AIX-MARSEILLE UNIVERSITE

FACULTE DES SCIENCES MEDICALES ET

PARAMEDICALES DE MARSEILLE

ECOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE

THESE

Présentée et publiquement soutenue devant

LA FACULTÉ DE MÉDECINE DE MARSEILLE

Le 04 Juillet 2019

Par Amadou Hamidou TOGO

Evaluation du microbiote du lait maternel

Pour obtenir le grade de Docteur d'Aix-Marseille Université

Biologie-Santé, Spécialité Maladies Transmissibles et Pathologies Tropicales

Membres du Jury de la Thèse:

Mr le professeur Jean-Christophe Lagier	Aix-Marseille Université	Président
Mme le professeur Marie Kempf	Université d'Angers	Rapporteur
Mr le professeur Bouréma Kouriba	Université de Bamako	Rapporteur
Mr le professeur Max Maurin	Université de Grenoble	Examinateur
Mr le professeur Didier Raoult	Aix-Marseille Université	Directeur de thèse

Institut Hospitalo-Universitaire Méditerranée Infection, Unité de Microbes

Evolution Phylogénie et Infections - MEPHI, IRD, AP-HM

Avant-propos

Le format de présentation de cette thèse correspond à une recommandation de la spécialité Maladies Infectieuses et Microbiologie du Master des Sciences de la Vie et de la Santé qui dépend de l'Ecole Doctorale des Sciences de la Vie de Marseille.

Le candidat est amené à respecter des règles qui lui sont imposées et qui comportent un format de thèse utilisé dans le Nord de l'Europe et qui permet un meilleur rangement que les thèses traditionnelles. Par ailleurs, la partie introduction et bibliographie est remplacée par une revue publiée dans un journal scientifique afin de permettre une évaluation extérieure de la qualité de la revue et de permettre à l'étudiant de commencer le plus tôt possible une bibliographie exhaustive sur le domaine de cette thèse. Par ailleurs, la thèse est présentée sur la base d'articles publiés, acceptés où soumis associés d'un bref commentaire donnant le sens général du travail. Cette forme de présentation parait plus en adéquation avec les exigences de la compétition internationale et permet de se concentrer sur des travaux qui bénéficieront d'une diffusion internationale.

Professeur Didier RAOULT

1

Remerciements

Sommaire

Résumé

Malgré les progrès technologiques dans l'exploration du microbiote humain et les nombreuses études menées sur le microbiote du tube digestif, le microbiote du colostrum et du lait maternel demeure un microbiote négligé. Cependant, les résultats de quelques études ont montré une grande diversité de bactéries commensales qui constituent le microbiote du colostrum et du lait maternel qui pourrait coloniser l'intestin du nourrisson. Selon les auteurs de ces études, le colostrum et le lait maternel pourraient être impliqués dans la transmission verticale des phénotypes associés au microbiote intestinal maternel.

Dans cette thèse, nous avons d'abord évalué la diversité bactérienne de 154 échantillons de colostrum et de lait maternel de 144 mères (France, 134 et Mali, 10) par culture à travers la «microbial culturomics» et par métagénomique ciblée à travers le séquençage du gène de L'ARN ribosomal 16S. Avec l'approche de la «culturomics microbienne», nous avons analysé 20 échantillons de colostrum et de lait (8 échantillons de colostrum et 12 échantillons de lait) et tous les échantillons ont été analysés par l'approche de métagénomique ciblée. Nous avons également cultivé 20 échantillons, dont 9 de colostrums et 11 de laits matures, pour l'isolement des archées méthanogènes qui jouent un rôle important dans la digestion de certains aliments. Dans la deuxième partie, nous avons procédé à la description taxonogénomique des nouvelles bactéries isolées à partir du colostrum, du lait maternel et des selles des patients obèses.

Nous avons observé une grande diversité bactérienne du colostrum et du lait maternel grâce aux approches de la «microbial culturomics» et de métagénomique ciblé. Nous avons pu cultiver deux espèces d'archée méthanogènes: *Methanobrevibacter smithii* qui est présent dans le tube digestif de la plupart des êtres humain et *Methanobrevibacter oralis* qui colonise la cavité orale humaine. Nous avons ainsi augmenté le répertoire microbien associé au colostrum et au lait maternel humain en découvrant et en décrivant plusieurs nouvelles espèces bactériennes.

Mot clés: *Microbiote*, *colostrum*, *lait maternel*, *Humain*, *culturomics*, *taxonogénomique*

Abstract

Despite technological advances in human microbiota exploration and numerous studies on digestive tract microbiota, colostrum and breast milk microbiota remain neglected microbiotes. However, the results of few studies have shown a significant diversity of commensal bacteria that constitute the microbiota of colostrum and breast milk that could colonise infant gut. According to the authors of these studies, colostrum and breast milk could be involved in the vertical transmission of phenotypes associated to mother gut microbiota.

In this thesis, we first evaluated the bacterial diversity of 154 colostrum and breast milk samples from 144 mothers (France, 134 and Mali, 10) using culture (culturomics) and targeted metagenomics (16S rRNA gene) approaches. With culturomics approach, we analysed 20 samples of colostrum and milk (8 colostrum and 12 milk) and all samples with targeted metagenomics. We also cultured 20 samples, including 9 colostrum and 11 mature milk, for the isolation of methanogenic archaea that play a key role in the digestion of certain foods. In the second part we proceeded to the taxonogenomic description of new bacteria isolated from colostrum, breast milk and stools of obese patients.

We observed a high bacterial diversity of colostrum and breast milk through culturomics and metagenomics approaches. We have been able to cultivate two species of methanogenic archaea; *Methanobrevibacter smithii* which is present in the digestive tract of most people and *Methanobrevibacter oralis* which colonizes the human oral cavity. We have thus increased the microbial repertoire associated with colostrum and human breast milk by discovering and describing several new bacterial species.

Keywords: *Microbiota*, *Colostrum*, *Breast Milk*, *Human*, *Culturomics*, *taxonogenomics*

Introduction

Le microbiote humain est l'ensemble des communautés microbiennes associé au corps humain. C'est un écosystème complexe qui se compose de virus, de bactéries, d'archées, de champignons, de parasites et d'autres micro-organismes. Le plus grand nombre de ces micro-organismes (10^{10} à 10^{14} bactéries) réside dans le tube digestif (1), d'autres micro-organismes résident sur d'autres sites anatomiques du corps humain tel que la peau, la bouche, les voies respiratoires mais aussi le tissu mammaire et le lait maternel. Le microbiote du lait est l'ensemble des communautés microbiennes associé au colostrum et au lait. Dans notre travail, nous nous sommes intéressés aux bactéries et aux archées méthanogènes.

Le lait maternel est un liquide biologique complexe, de couleur jaunâtre ou blanchâtre selon la période de lactation, sécrété par les glandes mammaires (2). Le lait maternel est considéré comme le meilleur nutriment pour la croissance du nouveau-né (3). Il est adapté aux besoins nutritionnels du nouveau-né. Un bébé allaité au sein est mieux protégé contre les maladies infectieuses par rapport à un bébé nourri au biberon (4–6).

Le colostrum et le lait maternel jadis considérés comme stériles s'avèrent être de sources importantes de bactéries pathogènes et commensales (7,8). Des études récentes ont démontré que le colostrum et le lait maternel abritent non seulement des cellules humaines et des protéines mais aussi un microbiote complexe (archées, bactéries, virus, parasites, et champignons) d'origine inconnue, pouvant coloniser le tube digestif du nouveau-né (9–11). Particulièrement, les espèces du genre *Streptococcus*, *Staphylococcus* et les bactéries lactiques (*Lactobacillus* et *Bifidobacterium*) qui sont des probiotiques typiques qu'on rencontre le plus souvent dans le lait maternel et les feces des nourrissons (10,12). Les bactéries lactiques (des lactobacilles et des bifidobactéries) sont souvent associées à la régulation du poids chez l'homme (13). En outre, l'allaitement maternel pourrait être critique dans la transmission verticale à travers le colostrum et le lait des caractéristiques phénotypes associés au microbiote intestinal, en particulier dans la régulation du poids et l'installation de l'obésité (11,14). Des études menées dans différentes zones géographiques, selon le poids de la mère, le mode de délivrance, le sexe de l'enfant et les périodes de lactations en utilisant différentes méthodes d'analyse, moléculaires en général, ont rapporté des résultats variés sur la composition microbienne du colostrum et du lait maternel (15–17). Cependant, la composition bactérienne viable du colostrum et du lait humain est assez peu étudiée par les approches de culture.

Nous avons entrepris une étude transversale pour évaluer la diversité bactérienne du colostrum et du lait maternel par les méthodes de culture par l'approche de «microbial

culturomics» et de métagénomique ciblée (séquençage du gène de l'ARN ribosomal 16S) chez les femmes allaitantes en état de bonne santé apparente du Mali et de la France. C'est la première fois, à notre connaissance que l'approche de «microbial culturomics» est appliqué à l'exploration du microbiote du colostrum et du lait maternel.

Nous avons commencé par faire un état des lieux du microbiote connu dans le tissu mammaire, le colostrum et le lait. L'analyse des résultats de ce travail nous a permis de calculer la valeur prédictive d'infection (VPI) pour chaque bactérie (voir article I). Une bactérie est considérée probiotique si sa VPI est inférieure à 0,10. Elle est considérée commensale si sa VPI est à 0,50 et est considérée comme potentiellement pathogène si cette valeur est supérieure à 0,50. Cette mesure nous a permis d'observer que les espèces *Staphylococcus aureus* et *Streptococcus agalactiae* sont les pathogènes les plus fréquemment rencontrées dans les infections mammaires et néonatales mais aussi les commensales les plus fréquemment rencontrées dans le tissu mammaire et le lait maternel avec des VPI de 0,5. Les espèces *Salmonella enterica*, *Burkholderia ambifaria* et *Mycolicibacterium gilvum* étaient associées aux infections avec de fortes VPI de 0,93 tandis que les espèces du genre *Bifidobacterium* et *Lactobacillus* sont associées à une absence d'infection avec des VPI presque nulles; ce sont les probiotiques les plus couramment rencontrés dans le colostrum et le lait maternel et les feces des nouveau-nés.

Afin d'acquérir de l'expérience dans le domaine de la culturomics, nous avons d'abord participé à des travaux déjà en cours au sein de notre laboratoire sur l'analyse du microbiote digestif. Cet travail portait sur l'étude du microbite digestif des patients obèses français avant et après traitement par la chirurgie bariatrique et de selles fraichement émises d'un jeune français en bon état de santé apparent. Ainsi, nous avons cultivé les selles avant et après chirurgie bariatrique par l'approche de «microbial culturomics». Au cours de ce travail préliminaire, nous avons isolé 13 nouvelles espèces bactériennes dont 11 isolées dans les selles des patients obèses et 2 dans la selle du jeune Français (Figure 1). Grâce à cette approche nous avons renommé 5 espèces déjà connues et mal classées (*Ruminococcus faecis*, *Ruminococcus gnavus*, *Ruminococcus lactaris*, *Ruminococcus torques* et *Clostridium glycyrrhizinilyticum*) dont certaines (*Ruminococcus gnavus*) jouent un rôle important dans la santé humaine telle que l'obésité. L'analyse phylogénétique du genre *Ruminococcus* basée sur des données de séquence de l'ARN ribosomal 16S (18) et une concaténation de 271 gènes orthologues provenant de 28 espèces (voir article VI) ont révélé deux clusters distincts et non apparentés. Le premier cluster est composé d'espèces appartenant à la famille *Ruminococcaceae* avec l'espèce type du genre: *Ruminococcus flavefaciens*, *Ruminococcus albus*, *Ruminococcus bromii*, *Ruminococcus*

callidus et *Ruminococcus champanellensis;* et le second cluster comprend: *Ruminococcus faecis*, *Ruminococcus gnavus*, *Ruminococcus lactaris*, *Ruminococcus torques*, et notre nouvelle espèce appartiennent à la famille *Lachnospiraceae*. Cela montre bien que l'approche de la «microbial culturomics» peut contribuer à faire progresser la taxonomie.

La «microbial culturomics» nous a également permis d'isoler et d'écrire selon l'approche taxonogénomique un nouveau genre de bactérie (*Fournierella*) proche de *Faecalibacterium prausnitzii* (19) productrice de butyrate. Les bactéries productrices de butyrate jouent un rôle importantes dans la santé du côlon et, leur absence contribue à l'émergence de certaines maladies, comme la recto-colite nécrosante et le diabète de type II. Cela pourrait s'explique par le fait que le butyrate est la source d'énergie préférentielle des cellules épithéliales (20,21).

Nous avons utilisé l'approche de la «microbial culturomics» couplée à l'analyse métagénomique ciblé pour explorer le microbiote du colostrum et du lait maternel des femmes allaitantes en bonne état de santé au Mali et en France. Pour cela, nous avons collecté 154 échantillons de colostrum et de lait auprès de 134 françaises et 10 maliennes. La composition microbienne (archée et bactérie) de ces échantillons a été analysée par les approches de la «microbial culturomics» et de métagénomique. La culture nous a permis d'isoler 137 espèces de bactéries. Sur les 137 espèces bactériennes, 113 ont été isolées dans les prélèvements français dont une nouvelle espèce bactérienne (*Veillonella massiliensis*) et 37 dans les prélèvements maliens dont quatre nouvelles espèces bactériennes (*Acidipropionibacterium timonense*, *Anaerolactibacter massiliensis*, *Galactobacillus timonensis* et *Lactimicrobium massiliensis*) (Figure 2). Nous avons isolé deux espèces d'archée méthanogène; *Methanobrevibacter oralis* et *Methanobrevibacter smithii* du colostrum et du lait maternel frais des femmes françaises par une technique de culture spécifique mis au point par au sein de notre laboratoire. Les archées méthanogènes sont des procaryotes anaérobies stricts et leur culture reste encore fastidieuse. Elles jouent un rôle important dans la santé humaine, notamment par *Methanobrevibacter smithii* qui est présent dans le tube digestif de 95% des Hommes (22). Elles agissent, sur la digestion bactérienne des polysaccharides alimentaires par la consommation de molécules dihydrogène produites par des bactéries productrices de butyrate telles que *Bacteroides thetaiotaomicron*, accélérant ainsi la récupération d'énergie par les cellules intestinales (23,24). Pour la première fois, à notre connaissance, nous avons pu isoler ces micro-organismes de culture fastidieuse et cela grâce au milieu de culture SAB (25) inventé dans notre laboratoire. *Methanobrevibacter smithii* a été isolée chez près de 30% des mamans contre 1% pour *Methanobrevibacter oralis*. Nous avons séquencé le génome d'une souche de

Methanobrevibacter smithii et de *Methanobrevibacter oralis* du lait et comparé avec le génome d'une souche du tube digestif pour *Methanobrevibacter smithii* et la souche de référence pour *Methanobrevibacter oralis*. Nous n'avons pas observé de différence entre les souches du lait et les souches du digestif et de la bouche. Nous avons également isolé la bactérie *Listeria monocytogenes* responsables de la listériose humaine. Cette bactérie est fréquemment retrouvée dans l'environnement (sols, végétaux, eaux usées), dans les aliments et dans le tube digestif de l'homme. La listériose se transmet par la consommation d'aliments contaminés (26). Chez les personnes immunocompétentes, elle provoque un trouble digestif à type de diarrhée de douleur abdominale accompagné de fièvre et de douleur musculaire. Chez les personnes plus vulnérables (enfants, personnes âgées, femmes enceintes et immunodéprimés) elle entraine des manifestations plus graves et ou le décès (27–29). *Listeria monocytogenes* a été isolée dans 9 échantillons sur dix (9/10, 90%) du Mali par la méthode de culturomics. Tous les 10 échantillons étaient positifs par les différentes techniques d'analyse testées cependant les 10 échantillons de France que nous avons analysés étaient tous négatifs. Nous avons confirmé ces résultats par la culture sur un milieu spécifique au Listeria; la gélose PALCAM (Oxoid Deutschland GmbH, Am Lippeglacis 4-8, 46483 Wesel, Germany) et par la PCR quantitative spécifique. Nous avons également procédé au séquençage génomique des différentes souches que nous avons isolées et comparées avec au génome de l'espèce type.

Notre thèse sera présentée sur deux axes principaux. La première est consacrée à l'étude du microbiote du colostrum et du lait maternel par l'approche de culture à travers la «microbial culturomics» et la metagénomique. La deuxième partie est consacrée à la description taxono-génomique de nouvelles espèces bactériennes.

Figure 1: Analyse phylogénétique moléculaire par la méthode de "Maximum Likelihood" des

Butyricimonas phoceinsis strain AT9 (LN881597)
Hugonella massiliensis strain AT8 (LN881601)
Anaerotruncus massiliensis strain AT3 (LN866995)
Marseillobacillus timonensis strain Marseille-P2399 (LT161891)
Fournierella massiliensis strain AT2 (LN846908)
Enterococcus massiliensis strain AM1 (LN833866)
Paenibacillus ihumii strain AT5 (LN881615)
Negativicoccus massiliensis strain Marseille-P2082 (NR 144729)
Bariatricus massiliensis strain AT12 (NR_144735)
Drancourtella timonensis strain AT6 (LN870296)
Mediterraneibacter massiliensis strain Marseille-P2086(LN881607)
Lachnolostridium bouchesdurhonense strain AT4 (LN881614)
Eisenbergiella massiliensis strain AT11 (NR 144731)

● Bactéries isolées dans les selles des patients obèses.
● Bactéries isolées dans les selles fraiches.

nouvelles bactéries isolées de la selle fraiche et des selles des patients obèses.

Anaerolactibacter massiliensis strain Marseille-P4302 (LT960585)
Lactimicrobium massiliense strain Marseille-P4301 (LT996090)
Galactobacillus timonensis strain Marseille-P4641 (LT934541)
Veillonella massiliensis strain Marseille-P3594 (LT714171)
Acidipropionibacterium timonense strain Marseille-P4482 (LT996089)
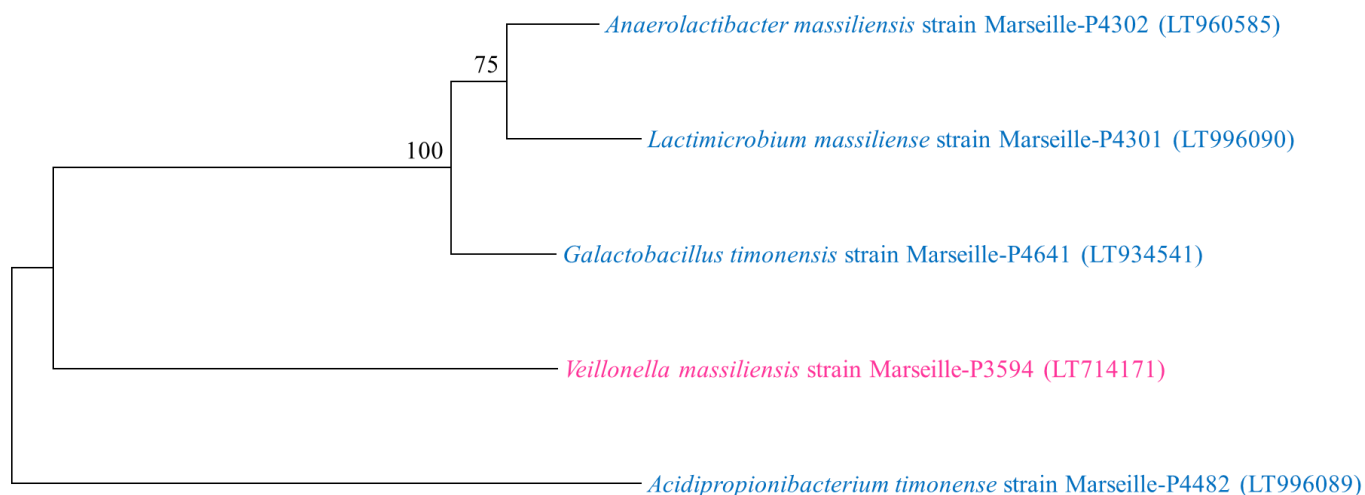
● Bactéries isolées du lait (Mali).
● Bactéries isolées du colostrum (France).

Figure 2: Analyse phylogénétique moléculaire par la méthode de "Maximum Likelihood" des nouvelles bactéries isolées de colostrum et du lait maternel.

Partie I

Caractérisation du microbiote du colostrum et du lait maternel, des femmes allaitantes en bonne santé en France et au Mali, par culturomics et metagénomique.

Avant-propos

Au cours de ce travail, nous avons fait le point sur la connaissance du microbiote (bactéries et archées) du tissu mammaire et du lait maternel. Pour ce faire, nous avons réalisé une recherche bibliographique avec les moteurs de recherche MEDLINE et Pubmed pour identifier l'ensemble des articles publiés sur le microbiote du tissu mammaire et du lait maternel de 1958 à 2018. À cet effet, nous avons utilisé la liste de 20660 bactéries et d'archées isolées au moins une fois chez l'être humain. Ainsi, nous avons formulé des requêtes avec cette liste de bactéries et d'archée et les termes « MESH » en rapport avec le tissu mammaire, le colostrum, le lait maternel, et l'allaitement. Cette requête nous a permis d'identifier 140 articles et 102 articles ont été ajoutés par recherche croisée. Un total de 242 articles a été retenus pour cette revue. Cette revue nous a permis d'établir un premier répertoire bactérien de 820 espèces appartenant à 17 phyla cultivés ou détectés par les techniques de biologie moléculaire dans le tissu mammaire, le colostrum et le lait maternel. Cette revue a fait l'objet d'un article publié dans le journal Future Microbiology (**Article I**).

Nous avons exploré le microbiote du colostrum et lait maternel des femmes allaitantes en bonne santé au Mali et en France. Nous avons analysé la composition microbienne (archées et bactéries) de ces échantillons par les approches de « microbial culturomics » et de métagénomique (**Article V**). La culture nous a permis d'isoler 139 espèces de bactéries dont cinq nouvelles bactéries, qui ont été décrites selon les approches taxonogénomique (**Figure 1, Articles VI et VII**) et deux espèces d'archée méthanogène (*Methanobrevibacter oralis* et *Methanobrevibacter smithii*) (**Article II**). Nous avons également isolé la bactérie *Listeria monocytogenes* (**Article III**). Nous avons testé certaines souches de bactéries (*Ihuella massiliensis*, *Alistipes senegalensis* et *Phascolarctobacterium faecium*) à la recherche d'effet anti-listeria connue chez certaines bactéries comme *Lactobacillus brevis*, *Enterococcus hirae* and *Pediococcus pentosaceus* (30,31) contre les différentes souches isolées du lait maternel (**Article IV**).

14

Article I: Revue

Repertoire of Human Breast and Milk Microbiota: A Systematic Review.

Amadou Togo, Jean-Charles Dufour, Jean-Christophe Lagier, Gregory Dubourg, Didier Raoult, Matthieu Million.

Repertoire of human breast and milk microbiota: a systematic review

Amadou Togo[1,2], Jean-Charles Dufour[3], Jean-Christophe Lagier[1,2], Gregory Dubourg[1,2], Didier Raoult[1,2] & Matthieu Million*[1,2]

[1]IHU-Méditerranée Infection, Marseille, France
[2]Aix Marseille Univ, IRD, APHM, MEPHI, Marseille, France
[3]Aix Marseille Univ, APHM, INSERM, IRD, SESSTIM, Hop Timone, BioSTIC, Marseille, France
*Author for correspondence: matthieumillion@gmail.com

Breastfeeding is a major determinant of human health. Breast milk is not sterile and ecological large-scale sequencing methods have revealed an unsuspected microbial diversity that plays an important role. However, microbiological analysis at the species level has been neglected while it is a prerequisite before understanding which microbe is associated with symbiosis or dysbiosis, and health or disease. We review the currently known bacterial repertoire from the human breast and milk microbiota using a semiauto-mated strategy. Total 242 articles from 38 countries, 11,124 women and 15,489 samples were included. Total 820 species were identified mainly composed of Proteobacteria and Firmicutes. We report variations according to the analytical method (culture or molecular method), the anatomical site (breast, colostrum or milk) and the infectious status (healthy control, mastitis, breast abscess, neonatal infection). In addition, we compared it with the other human repertoires. Finally, we discuss its putative origin and role in health and disease.

Breastfeeding has been associated with the health of the child and the mother, regardless of geography and socioeconomic level [1], child's growth and cognitive development [1,2]. Human studies suggest a protective effect against diarrhea, necrotizing enterocolitis, otitis media and respiratory infections in the short term and against leukemia, malocclusion, inflammatory bowel disease, malnutrition including kwashiorkor, obesity and diabetes in the long term [1,3,4]. Up to half of deaths caused by infections in children aged 6–23 months may be associated with the absence of adequate breastfeeding [1]. Breastfeeding has also been associated with a protective effect for mothers against breast cancer and possibly against ovarian cancer and diabetes [1].

In addition to nutrients, bioactive molecules such as human milk oligosaccharides, human maternal cells but also extracellular vesicles [5–8], milk contains many microbes and is therefore not sterile. The importance of the microbial diversity of human milk for the health of the offspring has been largely neglected to date. More recently, an increasing number of studies have reported an unsuspected diversity, including many health-promoting bacteria (probiotics) in the breast, colostrum or milk [9–11] but also bacteria usually considered pathogenic and frequently found in healthy controls [12]. The milk microbiota likely plays a critical role in the colonization of the child's digestive tract and in the development of its immunity [13–16]. Prolonged exclusive breastfeeding is associated with reduced diarrhea-related gut microbiota dysbiosis and microbiota differences that persist for life [8]. The disruption of this ecosystem has also been associated with maternal diseases, such as lactational mastitis and breast cancer [17–20]. Breast bacteria could also play a role in maintaining healthy breast tissue, including stimulating host immunity [17].

Many questions remain unanswered regarding the microbiome, such as the fact that microbes considered pathogenic are frequently found in controls, the mechanism by which the microbiota impacts immune development and how dysbiosis leads to gut inflammation [8]. Researchers suggested that future studies should employ metagenomics, metatranscriptomics and metabolomics approaches to understand the complete taxonomical, func-

Future Microbiol. (Epub ahead of print)

tional and metabolic profile [8]. However, all these approaches are useless if the microbes have not been previously cultured, isolated and identified.

This led us to set up microbial culturomics consisting of multiple culture conditions combined with the rapid identification of bacteria [21]. We devised that, by using different atmospheres, temperatures, pH, nutrients, minerals, antibiotics or phages, 'microbial culturomics' could provide comprehensive culture conditions simulating, reproducing or mimicking the entirety of selective constraints that have shaped each human microbiota for thousands of years [21]. Microbial culturomics is the best way to capture the functional and viable human-associated microbiota biodiversity of each human individual through large-scale isolation of microorganisms. The further sequencing of their genomes enables in addition to extract the deepest genetic biodiversity [22,23]. In many situations, culturomics can detect species that are not identified by metagenomics [24]. While culturomics is not adequate for quantification of the concentration or relative abundance of each species, it is the best method to establish the microbial repertoire in a microenvironment. This repertoire illuminates the microbial dark matter obtained by metagenomics [25]. Before deciphering mechanisms, the establishment of the most comprehensive microbial repertoire of each human niche is an indispensable prerequisite for the interpretation of the microbiota profile and of the normal or abnormal presence, concentration and relative abundance of any microbe.

Here, our objective was to establish a microbial species-level repertoire of the human breast and milk microbiota and its variations according to the method of detection, the anatomical site (breast or milk) and the infection status (i.e., breast abscess, mastitis and neonatal infection).

The repertoire of all bacteria and archaea isolated at least once from the human body has already been reported [26,27]. The repertoires of the anatomical niches have been published for the oral, respiratory, digestive, urinary tract and vagina [23,26–28]. Therefore, we used an automatic and comprehensive method to review all the literature that has used culture (isolation) and DNA-based (detection) methods from human breasts and milks of different geographical origins. We took this opportunity to compare methods of detection (culture or molecular), anatomical compartment (breast or milk) and infection status (healthy mother–child pair, mastitis, breast abscess and neonatal infection). In addition, we analyzed the first isolation site of each species and compare the repertoire of breast and milk with the other anatomic repertoires already reported (oral, respiratory, gut, urinary and vagina). It is already known that the human milk contains a diverse and varied microbiota, such as prokaryotes (bacteria, archaea), eukaryotes (fungi and protozoa) and viruses [19]. In this review, we have focused on bacteria and archaea only. As the probiotics effect of *Lactobacillus* was shown to be species specific [29–32], only studies with a species-level resolution were included.

Method & search engine

We conducted a systematic review following the PRISMA guidelines by using an automated search using the list of bacterial and archaeal species isolated at least once from humans, the List of prokaryotic names with Standing in the Nomenclature (http://www.bacterio.net) and the taxonomy of the National Centre for Biotechnology Information (http://www-ncbi-nlm-nih-gov.gate2.inist.fr/taxonomy/?term=) as previously reported [26,27]. The list thus established included 20,660 bacterial and archaeal species. These prokaryotes were searched in all studies indexed in PubMed from 1 January 1953 to 31 December 2017. The exact keywords and automatic search script are given in the additional data (additional file 1). The query pattern was designed to express, with several combinations of keyword and text words, four notions that were expected to be present together in the retrieved articles: prokaryote denomination at the species level; milk or breast; culture techniques and studies carried out on human being.

The eligibility criteria were as follows: a primary study with title and/or abstract reported in English (no language restriction on full text); concerns only human subjects; studies microbial diversity in breast milk, colostrum and breast tissue regardless of the detection method used and reports primary results at the bacterial and archaeal species level. All studies that did not meet these inclusion criteria, systematic reviews, personal opinions, letters to the editor and animal studies were excluded. 'Breast tissue' was defined as samples obtained from breast abscesses and mastectomy. 'Milk' was defined as samples obtained from milk expressed manually or with a breast pump. The primary objective was to determine the repertoire of bacteria and Archaea at the species level detected in the breast or human milk, regardless of the method (culture or molecular method), anatomical site (breast tissue or milk) or the infection status (healthy mother–child pair, mastitis, breast abscess or neonatal infection). We subsequently analyzed the variation of the repertoire according to different methods, anatomical site and infection status.

To achieve all these objectives, and because individual data are frequently lacking in published studies on this topic, we defined a study group as the combination of an article (PMID), a method, an anatomical site and an infection status. Thus, some articles included more than one study group (comparison of culture and molecular methods, case-control studies on mastitis, etc.). The study group was the unit of our systematic review. Accordingly, all proportions, frequencies and percentages are not %individuals but are %study groups. The predictive value of infection (PVI) for each microbe was calculated using the following formula: (number of study groups with infection positive for this microbe/[number of study groups with infection positive for this microbe + number of study groups of healthy controls positive for this microbe]). Accordingly, a probiotic was expected to have a low PVI ($<<0.5$), a pathogen was expected to have a high PVI ($>>0.5$).

The data collected were PMID, title, first author's name, year, country and size of each study group (number of mothers), analytical technique (culture or molecular), anatomical niche (breast or milk), infectious status (without infection: healthy control, sick mothers without infection (mainly breast cancer) – with infection: mastitis, abscess, neonatal infection). For each study group defined as above, the detected bacterial or archaeal species were identified regardless of their abundance or frequency in the study group. No threshold was applied for relative abundance as species detected in very low abundance may be of critical importance for human health [22].

The bacterial species of breast and human milk identified have been classified by taxonomy (branch, class, order, family and genus) according to the list of prokaryotic names with standing in the nomenclature (http://www.bacterio.net) [33,34] and the NCBI taxonomy website (http://www-ncbi-nlm-nih-gov.gate2.inist.fr/taxonomy/?term=).

We recently observed that the gut dysbiosis is associated with an enrichment of microbes tolerant to oxygen and abnormal redox potential [35,36]. In order to evaluate the proportion of microbial species tolerant to oxygen in the normal and abnormal (infected) breast microbiota and human milk, we assigned each species its oxygen tolerance status: strict anaerobic or aerotolerant, according to a previously published database [35]. We subsequently compare the proportion of strict anaerobes in the breast and milk repertoire in healthy controls and in infection.

Results

A total of 17,023 articles (additional file 2) were identified corresponding after elimination of duplicates (an article can be automatically included many times if it mentions many species), to 4812 unique articles recovered by automated search (Figure 1). A total of 4672 articles were excluded based on the title or after reading the summary, so that 140 articles were identified by automated search and 102 articles were added by cross reference search. These 242 articles included 142 for milk and 100 for breast tissue (Table 1 & additional file 3). Among included articles, 226 included only one study group (only one method (culture or molecular method) for only one anatomical site (breast or milk) and only one infection status (infection or absence of infection)). As ten articles reported both culture and molecular methods (three for breast tissue and seven for breast milk), and six reported both infected and noninfected microbiota, 258 study groups were included (Table 2). Total 101 study groups included healthy controls, 149 included a mother in the context of an infection (mastitis, breast abscess or neonatal infection), seven included mothers without infections but with another disease (mostly breast cancer). For one study, the method and the infectious status could not be determined (Table 2). No article analyzing both breast tissue and milk were identified. No unpublished documents were identified during this investigation.

The 242 included articles were published between 1953 and 2017 in 38 countries (Table 1 & additional file 3). Several designs were identified including interventional, observational and microbiological studies. Case reports account for 61% (61/100) of included studies on breast tissue and 18% (26/142) for breast milk. The 100 articles on breast tissue microbiota included 1941 women and 2006 samples, while the 142 articles on breast milk microbiota included 9183 women and 13,483 samples. Regarding the technique used, culture was the most frequently used (192 studies) while 14 studies used 16S rRNA gene amplicon sequencing and three used nontargeted (shotgun) metagenomics.

The human breast & milk repertoire

The 258 study groups from the 242 articles included yields 820 bacterial and archaeal species belonging to 17 phyla; Acidobacteria, Actinobacteria, Bacteroidetes, Cyanobacteria, Deferribacteres, Deinococcus–Thermus, Fibrobacteres, Firmicutes, Fusobacteria, Gemmatimonadetes, Planctomycetes, Proteobacteria, Spirochaetes, Synergistetes, Tenericutes, Thermotogae and Verrucomicrobia; 33 classes; 72 orders; 137 families and 303 genera identified by culture or by molecular methods (additional file 4). The ten most frequently reported bacterial species in breast tissue and breast milk were *Staphylococcus aureus* (26% study groups), *Staphylococcus epidermidis* (23%),
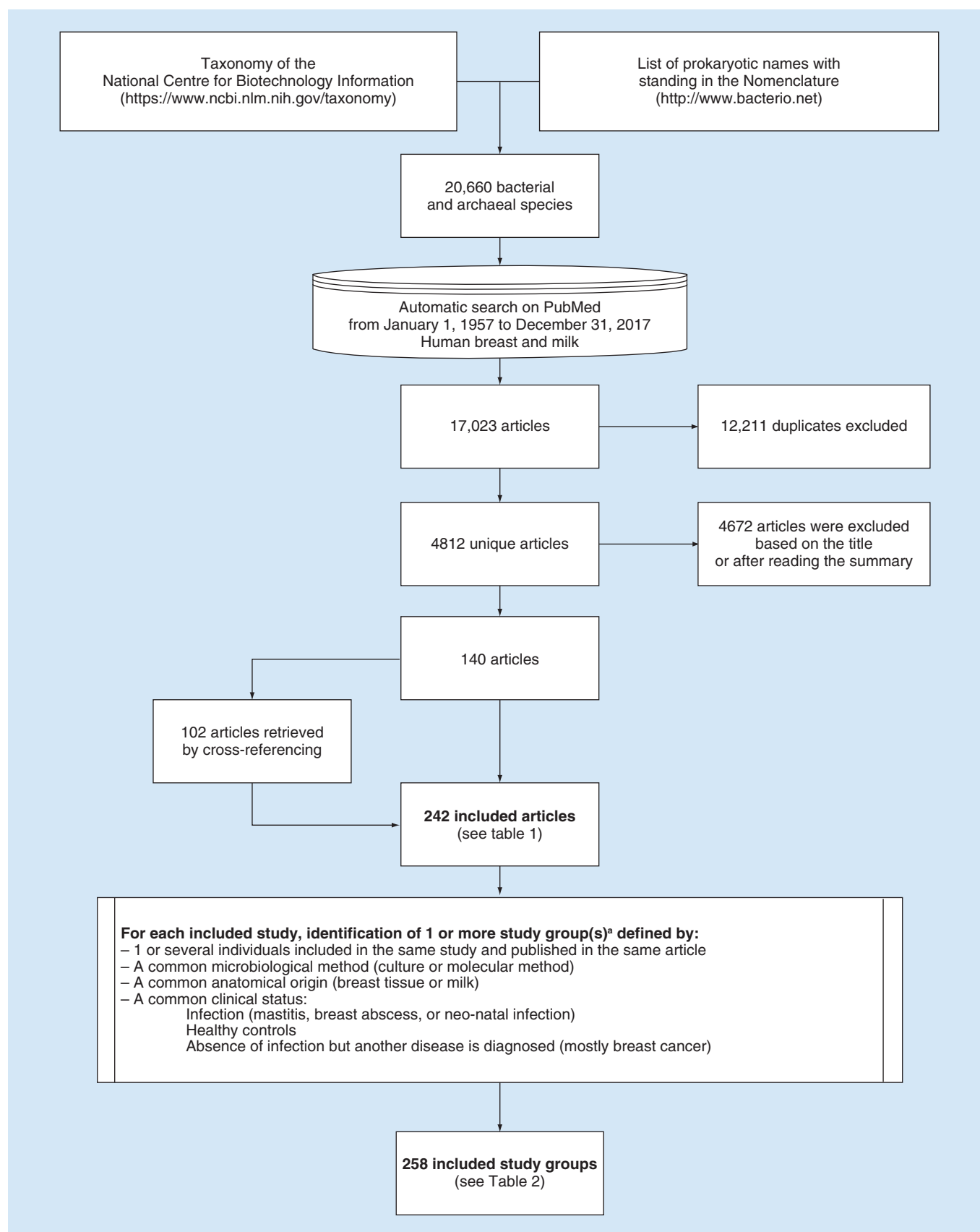
Figure 1.   Study flowchart.

