledgements

Author Contributions

H.A. performed the genomic analysis and phenotypic tests, R.A. and N.C. participated in the genomic analysis, A.F. did the electronic microscopy, D.R. 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.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-44343-8.

Competing Interests: The authors declare no competing interests.

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