Table 1. Summary of the 242 included studies.		
	Breast tissue (n = 100)	**Colostrum and milk (n = 142)**
Years of publication	1953–2017	1957–2017
Countries of publication[†]	31 countries USA (17 studies), India (13), Spain (10), UK (10), Canada (5), France, New Zealand (4), Argentina, Germany, Turkey (3), Australia, China, Japan, Kuwait (2), Belgium, Brazil, Colombia, Czech Republic, Denmark, Greece, Iran, Israel, Italy, Mexico, Pakistan, Qatar, Russia, Saudi Arabia, Singapore, Taiwan, The Netherlands (1)	38 countries Spain (36 studies)[‡], Italy (11), USA (10), France, Germany (8), UK (7), Brazil (6), Finland (5), Australia, India, Russia, Turkey (4), Canada, China, Taiwan, Czechoslovakia (year of publication 1987 & 1990), Japan, Nigeria, Switzerland (2), Argentina, Austria, Chile, Czech Republic, Denmark, Estonia, Iran, Ireland, Kenya, Korea, Malaysia, Mozambique[‡], Nepal, Norway, Saudi Arabia, Serbia, Slovenia, Sweden, Yugoslavia (1)
Design[†]		
Treatment studies (interventional studies)	0	4
Observational studies		
Descriptive		
Case reports	61	26
Case series	26	12
Analytical		
Cohort studies	3	5
Case-control studies	6	6
Cross-sectional study	0	56[§]
Microbiological studies[¶]	2	27[§]
Other or undetermined	2	7
Total numbers		
Individuals (women)	1941	9183
Samples (breast tissue or milk samples)	2006	13,483
Microbiological methods[†]		
Culture	85	107
16S rRNA gene amplicon sequencing	5	9
Nontargeted metagenomics (shotgun)	0	3
Others and not provided	13	26

[†] Some studies included more than one country, design or microbiological methods.
[‡] One study from Spain and Mozambique.
[§] One study includes a cross-sectional design and a microbiological study.
[¶] Microbiological study defined as a study focusing on microbes and not on individuals (for instance, description of genome of a bacterium isolated from human milk).

Streptococcus agalactiae (17%), *Cutibacterium acnes* (10%), *Enterococcus faecalis* (10%), *Bifidobacterium breve* (9%), *Escherichia coli* (9%), *Streptococcus sanguinis* (9%), *Lactobacillus gasseri* (8%) and *Salmonella enterica* (8%). Most of the bacterial species isolated in the breast and milk were members of four phyla bacteria: Proteobacteria (270 species), Firmicutes (268 species), Actinobacteria (203 species) and Bacteroidetes (47 species), respectively. The remaining 32 species were shared between the other 13 phyla mentioned above with one to eight species per phylum (Figure 2 & additional file 5).

Some species of these rare phyla seem particularly important. For instance, *Mucispirillum schaedleri* was the only species of the Deferribacteres phylum and was detected both by culture and molecular methods in both milk and breast tissue (additional file 4 & 6). This species was identified in termites, cockroaches, mice, turkeys, dogs, pigs and goats and specifically colonizes the gut mucus layer with a mucolytic activity and specific interrelationships with mucosal immunity [37,38]. Similarly, *Akkermansia muciniphila* was the only species of the phylum Verrucomicrobia to be identified in both breast and milk. This species, which also colonizes gut mucus with mucolytic activity, is considered a symbiont with an antiobesity effect in humans [39].

The repertoire of breast tissue and breast milk was dominated by facultative anaerobic and aerobic bacteria. After exclusion of four obligate intracellular, 22% (177) of the 816 species were strict anaerobes (23% [164/719] for species detected by molecular method, 14% (39/274) for species detected by culture). Overall, the proportion of strict anaerobes decreased with infection (40/241 [16%]) compared with healthy controls (124/485 [25%],

Table 2. Study groups included†.

Healthy controls	101
Breast tissue	4
–Culture	2
– Molecular methods	2
Breast milk	97
– Culture	78
– Molecular methods	19
Infection	**149**
Breast tissue	97
– Culture	90
– Molecular methods	7
Breast milk	52
– Culture	43
– Molecular methods	9‡
Cases without infection	**7**
Breast tissue	4
– Culture	1
– Molecular methods	3
Breast milk	3
– Culture	1
– Molecular methods	2
Unknown	**1§**

†Ten studies included two study groups.
‡Two study groups were atypical: *Borrelia burgdorferi* was detected by PCR in breast milk of mother with erythema migrans (PMID7648832), *Mycobacterium avium* was detected in the breast milk of three mothers with Crohn's disease (PMID10763975).
§*Mycobacterium leprae* was reported from human milk but full text, infectious status and technique could not be obtained (article dating back to 1967–PMID4864195).
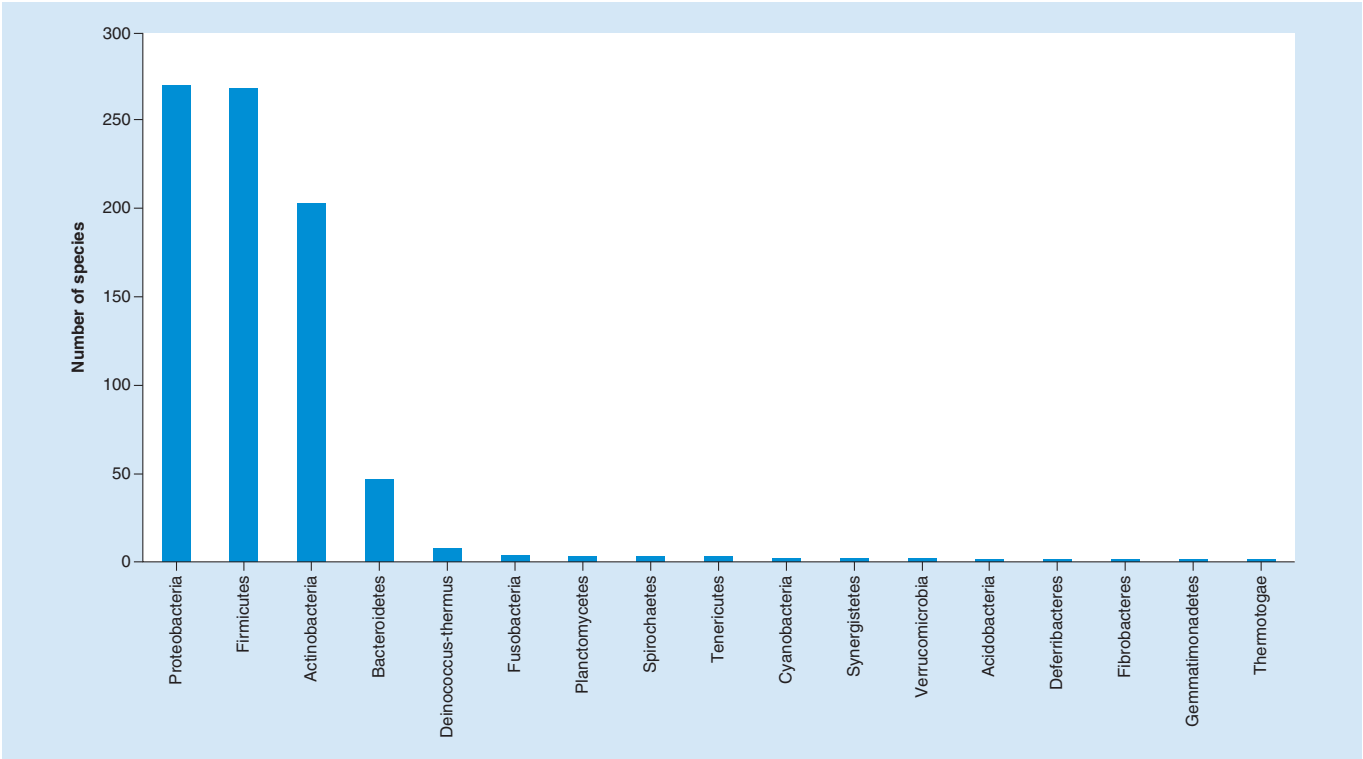
Figure 2.   Bacterial species distribution according to their phylum. Based on the National Center for Biotechnology Informatiotaxonomy and List of Prokaryotic names with Standing in Nomenclature classification reported in this repertoire.
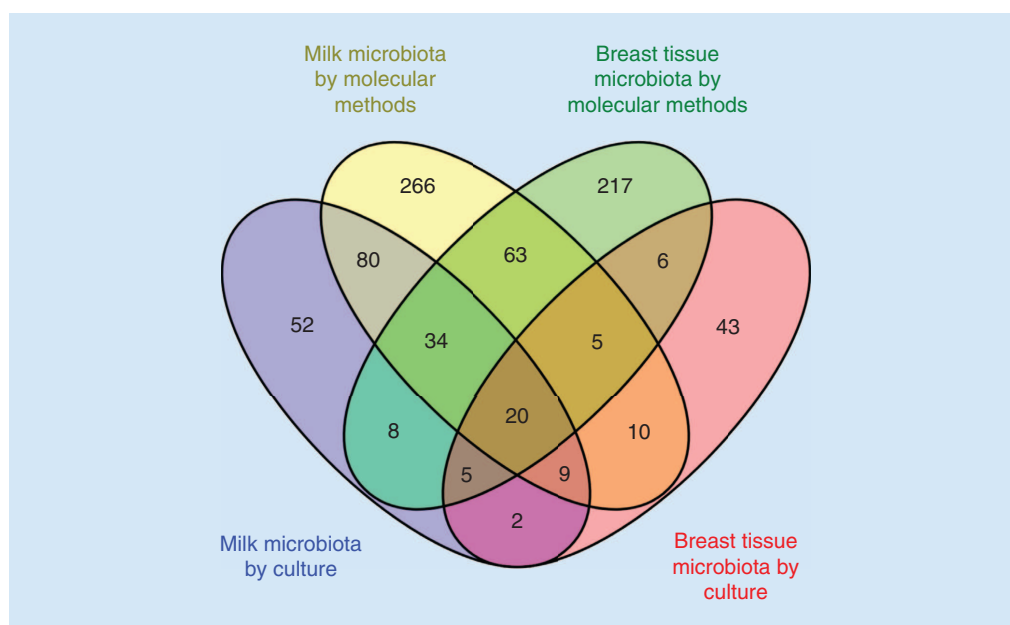
Figure 3.   Comparison of breast and milk repertoires according to culture or molecular technique. Venn diagram approach of methods (culture and molecular) used for bacterial identification in breast and breast milk.

two-sided Chi square test, p = 0.006). This difference was observed only for molecular methods (30/191 [16%] vs 122/437 [28%]) but not for culture (20/151 [13%] vs 26/209 [12%]).

The predominance of bacteria tolerant to oxygen contrasts with the predominance of strict anaerobes in the digestive microbiota. The anaerobic gut microbiota is characteristic of host maturation and health [40]. The important members of this healthy human mature anaerobic gut microbiota correspond to Akkermansiaceae (*A. muciniphila*) and Clostridium cluster IV and XIV like Ruminococcaceae (*Faecalibacterium prausnitzii*) and Lachnospiraceae (*Mediterraneibacter gnavus* (formerly named *Ruminococcus gnavus*) [41], *Blautia obeum*, *Roseburia faecis*) [40]. Even if present in low abundance, these species were detected in breast or milk by molecular method (additional file 6) [19,42–45]. However, none of these highly oxygen-sensitive bacteria have been cultured from breast tissue or human milk in the literature.

Archaea are among the most sensitive to oxygen prokaryotes and are common inhabitants of the human gut, being key members of the host–microbial mutualism [35,46,47]. Some metagenomic studies have detected archaeal sequences assigned at the genus level, but not at the species level, in human milk. Ward *et al.* 2013 reported the presence of methanogenic archaea in the metagenome of human breast milk [48]. Jiménez *et al.* 2015 reported the presence of archaea in the breast milk of healthy women with a depletion in mastitis [19]. No archaea have been identified either from the breast or from the breast milk by the culture method.

Complementarity of culture & molecular methods

The culture methods allowed isolating 24% (100/422) of the bacterial species detected in breast tissue and 38% (210/554) in breast milk, while molecular methods detected 85% (358/422) in breast tissue and 88% (487/554) in breast milk, respectively (Figure 3). Molecular methods seem more sensitive, but the two approaches are nevertheless complementary. Indeed, 76% (322/422) species from breast tissue and 62% (344/554) from breast milk were detected only by molecular methods, while 15% (64/422) from breast tissue and 12% (67/554) from breast milk were detected only by culture. The overlap between the two methods was therefore very limited as only 9% (36/422) and 26% (143/554) of the species were detected by both techniques for breast tissue and milk, respectively. This significant discrepancy between culture and molecular methods has already been observed in all our studies where we compared microbial culturomics and metagenomics with a usual overlap of about only 15% [21–23].

Faecalibacterium prausnitzii was detected by molecular method in four studies on breast milk and two studies on breast tissue but never in culture. *Akkermansia muciniphila* was detected in one breast tissue study and three breast milk studies but never in culture (additional file 6). Overall, future studies will need to determine whether

the many species detected only by molecular methods correspond to living microbes, in particular by using the microbial culturomics approach [21–23], and whether the few species detected only by culture correspond to species poorly detected by amplicon sequencing (relative abundance below the detection threshold of current machines ($\sim 10^6$ reads), primer bias, etc... [22]) or whether they are contaminants.

Comparison between the breast & milk repertoires

Of the 820 bacterial species, 554 were identified in breast milk and 422 in breast tissue (Figure 3 & additional file 6). The 554 species detected in human milk belonged to 13 phyla, 24 classes, 52 orders, 92 families and 178 genera. The 422 species of breast tissue belonged to 13 phyla, 27 classes, 62 orders, 119 families and 220 genera. The details of the taxonomic classification are presented in the additional file 4. We noted that breast tissue and breast milk shared only 156 bacterial species (Figure 3). This difference may result from changes in the mammary microbiota during pregnancy. Indeed, Koren *et al.* 2012 observed a drastic change in the intestinal microbiota between the first and third trimester of pregnancy with an increase in bacterial diversity [49].

Variation in the breast & milk microbiota during infection

The repertoire associated with healthy mother–child pairs without infection included 486 species while that associated with infection included 241 species (additional file 7). Total 112 species were common to both groups, 374 were found only in controls and 129 were found only in cases of infection.

Surprisingly, *S. aureus* and *S. agalactiae*, usually considered as pathogens, were not identified more frequently in infection (*S. aureus*: 34% in healthy controls vs 22% in infection, bilateral Chi-square test, p < 0.05 – *S. agalactiae*: 21 vs 14%, p = 0.18). We sought to clarify if the reason for *S. agalactiae* and *S. aureus* to be predominant was because of the interest around infection during the neonatal period and therefore enrichment culture media used to identify them which could eradicate the gram negatives. We therefore focus on molecular studies of healthy controls. Such studies would have not been biased by any interest around infection and enrichment culture media that could eradicate the gram negatives. We found 101 study groups of healthy controls analyzed by molecular studies (Table 2): The four most frequent species were *S. epidermidis* (36%), *S. aureus* (34%), *B. breve* (22%) and *S. agalactiae* (21%). This confirms that *S. aureus* and *S. agalactiae* are both among the most frequent commensals and pathogens of human breast and milk questioning the definition of such species 'pathogen' or 'commensal'. This is also observed by the fact that the PVI for these two species was very close to 0.5 (0.48 for *S. aureus*, 0.5 for *S. agalactiae* – additional file 6).

In contrast, *S. enterica* and *Burkholderia ambifaria* were the only two species detected only in case of infection (PVI = 1, Figure 4). Moreover, *Mycolicibacterium gilvum* had a very high PVI (0.93, p < 0.05).

In contrast, 48 species were significantly associated with the absence of infection and were never detected in infection (PVI = 0), including 11 *Bifidobacterium* and ten *Lactobacillus* (Figure 4). This is consistent with the *in vivo* efficacy of *Lactobacillus* strains for the treatment of mastitis [50] and for the prevention of gastrointestinal, respiratory, total infections and sepsis in infants [8,20].

As mentioned above, the most common bacteria in the breast and milk microbiota (*S. aureus, S. agalactiae, E. coli* – additional file 7) are also the bacteria that are the most often involved in maternal (mastitis) and/or infant (neonatal infection) infections (additional file 6 & 7). Environmental factors may alter the ecosystem of the microbiota of the breast and milk, the behavior of the microbes or the mucosal immunity of the mother and/or the newborn, leading from symbiosis to infection. A typical example is the association between cigarette smoking, recurrent breast abscesses and *C. kroppenstedtii* [51–55]. This species is a member of the healthy mammary microbiota (additional file 6) and is found in 4% of healthy mother–child pairs, 5% in mastitis and 12% in abscess (additional file 6). Cigarette smoking is probably able to cause an alteration of the mammary microenvironment that leads to recurrent infections, and stopping smoking is a key for definitive cure. In addition, it is possible that strains within a species usually present in the breast or milk microbiota may be particularly virulent.

By separately analyzing mastitis, breast abscesses and neonatal infections (Figure 5 & additional file 8), similar results were found. *Bifidobacterium breve* for example was never found in these three infections. *Lactobacillus gasseri* was never found in case of breast abscess or neonatal infection. In contrast, *S. enterica* and *B. ambifaria* were associated with breast abscesses and neonatal infections but not mastitis. *Corynebacterium kroppenstedtii* was specifically associated with breast abscess. Two other *Corynebacterium* (*Corynebacterium propinquum* and *Corynebacterium renale*) were associated with mastitis. Conversely no *Corynebacterium* were associated with neonatal infections. This suggests unspecific and general mechanisms to prevent infection (*B. breve*) but higher specificity

Species	Health (n = 101) (%)	Infection (n = 147) (%)	Predictive value for infection	p-value
Bifidobacterium breve	22	0		****
Bifidobacterium magnum	16	0		****
Lactobacillus pontis	12	0		****
Bifidobacterium bombi	11	0		****
Dechloromonas aromatica	8	0		***
Lactobacillus rossiae	7	0		**
Faecalicatena fissicatena	6	0		**
Bifidobacterium aquikefiri	5	0		*
Bifidobacterium choerinum	5	0		*
Faecalibacterium prausnitzii	5	0		*
Lactobacillus cellobiosus	5	0		*
Stenotrophomonas maltophilia	5	0		*
Akkermansia muciniphila	4	0		*
Bifidobacterium psychraerophil	4	0		*
Bifidobacterium thermophilum	4	0		*
Brevundimonas subvibrioides	4	0		*
Enterococcus faecium	4	0		*
Lactobacillus fermentum	4	0		*
Lactobacillus plantarum	4	0		*
Lactobacillus zeae	4	0		*
Mediterraneibacter torques	4	0		*
Streptococcus tigurinus	4	0		*
Veillonella dispar	4	0		*
Acinetobacter Iwoffii	3	0		*
Bacteroides uniformis	3	0		*
Bifidobacterium animalis	3	0		*
Bifidobacterium callitrichos	3	0		*
Bifidobacterium gallinarum	3	0		*
Bifidobacterium pseudolongum	3	0		*
Corynebacterium freneyi	3	0		*
Corynebacterium striatum	3	0		*
Enterococcus hirae	3	0		*
Haemophilus paraphrohaemoly	3	0		*
Hafnia alvei	3	0		*
Lactobacillus casei	3	0		*
Lactobacillus graminis	3	0		*
Lactobacillus johnsonii	3	0		*
Lactobacillus paracasei	3	0		*
Leuconostoc fallax	3	0		*
Neisseria elongata	3	0		*
Prevotella tannerae	3	0		*
Pseudomonas fragi	3	0		*
Ruminococcus bromii	3	0		*
Streptococcus gordonii	3	0		*
Truepera radiovictrix	3	0		*
Veillonella atypica	3	0		*
Weissella cibaria	3	0		*
Weissella confusa	3	0		*
Lactobacillus gastricus	17	1		****
Bifidobacterium aesculapii	8	1		**
Enterobacter agglomerans	7	1		*
Micrococcus roseus	7	1		*
Streptococcus peroris	7	1		*
Streptococcus pseudopneumoniae	10	1		**
Kytococcus schroeteri	5	1		*
Lactobacillus vaginalis	9	1		**
Lactobacillus gasseri	17	3		****
Streptococcus sanguinis	17	3		****
Enterococcus flavescens	8	1		*
Rothia mucilaginosa	10	2		*
Staphylococcus lugdunensis	12	3		**
Lactobacillus rhamnosus	9	2		*
Bacillus cereus	6	1		*
Gemmata obscuriglobus	6	1		*
Streptococcus parasanguinis	11	3		*
Staphylococcus haemolyticus	8	2		*
Streptococcus oralis	10	3		*
Escherichia coli	13	4		*
Staphylococcus hominis	13	4		*
Streptococcus mitis	13	4		*
Staphylococcus epidermidis	37	12		****
Cutibacterium acnes	15	6		*
Enterococcus faecalis	15	6		*
Staphylococcus aureus	34	22		*
Mycolicibacterium gilvum	1	10		*
Burkholderia ambifaria	0	6		*
Salmonella enterica	0	14		***

Figure 4.   Species associated with infection or absence of infection. Only species with a p-value <0.05 are shown (two-sided Chi-square test). Predictive value for infection was shown in green when a protective association was found (predictive value of infection significantly less than 0.5) and in red when positive association with infection was found (predictive value of infection significantly higher than 0.5). See methods for the calculation for predictive value of infection. *p < 0.05, **p < 0.005, ***p < 0.0005, ****p < 0.00005.
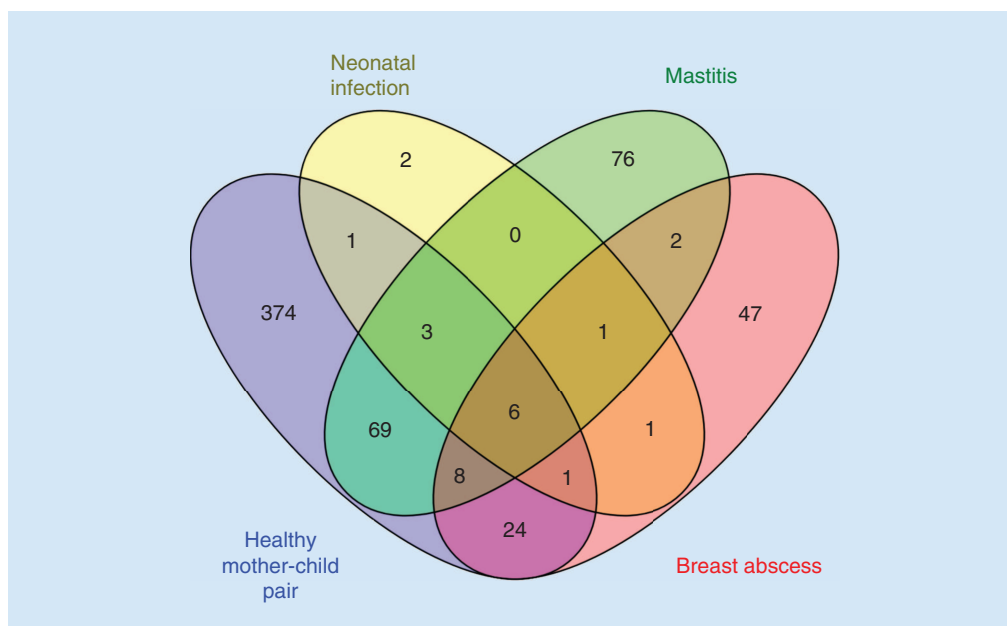
Figure 5.   Comparison of breast and milk repertoires according to absence of infection, mastitis, breast abscess or neonatal infection. Venn diagram approach of infectious status (healthy mother–child pair without infection, mastitis, breast abscess neonatal infection).

for infection. For instance, *Corynebacterium* spp. (lipid-requiring bacteria) infects the breast tissue which is rich in lipid but is not associated with neonatal infections. *Mycolicibacterium gilvum* is associated only with breast abscess.

Comparison of the repertoire of breast & milk with the other repertoires

In order to study the specificity of the milk microbiota and its relationship with other human, animal or environmental microbial niches, we looked in which sample each species included in this study was isolated for the first time. We found that 327 (40%) were first described in humans, including 98 (12%) in the gut, 103 (12%) in clinical samples (suspected infected persons), 47 (6%) in the respiratory tract, 45 (5%) in the oral cavity, 14 (2%) in the urinary tract, 12 (1%) in the skin and 7 (1%) in vagina (Figure 6 & additional file 4). Only one species (0.1%) was initially isolated from human milk; *Streptococcus lactarius* [56]. This suggests that the microbial repertoire of milk has been neglected compared with other human anatomical niches, such as the intestine or the respiratory tract.

Of the 493 (60%) bacterial species not isolated from humans at the time of their initial description, 277 (34%) were first described in the environment, 104 (13%) in animals, 53 (6%) in plants and 46 (6%) in food. For 13 of them (2%), we were unable to obtain the first isolation site. The fact that only one species was initially described from milk suggests that there is no strict anatomical specificity for human microbial repertoires and that a large proportion or even the entire human microbiota is shared between different anatomical niches and with its environment.

To identify the shared and specific microbiota according to the anatomical niche, we compared the breast and milk repertoire with the repertoires of the intestine, oral cavity, respiratory tract, urinary tract and vagina in humans. The human breast and milk microbiota shared 49% (401/820) of the species of its repertoire with the gut, 30% (245/820) with the vagina, 28% (230/820) with the urinary tract, 28% (229/820) with the respiratory tract and 21% (174/820) with the oral cavity (Figure 7 & additional file 9). This is consistent with the fact that it has already been demonstrated that *Lactobacillus*, *Bifidobacterium*, *Streptococcus* and *Staphylococcus* are shared between different anatomical niches of the human body such as the intestine and the oral cavity [11,13,43].

When comparing the breast and milk microbial repertoire with other anatomical niches in the human body, we observed that four *Lactobacillus* were shared by all the anatomical niches studied (milk, mouth, lung, intestine, urinary tract and vagina), namely *L. casei*, *L. fermentum*, *L. gasseri* and *L. rhamnosus* (additional file 10). Surprisingly and interestingly, no *Bifidobacterium* was detected in the respiratory tract, while four *Bifidobacterium* were found
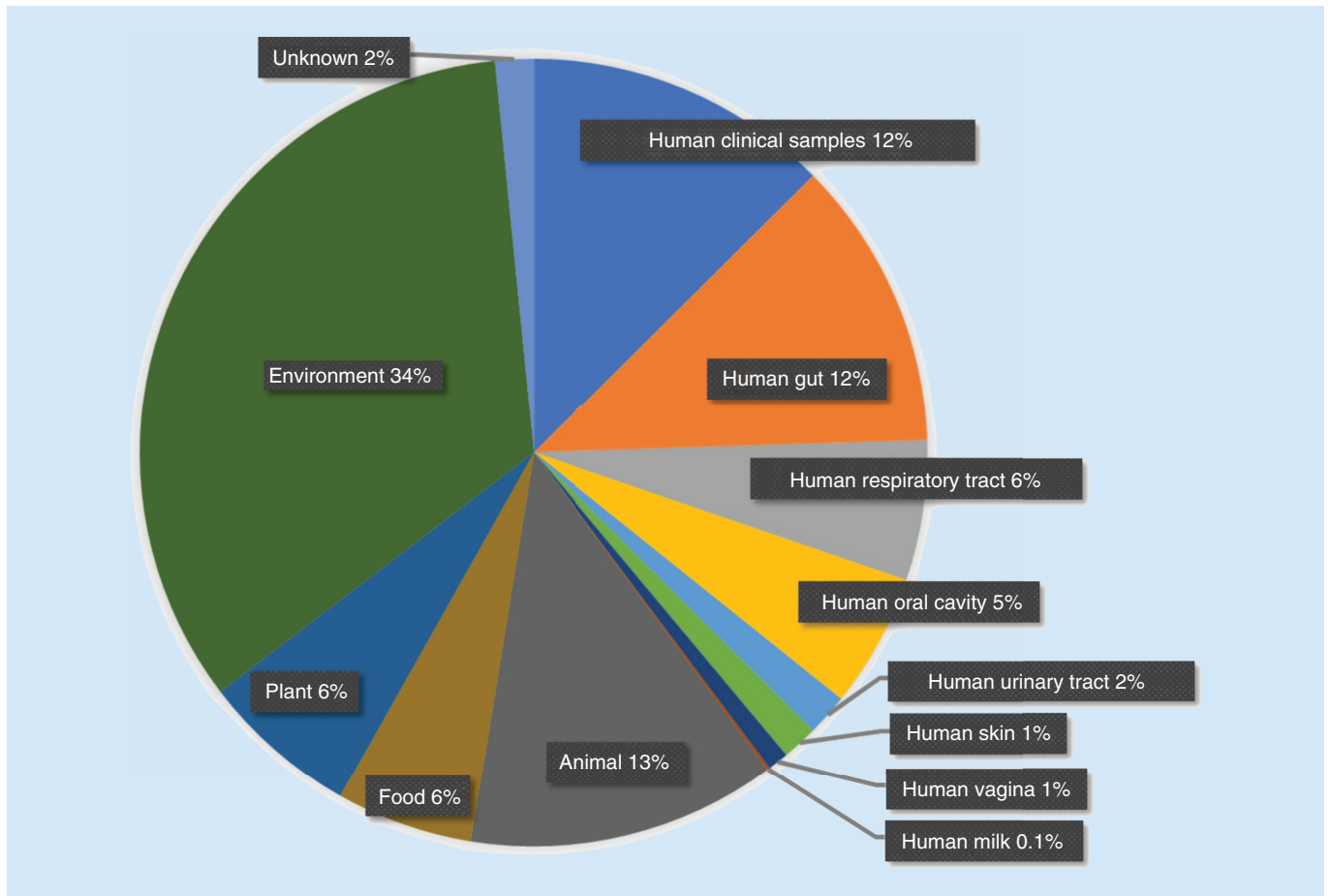
Figure 6.   Proportion of bacterial species identified from breast and breast milk according their first isolation source.

in the milk, mouth, intestine, urinary tract and vaginal repertoires, namely *B. breve, B. longum, B. dentium* and *B. scardovii* (additional file 10). This is consistent with the presence of anaerobic pockets in the mouth but not in the respiratory tract and the fact that *Bifidobacterium* is strictly anaerobic, while all *Lactobacillus* usually tolerate oxygen. This is related to a notable non-heme manganese catalase in *Lactobacillus* species. This suggests that human microbial niches can be connected to each other as a 'microbial network' and that the profile of each repertoire (richness and relative abundance of each species) is adapted to the microenvironment of the corresponding anatomical niche. Finally, *B. breve, B. longum* and *B. dentium* are the three *Bifidobacterium* associated with breastfeeding (vs formula) and babies (vs adults) in one of the largest human gut metagenomic studies (supplementary data of [57]).

Putative origin of breast milk

Bacterial transmission from mother to child through breastfeeding has been proven in several studies [13,14,45,58]. Martín *et al.* in 2003, observed that lactic acid bacteria strains were the same in breast milk, infant's mouth and infant's stools, but were different from those isolated from the mother's skin and vagina [11]. In 2010, Solís *et al.* had found the same genetic profiles by Random amplified polymorphic DNA (RAPD) among species of bifidobacteria in breast milk and stools at different sampling times [59]. These profiles were not shared between the different mother–child pairs included in the study. Jost *et al.* in 2014, found that the same strain of *Bifidobacterium breve* was found in the mother's stool, in breast milk and in the child's stool by culture [13]. Asnicar *et al.* 2017, observed the sharing of certain bacterial strains of *Bifidobacterium bifidum* and *Ruminococcus bromii* between mother and child by metagenomic, pangenomic (PanPhlAn) and genetic (StrainPhlAn) analysis, with comparison of single nucleotide variants (SNVs) [14]. All these results suggest a mammary microbiota that supports the vertical transmission of the mother's microbiota through milk, and not skin or vaginal colonization [11]. The possible transmission of the gut anaerobe *R. bromii* suggests that this vertical transmission is not limited to lactic acid bacteria such as
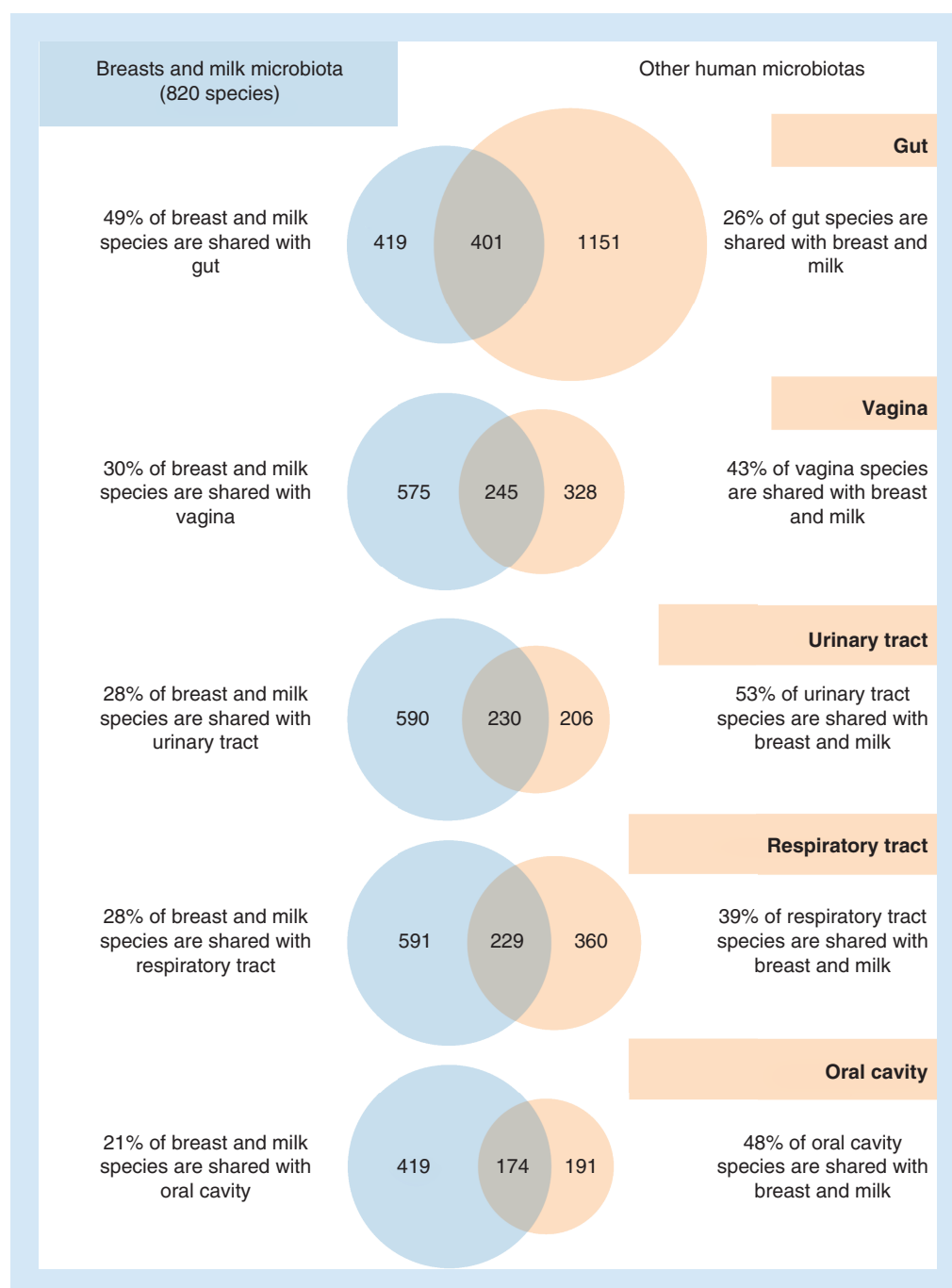
Figure 7.   Proportion of species from the human breast and milk shared with other anatomical niches. Proportional Venn diagrams were performed using the online service https://www.meta-chart.com/

Lactobacillus and *Bifidobacterium* but involve all the components of the human microbiome including the healthy mature anaerobic gut microbiota [40] that produces butyrate, a key molecule in symbiosis with the host [60].

The existence of a resident mammary flora was first suggested in 1988 by Thornton *et al.,* who cultured breast tissue obtained after plastic surgery in 30 women [61]. The flora identified was similar to the skin flora. This microbiota was composed of aerobic (coagulase-negative staphylococci, diphtheroids, lactobacilli, enterococci and α-hemolytic streptococci) and anaerobic (*C. acnes*, *Peptococcus* sp. and *Clostridium sporogenes*) bacteria. In 2014, Urbaniak *et al.* studied breast tissue of 81 women with and without cancer in Canada and Ireland [10]. They

identified 121 operational taxonomy units belonging to seven phyla (Proteobacteria, Firmicutes, Actinobacteria, Bacteroidetes, Deinococcus–Thermus, Verrucomicrobia and Fusobacteria), 57 genera and 25 species. In the present review, we found only seven species in culture from healthy breast control tissue from two studies (additional file 6): *C. acnes, Micrococcus luteus, Propionibacterium granulosum, S. epidermidis, Staphylococcus saprophyticus, S. agalactiae* and *Streptococcus oralis* [10,62]. *C. acnes, S. epidermidis* and *S. agalactiae* are among the most common bacteria in the breast and milk microbiota of healthy controls (additional file 6 & 7), which supports the hypothesis of a resident mammary microbiota.

Several members of the healthy human mature anaerobic intestinal microbiota, such as *Faecalibacterium, Roseburia, Ruminococcus* and *Coprococcus* [40], have been frequently detected in human breast and milk by the molecular tools mentioned above [19,42–45]. However, the mechanisms by which bacteria could cross the intestinal epithelium and reach the mammary glands are not yet understood. Bacteria in the human digestive tract could endogenously reach the mammary glands by translocation with the involvement of dendritic cells [15,63]. In 2015, Hashiguchi *et al.* observed that Peyer's patch innate lymphoid cells regulate the expansion of commensal bacteria outside the intestinal lumen [64]. *Lactobacillus fermentum* and *Lactobacillus salivarius*, when administered orally to the mother, can reach the mammary glands and are then excreted in milk [50].

Staphylococcus, Streptococcus, Cutibacterium and *Corynebacterium*, predominant in breast and milk studies, are also predominant on human skin [65]. Therefore, it is possible that the microbiota in the mother's skin may also be involved in the establishment of part of the microbiota in breast milk. The microbiota in breast milk probably has several origins; the resident mammary flora, the digestive flora, the cutaneous flora and the flora of the infant's oral cavity. Regurgitation during breastfeeding could lead to the colonization of breast milk by the flora of the infant oral cavity contributing to the formation of the microbiota in breast milk [42,66].

About 300 bacterial species (303) have been found only in the human breast and milk microbiota and not in the other human microbial repertoires at our disposal. Total 21 of these 303 species belonged to the *Bifidobacterium* genus (additional file 9). Exploring the origin of these species is a challenge for future studies conducted on the human microbiota. One hypothesis is that the resident mammary microbiota could establish as early as the first days of life in the newborn. This hypothesis is supported by the fact that maternal estrogens may result in bilateral physiologic breast hypertrophy in the full-term newborn with possible nipple discharge [67,68]. Moreover, neonatal mastitis and breast abscesses most often occur before 2 months and are usually associated with *S. aureus, S. agalactiae, E. coli* or *S. enterica* [67,68]. As reported here, these bacteria could be brought to the breast of the breastfed newborn from the first days of life by the entero-mammary pathway. In addition, *Bifidobacterium* (particularly *B. longum, B. breve* and *B. dentium*, found in all human repertoires except the respiratory tract [additional file 10]) are particularly abundant in the gut during the first weeks of life and subsequently decrease while Lachnospiraceae (including *Roseburia intestinalis*) and Ruminococcaceae (including *F. prausnitzii*) increased with age [40,57]. Studies are needed to clarify whether the hormonal context and profile of the neonatal oral and gut microbiota promotes specific and early colonization of the baby's breast in the first months of life.

Putative role of breast milk microbiota

Some authors have reported that it is during this period called the 'window of opportunity' [16] that the colonization of Peyer's patches by a resident flora including certain bacteria, such as *Alcaligenes* and *Ochrobactrum*, occurs [64,69,70]. This colonization could be a determining factor in triggering the maturation of the child's immune system. However, this remains a subject of debate. Here, we identified *Alcaligenes faecalis* as members of the breast and milk microbiota (additional file 4). Therefore, milk is a putative vector for the transmission of the resident flora of the child's Peyer patches.

Strains of lactic acid bacteria such as *Lactobacillus, Lactococcus, Bifidobacterium* isolated in human or animal milk have been used in biotherapy to prevent or treat lactational mastitis with similar or better efficacy than traditional antibiotic treatment in animals and humans [50]. This antipathogenic effect can be achieved through a direct (bacteriocins, reactive oxygen species) or indirect mechanism. In addition to their antipathogenic effect, bacteriocins are likely to modulate the global profile of the microbiota [71,72].

It has been shown that the same strains of *Lactobacillus* and *Bifidobacterium* were found in the mother's milk and in the baby's stool [11,59]. This suggests that breastfeeding and milk contribute to the vertical transmission of lactic acid bacteria, critical to modulate the child's microbiota profile [71,72]. However, recent studies suggest that breast milk can provide all or part of the child's microbiota including nonlactic acid bacteria [8], such as *Ruminococcus* [14]. In experimental models, crossfostering immediately after birth induced a microbiota shift shaped by the nursing

mother and not by the fostering one, which persists after weaning and for life [73]. If confirmed, breastfeeding could represent the transmission of the human microbiological memory, an epigenetic regulator that drives the balance between good health and metabolic disorders in offspring [74].

Conclusion

Our systematic, automated and comprehensive literature review confirms a high diversity of human breast and milk microbiota with more than 800 bacterial species (820 species). The most frequently detected species were, in decreasing order of frequency: *S. aureus, S. epidermidis, S. agalactiae, C. acnes, E. faecalis, B. breve, E. coli, S. sanguinis, L. gasseri* and *S. enterica*.

Most of the bacterial species detected in breast and milk were facultative anaerobic or strictly aerobic bacteria, but the strict anaerobic bacteria usually associated with the gut were also detected. Although archaeal DNA sequences have been detected in milk, no species of Archaea has been isolated either from breast or milk by culture method. Since Archaea are a key to host–microbiota mutualism [46,47], are detected in the gut of almost all healthy adults, are isolated from the adult intestine and oral cavity [75,76] but also from the newborn stomach [77], culture of Archaea from breast or milk represent an exciting challenge. Infection was associated with an increased proportion of aerotolerant bacteria suggesting that, as reported in the gut [35,40], the anaerobic breast and milk microbiota is associated with health. Overall, anaerobic culture of fresh samples has been neglected in the exploration of the human microbiome.

Analyzing predictive value for infection allow us to discriminate three groups of bacteria: the commensals-potential probiotics (PVI << 0.5), the commensals-potential pathogenic (PVI around 0.5) and the pathogens (PVI >> 0.5). Several species, in particular *Bifidobacterium* and *Lactobacillus* species, were associated with the absence of infection and were never detected in infection (PVI = 0), suggesting their probiotic potential in the context of breastfeeding for the prevention of mastitis, breast abscess and neo-natal infection. *Bifidobacterium* species were detected both in the breast tissue and milk and represent a large part of the species specifically found in the breast and milk microbiota and not in other human repertoires, confirming that these maternal probiotics are of particular importance to human health [35,40]. Commensals-potential pathogens (PVI around 0.5) challenge Koch's postulates because *S. aureus* and *S. agalactiae* are both among the most frequent bacteria in healthy controls and are also often found in pure culture in cases of breast and neonatal infection [78]. Differences in virulence between strains, a role of the microbial or immune context or stress causing a change in the behavior of these microbes could help understanding how these commensals are also associated with potentially severe infections for the mother and newborn. We confirm (*S. enterica*) but also evidenced (*B. cepacia/ambifaria*, *M. gilvum*) bacteria almost invariably associated with infection (PVI > 0.9). Future studies could test these bacteria in undocumented mastitis, breast abscess or neo-natal infection by specific culture and quantitative PCR. Noteworthy, *B. cepacia* has been associated with mastitis in animals [79] and several case report of breast implant mycobacterial infections have been reported [80].

The first limitation of our study is that we used only one search engine (PubMed) because the automatic script used was designed specifically for this database using the PMID (PubMed unique identifier). Future studies may use several scripts adapted to several search engines (Google, Google scholar, Embase, Ovid, EBSCOhost) but also search the open genomics/microbiome database in order to be more comprehensive. However, Pubmed currently remains the reference search engine in medicine and biology, tracking several databases (MEDLINE, Pre-MEDLINE, PMC, NCBI Bookshelf), with a very complete indexing (29,239,361 items as on 7 January 2019, records going back to 1966, selectively to the year 1865, and very selectively to 1809; about 500,000 new records are added each year) and particularly adapted for exhaustive searches based on complex combinations of MeSH (medical subject headings) terms.

Here, we performed a literature review. Another approach would be metagenomic meta-analysis. The Human Microbiome Project (https://portal.hmpdacc.org/ assessed on 7 February 2019 – data release 1.1 – 26 September 2017) include 48 primary sites and 32,036 samples, mainly from feces, buccal mucosa, vagina but no sample from the breast or milk. The sequence reads archive (https://www.ncbi.nlm.nih.gov/sra) include 1064 biosamples corresponding to 'human milk metagenome'. This open new way to characterize the repertoires of human breast and milk microbiota. Conversely, our literature review based on peer-reviewed published articulations ensures better quality of metadata (sample characteristics: anatomical site, infection) and includes culture data not available from metagenomic databases.

Our search criteria dates are large and reflect evolution of microbiological techniques which will make direct comparison of results difficult. In addition, certain taxa of bacteria would not have been isolated in earlier studies as it is only with molecular methods and better culture techniques that these have been identified. Another limitation of this review is that it does not consider the difference in quality between the included studies. It is possible that some members of the repertoire reported here are in fact contaminants. Indeed, all microbial identification methods, whether culture or molecular methods, are likely to produce false positives, by contamination by DNA sequences or by living microbes in the environment or introduced by the operator. Unfortunately, negative controls are rarely provided or difficult to interpret in metagenomic studies. Moreover, many reports do not identify storage conditions and storage media. We did not consider collection techniques, cleaning of the skin, fore versus hind milk. As we were primarily focused on the repertoire, we did not include any ecological analysis such as the study of the number of observed species per sample or of intra- or intersample diversity indices. In addition, we did not report the frequency of bacteria by individuals within included studies as we only considered the frequency among study groups. While individual meta-analysis is the gold standard, it is practically impossible because individual data are rarely provided in published studies. Future reviews on the milk microbiota should attempt to go beyond all these limitations.

Our automatic strategy allowed us to include 140 articles while 102 were added by crossreferencing (Figure 1). This suggests that automatic search helped but did not replace humans to perform systematic reviews. In the future, script could be improved by excluding taxa not identified as a species (for instance, excluding taxa that include the following terms: 'bacterium' or 'organism' – see additional file 2). Doing this, from 17,023 search results, 9613 could be excluded because they included the term 'bacterium', and 200 because they included the term 'organism' so that only 7210 search results remained corresponding to bacterial or archaeal species.

Future culturomics studies could extend the repertoire of the breast and human milk microbiota and thus change the paradigm: the human microbiome could be a connection of all human microbial niches such as the mouth, intestine, urinary tract, vagina, lung and breast. Breast milk could be the vector for the transmission of the entire microbiome from mother to child, including rare phyla as Deferribacteres or Verrucomicrobia and archaea. To clarify this point, future studies should confirm that the most oxygen-sensitive members of the human microbiome, such as *A. muciniphila*, *F. prausnitzii* or *M. smithii*, are alive in human milk.

Future perspective

This review opens a new perspective for the human microbiota. Beyond lactic acid bacteria, milk is probably the vector for the vertical transmission of the entire microbiota from mother to child. If this is confirmed, future artificial milks should be supplemented with a complete human milk microbiota and not only lactic ferments. Future studies may clarify whether the different vertical transmission pathways of the microbiota are redundant or whether they transmit specific microbes, for example by comparing the repertoire of breast and milk with that of the maternal oral flora, vagina and placenta, and by experimental models (e.g., breastfed axenic mice).

Summary points

- The current repertoire of the human breast and milk microbiota include 820 different species with a predominance of aerotolerant prokaryotes, Proteobacteria and Firmicutes.
- *Archaea, Faecalibacterium prausnitzii* and *Akkermansia muciniphila* were detected both in the breast tissue and milk by molecular method but not by culture prompting new anaerobic culture studies of fresh milk samples.
- Several Bifidobacterium and Lactobacillus species were associated with absence of infection while *Salmonella enterica, Burkholderia cepacia / ambifaria* and *Mycolicibacterium gilvum* were associated with infection.
- *Staphylococcus aureus* and *Streptococcus agalactiae*, commonly isolated in pure culture from infected breast and milk, were the most frequent species both in infection and in healthy controls, challenging the Koch's postulates.
- The human breast and milk microbiota shared 49% of the species of its repertoire with the gut, 30% with the vagina, 28% with the urinary tract, 28% with the respiratory tract and 21% with the oral cavity.
- Total 303 bacterial species were found only in the human breast and milk microbiota and not in other human microbial repertoires suggesting a specific breast microbiota.

References

Papers of special note have been highlighted as: ● of interest; ●● of considerable interest

1. Victora CG, Bahl R, Barros AJ *et al.* Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. *Lancet* 387(10017), 475–490 (2016).

●● **Breastfeeding is a major determinant of human health.**

2. Gartner LM, Morton J, Lawrence RA *et al.* Breastfeeding and the use of human milk. *Pediatrics* 115(2), 496–506 (2005).

3. Klement E, Cohen RV, Boxman J, Joseph A, Reif S. Breastfeeding and risk of inflammatory bowel disease: a systematic review with meta-analysis. *Am. J. Clin. Nutr.* 80(5), 1342–1352 (2004).

4. Williams CD. A nutritional disease of childhood associated with a maize diet. *Arch. Dis. Child.* 8(48), 423–433 (1933).

5. Newburg DS. Innate immunity and human milk. *J. Nutr.* 135(5), 1308–1312 (2005).

6. Ballard O, Morrow AL. Human milk composition: nutrients and bioactive factors. *Pediatr. Clin. North Am.* 60(1), 49–74 (2013).

7. Lönnerdal B, Erdmann P, Thakkar SK, Sauser J, Destaillats F. Longitudinal evolution of true protein, amino acids and bioactive proteins in breast milk: a developmental perspective. *J. Nutr. Biochem.* 41, 1–11 (2017).

8. Pannaraj PS, Li F, Cerini C *et al.* Association between breast milk bacterial communities and establishment and development of the infant gut microbiome. *JAMA Pediatr.* 171(7), 647–654 (2017).

9. Heikkilä MP, Saris PEJ. Inhibition of Staphylococcus aureus by the commensal bacteria of human milk. *J. Appl. Microbiol.* 95(3), 471–478 (2003).

10. Urbaniak C, Cummins J, Brackstone M *et al.* Microbiota of human breast tissue. *Appl. Environ. Microbiol.* 80(10), 3007–3014 (2014).

11. Martín R, Langa S, Reviriego C *et al.* Human milk is a source of lactic acid bacteria for the infant gut. *J. Pediatr.* 143(6), 754–758 (2003).

● **Demonstrates the vertical transmission of *Lactobacillus* at the strain level.**

12. Berardi A, Rossi C, Creti R *et al.* Group B streptococcal colonization in 160 mother-baby pairs: a prospective cohort study. *J. Pediatr.* 163(4), 1099.e1091–1104.e1091 (2013).

13. Jost T, Lacroix C, Braegger CP, Rochat F, Chassard C. Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding. *Environ. Microbiol.* 16(9), 2891–2904 (2014).

14. Asnicar F, Manara S, Zolfo M *et al.* Studying vertical microbiome transmission from mothers to infants by strain-level metagenomic profiling. *mSystems* 2(1), pii:e00164–16 (2017).

15. Perez PF, Doré J, Leclerc M *et al.* Bacterial imprinting of the neonatal immune system: lessons from maternal cells? *Pediatrics* 119(3), e724–e732 (2007).

16. Gensollen T, Iyer SS, Kasper DL, Blumberg RS. How colonization by microbiota in early life shapes the immune system. *Science* 352(6285), 539–544 (2016).

17. Ho NT, Li F, Lee-Sarwar KA *et al.* Meta-analysis of effects of exclusive breastfeeding on infant gut microbiota across populations. *Nat. Commun.* 9(1), 4169 (2018).

18. Xuan C, Shamonki JM, Chung A *et al.* Microbial dysbiosis is associated with human breast cancer. *PLoS ONE* 9(1), e83744 (2014).

19. Urbaniak C, Gloor GB, Brackstone M, Scott L, Tangney M, Reid G. The microbiota of breast tissue and its association with breast cancer. *Appl. Environ. Microbiol.* 82(16), 5039–5048 (2016).

20. Jiménez E, De Andrés J, Manrique M *et al.* Metagenomic analysis of milk of healthy and mastitis-suffering women. *J. Hum. Lact.* 31(3), 406–415 (2015).

21. Lagier JC, Armougom F, Million M *et al.* Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin. Microbiol. Infect.* 18(12), 1185–1193 (2012).

22. Lagier JC, Dubourg G, Million M *et al.* Culturing the human microbiota and culturomics. *Nat. Rev. Microbiol.* doi:10.1038/s41579-018-0041-0 540-550 (2018) (Epub ahead of print).

●● **Microbial culturomics offers an unprecedented opportunity to study and manipulate the human microbiome.**

23. Lagier J-C, Khelaifia S, Alou MT *et al.* Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat. Microbiol.* 1, 16203 (2016).

24. Seck E-H, Senghor B, Merhej V *et al.* Salt in stools is associated with obesity, gut halophilic microbiota and *Akkermansia muciniphila* depletion in humans. *Int. J. Obes.* doi: 10.1038/s41366-018-0201-3 2018) (Epub ahead of print).

25. Dickson I. Gut microbiota: Culturomics: illuminating microbial dark matter. *Nat. Rev. Gastroenterol. Hepatol.* 14(1), 3 (2017).

26. Hugon P, Dufour J-C, Colson P, Fournier P-E, Sallah K, Raoult D. A comprehensive repertoire of prokaryotic species identified in human beings. *Lancet Infect. Dis.* 15(10), 1211–1219 (2015).

27. Bilen M, Dufour J-C, Lagier J-C *et al.* The contribution of culturomics to the repertoire of isolated human bacterial and archaeal species. *Microbiome* 6(1), 94 (2018).

28. Fonkou M-D, Dufour J-C, Dubourg G, Raoult D. Repertoire of bacterial species cultured from the human oral cavity and respiratory tract. *Future Microbiol.* 13, 1611–1624 (2018).

29. Morand A, Cornu F, Dufour J-C, Tsimaratos M, Lagier JC, Raoult D. Human bacterial repertoire of urinary tract: a paradigm shift. *J. Clin. Microbiol.* doi:10.1128/JCM.00675-18 (2018) (Epub ahead of print).

30. Diop K, Dufour J-C, Levasseur A, Fenollar F. Exhaustive repertoire of human vaginal microbiota. *Hum. Microbiome J.* doi:10.1016/j.humic.2018.11.002 (2018) (Epub ahead of print).

31. Million M, Angelakis E, Maraninchi M *et al.* Correlation between body mass index and gut concentrations of *Lactobacillus reuteri*, Bifidobacterium animalis, *Methanobrevibacter smithii* and *Escherichia coli*. *Int. J. Obes. (Lond.)* 37(11), 1460–1466 (2013).

32. Million M, Angelakis E, Paul M, Armougom F, Leibovici L, Raoult D. Comparative meta-analysis of the effect of *Lactobacillus* species on weight gain in humans and animals. *Microb. Pathog.* 53(2), 100–108 (2012).

33. Million M, Maraninchi M, Henry M *et al.* Obesity-associated gut microbiota is enriched in *Lactobacillus reuteri* and depleted in *Bifidobacterium animalis* and *Methanobrevibacter smithii*. *Int. J. Obes. (Lond.)* 36(6), 817–825 (2012).

34. Million M, Raoult D. Species and strain specificity of *Lactobacillus* probiotics effect on weight regulation. *Microb. Pathog.* 55, 52–54 (2013).

35. Euzéby JP. List of bacterial names with standing in nomenclature: a folder available on the internet. *Int. J. Syst. Evol. Microbiol.* 47(2), 590–592 (1997).

36. Parte AC. LPSN – List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. *Int. J. Syst. Evol. Microbiol.* 68(6), 1825–1829 (2018).

37. Million M, Tidjani Alou M, Khelaifia S *et al.* Increased gut redox and depletion of anaerobic and methanogenic prokaryotes in severe acute malnutrition. *Sci. Rep.* 6, 26051 (2016).

38. Million M, Raoult D. Linking gut redox to human microbiome. *Hum. Microbiome J.* 10, 27–32 (2018).

39. Campbell C, Dikiy S, Bhattarai SK *et al.* Extrathymically generated regulatory T cells establish a niche for intestinal border-dwelling bacteria and affect physiologic metabolite balance. *Immunity* 48(6), 1245–1257 e1249 (2018).

40. Loy A, Pfann C, Steinberger M *et al.* Lifestyle and horizontal gene transfer-mediated evolution of *mucispirillum schaedleri*, a core member of the murine gut microbiota. *mSystems* 2(1), pii:e00171–16 (2017).

41. Everard A, Belzer C, Geurts L *et al.* Cross-talk between *Akkermansia muciniphila* and intestinal epithelium controls diet-induced obesity. *Proc. Natl Acad. Sci. USA* 110(22), 9066–9071 (2013).

42. Million M, Diallo A, Raoult D. Gut microbiota and malnutrition. *Microb. Pathog.* 106, 127–138 (2017).

43. Togo AH, Diop A, Bittar F *et al.* Description of *Mediterraneibacter massiliensis*, gen. nov., sp. nov., a new genus isolated from the gut microbiota of an obese patient and reclassification of *Ruminococcus faecis*, *Ruminococcus lactaris*, *Ruminococcus torques*, *Ruminococcus gnavus* and *Clostridium glycyrrhizinilyticum* as Mediterraneibacter faecis comb. nov., *Mediterraneibacter lactaris* comb. nov., *Mediterraneibacter torques* comb. nov., *Mediterraneibacter gnavus* comb. nov. and *Mediterraneibacter glycyrrhizinilyticus* comb. nov. *Antonie Van Leeuwenhoek* doi:10.1007/s10482-018-1104-y (2018) (Epub ahead of print).

44. Patel SH, Vaidya YH, Patel RJ, Pandit RJ, Joshi CG, Kunjadiya AP. Culture independent assessment of human milk microbial community in lactational mastitis. *Sci. Rep.* 7(1), 7804 (2017).

45. Biagi E, Quercia S, Aceti A *et al.* The bacterial ecosystem of mother's milk and infant's mouth and gut. *Front. Microbiol.* 8, 1214 (2017).

46.   Li S-W, Watanabe K, Hsu C-C *et al.* Bacterial composition and diversity in breast milk samples from mothers living in Taiwan and mainland China. *Front. Microbiol.* 8, 965 (2017).

47.   Milani C, Mancabelli L, Lugli GA *et al.* Exploring vertical transmission of bifidobacteria from mother to child. *Appl. Environ. Microbiol.* 81(20), 7078–7087 (2015).

48.   Samuel B-S, Gordon J-I. A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism. *Proc. Natl Acad. Sci. USA* 103(26), 10011–6 (2006).

49.   Samuel BS, Hansen EE, Manchester JK *et al.* Genomic and metabolic adaptations of Methanobrevibacter smithii to the human gut. *Proc. Natl Acad. Sci. USA* 104(25), 10643–8 (2007).

50.   Ward TL, Hosid S, Ioshikhes I, Altosaar I. Human milk metagenome: a functional capacity analysis. *BMC Microbiol.* 13, 116 (2013).

51.   Koren O, Goodrich Julia K, Cullender Tyler C *et al.* Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell* 150(3), 470–480 (2012).

52.   Arroyo R, Martín V, Maldonado A, Jiménez E, Fernández L, Rodríguez Juan M. Treatment of infectious mastitis during lactation: antibiotics versus oral administration of Lactobacilli isolated from breast milk. *Clin. Inf. Dis.* 50(12), 1551–1558 (2010).

●●   **Treatment of infectious mastitis by oral administration of *Lactobacilli.***

53.   Schafer P, Furrer C, Mermillod B. An association of cigarette smoking with recurrent subareolar breast abscess. *Int. J. Epidemiol.* 17(4), 810–813 (1988).

54.   Co M, Cheng VCC, Wei J *et al.* Idiopathic granulomatous mastitis: a 10-year study from a multicentre clinical database. *Pathology* doi:10.1016/j.pathol.2018.08.010 (2018) (Epub ahead of print).

55.   Bundred NJ, Dover MS, Coley S, Morrison JM. Breast abscesses and cigarette smoking. *Br. J. Sur.* 79(1), 58–59 (1992).

56.   Bharat A, Gao F, Aft RL, Gillanders WE, Eberlein TJ, Margenthaler JA. Predictors of primary breast abscesses and recurrence. *World J. Surg.* 33(12), 2582–2586 (2009).

57.   Gollapalli V, Liao J, Dudakovic A, Sugg SL, Scott-Conner CE, Weigel RJ. Risk factors for development and recurrence of primary breast abscesses. *J. Am. Coll. Surg.* 211(1), 41–48 (2010).

58.   Martin V, Manes-Lazaro R, Rodriguez JM, Maldonado-Barragan A. *Streptococcus lactarius* sp. nov., isolated from breast milk of healthy women. *Int. J. Syst. Evol. Microbiol.* 61(5), 1048–1052 (2011).

59.   Yatsunenko T, Rey FE, Manary MJ *et al.* Human gut microbiome viewed across age and geography. *Nature* 486(7402), 222–227 (2012).

●   ***Bifidobacterium breve, Bifidobacterium longum* and *Bifidobacterium dentium* are the three species particularly associated with breastfeeding and with a higher relative abundance in babies versus adults (see supplementary data of this article).**

60.   Jost T, Lacroix C, Braegger CP, Chassard C. New insights in gut microbiota establishment in healthy breast fed neonates. *PLoS ONE* 7(8), e44595 (2012).

61.   Solís G, De Los Reyes-Gavilan CG, Fernández N, Margolles A, Gueimonde M. Establishment and development of lactic acid bacteria and bifidobacteria microbiota in breast-milk and the infant gut. *Anaerobe* 16(3), 307–310 (2010).

62.   Million M, Tomas J, Wagner C, Lelouard H, Raoult D, Gorvel J-P. New insights in gut microbiota and mucosal immunity of the small intestine. *Human Microbiome J.* 7–8, 10 (2018).

63.   Thornton JW, Argenta LC, Mcclatchey KD, Marks MW. Studies on the endogenous flora of the human breast. *Ann. Plast. Surg.* 20(1), 39–42 (1988).

64.   Ransjo U, Asplund OA, Gylbert L, Jurell G. Bacteria in the female breast. *Scand. J. Plast. Reconstr. Surg.* 19(1), 87–89 (1985).

65.   Lelouard H, Fallet M, De Bovis B, Meresse S, Gorvel JP. Peyer's patch dendritic cells sample antigens by extending dendrites through M cell-specific transcellular pores. *Gastroenterology* 142(3), 592.e3–601.e3 (2012).

66.   Hashiguchi M, Kashiwakura Y, Kojima H, Kobayashi A, Kanno Y, Kobata T. Peyer's patch innate lymphoid cells regulate commensal bacteria expansion. *Immunol. Lett.* 165(1), 1–9 (2015).

67.   Nielsen ML, Raahave D, Stage JG, Justesen T. Anaerobic and aerobic skin bacteria before and after skin-disinfection with chlorhexidine: an experimental study in volunteers. *J. Clin. Pathol.* 28(10), 793–797 (1975).

68.   Delgado S, Arroyo R, Martín R, Rodríguez JM. PCR-DGGE assessment of the bacterial diversity of breast milk in women with lactational infectious mastitis. *BMC infect. Dis.* 8, 51 (2008).

69.   Donaire A, Guillen J, Rajegowda B. Neonatal breast hypertrophy: Revisited. *Pediatr. Thera.* 6(3), 2 (2016).

70.   Raveenthiran V. Neonatal mastauxe (breast enlargement of the newborn). *J. Neonatal Surg.* 2(3), 31 (2013).

71.   Kunisawa J, Kiyono H. Alcaligenes is commensal bacteria habituating in the gut-associated lymphoid tissue for the regulation of intestinal IgA responses. *Front. immunol.* 3, 65 (2012).

72.   Obata T, Goto Y, Kunisawa J *et al.* Indigenous opportunistic bacteria inhabit mammalian gut-associated lymphoid tissues and share a mucosal antibody-mediated symbiosis. *Proc. Natl Acad. Sci. USA* 107(16), 7419–7424 (2010).

73.   Riboulet-Bisson E, Sturme MH, Jeffery IB *et al.* Effect of *Lactobacillus salivarius* bacteriocin Abp118 on the mouse and pig intestinal microbiota. *PLoS ONE* 7(2), e31113 (2012).

74. Murphy EF, Cotter PD, Hogan A *et al.* Divergent metabolic outcomes arising from targeted manipulation of the gut microbiota in diet-induced obesity. *Gut* 62(2), 220–226 (2013).

75. Le Doare K, Holder B, Bassett A, Pannaraj PS. Mother's milk: a purposeful contribution to the development of the infant microbiota and immunity. *Front. Immunol.* 9, 361 (2018).

76. Devaux CA, Raoult D. The microbiological memory, an epigenetic regulator governing the balance between good health and metabolic disorders. *Front. Microbiol.* 9, 1379 (2018).

77. Ferrari A, Brusa T, Rutili A, Canzi E, Biavati B. Isolation and characterization of *Methanobrevibacter oralis* sp. nov. *Curr. Microbiol.* 29(1), 6 (1994).

78. Miller TL, Wolin MJ, Conway De Macario E, Macario AJ. Isolation of Methanobrevibacter smithii from human feces. *Appl. Environ. Microbiol.* 43(1), 227–232 (1982).

79. Grine G, Boualam MA, Drancourt M. *Methanobrevibacter smithii*, a methanogen consistently colonising the newborn stomach. *Eur. J. Clin. Microbiol. Infec. Dis.* 36(12), 2449–2455 (2017).

•   **Suggests that *Methanobrevibacter smithii*, the main human gut methanogenic Archaea, is transmitthed through breastfeeding.**

80. Desnues B, Al Moussawi K, Raoult D. Defining causality in emerging agents of acute bacterial diarrheas: a step beyond the Koch's postulates. *Future Microbiol.* 5(12), 1787–1797 (2010).

Article II:

Culture of Methanogenic Archaea from Human Colostrum and Milk: The Neglected Critical Commensals.

Amadou Hamidou Togo[†], Ghiles Grine[†], Saber Khelaifia, Clothilde Des Robert, Véronique Brevaut, Aurelia Caputo, Emeline Baptiste, Marion Bonnet, Anthony Levasseur, Michel Drancourt, Matthieu Million[†], Didier Raoult.

35

1 **Culture of Methanogenic Archaea from Human Colostrum and Milk: The Neglected**

2 **Critical Commensals**

3 Amadou Hamidou TOGO[1,2†], Ghiles GRINE[1,2†], Saber KHELAIFIA[1,2], Clothilde DES

4 ROBERT[3], Véronique BREVAUT[4], Aurelia CAPUTO[1,2], Emeline BAPTISTE[1,2], Marion

5 BONNET[1,2], Anthony LEVASSEUR[1,2], Michel DRANCOURT[1,2], Matthieu MILLION[1,2†],

6 Didier RAOULT[1,2*]

7 [1]IHU-Méditerranée Infection, Marseille, France.

8 [2]Aix Marseille Univ, IRD, APHM, MEPHI, Marseille, France.

9 [3]APHM, CHU Hôpital de la Conception, Service de médecine néonatale, F-13385 Marseille,

10 France.

11 [4]APHM, CHU Hôpital Nord, Service de médecine néonatale, Marseille, France.

12 [†]These authors have contributed equally to this work.

13 ***Corresponding author**: Prof. Didier RAOULT

14 **Tel**: 00 33 4 13 73 73 24 01

15 **Fax** : 00 33 4 12 12 73 23 07

16 **E-mail** : didier.raoult@gmail.com

ABSTRACT

Archaeal sequences have been detected but no studies have determined whether living

archaea were present in human colostrum and milk. Methanogenic archaea are neglected

because not detected by usual molecular and culture methods. Using improved DNA detection

protocol and microbial culture technique using antioxidant previously developed in our

center, we investigated the presence of methanogenic archaea in human colostrum and milk

by culture, specific *M. smithii* and *M. oralis* real time PCR and 16S amplicon sequencing and

association with clinical variables. *M. smithii* was isolated from 3 colostrum and 5 milk (day

10) samples. *M. oralis* was isolated from 1 milk sample. For 2 strains, genome was

sequenced, and rhizome was indistinguishable from strains previously isolated from human

mouth and gut. *M. oralis* was not detected by molecular methods. *M. smithii* was detected in

the colostrum or milk of 5/13 (38%), 37/127 (29%), 3/132 (2%) and 43/132 (33%) mothers by

culture, qPCR, 16S amplicon sequencing, and any of these 3 techniques respectively.

Detection of *M. smithii* was associated with absence of maternal obesity (12% in obese versus

35% in non-obese mothers, p < .05). Methanogenic archaea are alive in human colostrum and

milk and associated with maternal health. Early initiation of breastfeeding is critical to seed

the infant's microbiota with these neglected critical commensals.

INTRODUCTION

Breastfeeding is a major determinant of human health.[1] Breast colostrum and milk contain a very diverse bacterial microbiota that plays a key role in human health.[2] However, while archaeal sequences have been detected,[3,4] no studies have determined whether living archaea were present in human milk.[2] In the new view of the tree of life,[5] archaea represent a lineage very distinct from bacteria and candidate phyla radiation but very closed to eukaryotes and the human archaeome is increasingly recognized thanks to dedicated molecular methods.[6] Methanogenic archaea[7] are particularly adapted to the gut and key components of the human-archaeal-bacterial mutualism. They improve energy harvest by consumption of end-product of microbial fermentation such as acetate, dihydrogen and carbon dioxide while producing methane.[8,9] The genomic analysis of the main representatives of human methanogenic archaea, *M. smithii*,[10] showed an evolutive adaptation to the human gut[9] and is associated with weight regulation and health. Indeed, *M. smithii* was found in virtually all healthy human adults and was depleted in obesity and in severe acute malnutrition.[11-13] Accordingly, culture of methanogenic archaea from human colostrum and milk represent an exciting challenge and a missing piece in the puzzle of human microbiota vertical transmission underpinning metabolic phenotype inheritance.[2]

The culture of methanogenic archaea was, until recently, very tedious, expensive, and time-consuming and required heavy and rare technical means to exclude oxygen (anaerobic chambers, pressurization of Hungate culture tubes at 200 kPa every 4 days by injection of 80% H_2:20% CO_2). In this context, medical interest in these microbes was neglected, so to date, only few studies have reported the culture of methanogenic archaea from humans. To the best of our knowledge, 3 species (*Methanobrevibacter smithii, Methanobrevibacter oralis* and *Methanosphaera stadtmanae*) were reported in 9 studies from 4 other teams. Since 2012, our team has reported 6 species of methanogenic archaea unknown to humans

61   (*Methanomassiliicoccus luminyensis, Methanobrevibacter arboriphilicus,*

62   *Methanobrevibacter massiliense, Methanobrevibacter millerae, Haloferax alexandrines,* and

63   *Haloferax massiliensis*) with 3 new species (*M. luminyensis, M. massiliense,* and *H.*

64   *massiliensis*) including 1 halophilic species (*H. massiliensis*) in 10 studies. *M. oralis* was

65   isolated from brain abscesses, subgingival samples, saliva, and periodontitis; *M. smithii* from

66   saliva, periodontitis, muscle abscess, newborn stomachs, and feces; *M. luminyensis* from

67   feces; *M. massiliense* from periodontitis; and *M. millerae*, *M. stadtmanae, H. alexandrines*

68   and *H. massiliensis* from feces (see Supplementary Tables 1 & 2).

69   The first studies based on methane production, culture and molecular methods

70   reported a variable prevalence of human gut methanogens.[14-16] In the largest study to date (n =

71   700), we have previously demonstrated that the molecular detection of human gut

72   methanogenic archaea can be overlooked since it is highly dependent on the extraction and

73   primer biases. *M. smithii* 16S rRNA and rpoB genes were found in 90% and 66% of human

74   adult fecal samples, respectively, using an automatic DNA extraction kit, while a specific

75   improved extraction protocol using glass beads increased the detection frequency to 100%

76   and 98%, respectively,[10] suggesting that this methanogen is a ubiquitous commensal of the

77   human gut.[8,9]

78   The main limitations to culture most of human gut microbes including methanogenic

79   archaea is their extreme sensitivity to oxygen and reactive derivatives. In this context, we

80   recently set up a microbial culture technique using patented antioxidant mixture containing

81   high doses of ascorbic acid, glutathione[17] and uric acid[18-20] (see conflict of interest section).

82   We then aerobically cultured several methanogenic archaea strains from culture collections

83   using a dual chamber system. All methanogenic archaea tested could be cultured.[21]

84   Here, we investigated the presence of methanogenic archaea in human colostrum and

85   milk by culture, specific real time PCR (*M. smithii* and *M. oralis*) and 16S targeted

86 metagenomics (amplicon sequencing). We also compared clinical variables of mother-child

87 dyads positive or negative for *M. smithii*, identified in this study as the most prevalent

88 methanogenic archaea in human colostrum and milk. Identification of methanogenic colonies

89 obtained by culture was performed by 16S rRNA gene sequencing.

90

91 **RESULTS**

92 **Included mothers**

93 132 mothers were included who had provided at least 1 sample (colostrum and/or milk (day

94 10)). 141 samples including 121 colostrum and 20 milk samples were obtained and analyzed

95 at least by one technique (culture, qPCR and/or v3v4 16S gene amplicon sequencing). Full

96 clinical details and results are provided in Supplementary Dataset 1.

97

98 **Culture and isolation**

99 For culture of methanogenic archaea, 20 samples from 13 mothers were analyzed (Table 1 &

100 Supplementary Table 3). Culture was positive in 9 out of 20 samples, including 3 colostrum

101 and 6 milk samples. *M. smithii* was isolated from 8 samples corresponding to 5 mothers (both

102 colostrum and milk isolations in 3 mothers). Focusing on the 6 mothers for whom both

103 colostrum and milk (day 10) were available, the excretion of *M. smithii* in colostrum was

104 associated in all cases with excretion in milk on day 10 (3/3). In contrast, the three mothers

105 that produced culture-negative colostrum also produced culture-negative milk (Table 1). For

106 all positive samples in culture for methanogens, colonies appeared after 9 days of incubation

107 (Supplementary figure 1). *M. oralis* was isolated from 1 sample (milk, no colostrum available

108 from this mother).

109 For all 9 strains (8 *M. smithii* and 1 *M. oralis*), we determined the 16S rRNA gene

110 sequence. For 2 strains (*M. smithii* strain C2 CSUR P5816 isolated from the colostrum of

111   Mother_096), *M. oralis* strain M2 CSUR P5920 isolated from human milk (day 10) of

112   Mother_076, see Table 1)), the 16S rRNA gene sequence was extracted from the genome

113   sequencing (genome sequences deposited in public repository, see Data avaibility). For the 7

114   other *M. smithii* strains, a partial 16S rRNA gene sequence was obtained (see methods) and

115   deposited in public repository (see Data avaibility). All 9 sequences (provided in the

116   Supplementary data) were aligned and compared with the reference sequence of the type

117   strain for each species.

118   According to the list of prokaryotic names with standing in nomenclature

119   (www.bacterio.net), the reference sequence U55233 of the type strain of *M. smithii* (PS =

120   ATCC 35061= DSM 861 = OCM 144) was used to align and compare the sequences of our 8

121   *M. smithii* strains (Table 2). The sequence of 2 strains (M2 &M6) were identical (100%

122   similarity), 3 strains (C1, C3, M3) showed 1 mismatch (99.83%) and 3 strains (M1, M5, M7)

123   showed 2 mismatches (99.61 to 99.66%). The sequence HE654003.1 of the *M. oralis* type

124   strain (DSM 7256) was used to align and compare the sequence of our *M. oralis* strain. A

125   99.39% similarity was found with 9 mismatches (1462/1471) and 3 gaps (3/1471 (0.20%)).

126   Phylogenetic analysis of the 16S rRNA gene sequence confirmed that all isolated *M. smithii*

127   strains grouped together with strains previously described and did not form outliers (Figure

128   1). This suggest that the methanogenic archaeal strains of human colostrum and milk are not

129   different than strains previously isolated mainly from the gut.

130

131   **Genome sequencing**

132   For two strains, *M. smithii* strain C2 (colostrum) and *M. oralis* strain M2 (milk, Table 1), the

133   genome was sequenced and analyzed. To clarify whether these strains were similar to the

134   human digestive (*M. smithii*) or oral (*M. oralis*) strains, we analyzed the rhizome as

135   previously described.[22] Rhizome analysis evaluates possible sequence exchanges and their

136   phylogenetic origin. The hypothesis is that strains living in the same microbial environment

137   (human microbiota) share the same lateral sequence exchange profile as phylogenetic groups

138   present in the same ecological niche. Rhizome analysis by visual examination of the global

139   pattern of sequences shared with other prokaryotic species showed that these strains had an

140   indistinguishable profile from strains previously isolated from the human mouth and intestine

141   for *M. oralis* and *M. smithii*, respectively (Figure 2).

142

143   **Detection of *M. smithii* and *M. oralis* using real-time PCR (rt-PCR)**

144   We performed specific *M. smithii* and *M. oralis* real time PCR for 127 of the 132 included

145   healthy mothers (Supplementary Table 4). In total, 136 samples collected from these 127

146   mothers were analysed by real time PCR included 117 colostrum and 19 milk samples. 32/117

147   (27.3%) colostrum and 5/19 (26.3%) milk samples were positive for *M. smithii*, totaling 37

148   positive samples from 136 total samples (27.2%). Among the 37 positive samples, the cycle

149   thresholds relatively high (median 38.40, interquartile range [36.75-40.00], range 31.50-

150   40.80). Among the 127 included mothers in this analysis, 37 were positive on colostrum

151   and/or milk (29.1%). *M. oralis*-specific real time PCR was negative on all 136 tested samples

152   (and 127 mothers).

153

154   **Ineffectiveness of amplicon sequencing using universal archaea-bacteria primers**

155   **targeting the 16S ribosomal RNA gene to detect *M. smithii* in human colostrum and milk**

156   Out of 122 colostrum analysed, only 3 were positive (2, 3 and 9 reads, respectively - 2.4%

157   positive – Supplementary Table 5). Of the 20 milks analyzed, only 1 was positive (55 reads -

158   5% positive). Among the 142 samples analyzed by amplicon sequencing using universal

159   archaea-bacterial probes targeting the v3v4 region of the 16S rRNA gene, only 4 were

160 positive (2.8%). Among the 132 mothers with at least one sample analyzed by this technique,

161 only 3 were positive (2.3%).

162

163 **Mother-child dyads positive or negative for *M. smithii***

164 Because only two species were found by culture and *M. oralis* was detected only by culture in

165 only one milk, sample, we focused on *M. smithii* to mine associations between detection of

166 this species and clinical variables. Among 132 included mothers, 43 (32.6%) were positive for

167 *M. smithii* by culture and/or qPCR and/or v3v4 amplicon sequencing. This evidenced that *M.*

168 *smithii* is the most prevalent methanogenic archaea in human colostrum and milk.

169

170 **Comparison between positive and negative mother-child dyads for *M. smithii***

171 Detection of *M. smithii* in colostrum or milk was associated with the absence of maternal

172 obesity (2/16 (12.5%) in obese mothers versus 41/116 (35.3%) in the absence of obesity,

173 unilateral Fisher exact test, p = 0.033 – Table 3). No other clinical variables of the mother or

174 newborn was significantly different between positive and negative dyads.

175

176 **DISCUSSION**

177 Here, we have shown that methanogenic archaea are present in colostrum and human milk.

178 Two species have been identified: *M. smithii* and *M. oralis* and their presence was

179 demonstrated by culture and confirmed by genome sequencing. Only a few teams in the world

180 are able to cultivate archaea (Supplementary Table 1 and 2) and the development of our new

181 environment enriched with the 3 major human antioxidants (ascorbate, glutathione, uric acid),

182 which are also the 3 major antioxidants in human colostrum and milk,[23] has been decisive in

183 this success.[21] Our culture approach can grow most of human methanogenic archaea as

184 previously demonstrated[21], however *Methanosphaera* and *Methanomassiliicoccus* were not

185 detected. Further studies should clarify if *M. smithii* and *M. oralis* are the only representative

186   of methanogenic archaea in human colostrum and milk. While *M. oralis* was detected only by

187   culture and in a single sample, we observed that *M. smithii* was the most frequent

188   methanogenic archaea of colostrum and human milk. Indeed, in our study, about 30% of

189   mother-child pairs were positive for *M. smithii* and the detection of *M. smithii* was associated

190   with the absence of maternal obesity.

191   The use of 3 different techniques (culture, real time PCR and 16S targeted

192   metagenomics) has allowed us to show that 16S targeted metagenomics is not suitable for the

193   detection of methanogenic archaea in colostrum and human milk. The use of universal archa-

194   bacterial primers has been associated with lower sensitivity in the literature.[6] However, the

195   system used here has already been successfully used to highlight the absence of *M. smithii* in

196   the stools of children with severe acute malnutrition while *M. smithii* was detected by this

197   system in 40% (16/40) healthy children (p<0.05).[13] In this previous study, real time PCR was

198   more sensitive (0/40 in malnourished people compared to 30/40 (75%) in controls). It is likely

199   that the relative abundance of methanogenic archaea is lower in colostrum and milk than in

200   stool due to exposure to oxygen. We have recently shown that the colostrum and milk

201   repertoire is enriched with aerotolerant prokaryotes but probably contains all the anaerobic

202   members of the human microbiota in low concentration.[2] By analyzing the publicly available

203   16S metagenomes, we were able to observe that, even if metagenomics targeted on 16S was

204   not the most sensitive method, it detected human methanogenic archaea with the highest

205   prevalence for the digestive tract (13%) with sometimes very high relative abundances

206   exceeding 1% (Supplementary Figure 2). It is possible that current techniques may be limited

207   through depth bias[24] and future studies using new chemistry already on the market and

208   producing up to 30 billion reads will need to clarify this. The discrepancy between culture and

209   real time PCR could be explained by the presence of high concentrations of qPCR inhibitors

210 in breast milk samples (exopolysaccharides[25], proteinases[26], calcium ions,[27] lactoferrin[28] and

211 heparin[29]) and by the extreme sensitivity to oxygen of these critical commensals.

212   Human methanogenic arches are neglected critical commensals. They are keys to host-

213 microbiota mutualism[8] with a symbiotic adaptation to the human gut of their main

214 representative, M. smithii[9], who is a member of the indigenous human microbiota.[30,31] M.

215 smithii has been associated with gut microbiota maturation.[32] The exceptionally high

216 antioxidant capacity of colostrum[23] explains how methanogenic archaea could be transmitted

217 alive to the child through colostrum despite transient exposure to oxygen.

218   Our results are also consistent with early neonatal digestive colonization described by

219 Palmer[14], Dridi[10], Koenig[33], and Wampach[34]. Palmer *et al.* reported archaeal sequences in the

220 feces of 50% of newborns during the first days of life but almost no sequences after the fifth

221 week[14]. Dridi *et al*. detected *M. smithii* by molecular methods in 7/8 samples of newborn

222 stools and in 16/16 children under two years of age[10]. Koenig *et al*. detected Euryarcheaota

223 sequences in 60/60 (100%) feces samples from birth (meconium) to 2.5 years of age[33].

224 Wampach *et al*. detected archaea in 59/65 (91%) newborn feces samples (97% by vaginal

225 delivery versus 86% by cesarean section) with a maximum concentration on the first day of

226 life (ratio to the average of all samples, 5.5 on day 1 decreasing to 0.5 on day 365)[34]. The

227 decreased frequency in cesarean section is consistent with delayed initiation of breastfeeding

228 in this situation[35].

229   Moreover, digestive colonization by *M. smithii* is associated with breastfeeding

230 (versus bottle feeding) in 1-day-old newborns[36] and with the consumption of organic milk and

231 yogurt in children aged 6 to 10 years.[37] *M. smithii* was detected by molecular methods in 88%

232 of 900 tanker trucks analyzed, and *M. smithii*'s DNA concentration was higher in raw milk

233 (3.73 log10 copies of DNA/ml) than in pasteurized milk (2.63 log10 copies of DNA/mL).[37]

234 However, we did not find any published studies that cultured methanogenic archaea from

235 human or animal colostrum or milk. Initial inoculation of maternal methanogenic archaea

236 followed by persistence below the detection threshold is likely before redetection during the

237 development of the mature human anaerobic microbiota, and this possibility is supported by

238 qPCR and metagenomic results.[34]

239 For the first time, we showed that methanogenic archaea and particularly *M. smithii*

240 and *M. oralis* are alive in human colostrum. Why is this important? Early initiation (<1h) is

241 recommended by the World Health Organization (WHO) and has been associated with

242 improved health and survival in the newborn,[38] pointing out the critical importance of

243 colostrum.[39] A marked dose-response of increasing risk of neonatal mortality with increased

244 delay was observed.[40] In one study in Ghana, 22% of neonatal deaths could be saved if all

245 infants were breastfeeding started within the first hour.[40] Early initiation particularly

246 decreases newborn infections, sepsis, pneumonia, diarrhea and hypothermia, and facilitates

247 sustained breastfeeding.[39] Barriers to early initiation include cesarean section[35] but also

248 traditional feeding practices including discarding colostrum, prelacteal feeding and delayed

249 initiation (>1h) particularly in specific areas of South Asia.[39] Indeed, less than 50% of

250 newborns are breastfed within 1h of birth in South Asia, including in Pakistan, India,

251 Bangladesh and Nepal.[39] Our results are a new and strong pathophysiological argument for

252 early initiation of breastfeeding.

253 We have previously shown that *M. smithii* was associated with the absence of

254 obesity[11,12] and malnutrition.[13] Here, we found that *M. smithii* was detectable in 35% of non-

255 obese mothers. *M. smithii* was detected less frequently (12%) in colostrum or milk of obese

256 mothers. Although studies with larger sample sizes are needed to confirm this, our study

257 opens a new way to understand why breastfeeding has been associated with a decreased risk

258 of obesity[41] but confirms that the milk of obese mothers may present dysbiosis. While we had

259 previously shown an association between obesity and increase in Lactobacillus[11,12,42], another

260   team found that *Lactobacillus* in colostrum was associated with maternal obesity.[43] Here, we

261   confirm that the dysbiosis associated with obesity observed by stool analysis is also

262   observable by colostrum and milk analysis, suggesting a global microbial dysbiosis in obese

263   people. Future studies are needed to clarify whether colostrum dysbiosis associated with

264   obesity (increase in *Lactobacillus* and depletion in *M. smithii*) may promote obesity in

265   children. In the absence of new evidence, breastfeeding of obese mothers should remain

266   advisable because the preventive effect of breastfeeding against obesity has also been

267   observed in this population.[44]

268   Skin contamination is not ruled out, but probability is very low. Skin archaeal

269   sequences are mainly members of the *Thaumarchaeota* phylum while *Euryarchaeota* and

270   specifically *Methanobrevibacter* were very rarely detected on human skin (<5%), and are

271   much more specific of the gastrointestinal tract.[6,45] The presence of multiple strains of *M.*

272   *smithii* in human individuals has been described for the digestive microbiota (up to 15

273   genotypes for a single individual).[46] Future studies using multispacer typing or pangenome

274   analysis will clarify the diversity of *M. smithii* strains across samples from the same mother

275   and across different mothers. Non-methanogenic archaea could also be present in human

276   colostrum and milk.[3,4] In 16S amplicon sequencing datasets from present study, we found 3

277   reads of *Haloferax alexandrines*, a halophilic archaea that we recently isolated for the first

278   time from humans,[47] in a colostrum sample.

279   Our findings could result in a paradigm shift concerning vertical transmission and

280   maturation of the human microbiome: The entire microbiome could be acquired from the first

281   hours of life, but the profile is initially dominated by lactic acid bacteria and aerotolerant

282   species.[2] Methanogenic archaea could be neglected by current metagenomics techniques for

283   several reasons: extraction bias, primer bias, and depth bias.[24] Our results suggest that early

284   initiation of breastfeeding is critical to seed the infant's microbiota with methanogenic

285   archaea; the neglected critical commensals.

METHODS

Patients and samples

Healthy mothers aged over 18 years of age, with a full-term pregnancy and who have opted

for mixed or exclusive breastfeeding, were selected to participate in this study. The sampling

was done by a paediatrician. Before sampling, the paediatrician has washed their hands with

alcoholic solution and then the gloves were put on before touching the sampling site.

mammary and areolas are not cleaned. The samples collected from mothers on the second-

and tenth-day postpartum in sterile tubes by manual pressing at the neonatology unit of

«Hôpital de la Conception and Hôpital Nord» Marseille, France. The main inclusion criteria

of this study were breastfeeding and acceptance of participation in the study protocol. The

main criteria for exclusion were exclusively artificial breastfeeding, refusal of participation

and the presence of a disorder (mastitis or breast abscess) that may have an impact on the

microbiota of the study subject. Breast colostrum and milk samples were collected from

healthy women at the second- and tenth-day postpartum. All participants were informed and

gave their written consent before the samples were collected. The local IFR 48 ethics

committee under ascending number 2016-004 approved the consent and study protocol. The

authors certified that this study was not in opposition to the declaration of Helsinki and in

accordance with French laws respectively (certificates available on request).

Culture and isolation

A volume of 250 µL of breast milk or colostrum sample was seeded in ambient air in a sterile

Hungate tube (Dominique Dutscher, Brumath, France). The culture and isolation of

methanogenic archaea were performed according to the previously published protocol[21] under

aerobic conditions using coculture with *Bacteroides thetaiotaomicron*. Each Hungate tube

contained 5 mL of SAB broth 32 supplemented with ascorbic acid (1 g/L; VWR International,

311 Leuven, Belgium), uric acid (0.1 g/L) and glutathione (0.1 g/L; Sigma-Aldrich). The pH of

312 the culture media was adjusted to 7.5 with KOH (10 M). Five milliliters of SAB medium and

313 250 µL of milk or colostrum were inoculated with *B. thetaiotaomicron* (10^5 cells/mL) for

314 hydrogen production 15 at 37°C with agitation for seven days. The growth of any methanogen

315 was inferred from the production of methane (CH4) detected by gas chromatography as

316 previously described.[48] Subcultures were seeded on a Petri dish containing SAB medium

317 supplemented with 15 g/L agar and deposited in the upper chamber of a double-chamber. A

318 tube containing a non-inoculated SAB medium was used as negative control. For solid

319 medium culture, a non-inoculated agar Petri dish that has been used as a negative control.

320

DNA extraction and 16S rRNA gene sequencing

322 DNA was extracted using the E.Z.N.A.® Tissue DNA Kit (OMEGA bio-tek, Inc. 400

323 Pinnacle Way, Suite 450 Norcross, GA 30071, USA) according to the manufacturer's

324 instructions and the modified extraction protocol described by Dridi *et al*. 2009.[10] PCR was

325 performed with a PTC-200 automated thermal cycler (MJ Research, Waltham, USA) in 50 µL

326 of PCR mixture. The archaeal 16S rRNA gene was amplified using a 40-cycle program with

327 the archaeal primers SDArch0333aS15 (50-TCCAGGCCCTACGGG- 30) and

328 SDArch0958aA19 (5'-YCCGGCGTTGAMTCCAATT-3') (Eurogentec, Seraing, Belgium).

329 PCR products were purified and sequenced using the 3500xL genetic analyzer

330 (ThermoFisher, Waltham, MA USA) and the Big-Dye Terminator, version 1.1, cycle

331 sequencing kit DNA according to the manufacturer's instructions (Applied Biosystems, Foster

332 City, USA). The chromas Pro1.34 software (Technelysium Pty. Ltd) used for sequence

333 correction. BLASTn (nucleotide Basic Local Alignment Search Tool) searches were

334 performed against GenBank (http://blast.ncbi.nlm.nih.gov.gate1.inist.fr/Blast.cgi) for

335   taxonomic assignment. Two negative control samples consisting in master mix and RNAse

336   free water were introduced for every 5 samples tested.

337

338   **Genome sequencing**

339   Genomic DNA of the isolates C2 CSUR P5816 and M2 CSUR P5920 cultured from

340   colostrum and milk, respectively, were sequenced using MiSeq Technology (Illumina, Inc.,

341   San Diego CA 92121, USA) with a paired end and barcode strategy with 15 other projects

342   constructed according to the Nextera XT library kit (Illumina). The gDNA was quantified by

343   a Qubit assay with a high sensitivity kit (Life Technologies, Carlsbad, CA, USA) to 0.8 ng/µL

344   and diluted to require one ng of DNA as input. The "tagmentation" step fragmented the

345   genomic DNA. Then, PCR cycle amplification completed the tag adapters and introduced

346   dual-index barcodes. After purification on AMPure beads (Life Technologies), the libraries

347   were normalized on specific beads according to the Nextera XT protocol (Illumina).

348   Normalized libraries were pooled into a single library for sequencing via MiSeq. A pooled

349   single strand library was loaded onto the reagent cartridge and then onto the instrument along

350   with the flow cell. We performed automated cluster generation and paired-end sequencing

351   with dual index reads in a single 39-hour run in 2x251-bp. We obtained total information with

352   an 8.5 Giga base from an 899 K/mm2 cluster density with 94.9% (16,382,000 clusters) of the

353   clusters passing quality control filters. Within this pooled run, the index representation of the

354   isolate was determined to be 13.50%. The 2,212,330-paired end reads were filtered according

355   to the read qualities.

356

357   **Genome assembly and construction of rhizome**

358   The Illumina reads obtained for the 2 strains were assembled using SPAdes software

359   (http://bioinf.spbau.ru/spades) helped by GapFiller v2.1.1 35 to reduce the set. Subsequently,

51

360 the assembly was refined using manual finishing. Blastp was performed on all translated

361 coding sequences using the nr database. For each coding DNA sequence, the best blast hit was

362 determined from the max bit score. Only all hits related to the *Methanobrevibacter* genus after

363 data filtering were considered ORFans. The origin of all genes of the two strains C2 CSUR

364 P5816 (*M. smithii*) and M2 CSUR P5920 (*M. oralis*) were determined according to their

365 taxonomic affiliation. The rhizome representation was created using circos software 36 for the

366 two strains (*M. oralis* M2 CSUR P5920 and *M. smithii* C2 CSUR P5816) and the *M. oralis*

367 strain MBORA DSM 7256 isolated from the human mouth[49] and *M. smithii* strain WWM1085

368 isolated from the human gut.[50]

369

370 **Detection of *M. smithii* and *M. oralis* using qPCR**

371 Real-time PCR assays were performed with a MX3000TM system (Stratagene, Amsterdam,

372 The Netherlands) using the QuantiTect Probe PCR Kit (Qiagen, Courtaboeuf, France) with 5

373 pmol of each primer, probe labeled with FAM, and 5 ml of DNA in a final volume of 25 ml.

374 The PCR amplification program for *M. smithii* was 95° C for 15 min, followed by 42 cycles

375 of 95° C for 30 s and 60° C for 1min and for *M. oralis* amplification was 95° C for 15 min,

376 followed by 42 cycles of 95° C for 10s, 60° C for 45s and 45°C for 30s as previously

377 described[10,51]. The primers and probes used for *M. oralis* and *M. smithii* amplification were as

378 follows: *M. oralis*: M-cnp602F 5'-GCTGGTGTAATCGAACCTAAACG-3', cnp602R 5'-

379 CACCCATACCCGGATCCATA-3' FAM 5'-AGCAGTGCACCTGCTGATATGGAAGG-

380 3'; *M. smithii*: Smit.16S-740F, 5'-CCGGGTATCTAATCCGGTTC-3', Smit.16S-862R, 5'-

381 CTCCCAGGGTAGAGGTGAAA-3', Smit.16S FAM5'-

382 CCGTCAGAATCGTTCCAGTCAG-3'. We used calibration curves as previously

383 described[10,51] Two negative control samples consisting in master mix and RNAse free water

384 were introduced for every 5 samples tested.

Amplicon sequencing using universal archaeal-bacterial primers targeting the

hypervariable v3v4 region of the 16S rRNA gene

Amplicon sequencing was performed using 16S v3v4 amplification prior

to large-scale sequencing using an Illumina MiSeq Engine as previously described.[13]

Assignment was performed down to the species level.

Metabolic variables and groups

As we previously associated human methanogenic archaea and specifically M. smithii to

absence of obesity,[11,12] we sought to test if this association was found in the present study. A

body mas index (BMI) categorical variable was defined as follows: lean mothers (BMI < 19),

normal weight mothers for height (BMI $\geq$ 19 and $\leq$ 25), overweight mothers (BMI > 25 and <

30) and mothers who meet obesity criteria (BMI $\geq$ 30).

Statistical analysis

Clinical characteristics of the mothers and children were compared between *M. smithii*

positive and negative dyad. Gemellar pregnancy were excluded from the comparison for

characteristic of the newborn. Quantitative variables were analysed using the unpaired t-test

or Mann Whitney test according to the distribution of the data. Qualitative variables were

analysed using the two-sided Fisher exact test. Test used to test the previously reported

association between the depletion of *M. smithii* and obesity[11,12] were unilateral. A p-value <

0.05 was considered significant. GraphPad prism v8.1.1 (GraphPad software, San Diego, CA

USA) and XLSTAT19 (Addinsoft, Paris, France) were used for statistical analysis.

432   **Table 1. Summary of the results of molecular and culture analyses**

	Colostrum		Milk (around day 10)	
	Methane production	Strain identification	Methane production	Strain identification
Mother_018	NA	NA	–	ND
Mother_076	–	ND	+	*M. oralis* **strain M2 CSUR P5920[a]**
Mother_095	+	*M. smithii* **strain C1 CSUR P5920**	+	*M. smithii* **strain M5 CSUR P5919[c]**
Mother_096	+	*M. smithii* **strain C2 CSUR P5816[a]**	+	*M. smithii* **strain M6 CSUR P5818**
Mother_097	+	*M. smithii* **strain C3 CSUR P5922**	+	*M. smithii* **strain M7 CSUR P5820**
Mother_098	–	ND	–	NA
Mother_099	–	ND	ND	ND
Mother_100	–	ND	ND	ND
Mother_101	–	ND	ND	ND
Mother_102	–	ND	–	NA

Mother_103	–	ND	–	NA
Mother_104	NA	ND	+	*M. smithii* **strain M1 CSUR P5819**
Mother_105	NA	ND	+	*M. smithii* **strain M3 CSUR P5921**

433 Methane production detected by gas chromatography on Hungate tube with SAB medium and antioxidants inoculated with maternal colostrum or milk after transport in Ae-
434 Ana medium supplemented with antioxidants. PCR: Polymerase chain reaction detecting all archaea performed on colonies identified on Petri dishes from methane-positive
435 samples. Strain identification was performed by 16S rRNA gene sequencing., **+** = positive, - = negative, ND: not done, NA: sample not available. [a]Strains sequenced for
436 genome analysis.

Table 2. Alignment of the 16S rRNA gene sequences of 8 *M. smithii* strains isolated from human colostrum and milk with the reference sequence of the type strain *M. smithii* PS (= ATCC 35061= DSM 861 = OCM 144) - Sequence accession no.: U55233.

Strain	Origin	Sequence number	Length	Identities	Gaps
C2 CSUR P5816	Colostrum	Genome CAABOX000000000	1472bp	1343/1343 (100%)	0/1343 (0%)
M6 CSUR P5818	Milk (day 10)	LR584037	592bp	592/592 (100%)	0/592 (0%)
M3 CSUR P5921	Milk (day 10)	LR584041	608bp	604/605 (99.83%)	1/605 (0.16%)
C1 CSUR P5920	Colostrum	LR584035	604bp	603/604 (99.83%)	0/604 (0%)
C3 CSUR P5922	Colostrum	LR584038	594bp	593/594 (99.83%)	1/594 (0.17%)
M5 CSUR P5919	Milk (day 10)	LR584040	607bp	603/605 (99.66%)	2/605 (0.33%)
M7 CSUR P5820	Milk (day 10)	LR584039	597bp	595/597 (99.66%)	2/597 (0.33%)
M1 CSUR P5819	Milk (day 10)	LR584036	517bp	516/518 (99.61%)	2/518 (0.38%)

Bp: base pairs

Table 3. Comparison of dyads (mother-child) positive or negative for *M. smithii*

	Positive	Negative	p-value
Recruitment_center (Conception 1, Nord 2)	34 (79.0%)	71 (79.8%)	0.92[a]
Mother	n = 43	n = 89	
Mother_Age	31.1 ± 6.8	29.8 ± 6.1	0.27[b]
Mother_Weight (kg)	65 [57 - 73]	63 [54 - 74]	0.72[c]
Mother_height (cm)	163.7 ± 5.9	162.7 ± 6.6	0.40[b]
Mother_BMI	23.9 [21.4 - 26.3]	23.5 [20.7 - 28.3]	0.96[c]
Mother with BMI ≥ 30	2 (4.6%)	14 (15.7%)	0.033[d]
Underweight (BMI < 19)	3 (27%)	8 (73%)	
Lean (BMI ≥ 19 and ≤ 25)	23 (32%)	49 (68%)	
Overweight (BMI > 25 and < 30)	15 (45%)	18 (55%)[e]	
Obese (BMI ≥ 30)	2 (12%)[e]	14 (88%)	
Gestity	2 [1 - 4]	2 [2 - 4]	0.76[c]
Parity	1 [0 - 2]	1 [0 - 2]	0.86[c]
Diabetes	8 (18.6%)	13 (14.6%)	0.55[a]
Smoking (Y/N)	8 (18.6%)	9 (10.1%)	0.17[a]
Smoking (cigarette/day)			
Pregnancy			
Term_pregnancy (WA)	40 [38 - 41]	39 [38 - 40]	0.43[c]
Preterm	4 (9.3%)	6 (6.7%)	0.83[f]
Delivery_route (C-section)	14 (32.5%)	21 (23.6%)	0.27[a]
Gemellar_pregnancy	1 (2.3%)	3 (3.4%)	>0.99[f]
Newborn (twins excluded)	n = 42	n = 86	
Male	16 (38.1%)	46 (53.5%)	0.10[a]
Weight_Newborn (kg)	3232 ± 603	3169 ± 530	0.55[b]

Height_Children (cm)	48.75 [47 - 50]	48.50 [47 - 50]	0.78[c]
Body mass index	13.3 [12.3 - 15.2]	13.6 [12.6 - 14.7]	0.91[c]
Cranial_perimeter (cm)	34.6 ± 1.5	34.3 ± 1.7	0.31[b]
Breastfeeding (exclusive 1/mixed 0)	32 (76.2%)	69 (80.2%)	0.60[a]
Comparison to WHO standard curves[g]	**n = 39**	**n = 81**	
Weight_for_length z-score	0.56 [-0.92 to 1.43]	0.35 [-0.32 to 1.19]	0.81[c]
Weight_for_age z-score	-0.06 ± 1.24	-0.18 ± 1.04	0.57[b]
Length_for_age z-score	-0.39 ± 1.15	-0.47 ± 1.20	0.74[b]
BMI_for_age z-score	0.01 [-0.93 to 1.26]	0.18 [-0.51 to 1.05]	0.87[c]
BMI_for_age z-score > +2	3 (7.7%)	1 (1.2%)	0.20[f]
HeadCircumference_for_Age z-score	0.46 ± 1.19	0.20 ± 1.24	0.34[b]

441   [a]Two-sided chi square test, [b]Two sided unpaired t-test, [c]Two-sided unpaired Mann-Whitney test, [d]One-sided Fisher exact test (obesity previously

442   associated with depletion of M. smithii[11,12], see main text), [e]Exact Fisher test for 2 columns and 4 rows (two cells were significantly different :

443   overweight negative (<expected number) and obese positive (<expected number)), [f]Two-sided Fisher exact test, [g]Twins and preterm excluded.
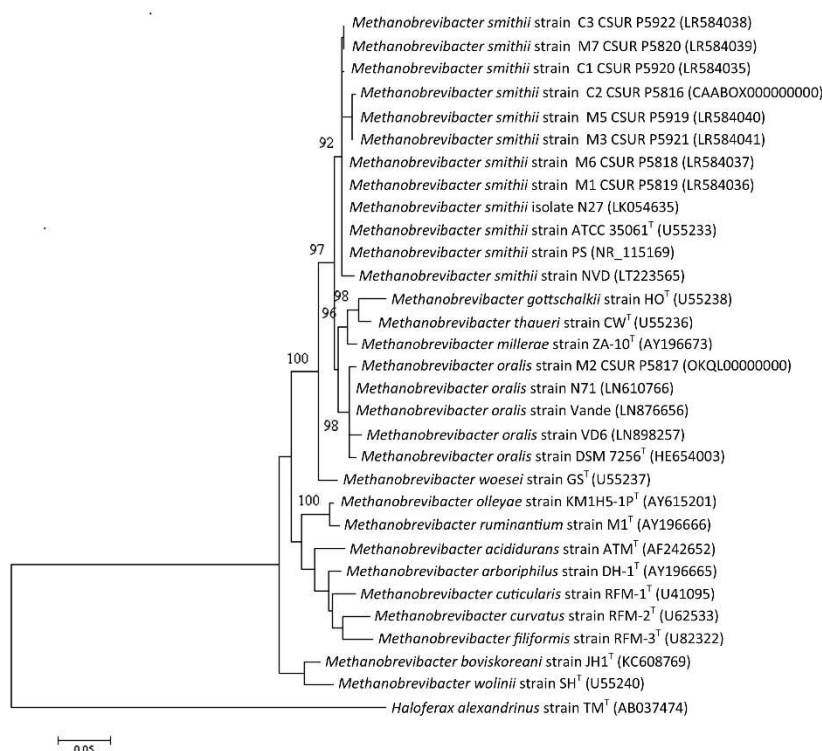
Figure 1: Molecular phylogenetic analysis by maximum likelihood method of the new isolates and their closest species.

Bootstrap values ≥ 90 % indicated at nodes. The evolutionary history was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model[52]. The tree with the highest log likelihood (-3941.91) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree for the heuristic search were obtained automatically by applying the Maximum Parsimony method. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.3801)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 18 nucleotide sequences. There were a total of 1490 positions in the final dataset. Evolutionary analyses were conducted in MEGA7[53].

a.

73.5% Archaea 77.2%
13.2% Bacteria 9.5%
0.4% Eukaryotes 0.3%
13.3% ORFans 12.6%
0.2% Undeter taxid 0.2%

M. oralis M2 CSUR P5920 from human's milk

M. oralis DSM7256 from human's oral cavity

b.

80.7% Archaea 79.1%
8.9% Bacteria 9.3%
0.3% Eukaryotes 0.4%
9.4% ORFans 10.9%
0.7% Undeter taxid 0.3%

M. smithii C2 CSUR P5816 from human's colostrum

M. smithii WWM1085 from human's feces
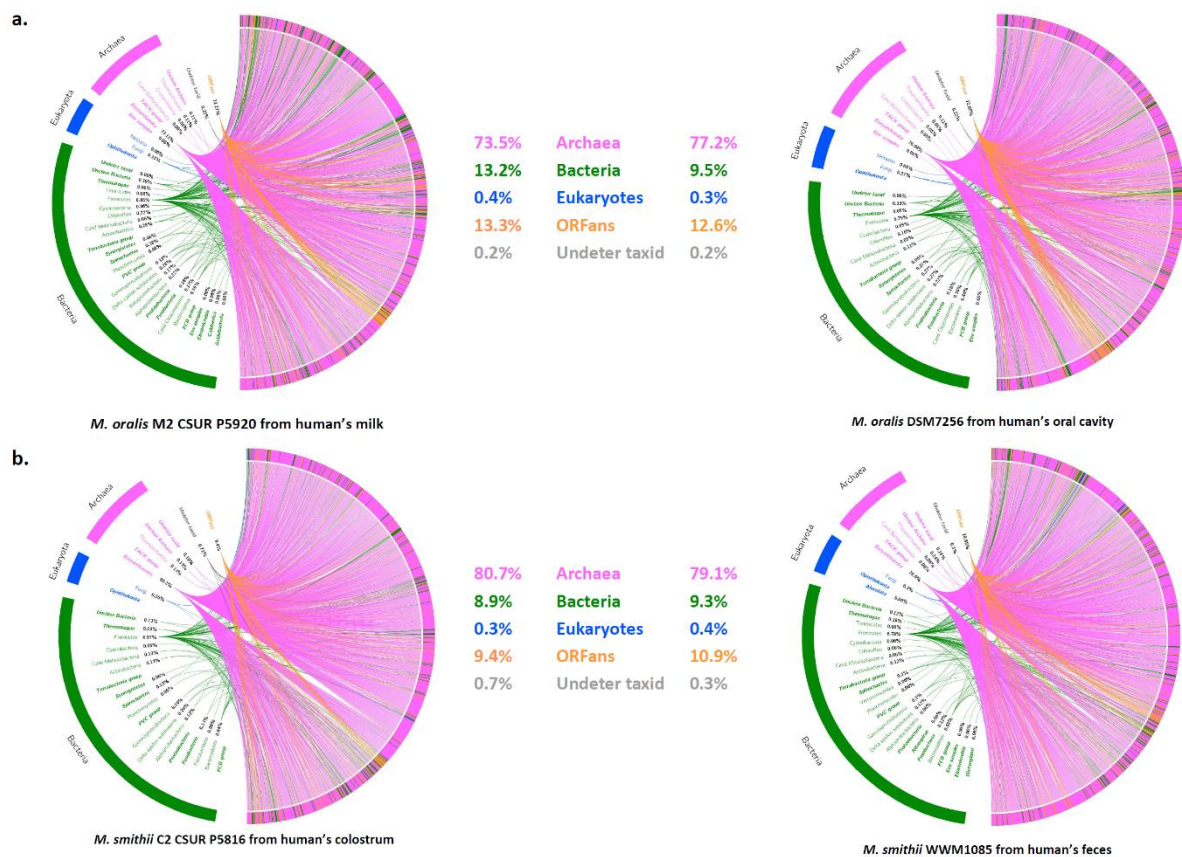
456 **Figure 2: Comparison of rhizomes of *Methanobrevibacter smithii* and**

457 ***Methanobrevibacter oralis* isolated from human milk with archaea previously described**

458 **in the digestive tract (*M. smithii*) and mouth (*M. oralis*).**

459 The *M. smithii* strain C2 CSUR P5816 was isolated from the colostrum of mother_2 (Table

460 1). Its genome (GenBank number: SAMEA104570327) was compared to the genome of a

461 strain isolated from human feces (=WWM1085, GenBank number: NQLD000000000000).

462 The *M. oralis* strain M2 CSUR was isolated from the milk (day 10 after delivery) of

463 mother_11. Its genome P5920 (GenBank number: SAMEA10457076) was compared to the

464 genome of the type strain of *M. oralis* strain ZR (genome number: NZ_LWMU00000000.1)

465 isolated from the human oral cavity (=DSM7256, =JCM 30027).

466   Reference List

467   1   Victora, C. G. *et al.* Breastfeeding in the 21st century: epidemiology, mechanisms, and

468   lifelong effect. *Lancet* **387**, 475-490, doi:10.1016/S0140-6736(15)01024-7 (2016).

469   2   Togo, A. *et al.* Repertoire of human breast and milk microbiota: a systematic review.

470   *Future Microbiol* (2019).

471   3   Ward, T. L., Hosid, S., Ioshikhes, I. & Altosaar, I. Human milk metagenome: a

472   functional capacity analysis. *BMC Microbiol.* **13**, 116, doi:10.1186/1471-2180-13-116

473   (2013).

474   4   Jiménez, E. *et al.* Metagenomic Analysis of Milk of Healthy and Mastitis-Suffering

475   Women. **31**, 406-415, doi:10.1177/0890334415585078 (2015).

476   5   Hug, L. A. *et al.* A new view of the tree of life. *Nat Microbiol* **1**, 16048 (2016).

477   6   Pausan, M. R. *et al. Exploring the archaeome: detection of archaeal signatures in the*

478   *human body.* (2019).

479   7   Thauer, R. K., Kaster, A. K., Seedorf, H., Buckel, W. & Hedderich, R. Methanogenic

480   archaea: ecologically relevant differences in energy conservation. *Nat Rev Microbiol*

481   **6**, 579-591, doi:10.1038/nrmicro1931 (2008).

482   8   Samuel, B. S. & Gordon, J. I. A humanized gnotobiotic mouse model of host-

483   archaeal-bacterial mutualism. *Proceedings of the National Academy of Sciences* **103**,

484   10011-10016, doi:10.1073/pnas.0602187103 (2006).

485   9   Samuel, B. S. *et al.* Genomic and metabolic adaptations of *Methanobrevibacter*

486   *smithii* to the human gut. *Proceedings of the National Academy of Sciences* **104**,

487   10643-10648, doi:10.1073/pnas.0704189104 (2007).

488   10   Dridi, B., Henry, M., El Khéchine, A., Raoult, D. & Drancourt, M. High Prevalence of

489   *Methanobrevibacter smithii* and *Methanosphaera stadtmanae* Detected in the Human

490   Gut Using an Improved DNA Detection Protocol. *PLoS ONE* **4**, e7063,

491   doi:10.1371/journal.pone.0007063 (2009).

492   11   Million, M. *et al.* Correlation between body mass index and gut concentrations of

493   *Lactobacillus reuteri*, *Bifidobacterium animalis*, *Methanobrevibacter smithii* and

494   *Escherichia coli*. *Int J Obes (Lond)* **37**, 1460-1466, doi:10.1038/ijo.2013.20 (2013).

495   12   Million, M. *et al.* Obesity-associated gut microbiota is enriched in *Lactobacillus*

496   *reuteri* and depleted in *Bifidobacterium animalis* and *Methanobrevibacter smithii*. *Int*

497   *J Obes (Lond)* **36**, 817-825, doi:10.1038/ijo.2011.153 (2012).

498   13   Million, M. *et al.* Increased gut redox and depletion of anaerobic and methanogenic

499   prokaryotes in severe acute malnutrition. *Scientific Reports* **6**, 26051,

500   doi:10.1038/srep26051 (2016).

501   14   Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. & Brown, P. O. Development

502   of the human infant intestinal microbiota. *PLoS biology* **5**, e177,

503   doi:10.1371/journal.pbio.0050177 (2007).

504   15   Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308**,

505   1635-1638, doi:10.1126/science.1110591 (2005).

506   16   Miller, T. L., Wolin, M. J., Conway de Macario, E. & Macario, A. J. Isolation of

507   *Methanobrevibacter smithii* from human feces. *Applied and Environmental*

508   *Microbiology* **43**, 227-232 (1982).

509   17   La Scola, B., Khelaifia, S., Lagier, J. C. & Raoult, D. Aerobic culture of anaerobic

510   bacteria using antioxidants: a preliminary report. *European journal of clinical*

511   *microbiology & infectious diseases : official publication of the European Society of*

512   *Clinical Microbiology* **33**, 1781-1783, doi:10.1007/s10096-014-2137-4 (2014).

513   18   Ames, B. N., Cathcart, R., Schwiers, E. & Hochstein, P. Uric acid provides an

514   antioxidant defense in humans against oxidant- and radical-caused aging and cancer: a

515   hypothesis. *Proc Natl Acad Sci U S A* **78**, 6858-6862 (1981).

516   19   Dione, N., Khelaifia, S., La Scola, B., Lagier, J. C. & Raoult, D. A quasi-universal

517   medium to break the aerobic/anaerobic bacterial culture dichotomy in clinical

518   microbiology. *Clinical microbiology and infection : the official publication of the*

519   *European Society of Clinical Microbiology and Infectious Diseases* **22**, 53-58,

520   doi:10.1016/j.cmi.2015.10.032 (2016).

521   20   Peden, D. B. *et al.* Uric acid is a major antioxidant in human nasal airway secretions.

522   *Proc Natl Acad Sci U S A* **87**, 7638-7642 (1990).

523   21   Khelaifia, S. *et al.* Aerobic culture of methanogenic archaea without an external

524   source of hydrogen. *European Journal of Clinical Microbiology & Infectious Diseases*

525   **35**, 985-991, doi:10.1007/s10096-016-2627-7 (2016).

526   22   Levasseur, A. *et al.* The Rhizome of Lokiarchaeota Illustrates the Mosaicity of

527   Archaeal Genomes. *Genome biology and evolution* **9**, 2635-2639,

528   doi:10.1093/gbe/evx208 (2017).

529   23   Marinkovic, V. *et al.* Antioxidative Activity of Colostrum and Human Milk: Effects of

530   Pasteurization and Storage. *Journal of pediatric gastroenterology and nutrition* **62**,

531   901-906 (2016).

532   24   Lagier, J. C. *et al.* Culturing the human microbiota and culturomics. *Nat Rev*

533   *Microbiol*, 540-550, doi:10.1038/s41579-018-0041-0 (2018).

534   25   Monteiro, L. *et al.* Complex polysaccharides as PCR inhibitors in feces: Helicobacter

535   pylori model. *J Clin Microbiol* **35**, 995-998 (1997).

536   26   Powell, H. A., Gooding, C. M., Garrett, S. D., Lund, B. M. & McKee, R. A.

537   Proteinase inhibition of the detection of *Listeria monocytogenes* in milk using the

538   polymerase chain reaction. *Lett. Appl. Microbiol.* **18** (1994).

539   27   Bickley, J., Short, J. K., McDowell, D. G. & Parkes, H. C. Polymerase chain reaction

540   (PCR) detection of *Listeria monocytogenes* in diluted milk and reversal of PCR

541   inhibition caused by calcium ions. *Lett Appl Microbiol* **22**, 153-158 (1996).

542   28   Al-Soud, W. A. & Radstrom, P. Purification and characterization of PCR-inhibitory

543   components in blood cells. *J Clin Microbiol* **39**, 485-493, doi:10.1128/JCM.39.2.485-

544   493.2001 (2001).

545   29   Vaughan-Shaw, P. G. *et al.* A simple method to overcome the inhibitory effect of

546   heparin on DNA amplification. *Cellular oncology* **38**, 493-495, doi:10.1007/s13402-

547   015-0250-8 (2015).

548   30   Miller, T. L., Weaver, G. A. & Wolin, M. J. Methanogens and anaerobes in a colon

549   segment isolated from the normal fecal stream. *Appl Environ Microbiol* **48**, 449-450

550   (1984).

551   31   Weaver, G. A., Krause, J. A., Miller, T. L. & Wolin, M. J. Incidence of methanogenic

552   bacteria in a sigmoidoscopy population: an association of methanogenic bacteria and

553   diverticulosis. *Gut* **27**, 698-704 (1986).

554   32   Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography.

555   *Nature* **486**, 222-227, doi:10.1038/nature11053 (2012).

556   33   Koenig, J. E. *et al.* Succession of microbial consortia in the developing infant gut

557   microbiome. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4578-4585,

558   doi:10.1073/pnas.1000081107 (2011).

559   34   Wampach, L. *et al.* Colonization and Succession within the Human Gut Microbiome

560   by Archaea, Bacteria, and Microeukaryotes during the First Year of Life. *Front*

561   *Microbiol* **8**, 738, doi:10.3389/fmicb.2017.00738 (2017).

562   35   Rowe-Murray, H. J. & Fisher, J. R. W. Baby friendly hospital practices: cesarean

563   section is a persistent barrier to early initiation of breastfeeding. *Birth* **29**, 124-131

564   (2002).

565   36   Grine, G., Boualam, M. A. & Drancourt, M. *Methanobrevibacter smithii*, a

566   methanogen consistently colonising the newborn stomach. *European Journal of*

567   *Clinical Microbiology & Infectious Diseases* **36**, 2449-2455, doi:10.1007/s10096-017-

568   3084-7 (2017).

569   37   van de Pol, J. A. *et al.* Gut Colonization by Methanogenic Archaea Is Associated with

570   Organic Dairy Consumption in Children. *Front Microbiol* **8**, 355,

571   doi:10.3389/fmicb.2017.00355 (2017).

572   38   Khan, J., Vesel, L., Bahl, R. & Martines, J. C. Timing of breastfeeding initiation and

573   exclusivity of breastfeeding during the first month of life: effects on neonatal mortality

574   and morbidity--a systematic review and meta-analysis. *Matern Child Health J* **19**,

575   468-479 (2015).

576   39   Sharma, I. K. & Byrne, A. Early initiation of breastfeeding: a systematic literature

577   review of factors and barriers in South Asia. *Int Breastfeed J* **11**, 17 (2016).

578   40   Edmond, K. M. *et al.* Delayed breastfeeding initiation increases risk of neonatal

579   mortality. *Pediatrics* **117**, e380-386 (2006).

580   41   Armstrong, J., Reilly, J. J. & Child Health Information, T. Breastfeeding and lowering

581   the risk of childhood obesity. *Lancet* **359**, 2003-2004, doi:10.1016/S0140-

582   6736(02)08837-2 (2002).

583 42 Armougom, F., Henry, M., Vialettes, B., Raccah, D. & Raoult, D. Monitoring

584   bacterial community of human gut microbiota reveals an increase in *Lactobacillus* in

585   obese patients and Methanogens in anorexic patients. *PLoS One* **4**, e7125,

586   doi:10.1371/journal.pone.0007125 (2009).

587 43 Cabrera-Rubio, R. *et al.* The human milk microbiome changes over lactation and is

588   shaped by maternal weight and mode of delivery. *Am J Clin Nutr* **96**, 544-551,

589   doi:10.3945/ajcn.112.037382 (2012).

590 44 Mayer-Davis, E. J. *et al.* Breast-feeding and risk for childhood obesity: does maternal

591   diabetes or obesity status matter? *Diabetes Care* **29**, 2231-2237, doi:10.2337/dc06-

592   0974 (2006).

593 45 Moissl-Eichinger, C. *et al.* Human age and skin physiology shape diversity and

594   abundance of Archaea on skin. *Sci Rep* **7**, 4039, doi:10.1038/s41598-017-04197-4

595   (2017).

596 46 Nkamga, V. D., Huynh, H. T. T., Aboudharam, G., Ruimy, R. & Drancourt, M.

597   Diversity of human-associated *Methanobrevibacter smithii* isolates revealed by

598   multispacer sequence typing. *Current Microbiology* **70**, 810-815, doi:10.1007/s00284-

599   015-0787-9 (2015).

600 47 Seck, E. H. *et al.* Salt in stools is associated with obesity, gut halophilic microbiota

601   and *Akkermansia muciniphila* depletion in humans. *Int J Obes (Lond)*,

602   doi:10.1038/s41366-018-0201-3 (2018).

603 48 Nkamga, V. D. & Drancourt, M. in *Bergey's Manual of Systematics of Archaea and*

604   *Bacteria* (eds William B. Whitman *et al.*) 1-7 (John Wiley & Sons, Ltd, 2016).

605 49 Ferrari, A., Brusa, T., Rutili, A., Canzi, E. & Biavati, B. Isolation and characterization

606   of *Methanobrevibacter oralis* sp. nov. *Current Microbiology* **29**, 6 (1994).

607 50 Jennings, M. E., Chia, N., Boardman, L. A. & Metcalf, W. W. Draft Genome

608 Sequence of *Methanobrevibacter smithii* Isolate WWM1085, Obtained from a Human

609 Stool Sample. *Genome Announc* **5**, doi:10.1128/genomeA.01055-17 (2017).

610 51 Bringuier, A., Khelaifia, S., Richet, H., Aboudharam, G. & Drancourt, M. Real-Time

611 PCR Quantification of Methanobrevibacter oralis in Periodontitis: Table 1. *Journal of*

612 *Clinical Microbiology* **51**, 993-994, doi:10.1128/JCM.02863-12 (2013).

613 52 Kimura, M. A simple method for estimating evolutionary rates of base substitutions

614 through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111-120 (1980).

615 53 Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics

616 Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**, 1870-1874,

617 doi:10.1093/molbev/msw054 (2016).

618

Article III:

Extremely Common Occurrence of Listeria monocytogenes in Human Milk in Mali: A Health Emergency.

Amadou Hamidou TOGO[1,2], Gregory DUBOURG[1], Aminata CAMARA[1,2], Salimata KONATE[1,2], Jeremy DELERCE[1], Claudia ANDRIEU[1], Abdoulaye DJIMDE[2], Mahamadou Ali THERA[2], Matthieu MILLION[1], Didier RAOULT[1,*]

.

1 **Extremely Common Occurrence of *Listeria monocytogenes* in Human Milk in Mali: A**

2 **Health Emergency**

3 Amadou Hamidou TOGO[1-3], Gregory DUBOURG[1,2], Aminata CAMARA[1-3] , Salimata

4 KONATE[1,-3], Jeremy DELERCE[1,2], Claudia ANDRIEU[1,2], Abdoulaye DJIMDE[3],

5 Mahamadou Ali THERA[3], Matthieu MILLION[1,2], Didier RAOULT[1,2*]

6 [1]Aix Marseille Univ, IRD, AP-HM, MEPHI, Marseille, France

7 [2]IHU-Méditerranée Infection, Marseille, France

8 [3]Malaria Research and Training Center, Department of Epidemiology of Parasitic Diseases,

9 University of Science, Techniques and Technologies of Bamako, Bamako, Mali

10 ***Corresponding author**: Prof. Didier RAOULT

11 **Tel**: 00 33 4 13 73 73 24 01

12 **Fax** : 00 33 4 12 12 73 23 07

13 **E-mail** : didier.raoult@gmail.com

14 **Keywords**: *Listeria monocytogenes*, colostrum, milk, microbiome, microbiota, culture,

15 human, breastfeeding, Mali

16 **Abstract word count:** 50

17 **Word count**: 1,479 words

18 **Ref**: 12

19   **ABSTRACT**

20   *Listeria monocytogenes* is a foodborne pathogen. While exploring human milk microbiota, we

21   serendipitously found *L. monocytogenes* in the 10 included Malian mothers but in none of the

22   144 mothers in France. We alert world health authorities to explore a possible African

23   hyperendemicity and to investigate *L. monocytogenes* transmission by breastfeeding.

INTRODUCTION

According to infectious diseases textbooks, *Listeria monocytogenes* is a foodborne pathogen, widely distributed in the environment and regularly recovered from soil, dust, decaying vegetation, water, sewage, plants and animal feed [1,2]. It is considered to be part of the fecal flora of many mammals. *L. monocytogenes* has been isolated from a wide variety of foods and food products, such as dairy products, raw or processed meat, smoked fish, seafood and raw vegetables. Up to 5% of healthy humans are asymptomatic gut carriers with higher rates reported from household contacts of patients with clinical infection [1,2]. Although one transmission from mother's milk to her baby with neonatal infection has been reported [3], it is not known to be part of the human milk microbiota repertoire [4].

Recently, the world's highest level of endemicity of listeriosis to date was observed in South Africa in 2017-2018, where more than 1011 cases and 193 deaths were reported (with 28% mortality among cases with known outcome [5]). The global burden due to this disease was estimated at 23,150 illnesses, 5,463 deaths and 172,823 disability-adjusted life-years in 2010 [6]. Perinatal listeriosis represented one fifth of all cases (21%) and has been associated with amnionitis, leading to respiratory symptoms, meningitis and death among newborns. While placental infection has been well documented, maternal milk has been neglected as a potential source of *L. monocytogenes* infection and prevalence in human milk has not been investigated [5,6].

Over the past 10 years, we have developed a polyphasic approach to the study of the human microbiota by combining culturomics and metagenomics [7], two approaches that exclude *a priori* hypotheses and identify unexpected outbreaks. In an ongoing work using this combined approach to decipher the human milk microbiota in Mali and France, we serendipitously isolated *L. monocytogenes* from healthy Malian mothers. This prompted us to point out a putative critical role of the human milk in listeriosis.

METHODS

Healthy lactating mothers were recruited in Mali (at Kalabancoro Vaccination and Healthy

Children Surveillance Unit, Kalabancoro (a suburb of Bamako), which is an area of severe

acute malnutrition, n=10) and in France (Marseille, neonatology unit of the Hôpital de la

Conception and Hôpital Nord, n=144). Milk samples (250-1000µL) were collected in sterile

tubes by manual pressing after standard hygiene. The main inclusion criteria of this

observational study were breastfeeding and acceptance of participation in the study protocol.

A written consent was obtained from each mother before sampling, in accordance with the

Helsinki declaration and Council for International Organizations of Medical Sciences 2016.

Study and consent procedure have been approved by the ethics committee of "Institut

Fédératif de Pathologies Transmissibles et Pathologies Infectieuses Tropicales 48", under the

consent number 2016-004 and "Faculté de Médecine, de Pharmacie et d'Odonto-Stomatologie

du Mali" Institutional Ethics Committee under Number 20l4/46/CE/FMPOS. The samples

were transferred from Mali to France in accordance with the Nagoya protocol.

In the present study, two conditions were used in relation to the small volume

(<1000µL), including incubation in blood culture bottle with rumen and sheep blood, in

aerobic and anaerobic atmosphere at 37°C, as previously reported [7]. Briefly, the milk

samples were inoculated in blood culture bottles at 37°C. One mL was collected from the

blood culture bottle using a sterile syringe at time 24h, 72h, 7 days, 10 days, 15 days, 21 days

and 30 days. Serial dilutions were performed up to 10^{-10} and inoculated on Columbia agar

with 5% sheep blood (bioMérieux, Marcy-l'Etoile, France) in aerobic or anaerobic

atmosphere. Phenotypic analysis was performed by matrix assisted laser desorption

ionization–time of flight (MALDI-TOF). All samples were also analyzed by v3v4 16S

metagenomics as previously reported [8].

74　　　　To confirm the detection of *L. monocytogenes*, we used specific PCR and targeted

75　　culture. The specific *L. monocytogenes* PCR targeting the *hly* gene was directly performed on

76　　specimen as previously described [9]. In parallel, specimens were enriched in Fraser broth

77　　and incubated during 24h at 37°C and 48h at 4°C. Broths were then subcultured on Columbia

78　　agar and PALCALM medium at 37°C. Specimen were also directly inoculated onto solid

79　　PALCAM medium (at 37°C and 4°C). Suspected colonies were subjected to MALDI-TOF

80　　analysis.

81　　　　　Species identification was confirmed by genome sequencing and analysis of the 16S

82　　and rpoB gene, digital DNA-DNA hybridization, and average nucleotide identity. Genomic

83　　DNA was extracted, after a 30 min lysozyme incubation at 37°C, on the EZ1 advanced XL

84　　biorobot (Qiagen, Hilden, Germany) using the EZ1 DNA tissues kit and quantified by a Qubit

85　　assay with the high sensitivity kit (Thermofisher Scientific, Waltham, MA, USA). It was then

86　　sequenced on the MiSeq technology (Illumina, San Diego, CA, USA) with the paired end

87　　strategy by the use of Nextera XT DNA library prep kit (Illumina, San Diego, CA, USA).

88　　Sequencing with dual index reads was performed over a single 39-hour run in 2x250-bp with

89　　a Miseq reagent Kit (V2-500 cycles) (Illumina, San Diego, CA, USA). The assembly of the

90　　strains was performed with Spades and annotated with Prokka. Average nucleotide identity

91　　was calculated using orthoANI [10]. The pangenome of the 8 strains isolated in this study and

92　　the representative genome of *L. monocytogenes* (National Center for Biotechnological

93　　Information) was performed according to the presence/absence of genes using Roary.

94

95　**RESULTS**

96　*L. monocytogenes* was detected in all the 10 included mothers in Mali but in none of the 144

97　French mothers (Supplementary Data). Using non-specific methods, 9/10 were positive by

98　culturomics and 6/10 by v3v4 targeted metagenomics. We confirmed the presence of viable L.

99   monocytogenes in milk samples from Mali using specific culture and PCR. 4/10 samples were

100   positive using Fraser broth medium subsequently seeded on PALCAM medium, 4/10 with

101   direct inoculation on PALCALM medium, and 9/10 by quantitative PCR.

102   Two strains isolated by culturomics and six strains isolated by specific culture were

103   available for MALDI-TOF and genomic analysis. MALDI-TOF analysis suggested that the

104   strains isolated in this study were closed but formed a cluster that do not include the type

105   strain (Figure 1a). Genomic analysis was performed, and the digital DNA-DNA hybridization

106   (dDDH) showed that the closest species was *L. monocytogenes* with an 87.3% dDDH value.

107   Average nucleotide identity (OrthoANI values) between our strains ranged from 99.9 to 100%

108   and 98.52-98.58% with *L. monocytogenes* (Figure 1b-c). OrthoANI values were much lower

109   for other *Listeria* species: 89.78% with *L. marthii*, 88.2% with *L. innocua* and 82.8% with *L.*

110   *ivanovii*. Therefore, the genomic analysis demonstrated that the 8 strains recovered in this

111   study were closely related and corresponded to the *L. monocytogenes* species.

112   The pangenome analysis including our 8 strains and the type strain confirmed that the

113   type strain formed an outgroup. We therefore investigated if all the strains isolated here

114   correspond to a clonal dissemination. All but two strains (P9669 and P9726) harbored a

115   plasmid (59 genes, 62,000 nucleotides), and the results of the accessory genome analysis were

116   in support of polyclonality (Figure 1d).

117

118   **DISCUSSION**

119   Here, we show that *L. monocytogenes* could be part of the human milk microbiota, notably in

120   developing countries in tropical areas endemic for severe acute malnutrition. This was

121   confirmed by the culture of several strains, specific culture, specific PCR, genome sequencing

122   and most recent taxonomic methods (average nucleotide identity of orthologous genes). In a

123   recent comprehensive literature analysis, we found only one case reporting *L. monocytogenes*

124   in human milk associated with neonatal infection [3,4]. Transmission of *L. monocytogenes* in

125   mammal milk is well known but has not been investigated in humans. It is likely that

126   anthropocentrism and disconnection between humans and other animals have slowed medical

127   progress.

128   The fact that human milk could be a source of *L. monocytogenes* for newborns and

129   infants in developing countries is critical since breastfeeding is largely used and

130   recommended in this context for several months and years. This a major public health issue

131   because it represents a neglected source of contamination and an unexpected opportunity to

132   prevent and/or early treat neonatal listeriosis, chronic diarrhea and severe acute malnutrition

133   in children.

134   The prevalence of *L. monocytogenes* is probably neglected in developing countries

135   endemic for chronic diarrhea and malnutrition because *L. monocytogenes* is not systematically

136   sought in this context. However, in one of the largest population-based studies including 6785

137   stool samples, *L. monocytogenes* was associated with chronic non-febrile diarrhea without

138   nausea in patients without antibiotics or recent travel [11]. Since severe acute malnutrition

139   and chronic diarrhea are intimately intricated [12], it is likely that digestive contamination

140   through breastfeeding in newborns and infants can contribute to neonatal listeriosis, chronic

141   diarrhea and malnutrition in children.

142   Assessment of the geographical distribution of this pathogen in Africa and developing

143   countries and control of this public health problem should be given top priority and

144   correspond to the WHO Millennium Development Goals 4 and 5: reduce child mortality and

145   improve maternal health. Here, we ask global health authorities to deploy the necessary

146   resources to explore a possible African hyperendemicity of *L. monocytogenes* and to

147   investigate the transmission of this pathogen by human milk.

160 **Competing interests**

161 The authors declare no competing interests.

162

163 **Ethical consideration**

164 Written consent was obtained from each mother before sampling in accordance with the

165 Helsinki declaration and CIOMS 2016. The study and consent procedure were approved by

166 the ethics committee of IFR 48 under consent number 2016-004 and by FMPOS Institutional

167 Ethics Committee (Mali, CE-FMPOS) under number 20l4/46/CE/FMPOS as of May 22, 2014

168 (available on request). The material transfer agreement is available upon request. The samples

169 were transferred from Mali to France in accordance with the Nagoya protocol.

170

171 **Data deposition**

172 The 8 strains were deposited in our culture collection (Collection de Souches de l'Unité des

173 ricketssies) under numbers P9661, P9662, P9663, P9664, P9668, P9669, P9725, and P9726.

174 The 8 corresponding genomes were deposited under the Bioproject PRJEB32287.

References

1. Prod'hom G, Bille J. *Listeria Monocytogenes*. In: Section 8 Clinical Microbiology: Bacteria Aerobic Gram-Positive bacilli. In: Cohen J, Powderly W.G. and Opal S.M. Infectious Diseases, Fourth Edition. Vol 2. Elsevier; **2017**: 1537–1540.

2. Bennett L. *Listeria monocytogenes*. In: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. Seventh Edition. Vol 2. Philadelphia, PA: Churchill Livingstone Elsevier; **2010**: 2707–2714

3. Svabić-Vlahović M, Pantić D, Pavićić M, Bryner JH. Transmission of *Listeria monocytogenes* from mother's milk to her baby and to puppies. Lancet **1988**; 2:1201.

4. Togo A, Dufour JC, Lagier JC, Dubourg G, Raoult D, Million M. Repertoire of human breast and milk microbiota: a systematic review. Future Microbiol. 2019 Apr 26. In press.

5. Dramowski A, Lloyd LG, Bekker A, et al. Neonatal listeriosis during a countrywide epidemic in South Africa: A tertiary hospital's experience. S Afr Med J **2018**; 108:818–27.

6. de Noordhout CM, Devleesschauwer B, Angulo FJ, et al. The global burden of listeriosis: a systematic review and meta-analysis. Lancet Infect Dis **2014**; 14:1073–1082.

7. Lagier JC, Dubourg G, Million M, et al. Culturing the human microbiota and culturomics. Nat Rev Microbiol **2018**; 16:540–550.

8. Million M, Tidjani Alou M, Khelaifia S, et al. Increased Gut Redox and Depletion of Anaerobic and Methanogenic Prokaryotes in Severe Acute Malnutrition. Sci Rep. **2016**; 6:26051.

198   9. Morel AS, Dubourg G, Prudent E, et al. Complementarity between targeted real-time

199   specific PCR and conventional broad-range 16S rDNA PCR in the syndrome-driven

200   diagnosis of infectious diseases. Eur J Clin Microbiol Infect Dis. **2015**; 34:561-70.

201   10. Lee I, Ouk Kim Y, Park SC, Chun J. OrthoANI: An improved algorithm and software

202   for calculating average nucleotide identity. Int J Syst Evol Microbiol. **2016**; 66:1100-

203   1103.

204   11. Schlech WF 3rd, Schlech WF 4th, Haldane H, et al. Does sporadic *Listeria*

205   gastroenteritis exist? A 2-year population-based survey in Nova Scotia, Canada. Clin

206   Infect Dis **2005**; 41:778–84.

207   12. Guerrant RL, Schorling JB, McAuliffe JF, de Souza MA. Diarrhea as a cause and an

208   effect of malnutrition: diarrhea prevents catch-up growth and malnutrition increases

209   diarrhea frequency and duration. Am J Trop Med Hyg. **1992**; 47:28-35.
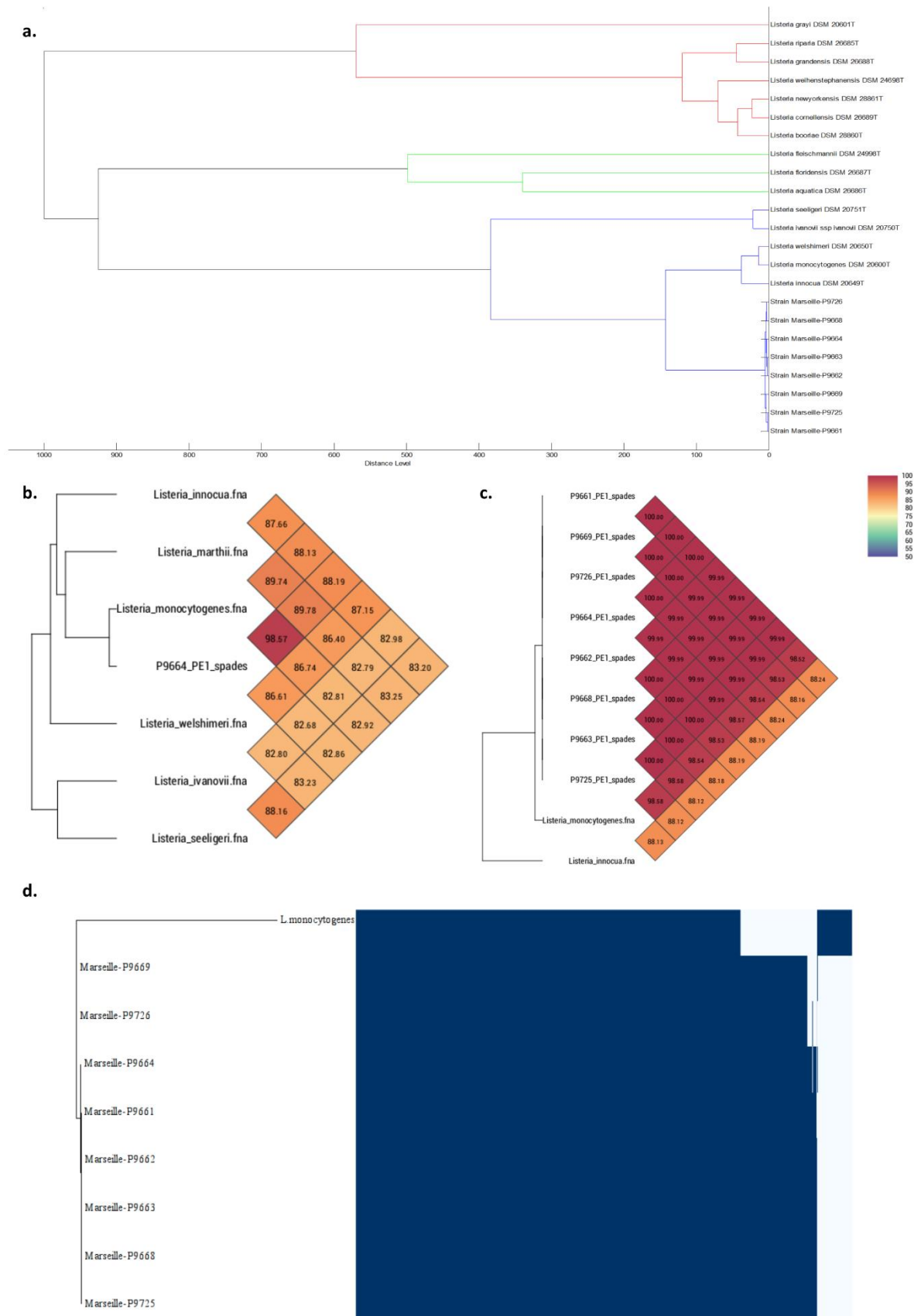
Figure 1. Phenotypic and genomic analyses of *L. monocytogenes* isolated from human milk in Mali

212 **a.** Phenotypic analysis by matrix assisted laser desorption ionization–time of flight (MALDI-

213 TOF) identified a cluster with the 8 present strains and *L. monocytogenes*, *L. innocua* and *L.*

214 *welshimeri*. The technique was unable to discriminate among these species. Main spectra

215 library dendrogram. **b. and c.** Genomic analysis confirmed that all 8 strains were closely

216 related and corresponded to *L. monocytogenes*. **d.** Representation of the pangenome of the 8

217 strains isolated in this study and the representative genome of *L. monocytogenes* (National

218 Center for Biotechnological Information) according to the presence/absence of genes using

219 Roary with 3449 gene clusters. All but two strains (P9669 and P9726) harbored a plasmid (59

220 genes, 62,000 nucleotides), and the results of the accessory genome analysis were in support

221 of polyclonality.

Article IV:

Culturomics provide critical prokaryotes strains for anti-Listeria and anticancer probiotics

Jean-Christophe Lagier, Matthieu Million, **Amadou H. Togo**, Saber Khelaifia and Didier Raoult·

Accepted Manuscript

Culturomics provide critical prokaryotes strains for anti-Listeria and anti-cancer probiotics

Jean-Christophe Lagier , Matthieu Million , Amadou H. Togo , Saber Khelaifia , Didier Raoult

Please cite this article as: Jean-Christophe Lagier , Matthieu Million , Amadou H. Togo , Saber Khelaifia , Didier Raoult , Culturomics provide critical prokaryotes strains for anti-Listeria and anti-cancer probiotics, *International Journal of Antimicrobial Agents* (2019), doi: https://doi.org/10.1016/j.ijantimicag.2019.05.017

Culturomics provide critical prokaryotes strains for anti-Listeria and anti-cancer probiotics

Jean-Christophe LAGIER[1,2*], Matthieu MILLION[1,2*], Amadou H. TOGO[1,2], Saber KHELAIFIA[1,2] and Didier RAOULT[1,2#]

1. Aix Marseille Univ, IRD, Microbes, Evolution, Phylogeny and Infection (MEPHI), Marseille, France

2. IHU Méditerranée Infection, Marseille, France

*contributed equally

#To whom correspondence should be addressed.

Prof. Didier RAOULT

IHU Méditerranée Infection
19-21 boulevard Jean Moulin
13385 Marseille cedex 5
France
Tel: +33 (0)4 13 73 24 01
Fax: +33 (0)4 13 73 24 02

85

The recent identification of a consortium of 11 human bacterial strains capable of stimulating anti-infectious and anti-cancer immunity using a specific action by colonization of the cecal and colic mucus and activation of CD103[+] dendritic cells of the gut mucosa provides a significant opportunity for probiotic-based interventions (1). This activation has proven to be a highly specific and non-inflammatory process, mounting a highly effective mucosal but also systemic IFNγ[+] CD8[+] T cells immune response. This immune response is critical against intracellular bacteria such as *Listeria monocytogenes* and cancer, particularly in the context of treatment by immune checkpoint inhibitors. Colonization of axenic mice by this consortium allowed a persistent response in the long term (up to 6 months). For *L. monocytogenes* infection, colonization by this consortium allowed an evolution towards cure, reversible by anti-CD8 monoclonal antibodies. The anti-tumor effect was observed even in the absence of anti-PD1 treatment in favor of a direct anti-cancer effect, even if the synergistic effect of anti-PD1 associated with the microbial consortium was maximal and prevented completely the appearance of the syngeneic tumor after inoculation. This anti-cancer effect was also totally reversible if anti-CD8 antibodies were administered. This illustrates the potential of culturomics in investigating the role of microbiota in immunity and particularly in providing strains for study and treatment (2).

This consortium of 11 human bacterial strains was composed of 4 strains considered as effectors, and 7 strains acting as potentiators. We were pleased to observe that among these 11 strains, 2 (18%) corresponded to strains isolated recently by culturomics (2). *Alistipes senegalensis,* isolated in 2012, is one of the 7 potentiators species and we named *Ihuella massiliensis* ("*Ruminococcaceae* bacterium cv2"), one of the bacterial effectors, isolated in 2015. These bacterial strains are available in our international collection of strains (Collection de Souche de l'Unité des Rickettsies = CSUR) under the number CSUR P156 and CSURP1486 respectively, and available for study and treatment for any worldwide researcher

and medical doctor. Nevertheless, caution must be maintained in these experimental models because the test of anti-*Listeria* activity of 3 of these strains (*Alistipes senegalensis, Ihuella massiliensis and Phascolactobacterium faecium*) against 6 different strains of *L. monocytogenes* was not demonstrated in agar plates (**Figure 1**), also suggesting that an indirect host-immunity-mediated effect cannot be ruled out.

We were also surprised to see that this interferon response of the digestive mucosa had also been described for the probiotic effect of *Lactobacillus casei* BL23 and *Lactobacillus paracasei* CNCM I-3689 (4). However, according to Occam's razor principle, which postulates that a simpler explanation is more likely to be true, this probiotic effect against **oral administered** *Listeria monocytogenes,* **as in the experiment by Tanoue et al.,** is more likely related to the role of bacteriocins, which are a characteristic feature of lactic acid bacteria (4). This is worth future investigations since although *Listeria* is no longer a public health problem in Europe and the USA, an epidemic of more than 800 cases and 80 deaths is still ongoing in South Africa (5).

Whatever the physiopathological mechanism, culturomics, **providing new cultured strains**, is becoming a key partner in the discovery of anti-cancer probiotics, **but also in other preventive or therapeutic fields**. These strains, **associated with an immune checkpoint inhibitor promoting effect,** belong to *Bifidobacterium* species *(B. longum, B. breve, B. adolescentis)*, *Enterococcus hirae, Bacteroides fragilis, Bacteroides thetaiotaomicron, Barnesiella intestinihominis* or *Akkermansia muciniphila* (6-9).

The availability of 3 other strains, very similar to *Ihuella massiliensis (Ruminococcaceae* bacterium cv2), available in our collection, thanks to culturomics, expands the opportunities to reproduce the effect or identify a strain-dependent effect (10,11). Two new strains very close to *Ruminococcacea* bacterium cv2 have more recently been described by another Russian team, probably by work inspired by culturomics, and have led to the

description of a new species but their anti-cancer and anti-*Listeria* effect is worth investigating (12). These 6 strains offer an unparalleled opportunity to test and develop new anti-cancer probiotics. This is proof of concept that culturomics offers the probiotics of tomorrow.

The paradigm has changed (Figure). The very first probiotic strain isolated and widely marketed (*L. casei* strain shirota) was described in the 1930s by Minoru SHIROTA using a targeted human stool culture approach. The approach was selective on a bacterium resistant to gastric juice and able to reach alive the digestive tract. We propose a new model for the discovery of new anti-Listeria and anti-cancer probiotics (**Figure 2**):

1. ***The unsupervised extension of the repertoire***: Culturomics aim at the unsupervised extension of the human microbial repertoire. 500+ new human species have been isolated in our center thanks to this high throughput approach based on the ultra-fast identification by MALDI-TOF and spectral clustering. New species are rapidly identified, genomes are sequences and deposited in worldwide databases available for the community. No attention is paid to the clinical relevance of any strain or species. The goal is to discover all the human microbes whatever they do.

2. ***Clinical association***: Clinical studies on different diseases using untargeted strategies (culturomics or metagenomics) identify species / strains with clinical relevance. This is best done with case-control studies with microbiological repertoires updated thank to culturomics. For instance, the Silva_132 (latest release) database includes more and more culturomic species. In the study of Tanoue et al. (1), this has been done by including 6 human donors and discriminating those with high or low IFNgamma[+] CD8[+] T cell response in transplanted axenic mice.

3. *Experimental evidence*: Experimental animal studies confirm a potential role of species / strain identified in clinical association studies. These experimental studies use strains isolated in the previous step or available in the culturomic collection.

4. *The interventional evidence and therapeutic use*: Phase I to IV interventional studies confirmed the therapeutic effect. Big pharma can use strain(s) available in the culturomic collection.

This work once again highlights the need to obtain microorganisms in pure culture as a starting point for studies investigating the relationships between human microbiota and human diseases. Culturomics has been striving for nearly 10 years to explore the gut microbiota in depth. Indeed, culturomics has made it possible to isolate more than 1200 bacterial species from the human gut, including 501 new bacterial species (3). This study should prompt more teams to develop culturomics.

References

1. Tanoue T, Morita S, Plichta DR, Skelly AN, Suda W, Sugiura Y, Narushima S, Vlamakis H, Motoo I, Sugita K, Shiota A, Takeshita K, Yasuma-Mitobe K, Riethmacher D, Kaisho T, Norman JM, Mucida D, Suematsu M, Yaguchi T, Bucci V, Inoue T, Kawakami Y, Olle B, Roberts B, Hattori M, Xavier RJ, Atarashi K, Honda K. A defined commensal consortium elicits CD8 T cells and anti-cancer immunity**. Nature. 2019** Jan;565(7741):600-605.

2. Lagier JC, Dubourg G, Million M, Cadoret F, Bilen M, Fenollar F, Levasseur A, Rolain JM, Fournier PE, Raoult D. Culturing the human microbiota and culturomics. **Nat Rev Microbiol. 2018** May 1:540-550.

3. Lagier JC, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P, Caputo A, Cadoret F, Traore SI, Seck EH, Dubourg G, Durand G, Mourembou G, Guilhot E, Togo A, Bellali S, Bachar D, Cassir N, Bittar F, Delerce J, Mailhe M, Ricaboni D, Bilen M, Dangui Nieko NP, Dia Badiane NM, Valles C, Mouelhi D, Diop K, Million M, Musso D, Abrahão J, Azhar EI, Bibi F, Yasir M, Diallo A, Sokhna C, Djossou F, Vitton V, Robert C, Rolain JM, La Scola B, Fournier PE, Levasseur A, Raoult D. Culture of previously uncultured members of the human gut microbiota by culturomics. **Nat Microbiol. 2016** Nov 7;1:16203.

4. Million M, Angelakis E, Drissi F, Raoult D. Occam's razor and probiotics activity on Listeria monocytogenes. **Proc Natl Acad Sci 2013** Jan 2;110(1):E1. doi: 10.1073/pnas.1218418110. Epub 2012 Dec 27.

5. Dramowski A, Lloyd LG, Bekker A, Holgate S, Aucamp M, Reddy K et al. Neonatal listeriosis during a countrywide epidemic in South Africa: A tertiary hospital's experience. **S Afr Med J. 2018** Oct 2;108(10):818-827. doi: 10.7196/SAMJ.2018.v108i10.13207.

6. Daillère R, Vétizou M, Waldschmitt N, Yamazaki T, Isnard C, Poirier-Colame V, et al. Enterococcus hirae and Barnesiella intestinihominis Facilitate Cyclophosphamide-Induced Therapeutic Immunomodulatory Effects. **Immunity. 2016** Oct 18;45(4):931-943. doi: 10.1016/j.immuni.2016.09.009. Epub 2016 Oct 4.

7. Routy B, Le Chatelier E, Derosa L, Duong CPM, Alou MT, Daillère R, et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. Science. 2018 Jan 5;359(6371):91-97. doi: 10.1126/science.aan3706. Epub 2017 Nov 2.

8. Sivan A, Corrales L, Hubert N, Williams JB, Aquino-Michaels K, Earley ZM et al. Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. **Science 2015** ; 350, 1084–1089.

9. Vétizou M, Pitt JM, Daillère R, Lepage P, Waldschmitt N, Flament C, et al. Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. **Science. 2015** Nov 27;350(6264):1079-84.

10. Million M, Raoult D. Species and strain specificity of Lactobacillus probiotics effect on weight regulation. **Microb Pathog. 2013** Feb;55:52-4. doi: 10.1016/j.micpath.2012.09.013.

11. Fåk F, Bäckhed F. Lactobacillus reuteri prevents diet-induced obesity, but not atherosclerosis, in a strain dependent fashion in Apoe-/- mice. **PLoS One**. 2012;7(10):e46837. doi: 10.1371/journal.pone.0046837. Epub 2012 Oct 9.

12. Shkoporov AN, Chaplin AV, Shcherbakova VA, Suzina NE, Kafarskaia LI, Bozhenko VK, et al. Ruthenibacterium lactatiformans gen. nov., sp. nov., an anaerobic, lactate-producing member of the family Ruminococcaceae isolated from human faeces. **Int J Syst Evol Microbiol. 2016** Aug;66(8):3041-9. doi: 10.1099/ijsem.0.001143. Epub 2016 May 6.

91

Figure Legends

Alistipes senegalensis

Phascolarctobacterium faecium

Ihuella massiliensis
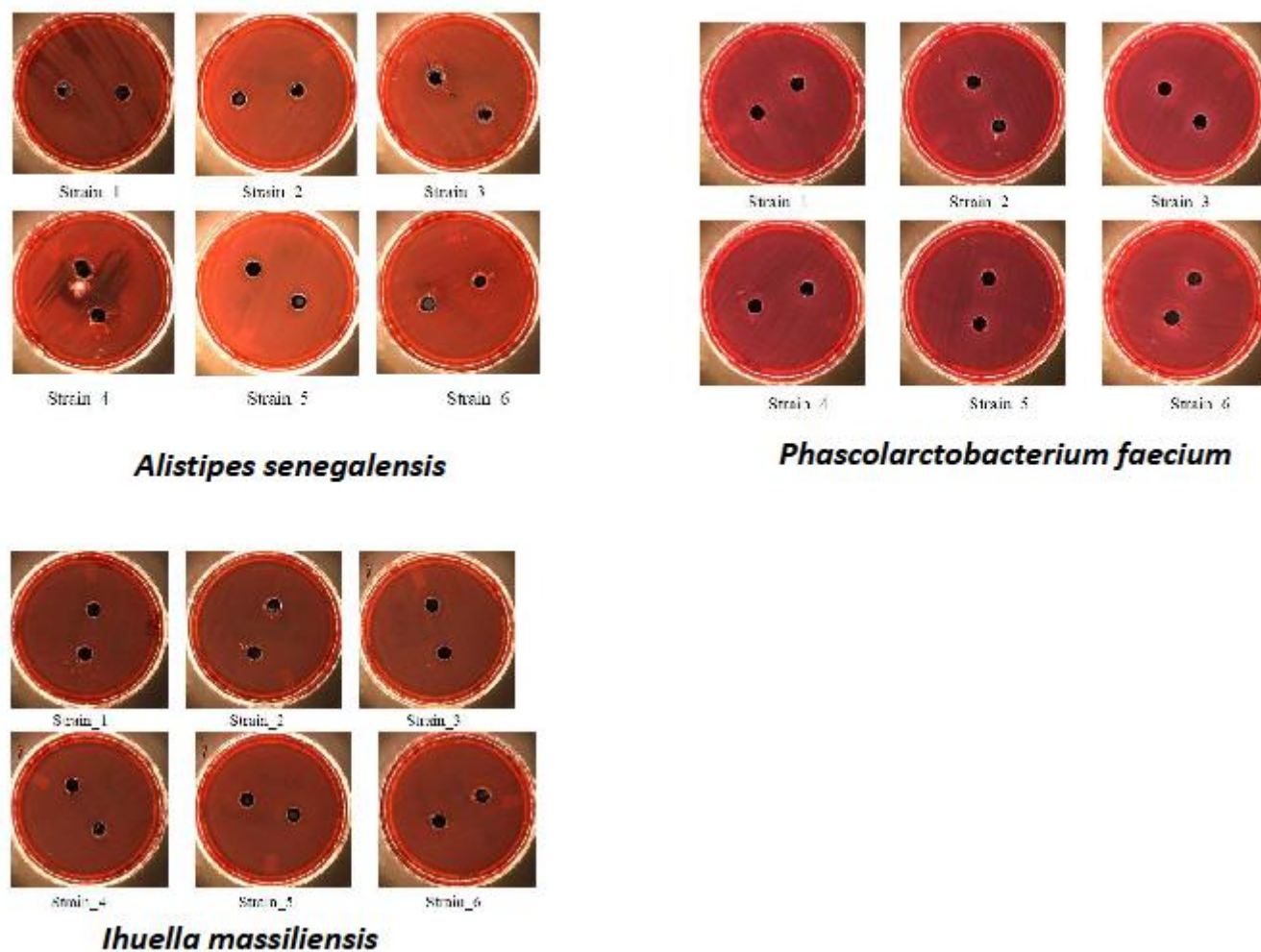
Figure 1: summary diagram of the extension of the repertoire to the potential modification of

the microbiota by probiotics
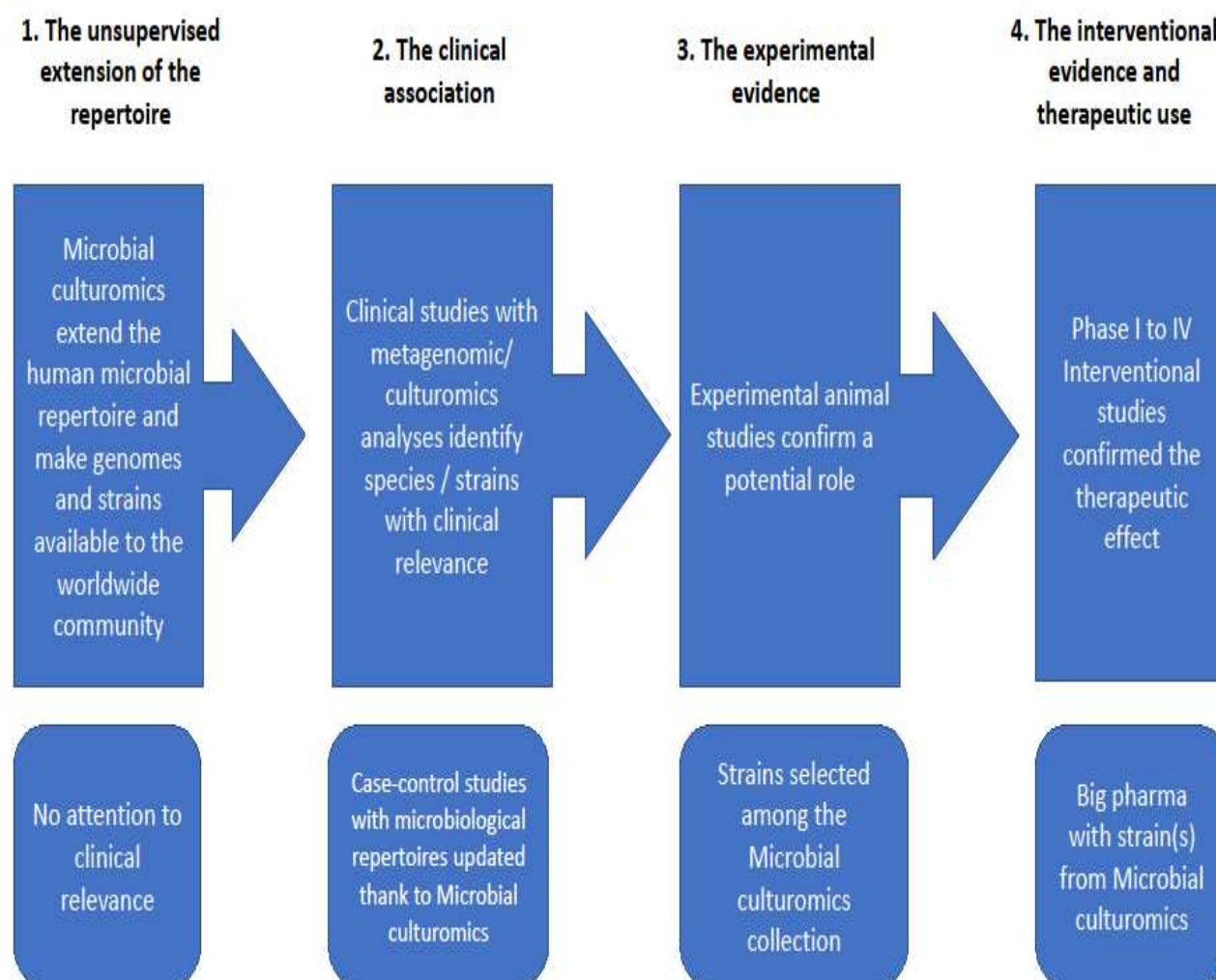
Figure 2: In vitro activity of *Alistipes senegalensis*, *Phascolarctobacterium faecium*, and *Ihuella massiliensis* against 6 *Listeria monocytogenes* strains

93

Article V:

Assessment of colostrum and breast milk microbiota from healthy breastfeeding mothers by culturomics and 16S rRNA Genes amplicon sequencing.

Amadou Hamidou Togo, Véronique Brevaut, Clotilde Des Robert, Aminata Camara, Salimata Konate, Dipankar Bachar, Catherine Robert, Mahamadou A Thera, Claude D'Ercole, Matthieu Million, Didier Raoult

En cours d'écriture

Article VI:

«*Veillonella massiliensis*», a new anaerobic species isolated from human colostrum.

Amadou Hamidou Togo, Clotilde Des Robert, Marion Bonnet, Pierre-Edouard Fournier, Didier Raoult and Matthieu Million.

95

«*Veillonella massiliensis*», a new anaerobic species isolated from human colostrum

CrossMark

Amadou Hamidou Togo [a], Clotilde Des Robert [b], Marion Bonnet [a], Pierre-Edouard Fournier [a], Didier Raoult [a], Matthieu Million [a,*]

[a] *Aix-Marseille Université, URMITE, UM63, CNRS7278, IRD198, Inserm 1095, Institut Hospitalo-Universitaire Méditerranée-Infection, 19-21 Boulevard Jean Moulin, 13005 Marseille, France*
[b] *APHM, CHU Hôpital de la Conception, Service de médecine néonatale, F-13385 Marseille, France*

ARTICLE INFO

ABSTRACT

We report here the main characteristics of "*Veillonella massiliensis*" strain Marseille-P3594[T] (CSUR P3594[T]) that was isolated from two human colostrum samples from two different mothers.

© 2017 Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons. org/licenses/by-nc-nd/4.0/).

In 2016, during an exploratory study of human breast milk by culturomics [1], two anaerobic isolates were isolated from the milk sampled from two French primigravida women, and a new species is proposed to introduce the strain as member of the genus *Veillonella*.

These bacterial strains could not be identified by our systematic matrix-assisted laser desorption-ionization time-of-flight screening on a Microflex spectrometer (Bruker Daltonics, Brenen, Germany) but both matched perfectly using this technique. Only the strain isolated first was described here. The patients gave a signed and informed consent and the study was validated by the ethics committee of the Institut Federatif de Recherche IFR48 under number 2016-004. The initial growth was obtained after three days of culture in a blood culture bottle supplemented with 5% of sheep blood and 5% of rumen fluid in anaerobic conditions. Agar-grown colonies were translucent with a mean diameter inferior to 0.5 mm. Bacterial cells were Gram-negative, coccoid, ranging in diameter from 1.5 to 3.9 μm with a mean of 2.4 μm. Strain Marseille-P3594[T] was catalase-negative and oxidase-negative. The 16S rRNA gene was sequenced using fD1-rP2 primers as previously described and a 3130-XL sequencer (Applied Biosciences, Saint Aubin, France). Strain Marseille-P3594[T] exhibited a 96.7% *sequence* similarity with *Veillonella ratti* strain JCM 6512[T], the phylogenetically closest species with standing in nomenclature (Fig. 1), which putatively classifies it as a member of the genus *Veillonella* within the family *Veillonellaceae* in the *Firmicutes* phylum. Species of the genus *Veillonella* are anaerobic, Gram-negative cocci and known for their ability to ferment lactate into propionate and acetate [2]. They are found in the mouth, the upper respiratory tract, the intestine, and the vagina as normal flora and are frequently isolated from human breast milk [3]. Accordingly, *Veillonella* species may be important for the development of the healthy mature anaerobic gut microbiota [4,5]. The species *Veillonella parvula* is often implicated in infectious endocarditis and osteomyelitis in humans. Strain Marseille-P3594[T] exhibited a 16S rRNA sequence divergence >1.3% (3.3%) with its phylogenetically closest species with standing in nomenclature, thus we propose the creation of the new *Veillonella* species named *V. massiliensis* (mas.si.li.en'sis. L. fem. adj. *massiliensis*, of Massilia, the Latin name of Marseilles, where the strain was isolated). Strain Marseille-P3594[T] is the type strain of the *Veillonella massiliensis* sp nov.

Nucleotide sequence accession number

The 16S r RNA gene sequence was deposited in Genbank under Accession number LT714171.

* Corresponding author at: URMITE, Aix Marseille Université, UM63, CNRS 7278, IRD 198, INSERM 1095, IHU – Méditerranée Infection, 19-21 Boulevard Jean Moulin, 13005 Marseille, France.

E-mail address: matthieumillion@gmail.com (M. Million).
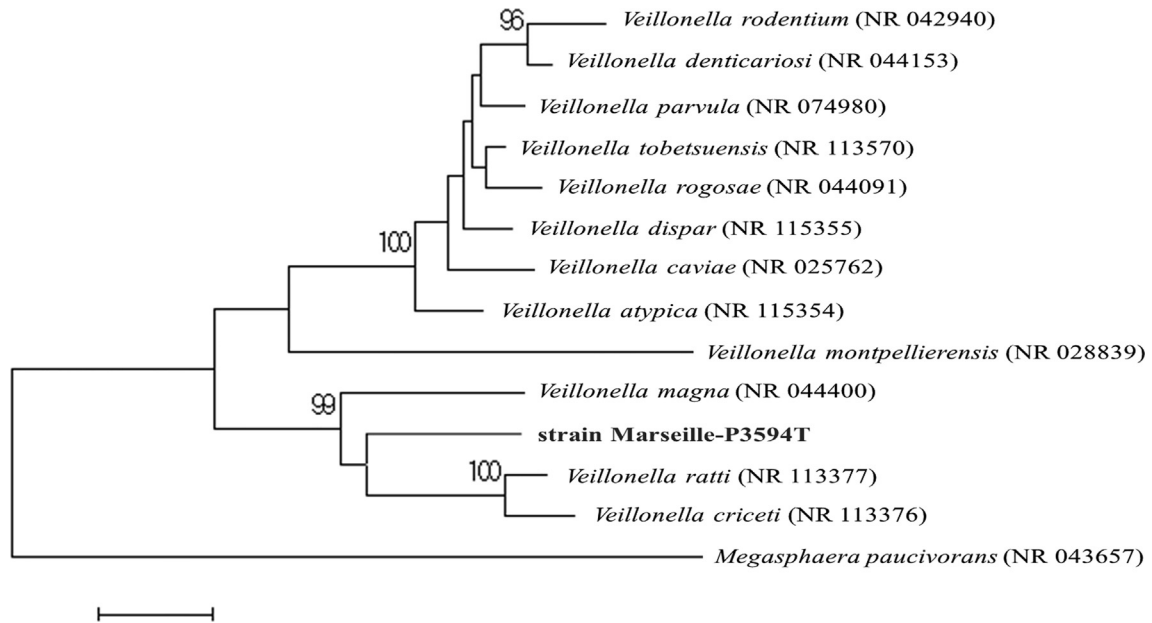
Fig. 1. Phylogenetic tree showing the position of "*Veillonella massiliensis*" strain Marseille-P3594[T] relative to the phylogenetically closest species. Sequences were aligned using CLUSTALW, and phylogenetic inferences obtained using the maximum-likelihood method within the MEGA software. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 500 times to generate a majority consensus tree. Only the bootstraps scores superior to 75 were retained. The scale bar indicates a 1% nucleotide sequence divergence.

Deposit in a culture Collection

Strain Marseille-P3594[T] was deposited in the Collection de Souches de l'Unité des Rickettsies (CSUR) under number P3594[T].

Conflict of interest

No conflict of interest to declare.

Funding sources

Acknowledgement

References

[1] Lagier JC, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P, et al. Culture of previously uncultured members of the human gut microbiota by culturomics. Nat Microbiol 2016;1:16203.
[2] Mays TD, Holdeman LV, Moore WEC, Rogosa M, Johnson JL. Taxonomy of the genus *Veillonella* Prévot. Int J Syst Evol Microbiol 1982;32:28–36.
[3] Cabrera-Rubio R, Collado MC, Laitinen K, Salminen S, Isolauri E, Mira A. The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. Am J Clin Nutr 2012;96:544–51.
[4] Million M, Diallo A, Raoult D. Gut microbiota and malnutrition. Microb Pathog 2016. http://dx.doi.org/10.1016/j.micpath.2016.02.003. pii: S0882-4010(15) 30212-6.
[5] Million M, Tidjani Alou M, Khelaifia S, Bachar D, Lagier JC, Dione N, et al. Increased gut redox and depletion of anaerobic and methanogenic prokaryotes in severe acute malnutrition. Sci Rep 2016;6:26051.

Article VII:

***Lactimicrobium massiliense* gen. nov., sp. nov.; *Anaerolactibacter massiliensis* gen. nov., sp. nov.; *Galactobacillus timonensis* gen. nov., sp. nov. and *Acidipropionibacterium timonense* sp. nov. isolated from Breast Milk from African Healthy Breastfeeding Women.**

Amadou Hamidou Togo, Awa Diop, Aminata Camara, Edmond Kuete, Salimata Konate, Véronique Brevaut, Clotilde Des Robert, Jeremy Delerce, Nicholas Armstrong, Yanis Roussel, Pierre-Edouard, Fournier, Mahamadou Ali Thera, Didier Raoult and Matthieu Million.

Lactimicrobium massiliense gen. nov., sp. nov.; *Anaerolactibacter massiliensis* gen. nov., sp. nov.; *Galactobacillus timonensis* gen. nov., sp. nov. and *Acidipropionibacterium timonense* sp. nov. isolated from breast milk from healthy breastfeeding African women

A. H. Togo[1,2], A. Diop[1,2], A. Camara[1,2], E. Kuete[1,2], S. Konate[1,2], V. Brevaut[3], C. Des Robert[4], J. Delerce[1,2], N. Armstrong[1,2], Y. Roussel[1,2], P.-E. Fournier[1,2], M. A. Thera[5], D. Raoult[1,2] and M. Million[1,2]

1) *Aix Marseille Univ, IRD, AP-HM, MEPHI,* 2) *IHU-Méditerranée Infection,* 3) *APHM, CHU Hôpital Nord, Service de médecine néonatale,* 4) *APHM, CHU Hôpital de la Conception, Service de médecine néonatale, F-13385, Marseille, France* and 5) *Malaria Research and Training Center, Department of Epidemiology of Parasitic Diseases, FMOS-FAPH, University of Science, Techniques and Technologies, Bamako, Mali*

Abstract

Four strains isolated by microbial culturomics from breast milk of healthy mothers from Mali were not identified and characterized by taxono-genomics. This led us to propose the new genera and species *Lactimicrobium massiliense*, *Anaerolactibacter massiliensis* and *Galactobacillus timonensis* containing type strain Marseille-P4301[T] (CSUR P4301[T]), Marseille-P4302[T] (CSUR P4302[T]) and Marseille-P4641[T] (CSUR P4641[T]), respectively. The strain Marseille-P4482 represents a novel species, *Acidipropionibacterium timonense,* in a previously known genus with type strain being Marseille-P4482[T] (CSUR P4482[T]).
© 2019 Published by Elsevier Ltd.

Corresponding author: M. Million, Aix Marseille Université, Institut Hospitalier Universitaire Méditerranée-Infection, 19-21 Boulevard Jean Moulin, 13005, Marseille, France.
E-mail: matthieumillion@gmail.com

Introduction

Human breast milk is a complex biological fluid produced by the mammary glands. Breast milk not only provides the essential nutrients for growth and development in new-borns, it also protects against different infectious diseases [1−5]. Several studies have reported unsuspected diversity, including many bacteria that promote maternal and child health in breast, colostrum and milk [6−9]. The breast-milk microbiota plays a key role in the colonization of the new-born's digestive tract and in the development of its immunity [10−12]. However, little is known about the composition of the human milk microbiota, and most published studies are limited by the use of metagenomics which does not differentiate between the DNA sequences of live bacteria and dead bacteria [13−16]. To date, to our knowledge, only one bacterial species, *Streptococcus lactarius*, has been officially described with breast milk as the first source of isolation [17]. This suggests that the human milk microbiota is neglected and remains largely unexplored.

We have therefore studied the microbiota of colostrum and breast milk of healthy mothers from France and Mali using the culturomics approach, an approach developed and applied in our laboratory over the past 10 years [18] to decipher the bacterial diversity of colostrum and breast milk. As part of this work, we isolated four new bacterial species from breast milk.

Here, we describe the isolation and taxonogenomic characterization of strain Marseille-P-4301[T], strain Marseille-P4302[T] and strain Marseille-P4641[T] as type strains of *Lactimicrobium massiliense* gen. nov., sp. nov. (CSUR 4301), *Anaerolactibacter massiliensis* gen. nov., sp. nov. (CSUR 4302) and *Galactobacillus*

timonensis gen. nov., sp. nov. (CSUR 4641), close to *Solobacterium moorei* strain JCM 10645 [19] and strain Marseille-P4482[T] as type strain of *Acidipropionibacterium timonense* sp. nov. (CSUR 4482) close to *Cutibacterium granulosum*. The four new bacterial species were isolated from a sample of breast milk from four healthy lactating Malian mothers.

Materials and methods

Sample collection

Milk sample were collected from healthy breastfeeding mothers in the suburban area of Bamako (Kalabankoro), Mali, between November 26 and December 1st, 2016. Approximately 20 mL of breast milk were collected aseptically in 50-mL sterile polypropylene conical tubes (Industrial Falcon, Reynosa, Mexico) containing 1 mL transport medium made with antioxidants, after breast cleaning by manual expression. Samples were collected between 5 and 19 months after delivery. All the donors had full-term pregnancies and their children were apparently healthy. Samples were stored at −20°C before being sent to our laboratory (IHU-Méditerranée Infection, Marseille, France) for analysis. Written consent was obtained from each mother before sampling, in accordance with the Helsinki declaration and CIOMS 2016. Study and consent procedures were approved by the ethics committee of IFR 48, under the Consent number 2016-004 and FMPOS Institutional Ethics Committee (Mali, CE-FMPOS) under Number 20l4/46/CE/FMPOS as at May 22, 2014 (available on request). The material transfer agreement (MTA) has been signed between IHU-Méditerranée Infection and Université des Sciences Technique et Technologique de Bamako (USTTB) and is available on request. The samples were transferred from Mali to France in accordance with the Nagoya protocol.

Strains isolation and identification

The first growth of these four strains occurred in May 2017. Approximately 2 mL of milk samples were preincubated under anaerobic conditions in blood-culture bottles enriched with 5% sheep blood and 5% rumen fluid (sterilised by filtration through a 0.2-μm diameter filter) and later inoculated onto sheep blood Columbia agar (bioMerieux, Marcy l'Etoile, France) as described elsewhere [18,20]. The identification procedure was conducted as previously described [20].

Phylogenetic analysis

The 16S rRNA gene amplification PCR and sequencing were performed as previously described [21]. Taxonomic assignation was performed as described elsewhere [20]. Phylogenetic analysis was performed by ClustalW alignment and the maximum likelihood method using MEGA7.0.26 software. The sequences from type strains were downloaded from the website https://www.ncbi.nlm.nih.gov.

Phenotypic, biochemical and chemotaxonomic analysis

Temperature range and atmosphere, pH and salinity for growth were assessed as previously described. Biochemical analysis using various strips (API® ZYM, API® 20 A, API® 50 CH and API Rapid ID 32 A) (bioMérieux) and oxidase and catalase tests (bioMérieux) were done according to the manufacturer's instructions. Analyses were performed as previously described [22]. Motility assay, Gram-staining, transmission electron microscopy and sporulation assay were also performed as describe elsewhere [23]. Cellular fatty acid methyl ester (FAME) and metabolic end products analysis were performed as previously described [20,24].

Genomic analysis

Genome sequencing, assemblage, annotation and comparison were performed as previously described [22,23,25]. The genomes of *Solobacterium moorei* strain RCA59-74[T] (NZ_AUKY00000000) [19], *Bulleidia extructa* strain W 1219[T] (NZ_ADFR00000000) [26], *Anaerorhabdus furcosa* strain VPI 3253[T] (NZ_FUWY00000000) [27], *Holdemania filiformis* strain J1-31B-1[T] (NZ_ACCF01000000) [28] and *Holdemania massiliensis* strain AP2[T] (CALK01000000) [29] were used for genome comparison of the strains Marseille-P4301, Marseille-P4302 and Marseille-P4641. The genomes of *Cutibacterium acnes* ATCC6919 (NZ_CP023676) [30,31], *Cutibacterium avidum* ATCC 25577 (NZ_AGBA01000000) [32], *Cutibacterium granulosum* DSM 20700 (NZ_AOSS00000000), *Pseudopropionibacterium propionicum* F0230a (NC_018142), *Propionibacterium acidifaciens* strain C3M 31 (NZ_AUFR00000000), *Acidipropionibacterium thoenii* strain NCFB 568 (NZ_AUHZ01000000) and *Acidipropionibacterium acidipropionici* strain NCFB 563 (NZ_ATYU01000000) were used for genome comparison of strain Marseille-P4482.

Results

Strain isolation and identification

The strains were first isolated after 7 days (Marseille-P4301 and Marseille-P4302 strain) and 10 days (Marseille-P4641 strain Marseille-P4482 strain) of preincubation of breast milk samples in an anaerobic blood-culture bottle enriched with 5% rumen fluid sterilized by filtration at 0.2 μm and 5% sheep blood and seeded on 5% sheep-blood Columbia agar (bioMérieux) under anaerobic condition at 37°C. The strains were not identifiable
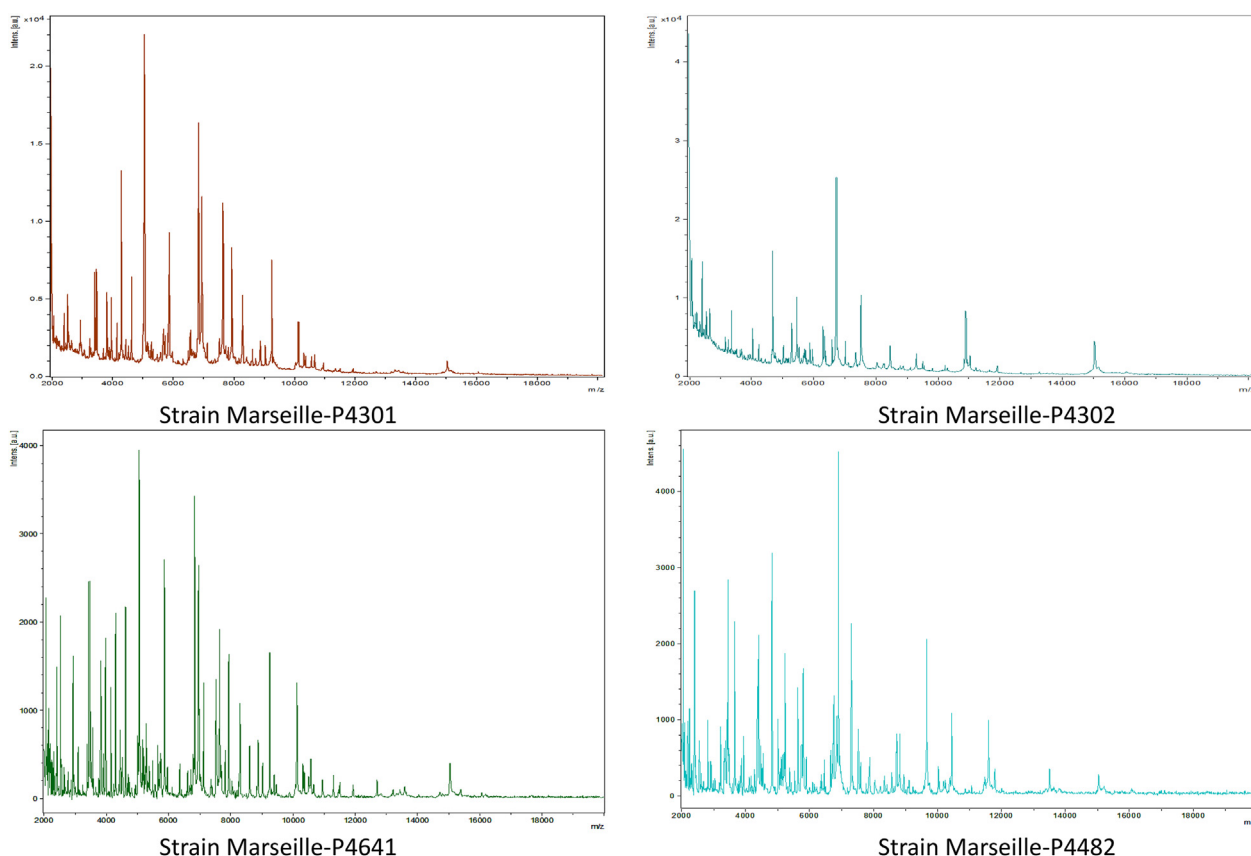
FIG. 1. Matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) reference mass spectrum from strains Marseille-P4301, Marseille-P4302, Marseille-P4641 and Marseille-P4482.
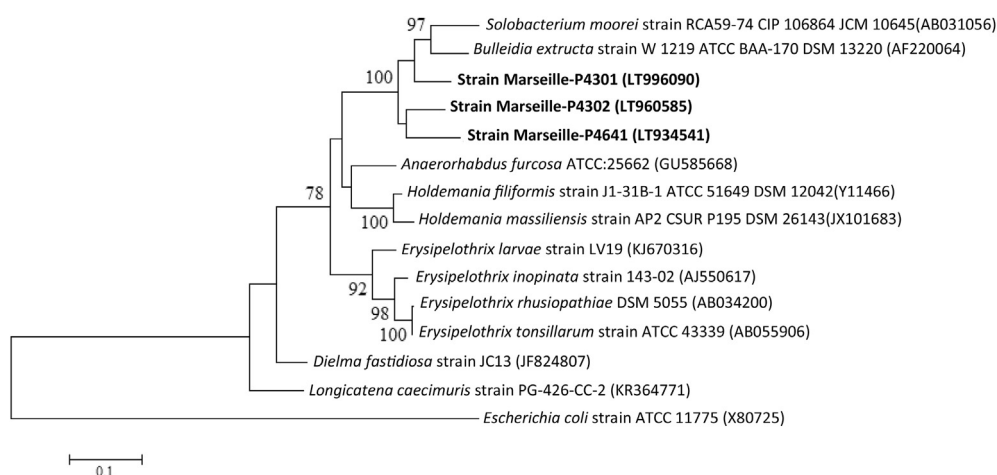
FIG. 2. Maximum likelihood phylogenetic tree highlighting the position of strains Marseille-P4301, Marseille-P4302, and Marseille-P4641 against most closely related species. The evolutionary history was inferred by using the maximum likelihood method based on the Kimura 2-parameter model. A discrete γ distribution was used to model evolutionary rate differences among sites (five categories (+G, parameter = 0.2353)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 15 nucleotide sequences. In total there were 1628 positions in the final dataset. The scale bar represents a 1% nucleotide sequence divergence.
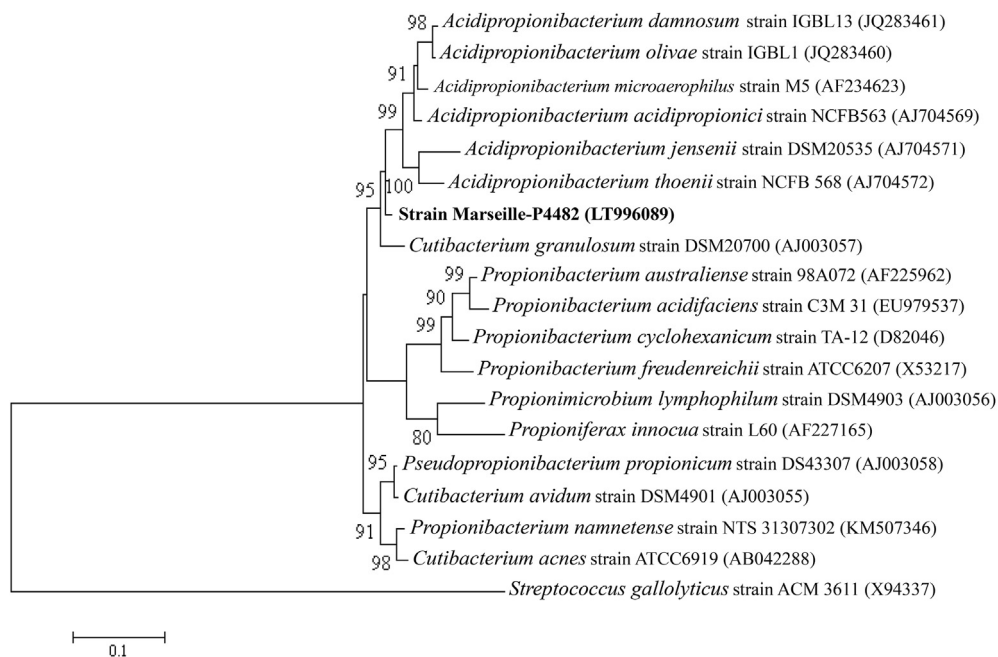
FIG. 3. Maximum likelihood phylogenetic tree highlighting the position of strain Marseille-P4482 against other most closely related species. The evolutionary history was inferred by using the maximum likelihood method based on the Kimura 2-parameter model. A discrete γ distribution was used to model evolutionary rate differences among sites (five categories (+G, parameter = 0.2353)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 15 nucleotide sequences. In total there were 1628 positions in the final dataset. The scale bar represents a 1% nucleotide sequence divergence.

using matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS). The spectra from these strains did not match any of the spectra in our database, either at the species level (strain Marseille-P4482) or at the genus level (strains Marseille-P4301, Marseille-P4302 and Marseille-P4641) (Fig. 1). In an attempt to identify these four strains, their 16S rRNA gene was sequenced, and the sequences obtained showed a similarity of 91.6%, 89.8% and 88.9% with *Solobacterium moorei* strain RCA59-74 (= CIP 106864[T] = JCM 10645[T]) for strains Marseille-P4301, Marseille-P4302, and Marseille-P4641, respectively, 96.28% with *Cutibacterium granulosum* strain ATCC 25564[T] (= CCUG 32987[T] = CIP 103262[T] = DSM 20700[T] = JCM 6498[T] = LMG 16726[T] = NCTC 11865[T]) for the strain Marseille-P4482 (Fig. 2, Fig. 3, Table 1, Table 2), the closest phylogenetically validated species with standing in nomenclature. The 16S rRNA gene sequences of these stain were deposited in EMBL-EBI under accession number: LT996090, LT960585, LT934541 and LT996089 (Strain Marseille-P4301, Marseille-P4302, Marseille-P4641 and Marseille-P4482 respectively).

TABLE 1. Pairwise comparison of strains Marseille-P4301, Marseille-P4302 and Marseille-P4641 for 16S rRNA sequence similarity with closely related species. Value (%)

Species	1	2	3	4	5	6	7	8
Strain Marseille-P4301	100	92.5	90.8	91.6	91.9	88.5	87.8	89.2
Strain Marseille-P4302		100	91.9	89.8	90.8	89.3	88.4	89.1
Strain Marseille-P4641			100	88.9	89.5	88.6	87.6	87.7
S. moorei strain RCA59-74[T]				100	92.4	87.8	87.1	88.9
B. extructa strain W 1219[T]					100	88.4	87.0	89.3
H. filiformis strain J1-31B-1[T]						100	97.1	91.7
H. massiliensis strain AP2[T]							100	91.0
A. furcosa strain VPI 3253[T]								100

1, strain Marseille-P4301; 2, Strain Marseille-P-4302; 3, Strain Marseille-P4641; 4, *Solobacterium moorei* strain RCA59-74[T]; 5, *Bulleidia extructa* strain W 1219[T]; 6, *Holdemania filiformis* strain J1-31B-1[T]; 7, *Holdemania massiliensis* strain AP2[T]; 8, *Anaerorhabdus furcosa* strain VPI 3253[T]. Identity was obtained using blastn suite-2 sequences (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&LINK_LOC=align2seq) with the following sequences (LT996090 for strain P4301, LT960585 for P4302, LT934541 for P4641, AB031056 for *S. moorei* strain RCA59-74[T], AF220064 for *B. extructa* strain W 1219[T], Y11466 for *H. filiformis* strain J1-31B-1[T], JX101683 for *H. massiliensis* strain AP2[T] and GU585668 for *A. furcosa* strain VPI 3253[T]).

TABLE 2. Pairwise comparison of strains Marseille-P4482 for 16S rRNA sequence similarity compared with closely related species. Value (%)

Species	1	2	3	4	5	6	7	8
Strain Marseille-P4482	100	95.0	95.8	96.3	95.6	92.2	94.9	95.6
Cutibacterium acnes **ATCC6919**		100	96.1	93.9	96.1	92.0	92.0	92.0
Cutibacterium avidum **ATCC 25577**			100	94.7	99.2	91.8	92.8	94.0
Cutibacterium granulosum **DSM 20700**				100	94.6	90.6	92.8	94.6
Pseudopropionibacterium propionicum **F0230a**					100	91.8	92.8	94.0
Propionibacterium acidifaciens **strain C3M 31**						100	90.3	91.0
Acidipropionibacterium thoenii **strain NCFB 568**							100	95.5
Acidipropionibacterium acidipropionici **strain NCFB 563**								100

1, Strain Marseille-P4301; 2, *Cutibacterium acnes* strain ATCC6919[T]; 3, *Cutibacterium avidum* strain DSM 4901[T]; 4, *Cutibacterium granulosum* strain DSM 20700[T]; 5, *Pseudopropionibacterium propionicum* strain F0230a[T]; 6, *Propionibacterium acidifaciens* strain C3M 31[T]; 7, *Acidipropionibacterium thoenii* strain NCFB 568[T]; 8, *Acidipropionibacterium acidipropionici* strain NCFB 563[T]. Identity was obtained using blastn suite-2 sequences (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&LINK_LOC=align2seq) with the following sequences (LT996089 for strain, AB042288 for *C. acnes*, AJ003055 for *C. avidum*, AJ003057 for C. granulosum, AJ003058 for *P. propionicum*, EU979537 for *P. acidifaciens*, AJ704572 for *P. thoenii* and AJ704569 for *Acidipropionibacterium acidipropionici*).

Phenotypic and biochemical characterization

Cells from the strains Marseille-P4301, Marseille-P4302 and Marseille-P4641 are Gram-negative staining, non-motile, non-spore-forming, strictly anaerobic rods. Those from strain Marseille-P4482 are Gram-positive staining, non-motile, non-spore forming and facultatively anaerobic coccobacilli. Strains Marseille-P4301, Marseille-P4302, Marseille-P4641 and Marseille-P4482 measure 0.5/1.5, 0.5/2, 0.4/0.8 and 0.8/1.2 μm width/length respectively by electron microscopy (Fig. 4). The four strains have no catalase or oxidase activity. Strain growth occurred between 28 and 45°C, but optimal growth was observed at 37°C after 24 or 48 h incubation in an anaerobic

Strain Marseille-P4301

Strain Marseille-P4302

Strain Marseille-P4641

Strain Marseille-P4482
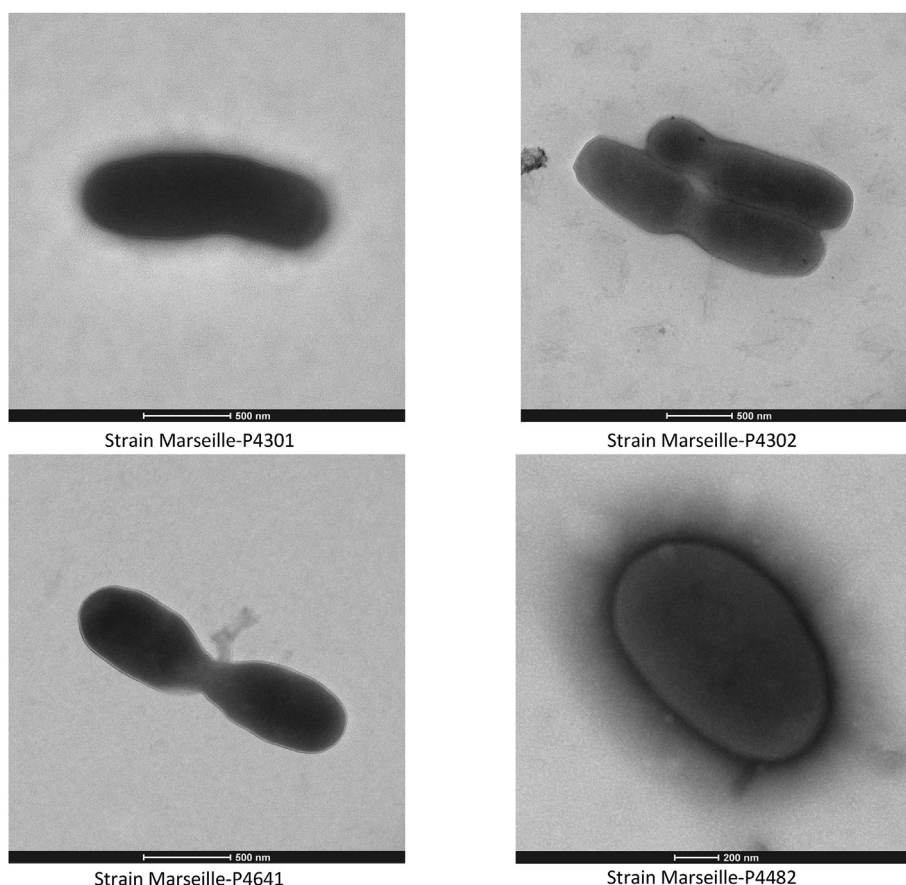
FIG. 4. Transmission electron microscopy of strains Marseille-P4301, Marseille-P4302, Marseille-P4641 and Marseille-P4482 using Tecnai G20 (FEI Company) at operating voltage of 60 kV. Scale bar = 500 nm (Strain marseille-P4301, Maeseille-P4302 and marseille-P4641) and 200nm (Strain Marseille-P4482).

atmosphere on 5% sheep-blood Columbia agar (bioMérieux). The strain Marseille-P4482 is able to growth under aerobic conditions but the strains Marseille-P4301, Marseille-P4302 and Marseille-P4641 are not. Colonies from strain Marseille-P4301 and Marseille-P4641 were translucent, non-haemolytic, regular and umbilicate with a mean diameter from 1 to 1.5 mm. Colonies from strain Marseille-P4302 were grey, regular, with a mean of 1-2 mm, the agar plate looks like blood burnt after 48 of incubation. Colonies from strain Marseille-P4482 were cream-coloured, regular and non-haemolytic with a mean diameter of 3−5 mm. No growth was observed beyond 10 g/L of NaCl concentration on Schaedler agar (bioMérieux) for the strains Marseille-P4301, Marseille-P4641 and Marseille-P4482, but growth was observed up to 50 g/L for strain Marseille-P4302. These strains were able to grow at pH levels ranging from 6.5 to 8, but the optimum was observed at pH 7.5. Using API strip; aesculin and gelatine are hydrolysed, indole is produced, but none of the four species produces urease or reduces nitrate. Cellobiose, maltose, sucrose and trehalose are fermented while arabinose, rhamnose, sorbitol and xylose are not for any of the four species. All strains exhibited acid phosphatase, naphthol-AS-BI-phosphohydrolase, esterase (C-4), esterase lipase (C-8), β-galactosidase and leucine arylamidase activity but not cystine arylamidase, lipase (C-14), glutamic acid decarboxylase, trypsin, α-chymotrypsin and β-glucuronidase activity. Table 3 displays the phenotypic and chemical characteristics of the four strains. The major cellular fatty acid are: $C_{16:0}$ (45.6, 39.3 and 45.9%), $C_{18:1n9}$ (21.8, 27.6 and 26.6%) and $C_{18:1n7}$ (20.2, 19.4 and 12.3%) for the strains Marseille-P4301, Marseille-P4302 and Marseille-P4641 respectively and iso-$C_{15:0}$ (62.5%) and anteiso-$C_{15:0}$ (18.4%) for the strain Marseille-P4482. Cellular fatty composition of the four species is shown in Table 4.

Genomic analysis

Draft genomes of these strains were deposited in EMBL-EBI under accession numbers OEPX00000000, OLMH00000000, OUNG00000000 and UWPF00000000 (Fig. 5). Genomes are 2 457 574 bp, 3 334 468 bp, 2 581 777 bp and 2 816 504 bp with 47%, 48.5%, 50% and 69% G+C content for the strains Marseille-P4301 Marseille-P4302, Marseille-P4302, and Marseille-P4482 respectively. At final assembly, the genomes are composed of 7, 2, 10 and 15 scaffolds for strains Marseille-P4301, Marseille-P4302, Marseille-P4641 and Mareille-P4482 respectively. The coding gene contents were 2346, 3121, 2370, and 2592, including 48, 53, 57 and 54 RNA genes for strains Marseille-P4301, Marseille-P4302, Marseille-P4641 and Marseille-P4482 respectively; the distribution of genes into

clusters of orthologous groups (COG) functional categories is presented in Table 5.

The genomic characteristics of the strains were compared to those of the other closest species for which the genomes are available. The distribution of genes into COG categories is similar for all species compared, with the exception of the

TABLE 3. Differential characteristics of the new strains

Properties	1	2	3	4
Cells size (μm)	0.5/1.5	0.5/2	0.4/0.8	0.8/1.2
Gram stain	—	—	—	+
Motility	—	—	—	—
Urease	—	—	—	—
Aesculin hydrolase	+	+	+	+
Gelatine hydrolase	—	+	+	+
Indole production	+	+	+	—
Nitrates reduction	—	—	—	—
Acid from:				
Arabinose	—	—	—	—
Cellobiose	+	+	+	+
Glucose	—	+	+	+
Glycerol	—	—	+	+
Lactose	+	—	+	+
Maltose	+	+	+	+
Mannitol	—	—	+	+
Mannose	—	—	v	+
Melezitose	—	—	+	+
Raffinose	—	—	—	+
Rhamnose	—	—	—	—
Saccharose	+	+	+	+
Salicin	—	+	+	+
Sorbitol	—	—	—	—
Trehalose	+	+	+	+
Xylose	—	—	—	—
Acid phosphatase	+	+	+	+
Alkaline phosphatase	v	—	—	+
N-Acetyl-β-glucosaminidase	+	—	—	+
Naphthol-AS-BI-phosphohydrolase	+	+	+	+
Esterase (C-4)	+	+	+	+
Esterase lipase (C-8)	+	+	+	+
Lipase (C-14)	—	—	—	—
Arginine dihydrolase	+	+	—	—
Alanine arylamidase	—	—	—	+
Arginine arylamidase	+	—	+	+
Cystine arylamidase	—	—	—	—
Glutamyl glutamic acid arylamidase	—	—	+	+
Glycine arylamidase	—	—	+	+
Histidine arylamidase	—	—	+	+
Leucine arylamidase	+	+	+	+
Leucyl glycine arylamidase	—	—	—	+
Phenylalanine arylamidase	—	—	—	+
Proline arylamidase	—	—	—	+
Pyroglutamic acid arylamidase	—	—	—	+
Serine arylamidase	—	—	—	+
Tyrosine arylamidase	—	—	—	+
Valine arylamidase	—	—	—	+
Glutamic acid decarboxylase	—	—	—	—
α-Arabinosidase	—	—	+	+
Trypsin	—	—	—	—
α-Chymotrypsin	—	—	—	—
α-Fucosidase	—	+	—	—
β-Galactosidase	+	+	+	+
β-Galactosidase 6 phosphate	+	—	—	—
α-Glucosidase	+	—	—	+
β-Glucosidase	+	—	+	+
β-Glucuronidase	—	—	—	—
α-Mannosidase	—	—	—	+
α-Galactosidase	+	—	—	+
Genome G+C %	47	48.5	50	69
Isolated from	Human milk	Human milk	Human milk	Human milk

+, Positive reaction; − negative reaction; v, variable reaction; 1, strain Marseille-P4301; 2, strain Marseille-P-4302; 3, strain Marseille-P4641; 4, strain Marseille-P4482.

TABLE 4. Cellular fatty acid composition (%) of (1) strain Marseille-P4301, (2) strain Marseille-P-4302, (3) strain Marseille-P4641, and (4) strain Marseille-P4482

Fatty acids	1	2	3	4
$C_{16:0}$	**45.6**	**39.3**	**45.9**	7.4
$C_{18:1n9}$	**21.8**	**27.6**	**26.6**	2.6
Iso-$C_{15:0}$	ND	ND	ND	62.5
Anteiso-$C_{15:0}$	ND	ND	ND	18.4
$C_{18:1n7}$	**20.2**	**19.4**	**12.3**	ND
$C_{18:0}$	9.2	7.8	8.7	1.8
$C_{14:0}$	1.7	2.9	3.8	<1
$C_{15:0}$	<1	1.3	1.1	1.1
Iso-$C_{5:0}$	ND	ND	ND	1.4
$C_{16:1n7}$	<1	<1	<1	ND
$C_{17:0}$	<1	<1	<1	<1
$C_{12:0}$	<1	<1	<1	ND
$C_{18:2n6}$	<1	<1	<1	<1
$C_{17:0\ iso}$	ND	ND	ND	3.0
Anteiso-$C_{17:0}$	ND	ND	ND	<1
Iso-$C_{13:0}$	ND	ND	ND	<1

ND, not detected.

presence of the RNA processing and modification gene for the strain Marseille-P4482, the absence of the chromatin structure and dynamics gene for the strains Marseille-P4301 and Marseille-P4302 and the extracellular structures gene for the strain Marseille-P4641 (Table 5).

The digital DNA–DNA hybridization (dDDH) values ranged from 17.4% between *S. moorei* and *A. furcosa* to 50% between strain Marseille-P4301 and strain Marseille-P4302, and 67.9% between strain Marseille-P4302 and *S. moorei* (Table 6). These values are certainly high but remain below the 75% threshold for defining whether two strains are of the same species. This value ranges from 19.7% between *P. propionicum* and strain Marseille-P4482 to 22.7% between *C. granulosum* and strain Marseille-P4482 when strain Marseille-P4482 is compared with its closest neighbours (Table 7).

A) Strain Marseille-P4301

B) Strain Marseille-P4302

C) Strain Marseille-P4641

D) Strain Marseille-P4482
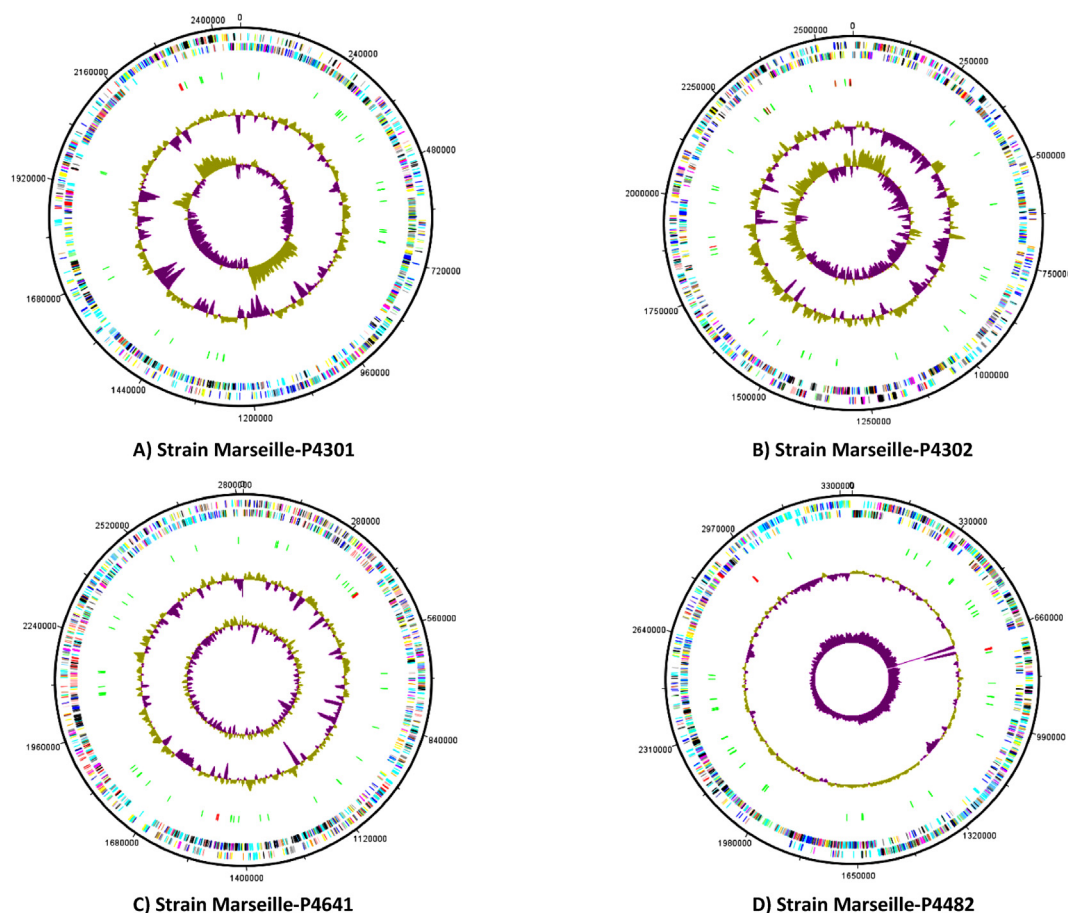
FIG. 5. Graphical circular map of the genome of (a) strain Marseille-P4301, (b) strain Marseille-P4302, (c) strain Marseille-P4641, and (d) strain Marseille-P4482. From outside to the centre: Contigs (red/grey), clusters of orthologous groups (COGs) category of genes on the forward strand (three circles), genes on forward strand (blue circle), genes on the reverse strand (red circle), COGs category on the reverse strand (three circles), G+C content.

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

Description	1	2	3	4
Translation, ribosomal structure and biogenesis	187	218	198	190
RNA processing and modification	0	0	0	1
Transcription	161	220	231	168
Replication, recombination and repair	134	171	152	108
Chromatin structure and dynamics	0	0	1	1
Cell cycle control, cell division, chromosome partitioning	35	48	57	39
Nuclear structure	0	0	0	0
Defence mechanisms	113	118	81	87
Signal transduction mechanisms	92	148	89	106
Cell wall/membrane/envelope biogenesis	115	154	128	122
Cell motility	17	20	13	11
Cytoskeleton	1	2	2	2
Extracellular structures	8	14	0	8
Intracellular trafficking, secretion, and vesicular transport	19	33	32	25
Posttranslational modification, protein turnover, chaperones	74	86	77	94
Mobilome: prophages, transposons	164	230	45	97
Energy production and conversion	100	129	115	127
Carbohydrate transport and metabolism	187	254	116	240
Amino acid transport and metabolism	166	185	144	203
Nucleotide transport and metabolism	73	94	58	82
Coenzyme transport and metabolism	71	76	93	135
Lipid transport and metabolism	57	67	59	76
Inorganic ion transport and metabolism	83	102	92	107
Secondary metabolites biosynthesis, transport and catabolism	22	19	21	30
General function prediction only	195	249	154	201
Function unknown	112	124	124	122
Hypothetical protein	433	714	541	513

1, Marseille-P4301; 2, Marseille-P4302; 3, Marseille-P46413; 4, Marseille-P4482.

TABLE 6. **Pairwise comparison of strains Marseille-P4301, Marseille-P4302 and Marseille-P46413 with other species using the genome-to-genome distance calculator (GGDC), formula 2 (digital DNA–DNA hybridization (dDDH) estimates based on identities/high-scoring segment pairs (HSP) length)[a]**

	1	2	3	4	5	6	7	8
Marseille-P4301	100%	50.0%±2.7	24.8%±2.4	28.0%±2.4	27.3%±2.5	26.1%±2.4	19.0%±2.3	27.1%±2.5
Marseille-P4302		100%	38.9%±2.3	67.9%±3	21.8%±2.4	26.2%±2.4	16.3%±2.2	23.4%±2.4
Marseille-P4641			100%	20.3%±2.3	34.0%±2.3	18.9%±2.3	18.9%±2.3	26.5%±2.4
S. moorei				100%	24.5%±2.4	26.4%±2.5	24.6%±2.4	17.4%±2.2
B. extructa					100%	29.5%±2.5	27.7%±2.4	26.4%±2.5
H. filiformis						100%	24.5%±2.4	28.6%±2.5
H. massiliensis							100%	28.8%±2.4
A. furcosa								100%

1, Marseille-P4301; 2, Marseille-P4302; 3, Marseille-P4641; 4, *Solobacterium moorei* strain RCA59-74[T]; 5, *Bulleidia extructa* strain W 1219[T]; 6, *Anaerorhabdus furcosa* strain VPI 3253[T]; 7, *Holdemania filiformis* strain J1-31B-1[T]; 8, *Holdemania massiliensis* AP2[T]
[a]Confidence intervals indicate inherent uncertainty in estimating DDH values from intergenomic distances based on models derived from empirical test data sets. These results are consistent with the 16S rRNA and phylogenomic analyses as well as the GGDC results.

TABLE 7. **Pairwise comparison of trains Marseille-P4482 with closest species using the genome-to-genome distance calculator (GGDC), formula 2 (digital DNA–DNA hybridization (dDDH) estimates based on identities/high-scoring segment pairs (HSP) length)[a]**

	1	2	3	4	5	6	7	8
Strain Marseille-P4482	100%	21.5%±2.4	21.2%±2.3	21.6%±2.4	21.40%±2.3	22.7%±2.4	19.8%±2.4	19.7%±2.4
A. acidipropionici		100%	20.8%±2.4	23.0%±2.4	20.0%±2.3	20.9%±2.4	19.3%±2.3	18.8%±2.3
C. avidum			100%	20.7%±2.4	23.6%±2.4	23.3%±2.4	18.8%±2.3	20.4%±2.4
A. thoenii				100%	20.7%±2.4	21.1%±2.4	19.5%±2.3	20.3%±2.4
C. acnes					100%	22.2%±2.4	20.1%±2.3	22.1%±2.4
C. granulosum						100%	19.1%±2.3	20.5%±2.4
P. acidifaciens							100%	18.4%±2.3
P. propionicum								100%

1, Marseille-P4482[T]; 2, *Acidipropionibacterium acidipropionici* strain DSM4900[T]; 3, *Cutibacterium avidum* strain ATCC25577[T]; 4, *Acidipropionibacterium thoenii* strain DSM20276[T]; 5, *Cutibacterium acnes* strain ATCC6919[T]; 6, *Cutibacterium granulosum* strain DSM20700[T]; 7, *Propionibacterium acidifaciens* strain C3M_31[T]; 8, *Pseudopropionibacterium propionicum* strain F0230a[T].
[a]Confidence intervals indicate inherent uncertainty in estimating DDH values from intergenomic distances based on models derived from empirical test data sets. These results are consistent with the 16S rRNA and phylogenomic analyses as well as the GGDC results.

Conclusion

Considering the specific phenotypic, biochemical, genomic and phylogenetic characteristics of new bacteria, we propose the creation of three new genera named:

Lactimicrobium, with the type species *Lactimicrobium massiliense*, type strain Marseille-P4301[T] (=CSUR P4301[T]); *Anaerolactibacter* with the type species *Anaerolactibacter massiliensis*, type strain Marseille-P4302[T] (=CSUR P4302[T]); *Galactobacillus*, with the type species *Galactobacillus timonensis*, type strain Marseille-P4641[T] (=CSUR P4641[T]). The main characteristics of this new species have been previously published with the former name *Lactomassilus timonensis* [33]. The name was changed following the advice of a world expert in taxonomy (we thank Professor A. Oren). We also proposed the creation of a new species: *Acidipropionibacterium timonense*, with type strain Marseille-P4482[T] (= CSUR P4482[T]).

Taxonomic and nomenclatural proposals

Description of Lactimicrobium *gen. nov.. Lactimicrobium* (Lac.ti.-mi.cro'bi.um. L. masc. n. *lac, lactis* milk; N.L. neut. n. *microbium* a microbe; N.L. neut. n. *Lactimicrobium* a microbe from milk). Cells are Gram-positive, non-motile, non-spore-forming and anaerobic rod-shaped bacteria. They are mesophilic and do not require NaCl for growth. pH tolerance ranges from pH 6.5 to pH 8. Cells do not produce catalase or oxidase activity and measure approximately 0.5/1.5μm width/length. The type species is *Lactimicrobium massiliense*. The taxonomic classification is *Bacteria*; *Terrabacteria* group; *Firmicutes*; *Erysipelotrichia*; *Erysipelotrichales*; *Erysipelotrichaceae*; *Lactimicrobium*.

Description of Lactimicrobium massiliense *sp. nov.. Lactimicrobium massiliense* (mas.si.li.en'se. L. neut. adj. *massiliense* of Massilia, the Latin name for Marseille). Colonies grown on 5% sheep blood Columbia agar plates (bioMérieux) after 48 h of incubation under anaerobic conditions are regular, umbilicate, translucent, non-haemolytic, around 1−1.5 mm in diameter. Using API strip (ZYM, 20 A and Rapid ID 32 A), indole is produced but urea is not. Aesculin is hydrolysed but gelatine is not. Cellobiose, lactose, maltose, saccharose and trehalose are fermented. Acid phosphatase, N-acetyl-β-glucosaminidase, naphthol-AS-BI-phosphohydrolase, esterase (C-4), esterase lipase (C-8), arginine dihydrolase, arginine arylamidase, leucine arylamidase, β-galactosidase, β-galactosidase 6-phosphate, α-glucosidase, β-glucosidase and α-galactosidase activity were positive. Their major fatty acid are $C_{16:0}$, $C_{18:1n9}$ and $C_{18:1n7}$. The DNA G+C content of the type strain is 47 % (genome sequence). The type strain is Marseille-P4301[T] (CSUR P4301[T]) isolated from a milk sample from a healthy lactating Malian mother.

Description of Anaerolactibacter *gen. nov.. Anaerolactibacter* (An.ae.ro.lac.ti.bac'ter. Gr. pref. *an* not; Gr. masc. or fem. n. *aer* air; L. masc. n. *lac, lactis* milk; N.L. masc. n. *bacter* a rod; N.L. masc. n. *Anaerolactibacter* an anaerobic rod from milk). Cells are Gram-positive, non-motile, non-spore-forming and anaerobic rod-shaped bacteria. They are mesophilic and do not need NaCl for their growth but tolerate up to 50 g/L of salt. pH tolerance ranges from pH 6.5 to pH 8. Cells do not produce catalase or oxidase activity and measure approximately 0.5/2μm width/length. The type species is *Anaerolactibacter massiliensis*. The taxonomic classification is *Bacteria*; *Terrabacteria* group; *Firmicutes*; *Erysipelotrichia*; *Erysipelotrichales*; *Erysipelotrichaceae*; *Anaerolactibacter*.

Description of Anaerolactibacter massiliensis *sp. nov.. Anaerolactibacter massiliensis* (mas.si.li.en'sis. L. masc. adj. *massiliensis* of Massilia, the Latin name for Marseille). Colonies grown on 5% sheep-blood Columbia agar plates (bioMérieux) after 48 h incubation under anaerobic conditions are grey, circular, around 1−2 mm in diameter; the agar plate takes the colour of burnt blood after 48 h of incubation. Using an API strip (ZYM, 20 A and Rapid ID 32 A), indole is produced while urease is not. Aesculin and gelatine are hydrolysed. Cellobiose, glucose, maltose, saccharose, salicin and trehalose are fermented. acid phosphatase, naphthol-AS-BI-phosphohydrolase, esterase (C-4), esterase lipase (C-8), arginine dihydrolase, leucine arylamidase, α-fucosidase and β-galactosidase activities were positive. Their major fatty acids are $C_{16:0}$, $C_{18:1n9}$ and $C_{18:1n7}$. The DNA G+C content of the type strain is 48.5 % (genome sequence). The type strain is Marseille-P4302[T] (CSUR P4302[T]) isolated from a milk sample from a healthy lactating Malian mother.

Description of Galactobacillus *gen. nov.. Galactobacillus* (Ga.lac.to.ba.cil'lus. Gr. neut. n. *gala, galaktos* milk; L. masc. n. *bacillus*, a small rod. N.L. masc. n. *Galactobacillus* a rod from milk). Cells are Gram-positive, non-motile, non-spore-forming, anaerobic rod-shaped bacteria. They are mesophilic and do not require NaCl for growth. pH tolerance ranges from pH 6.5 to pH 8. Cells do not produce catalase or oxidase activity and measure approximately 0.4/0.8μm width/length. The type species is *Galactobacillus timonensis*. The taxonomic classification is *Bacteria; Terrabacteria* group; *Firmicutes; Erysipelotrichia; Erysipelotrichales; Erysipelotrichaceae; Galactobacillus*.

Description of Galactobacillus timonensis *sp. nov.. Galactobacillus timonensis* (ti.mon.en'sis. N.L. masc. adj. *timonensis* the name of quarter La Timone where the strain was isolated). Colonies grown on 5% sheep blood Colombia agar plat (bioMérieux) after 48 h incubation under anaerobic conditions are regular and umbilicate, translucent, non-haemolytic, around 1−1.5 mm

in diameter. Using an API strip (ZYM, 20 A and Rapid ID 32 A), indole is produced but urease is not. Gelatine and aesculin are hydrolysed. Cellobiose, glucose, glycerol, lactose, maltose, mannitol, melezitose, saccharose, salicin and trehalose are fermented. Acid phosphatase, naphthol-AS-BI-phosphohydrolase, esterase (C-4), esterase lipase (C-8), arginine arylamidase, glutamyl glutamic acid arylamidase, glycine arylamidase, histidine arylamidase, leucine arylamidase, α-arabinosidase, β-galactosidase, β-glucosidase activity are positive. Their major fatty acid are $C_{16:0}$, $C_{18:1n9}$ and $C_{18:1n7}$. The DNA G+C content of the type strain is 50 % (genome sequence). The type strain is Marseille-P4641^T (CSUR P4641^T) isolated from a milk sample from a healthy lactating Malian mother.

Description of Acidipropionibacterium timonense *sp. nov.. Acidipropionibacterium timonense* (ti.mon'ense. N.L. neut. adj. *timonense* of quarter La Timone where the strain was isolated). Colonies grown on 5% sheep blood Colombia agar plat (bioMérieux) after 24 h incubation under anaerobic or aerobic conditions are creamy, non-haemolytic, circular, around 3−5 mm in diameter. Cells are Gram-positive, non-motile, non-spore-forming, facultatively anaerobic coccobacilli 0.8/1.2μm in width/length. pH tolerance ranges from pH 6.5 to pH 8. Using API strip (ZYM, 20 A and Rapid ID 32 A), gelatine and aesculin are hydrolysed but indole and urease are not produced. Cellobiose, glucose, glycerol, lactose, maltose, mannitol, mannose, melezitose, raffinose, saccharose, salicin and trehalose are fermented. Acid phosphatase, alkaline phosphatase, N-acetyl-β-glucosaminidase, naphthol-AS-BI-phosphohydrolase, esterase (C-4), esterase lipase (C-8), alanine arylamidase, arginine arylamidase, glutamyl glutamic acid arylamidase, glycine arylamidase, histidine arylamidase, leucine arylamidase, leucyl glycine arylamidase, phenylalanine arylamidase, proline arylamidase, pyroglutamic acid arylamidase, serine arylamidase, tyrosine arylamidase, valine arylamidase, α-arabinosidase, β-galactosidase, α-glucosidase, β-glucosidase, α-mannosidase and α-galactosidase activities were positive. The major cellular fatty acids of strain Marseille-P4482 are $C_{15:0}$ iso and anteiso-$C_{15:0}$.

The type strain is Marseille-P4482^T (CSUR P4482^T) isolated from a milk sample from a healthy Malian mother.

Transparency declaration

The authors have no conflicts of interest to declare. This work received financial support from the French Government through the Agence Nationale pour la Recherche (ANR), including the 'Programme d'Investissement d'Avenir' under the reference Méditerranée Infection 10-IAHU-03. This work was supported by the Région Provence-Alpes-Côte d'Azur and European funding FEDER PRIMMI (Fonds Européen de Développement Régional—Plateformes de Recherche et d'Innovation Mutualisées Méditerranée Infection).

References

[1] Newburg DS. Bioactive components of human milk. Springer Science & Business Media; 2012.

[2] Ballard O, Morrow AL. Human milk composition: nutrients and bioactive factors. Pediatr Clin North Am 2013;60:49−74.

[3] Arifeen S, Black RE, Antelman G, Baqui A, Caulfield L, Becker S. Exclusive breastfeeding reduces acute respiratory infection and diarrhea deaths among infants in Dhaka slums. Pediatrics 2001;108:E67.

[4] Duijts L, Jaddoe VWV, Hofman A, Moll HA. Prolonged and exclusive breastfeeding reduces the risk of infectious diseases in infancy. Pediatrics 2010;126:e18−25.

[5] Gartner LM, Morton J, Lawrence RA, Naylor AJ, O'Hare D, Schanler RJ, et al. American academy of pediatrics section on breastfeeding. Breastfeeding and the use of human milk. Pediatrics 2005;115: 496−506.

[6] Heikkilä MP, Saris PEJ. Inhibition of *Staphylococcus aureus* by the commensal bacteria of human milk. J Appl Microbiol 2003;95:471−8.

[7] Martín R, Langa S, Reviriego C, Jiménez E, Marín ML, Xaus J, et al. Human milk is a source of lactic acid bacteria for the infant gut. J Pediatr 2003;143:754−8.

[8] Patel SH, Vaidya YH, Patel RJ, Pandit RJ, Joshi CG, Kunjadiya AP. Culture independent assessment of human milk microbial community in lactational mastitis. Sci Rep 2017;7.

[9] Li S-W, Watanabe K, Hsu C-C, Chao S-H, Yang Z-H, Lin Y-J, et al. Bacterial composition and diversity in breast milk samples from mothers living in Taiwan and mainland China. Front Microbiol 2017;8:965.

[10] Asnicar F, Manara S, Zolfo M, Truong DT, Scholz M, Armanini F, et al. Studying vertical microbiome transmission from mothers to infants by strain-level metagenomic profiling. mSystems 2017;2.

[11] Milani C, Mancabelli L, Lugli GA, Duranti S, Turroni F, Ferrario C, et al. Exploring vertical transmission of *Bifidobacteria* from mother to child. Appl Environ Microbiol 2015;81:7078−87.

[12] Jost T, Lacroix C, Braegger CP, Rochat F, Chassard C. Vertical mother−neonate transfer of maternal gut bacteria via breastfeeding. Environ Microbiol 2014;16:2891−904.

[13] Biagi E, Quercia S, Aceti A, Beghetti I, Rampelli S, Turroni S, et al. The bacterial ecosystem of mother's milk and infant's mouth and gut. Front Microbiol 2017;8.

[14] Jiménez E, de Andrés J, Manrique M, Pareja-Tobes P, Tobes R, Martínez-Blanch JF, et al. Metagenomic analysis of milk of healthy and mastitis-suffering women. J Hum Lact 2015;31:406−15.

[15] Jiménez E, Delgado S, Fernández L, García N, Albújar M, Gómez A, et al. Assessment of the bacterial diversity of human colostrum and screening of staphylococcal and enterococcal populations for potential virulence factors. Res Microbiol 2008;159:595−601.

[16] Martín R, Heilig HGHJ, Zoetendal EG, Jiménez E, Fernández L, Smidt H, et al. Cultivation-independent assessment of the bacterial diversity of breast milk among healthy women. Res Microbiol 2007;158:31–7.

[17] Martin V, Manes-Lazaro R, Rodriguez JM, Maldonado-Barragan A. *Streptococcus lactarius* sp. nov., isolated from breast milk of healthy women. Int J Syst Evol Microbiol 2011;61:1048–52.

[18] Lagier J-C, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 2012;18:1185–93.

[19] Kageyama A, Benno Y. Phylogenic and phenotypic characterization of some *Eubacterium*-like isolates from human feces: description of *Solobacterium moorei* Gen. Nov., Sp. Nov. Microbiol Immunol 2000;44: 223–7.

[20] Togo AH, Durand G, Khelaifia S, Armstrong N, Robert C, Cadoret F, et al. *Fournierella massiliensis*, gen. nov., sp. nov., a new human-associated member of the family *Ruminococcaceae*. Int J Syst Evol Microbiol 2017;67:1393–9.

[21] Drancourt M, Bollet C, Carlioz A, Martelin R, Gayral JP, Raoult D. 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. J Clin Microbiol 2000;38:3623–30.

[22] Diop K, Diop A, Michelle C, Richez M, Rathored J, Bretelle F, et al. Description of three new *Peptoniphilus* species cultured in the vaginal fluid of a woman diagnosed with bacterial vaginosis: *Peptoniphilus pacaensis* sp. nov., *Peptoniphilus raoultii* sp. nov., and *Peptoniphilus vaginalis* sp. nov. Microbiologyopen 2018:e00661.

[23] Alou MT, Rathored J, Michelle C, Dubourg G, Andrieu C, Armstrong N, et al. *Inediibacterium massiliense* gen. nov., sp. nov., a new bacterial species isolated from the gut microbiota of a severely malnourished infant. Antonie Van Leeuwenhoek 2017;110:737–50.

[24] Tidjani Alou M, Khelaifia S, Michelle C, Andrieu C, Armstrong N, Bittar F, et al. *Anaerococcus rubiinfantis* sp. nov., isolated from the gut microbiota of a Senegalese infant with severe acute malnutrition. Anaerobe 2016;40:85–94.

[25] Togo AH, Diop A, Bittar F, Maraninchi M, Valero R, Armstrong N, et al. Description of *Mediterraneibacter massiliensis*, gen. nov., sp. nov., a new genus isolated from the gut microbiota of an obese patient and reclassification of *Ruminococcus faecis, Ruminococcus lactaris, Ruminococcus torques, Ruminococcus gnavus* and *Clostridium glycyrrhizinilyticum* as *Mediterraneibacter faecis* comb. nov., *Mediterraneibacter lactaris* comb. nov., *Mediterraneibacter torques* comb. nov., *Mediterraneibacter gnavus* comb. nov. and *Mediterraneibacter glycyrrhizinilyticus* comb. nov. Antonie Van Leeuwenhoek 2018;111:2107–28.

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

[27] Shah HN, Collins MD. Reclassification of *Bacteroides furcosus* Veillon and Zuber (Hauduroy, Ehringer, Urbain, Guillot and Magrou) in a new genus *Anaerorhabdus*, as *Anaerorhabdus furcosus* comb. Nov. Syst Appl Microbiol 1986;8:86–8.

[28] Willems A, Moore WEC, Weiss N, Collins MD. Phenotypic and phylogenetic characterization of some *Eubacterium*-like isolates containing a novel type B wall murein from human feces: description of *Holdemania filiformis* gen. nov., sp. nov. Int J Syst Bacteriol 1997;47: 1201–4.

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

[30] Cummins CS, Johnson JL. *Corynebacterium parvum*: a synonym for *Propionibacterium acnes*? J Gen Microbiol 1974;80:433–42.

[31] Douglas HC, Gunter SE. The taxonomic position of *Corynebacterium acnes*. J Bacteriol 1946;52:15–23.

[32] Eggerth AH. The Gram-positive non-spore-bearing anaerobic bacilli of human feces. J Bacteriol 1935;30:277–99.

[33] Togo AH, Camara A, Konaté S, Doumbo OK, Raoult D, Million M. "*Lactomassilus timonensis*," a new anaerobic bacterial species isolated from the milk of a healthy African mother. New Microbe New Infect 2018;21:122–4.

Partie II:

La microbial culturomics comme outil pour faire progresser la taxonomie bactérienne.

Avant-propos

La «microbial culturomics» est une technique basée sur la diversification des conditions de culture (en variant l'atmosphère, la température d'incubation, la durée et la composition du milieu de culture), en utilisant d'une part des conditions permettant de faciliter la croissance des espèces fastidieuses et d'autre part, d'utiliser des inhibiteurs de croissance des espèces majoritaires, ce qui permet de sélectionner les espèces minoritaires afin de les identifier (32,33). Tous ceux-ci, dans le but de reproduire le plus fidèlement possible les conditions naturelles des milieux de croissance des bactéries. Les isolats sont identifiés à l'aide de la spectrométrie de masse de type MALDI-TOF (matrix assisted laser desorption ionisation/time of flight). Si l'isolat n'est pas identifié par cette technique; il est identifié par le séquençage de son gène de l'ARN ribosomal 16S. L'isolat est considéré comme nouveau genre ou nouvelle espèce selon le pourcentage de similarité de séquence du gène de son ARN 16S avec l'espèce la plus proche connu et valide selon les règles de la classification taxonomique officielle appliquées aux procaryotes. L'isolat une fois identifié, est décrit selon le schéma taxonogénomique (34,35). Au départ, les concepteurs avaient utilisé 212 combinaisons distinctes de conditions de culture, qui ont permis d'identifier 340 espèces bactériennes, dont 31 putatives nouvelles espèces. Le résultat de ce travail a permis d'en retenir 70 meilleures conditions sur les 212 conditions de départ (36). Finalement, après plusieurs études successives, 18 conditions ont été retenues pour l'utilisation de routine dans un laboratoire de recherche ou d'analyse médicale (32,35,37).

La taxonomie provient du grec τάξις (taxis) « placement », « classement », « ordre » et de νομός (nomos) qui signifie « loi », « règle ». La taxonomie est une branche de la biologie, qui a pour objet de décrire les organismes vivants et de les regrouper en entités appelées taxons afin de les identifier, les nommer et les classer via des clés de détermination. Dans ce travail, nous nous sommes intéressés à la taxonomie appliquée aux procaryotes ou taxonomie microbienne qui peut être définie comme l'étude de la diversité des micro-organismes dans le but de les organiser dans une hiérarchie ordonnée (38,39). Cette taxonomie a évolué avec le temps, depuis l'introduction de nouveaux outils; le séquençage de l'ARN 16S (40), l'hybridation ADN-ADN (41), le MALDI-TOF MS (42) et le séquençage génomique (34) dans l'identification des micro-organismes (35). Elle est devenue taxono-génomique avec l'association de la génomique et la taxonomie. La taxono-génomique est une approche polyphasique intégrant à la fois les données phénotypiques et génotypiques pour la description de nouvelles espèces bactériennes (34). La «microbial culturomics» couplée à la taxonomie a

permis d'augmenter de manière significative le répertoire bactérien associé à l'homme. Plus de 500 nouvelles espèces bactériennes ont été découverte par la «microbial culturomics» (43).

Au total 13 des 18 nouvelles bactéries que nous avons isolées au cours de notre travail de thèse ont été décrites selon l'approche taxono-génomique *Fournierella massiliensis* (**Article IX**), *Mediterraneibacter massiliensis* (**Article VIII**), *Acidipropionibacterium timonense*, *Anaerolactibacter massiliensis*, *Galactobacillus timonensis* et *Lactimicrobium massiliense* isolés du lait maternel (**Article VII**), *Anaerotruncus massiliensis* (**Article X**), *Butyricimonas phoceensis* (**Article XI**), *Paenibacillus ihumii* (**Article XII**), *Eisenbergiella massiliensis* (**Article XIII**), *Hugonella massiliensis* (**Article XIV**), *Enterococcus massiliensis* (**Article XV**) et *Negativicoccus massiliensis* (**Article XVI**). Deux de ces nouvelles bactéries (*Fournierella massiliensis* et *Paenibacillus ihumii*) ont été officiellement validé par le comité chargé de la nomenclature et la classification des bactéries et des archées (http://www.bacterio.net). Il faut noter que *Fournierella massiliensis* est la première espèce de la «microbial culturomics» publiée dans un journal de référence dans la taxonomie de bactérie et d'archée. Trois bactéries ont été décrite selon l'approche plus courte dite New Species Announcement (NSA): *Bariatricus massiliensis* (**Article XVIII**) *Lachnoclostridium bouchesdurhonense* (**Article XIX**) et *Veillonella massiliensis* (**Article VI**). Deux n'ont pas été décrites (*Drancourtella timonensis* et *Marseillibacillus timonensis*).

Article VIII

Description of *Mediterraneibacter massiliensis*, gen. nov., sp. nov., a new genus isolated from the gut microbiota of an obese patient and reclassification of *Ruminococcus faecis*, *Ruminococcus lactaris*, *Ruminococcus torques*, *Ruminococcus gnavus* and *Clostridium glycyrrhizinilyticum* as *Mediterraneibacter faecis* comb. nov., *Mediterraneibacter lactaris* comb. nov., Mediterraneibacter torques comb. nov., *Mediterraneibacter gnavus* comb. nov. and *Mediterraneibacter glycyrrhizinilyticus* comb. nov.

Amadou Hamidou Togo, Awa Diop, Fadi Bittar, Marie Maraninchi, René Valero, Nicholas Armstrong, Grégory Dubourg Noémie Labas, Magali Richez, Jeremy Delerce, Anthony Levasseur, Pierre-Edouard Fournier, Didier Raoult and Matthieu Million

CrossMark

Description of *Mediterraneibacter massiliensis*, gen. nov., sp. nov., a new genus isolated from the gut microbiota of an obese patient and reclassification of *Ruminococcus faecis*, *Ruminococcus lactaris*, *Ruminococcus torques*, *Ruminococcus gnavus* and *Clostridium glycyrrhizinilyticum* as *Mediterraneibacter faecis* comb. nov., *Mediterraneibacter lactaris* comb. nov., *Mediterraneibacter torques* comb. nov., *Mediterraneibacter gnavus* comb. nov. and *Mediterraneibacter glycyrrhizinilyticus* comb. nov.

Amadou Hamidou Togo · Awa Diop · Fadi Bittar · Marie Maraninchi ·
René Valero · Nicholas Armstrong · Grégory Dubourg · Noémie Labas ·
Magali Richez · Jeremy Delerce · Anthony Levasseur · Pierre-Edouard Fournier ·
Didier Raoult · Matthieu Million

Abstract An anaerobic isolate, strain AT7^T, was cultivated from a stool sample of a morbidly obese French woman using a microbial culturomics approach. The 16S rRNA gene sequence analysis showed that strain AT7^T exhibited 96% nucleotide sequence similarity with *Ruminococcus torques* strain JCM 6553^T (= ATCC 27756^T = VPI B2-51^T), currently the closest related species with a validly published name. The strain was observed to be a Gram-stain positive, non-motile, asporogenous and coccobacillary-shaped bacterium. It was found to be catalase positive and oxidase negative. Its major fatty acids were identified as C$_{16:0}$ (54%) and C$_{18:1n9}$ (30%). The draft genome of strain AT7^T is 3,069,882 bp long with 42.4% G+C content. 2925 genes were predicted, including 2867 protein-coding genes and 58 RNAs. Based on phenotypic, biochemical, phylogenetic and genomic evidence, we propose the creation of the new

A. H. Togo · N. Armstrong · G. Dubourg ·
M. Richez · J. Delerce · A. Levasseur ·
D. Raoult · M. Million (✉)
Aix Marseille Univ, IRD, MEPHI, IHU-Méditerranée
Infection, 19-21 Boulevard Jean Moulin, 13005 Marseille,
France
e-mail: matthieumillion@gmail.com

A. Diop · F. Bittar · N. Labas · P.-E. Fournier
Aix Marseille Univ, IRD, VITROME, IHU-Méditerranée
Infection, 19-21 Boulevard Jean Moulin, 13005 Marseille,
France

M. Maraninchi · R. Valero
NORT "Nutrition, Obesity and Risk of Thrombosis",
INSERM1062, INRA1260, Aix Marseille Université,
13385 Marseille, France

R. Valero
CHU Hôpital de la Conception, Service Nutrition,
Maladies Métaboliques et Endocrinologie, APHM,
13385 Marseille, France

Ⓐ Springer

114

genus *Mediterraneibacter* and species, *Mediterraneibacter massiliensis*, that contains strain AT7[T] (= CSUR P2086[T] = DSM 100837[T]), and the reclassification of *Ruminococcus faecis*, *Ruminococcus lactaris*, *Ruminococcus torques*, *Ruminococcus gnavus*, *Clostridium glycyrrhizinilyticum* as *Mediterraneibacter faecis* comb. nov., with type strain Eg2[T] (= KCTC 5757[T] = JCM15917[T]), *Mediterraneibacter lactaris* comb. nov., with type strain ATCC 29176[T] (= VPI X6-29[T]), *Mediterraneibacter torques* comb. nov., with type strain ATCC 27756[T] (= VPI B2-51[T]), *Mediterraneibacter gnavus* comb. nov., with type strain ATCC 29149T (= VPI C7-9T) and *Mediterraneibacter glycyrrhizinilyticus* comb. nov., with type strain ZM35[T] (= JCM 13368[T] = DSM 17593[T]), respectively.

Abbreviations

AGIOS	Average of genomic identity of orthologous gene sequences
COG	Clusters of orthologous groups
CSUR	Collection de souches de l'Unité des Rickettsies
DDH	DNA–DNA hybridization
DSM	Deutsche Sammlung von Mikroorganismen
EUCAST	European Committee on antimicrobial susceptibility testing
FAME	Fatty acid methyl ester
GC/MS	Gas chromatography/mass spectrometry
GGDC	Genome-to-genome distance calculator
IUPAC	International Union of Pure and Applied Chemistry
ORF	Open reading frame
MALDI-TOF	Matrix-assisted laser-desorption/ionization time-of-flight

Introduction

Obesity is a major public health problem and the global obesity rate has doubled since 1980. In 2014, more than 1.9 billion adults were overweight and 600 million were obese (Ng et al. 2014). In France, the prevalence of obesity was 15.8% for men and 15.6% for women in 2016. Excess weight concerns nearly half of the French population (Matta et al. 2016). The treatment of obesity is a great challenge for health professionals. Bariatric surgery is currently the most effective treatment for morbid obesity. It is currently known that bariatric surgery leads to a lasting weight loss and reduces complications related to obesity. It has also been associated with an increase in the richness of the gut microbiota (Zhang et al. 2009; Kong et al. 2013). Bariatric surgery is a surgery that consists of gastric restriction (calibrated vertical gastroplasty, adjustable gastroplasty with adjustable rings and longitudinal gastrectomy) that reduces the amount of food to be ingested during a meal. It can be implemented in the form of a mixed system that combines gastric restriction with the bypass short-circuit (Roux-en-Y by-pass) to reduce the absorption of nutrients.

A new anaerobic bacterial species, strain AT7[T] = CSUR P2086 = DSM 100837, was isolated by a'microbial culturomics' approach from the faeces of a morbidly obese patient before bariatric surgery. The goal of culturomics was to set up a collection of all human-associated microbes using different bacterial growth conditions to mimic natural conditions (Lagier et al. 2012, 2016). The conventional approaches for bacterial delineation have been based on phenotypic characteristics, the 16S RNA gene sequences similarity (Kim et al. 2014), phylogenetic relationship (Stackebrant and Ebers 2006), the G+C content of the genomic sequence and DNA-DNA hybridization (DDH) (Rosselló-Mora 2006; Meier-Kolthoff et al. 2014). However, these tools have some limitations. We proposed to include genomic and spectrometric data in a polyphasic approach to describe new bacterial taxa. This new method of delineation was named taxono-genomics (Ramasamy et al. 2014; Fournier et al. 2015). This approach combines the phenotypic, biochemical characteristics, the MALDI-TOF spectra, genomic analysis and phylogenetic comparison to delineate new bacterial taxa.

The bacterial strain isolated in this study clustered in phylogenetic analyses with some species of the genus *Ruminococcus*, which was first described in Antonie Van Leeuwenhoek with *Ruminococcus flavefaciens* as the type species (Sijpesteijn 1949). The genus is composed of Gram-positive bacteria and currently contains ten species as reported in the'List of

prokaryotic names with standing in nomenclature' (http://www.bacterio.net/ruminococcus.html). Of eight other species originally identified as belonging to the genus *Ruminococcus*, six species have been reclassified in the genus *Blautia* (Liu et al. 2008; Lawson and Finegold 2015) and two as *Trichococcus* (Liu 2002). However, the remaining members of the genus *Ruminococcus* form two distinct phylogenetic groups in two different families, as previously described (Rainey and Janssen 1995; Willems and Collins 1995; Rainey 2010; Lawson and Finegold 2015). the family *Ruminococcaceae* contains the *Ruminococcus* type species *Ruminococcus flavefaciens*, along with *Ruminococcus albus*, *Ruminococcus bromii*, *Ruminococcus callidus* and *Ruminococcus champanellensis* (*Ruminococcus* sensu stricto; Rainey 2010; Chassard et al. 2012), whereas *Ruminococcus faecis*, *Ruminococcus gnavus*, *Ruminococcus lactaris* and *Ruminococcus torques* cluster with members of the family *Lachnospiraceae*. This separation of members of the genus *Ruminococcus* into two distinct families suggested that taxonomy of the current *Ruminococcus* species should be clarified.

Here, we describe the main phenotypic, phylogenetic and genotypic features of strain AT7[T] (= CSUR P2086 = DSM 100837) and propose the creation of a new genus, *Mediterraneibacter* gen. nov., that contains strain AT7[T] as the type strain of *Mediterraneibacter massiliensis* sp. nov. Furthermore, creation of this new genus resolves most of the inconsistencies observed in the taxonomy of the genus *Ruminococcus*.

Materials and methods

Sample collection

Stool samples were collected for a study comparing the microbiota of subjects suffering from morbid obesity before and after surgery. The patients gave a written informed consent and the study was validated by the ethics committee of the Institut Federatif de Recherche IFR48 under agreement number 09-022, 2010. The stool sample containing the bacterium described here was collected from a 37-year-old obese French woman (BMI 44.75 kg/m^2; 116 kg, 1.61 m) in July 2012. The samples were aliquoted and stored at − 80 °C degrees before analysis.

Strain isolation and growth conditions

The strain was grown in May 2015. The stool sample of the patient was pre-incubated in blood culture bottles enriched with 10% filter-sterilised rumen fluid and 10% sheep blood, as described elsewhere (Lagier et al. 2016). The growth and monitoring procedures, colony identification and purification procedures were similar to those described elsewhere (Togo et al. 2017). The isolated colonies were then identified by MALDI-TOF-mass spectrometry, as previously described (Seng et al. 2009). The current Bruker and local "culturomics" database contains 8687 reference spectra of bacterial and fungal species.

Phenotypic and biochemical characterisation

Different growth temperatures (25, 28, 37, 45 and 55 °C) were tested on 5% sheep blood-enriched Columbia agar (bioMérieux, Marcy l'Etoile, France). Growth of strain AT7[T] was tested under anaerobic atmosphere with the GENbag anaer system (bioMérieux), under microaerophilic atmosphere with the GENbag microaer system (bioMérieux) and under aerobic atmosphere, with or without 5% CO_2. Salt tolerance of the strain was tested using a 5–100 g/L NaCl concentration range on 5% sheep blood-enriched Schaedler agar (bioMérieux) under anaerobic atmosphere.

A fresh colony was observed between slides and slats using a Leica DM 1000 photonic microscope (Leica Microsystems, Nanterre, France) at 40× to assess bacterial motility. Transmission electron microscopy, using a Tecnai G20 microscope (FEI Company, Limeil-Brevannes, France) at an operating voltage of 60 kV was performed to observe strain AT7[T] after negative coloration. Gram staining was performed using a Gram staining kit (bioMérieux) and observed using a photonic microscope Leica DM 2500 (Leica Microsystems, Nanterre, France) with a 100× oil-immersion objective lens. Thermal shock at 80 °C for 20 min was carried out to test for sporulation.

Biochemical assays were performed in triplicate using API Gallery systems: API® ZYM (bioMérieux), API® 20A (bioMérieux) and API®50 CH (bioMérieux) according to the manufacturer's instructions. Detection of catalase and oxidase activity (Becton, Dickenson and Company, Le Pont de Claix, France) was also performed.

The antibiotic susceptibility of strain AT7[T] was tested following EUCAST recommendations (Citron et al. 1991; Matuschek et al. 2014). E-test strips for amikacin (0.016–256 μg/mL), vancomycin (0.016–256 μg/mL), imipenem (0.002–32 μg/mL), ceftriaxone (0.016–256 μg/mL), rifampicin (0.002–32 μg/mL), benzyl penicillin (0.002–32 μg/mL), amoxicillin (0.016–256 μg/mL), cefotaxime (0.002–32 μg/mL), metronidazole (0.016–256 μg/mL), minocycline (0.016–256 μg/mL), teicoplanin (0.016–256 μg/mL), erythromycin (0.016–256 μg/mL) and daptomycin (0.016–256 μg/mL) (bioMérieux) were deposited manually and the plates were incubated under anaerobic conditions for 48 h. Around the strip, elliptic zones of inhibition appeared and the intersection with the strip indicated the MIC (Citron et al. 1991). MICs were interpreted according to the EUCAST recommendations (http://www.eucast.org).

Fresh colonies were collected from 5% sheep blood-enriched Columbia agar (bioMérieux) after 48 h of incubation at 37 °C in an anaerobic atmosphere for cellular fatty acid methyl ester (FAME) analysis. The analysis was performed by Gas Chromatography/Mass Spectrometry (GC/MS), as described by Sasser (2006). GC/MS analyses were carried out as described by Dione et al. (2016). Metabolic end products were measured with a Clarus 500 chromatography system connected to a mass spectrometer (Perkin Elmer, Courtaboeuf, France), as detailed previously (Zhao et al. 2006), with some modifications. Acetic, propanoic, isobutanoic, butanoic, isopentanoic, pentanoic, isohexanoic, hexanoic and heptanoic acids were purchased from Sigma Aldrich (Lyon, France). A stock solution was prepared in water/methanol (50% v/v) at a final concentration of 50 mmol/L and then stored at − 20 °C. Calibration standards were freshly prepared in acidified water (pH 2–3 with 37% HCl) from the stock solution at the following concentrations: 0.5; 1; 5; 10 mmol/L. Short chain fatty acids were analysed from 3 independent culture bottles with BD Bactec[TM] Lytic/10 anaerobic/F culture vials media (Becton, Dickenson and Company); both blank and samples were analysed as described in previously (Togo et al. 2017).

Genomic characteristics

Sequencing and assembly

Genomic DNA (gDNA) of strain AT7[T] was sequenced with the MiSeq technology (Illumina Inc, San Diego, CA, USA) using the mate pair strategy. It was barcoded in order to be mixed with 11 other projects using the nextera mate pair sample prep kit. Qubit assay with the high sensitivity kit (Thermo Fisher Scientific, Waltham, MA, USA) were used to quantify the gDNA of the strain at a concentration of 130 ng/μl. The nextera mate pair Illumina guide was used to prepare the mate pair library with 1.5 μg of gDNA. The 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. The DNA fragments ranged from 1.5 kb up to 11 kb with an optimal size at 7.3 kb. No size selection was performed and 600 ng of tagmented fragments were circularised.

The circularised DNA was mechanically sheared to small fragments with an optimal size at 1336 bp on a Covaris device S2 in T6 tubes (Covaris, Woburn, MA, USA). The library profile was visualised on a high sensitivity bioanalyzer LabChip (Agilent Technologies Inc, Santa Clara, CA, USA) and the final concentration of the library was measured as 13.9 nmol/L. The libraries were normalised and pooled at 2 nM. After a denaturation step and dilution to 15 pM, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and sequencing run were performed in a single 39-h run in a 2 × 151-bp. Total information of 8.9 Giga bases was obtained from a 1009 K/mm[2] cluster density with a cluster passing quality control filters of 91.5% (17,486,000 passing filter paired reads). Within this run, the index representation for strain AT7[T] was determined to be of 8.4%. The 1,470,265 paired reads were trimmed and then assembled into 5 scaffolds using the SPAdes software (Bankevich et al. 2012).

Annotation and comparison

Open Reading Frames (ORFs) were predicted using Prodigal (Hyatt et al. 2010) with default parameters.

Nevertheless, the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched against the GenBank and Clusters of Orthologous Groups (COG) databases using BLASTP (Evalue 1e-03, coverage 70% and identity percent 30%). The tRNAs and rRNAs were predicted using the tRNA Scan-SE and RNAmmer tools, respectively (Lowe and Eddy 1997). SignalP and TMHMM were used to identify signal peptides and the number of transmembrane helices, respectively (Krogh et al. 2001; Bendtsen et al. 2004). Mobile genetic elements were predicted using PHAST and RAST (Zhou et al. 2011; Overbeek et al. 2014). ORFans were identified if their BLASTP E-value was lower than 1e-03 for an alignment length greater than 80 amino acids. If alignment lengths were smaller than 80 amino acids, we used an E-value of 1e-05. Artemis and DNA Plotter were used for data management and visualisation of genomic features, respectively (Carver et al. 2009, 2012). Genomes were automatically retrieved from the 16S rRNA tree using XEGEN software (Phylopattern) (Gouret et al. 2009). For each selected genome, complete genome sequence, proteome genome sequence and orfeome genome sequence were retrieved from the FTP of NCBI. All proteomes were analysed with proteinOrtho (Lechner et al. 2011). Then, for each pair of genomes, a similarity score of the average genomic identity of orthologous gene sequences (AGIOS) was computed. This score is the mean value of nucleotide similarity between all pairs of orthologous proteins for the two genomes studied (Ramasamy et al. 2014). For the evaluation of genomic similarity, digital DDH (dDDH) values were estimated using GGDC formula 2 (Meier-Kolthoff et al. 2013b). The average amino acid identity (AAI) was also calculated, based on the overall similarity between two genomic datasets of proteins, (Konstantinidis and Tiedje 2005; Rodriguez-R and Konstantinidis 2014) and is available at http://enve-omics.ce.gatech.edu/aai/index.

For the genomic comparison of strain AT7^T, the genomes of *R. lactaris* strain ATCC 29176^T = VPI X6-29^T (ABOU00000000) (Moore et al. 1976), *R. torques* strain ATCC 27756^T = VPI B2-51^T (GCA0001153925) (Holdeman and Moore 1974), *R. faecis* strain Eg2^T = KCTC 5757^T = JCM15917^T (BBDW0100000) (Kim et al. 2011), *Clostridium glycyrrhizinilyticum* strain ZM35^T = JCM 13368^T = DSM 17593^T (BBAB0100000) (Sakuma et al.

2006), *R. gnavus* strain ATCC 29149^T = VPI C7-9^T (PUEL00000000) (Moore et al. 1976), *Ruminococcus gauvreauii* strain CCRI-16110^T = NML 060141^T = CCUG 54292^T = JCM 14987^T (AUDP00000000) (Domingo et al. 2008), *R. albus* strain 7^T = ATCC 27210^T = DSM 20455^T = JCM 14654^T (CP002403) (Hungate 1957), *R. bromii* strain V.P.I. 6883^T = ATCC 27255^T (FMUV00000000) (Moore et al. 1972), *R. callidus* strain ATCC 27760^T = VPI S7-31^T (AWVF00000000) (Holdeman and Moore 1974), *R. champanellensis* strain 18P13^T = DSM 18848^T = JCM 17042^T (FP929052) (Chassard et al. 2012) *Coprococcus comes* strain ATCC 27758^T = VPI C1-38^T (ABVR00000000) (Holdeman and Moore 1974) and *R. flavefaciens* strain C94^T = ATCC 19208^T (JAEF00000000) (Sijpesteijn 1949) were used.

An annotation of the entire proteome was performed to define the distribution of functional classes of predicted genes according to the Clusters of Orthologous Groups of proteins (by using the same method as for genome annotation). The genome of this AT7^T was locally aligned pairwise using the BLAST algorithm against each of the selected genomes (Kent 2002; Auch et al. 2010).

Phylogenetic analysis

To clarify the taxonomic inconsistencies among *Ruminococcus* species, we have achieved the most robust strategy to date based on a phylogenetic tree based on 271 orthologous genes from the genomes of 27 closely related species and 1 outgroup (*Escherichia coli*). All 28 genomes were downloaded from NCBI (www.ncbi.nlm.nih.gov). For orthologue detection, we applied Proteinortho with default values (Lechner et al. 2011). All orthologous genes were aligned using Muscle (Edgar 2004) and then concatenated. Phylogenetic reconstruction was performed using the maximum likelihood method with the Kimura 2 parameter model and bootstrap value of 100.

Results

MALDI-TOF analysis

The spectrum generated from strain AT7^T (Fig. 1) did not match with that of any reference strain in the

Bruker plus culturomics database. Accordingly, this strain was suspected to correspond to a new species so that phenotypic and chemotaxonomic characteristics were determined, and genome sequencing was performed.

Phenotypic and biochemical characterisation

Strain AT7[T] was observed to be non-motile, asporogenous, coccobacillary-shaped, Gram-strain positive (Fig. S1) and anaerobic. The strain exhibits catalase activity but not oxidase activity. Growth was observed on 5% sheep blood Colombia agar plates between 28 and 45 °C, with optimal growth observed at 37 °C after 48 h of incubation under anaerobic atmosphere. The colonies were observed to be small (about 0.5–1 mm in diameter), translucent, punctiform and not haemolytic on 5% sheep blood Columbia agar. No growth of this bacterium was observed using 10–100 g/L of NaCl concentration on 5% sheep blood Schaedler agar plates. Strain AT7[T] was observed to grow at pH ranging from 6.5 to 8.5, with optimal growth at 7.2. Cells were determined to be 0.2–0.4 wide and 1–1.4 μm long under electron microscopy (Fig. S2). The phenotypic characteristics of strain AT7[T] were compared with those of its close phylogenetic neighbours, as shown in Table 1.

Using the API® ZYM test system, positive reactions were observed with trypsin, α-chymotrypsin, naphtol-AS-BI-phosphohydrolase and β-glucuronidase but negative reactions were observed with phosphatase alkaline, esterase, esterase lipase, lipase, leucine arylamidase, valine arylamidase, cystine arylamidase, phosphatase acid, α-galactosidase, β-galactosidase, α-glucosidase, β-glucosidase, *N*-acetyl-β-glucosaminidase, α-mannosidase and α-fructosidase. The API® 50 CH test system revealed that strain AT7[T] exhibits positive reactions for aesculin, arbutine, D-arabinose, D-cellobiose, D-fructose, D-galactose, D-glucose, D-lactose, D-maltose, D-mannitol, D-mannose, D-melibiose, D-saccharose, D-trehalose, D-xylose, dulcitol, gentiobiose, inositol, L-arabinose, L-sorbose, L-xylose, methyl-α D-glucopyranoside, potassium 2-cetogluconate, salicin and xylitol. Negative reactions were obtained with adonitol, amygdalin, D-arabitol, D-fucose, D-lyxose, D-melezitose, D-raffinose, D-ribose, D-sorbitol, D-tagatose, D-turanose, erythritol, glycerol, inulin, L-rhamnose, methyl-αD-mannopyranoside, methyl-β D-xylopyranoside, *N*-acetyl-glucosamine, glycogen, L-arabitol, L-fucose, potassium gluconate, potassium 5-cetogluconate and starch. Using the API® 20A test system, positive reactions were observed with aesculin, D-cellobiose, D-glucose, D-lactose, D-maltose, D-mannitol, D-mannose, D-saccharose, D-xylose,
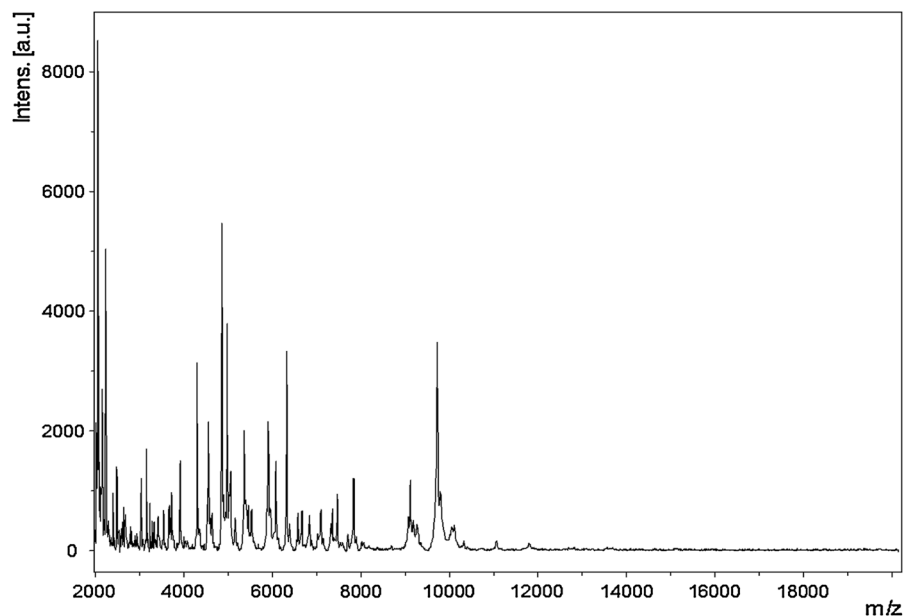
Fig. 1 Reference mass spectrum from strain AT7[T]. Spectra from 12 individual colonies were compared and a reference spectrum was generated

Table 1 Differential characteristics of strain AT7[T] compared to those of closely related species. (1) Strain AT7[T]; (2) *M. faecis* strain Eg2[T] (= KCTC 5757[T] = JCM 15917[T]); (3) *M. lactaris* strain ATCC 29176[T;] (= VPI X6-29[T]); (4) *M. torques* strain ATCC 27756[T] (= VPI B2-51[T]); (5) *M. glycyrrhizinilyticus* strain ZM35[T] (= JCM 13368[T] = DSM 17593[T]); (6) *M. gnavus* strain ATCC 29149[T] (= VPI C7-9[T]); (7) *Co. comes* strain ATCC 27758[T] (= VPI C1-38[T]), (8) *R. gauvreauii* strain CCRI-16110[T] (= NML 060141[T-] = CCUG 54292[T] = JCM 14987[T]); (9) *R. albus* strain 7[T] (= ATCC 27210[T] = DSM 20455[T] = JCM 14654[T]); (10) *R. bromii* strain V.P.I. 6883[T] (= ATCC 27255[T]); (11) *R. callidus* strain ATCC 27760[T] (= VPI S7-31[T]); (12) *R. champanellensis* strain 18P13[T] (= DSM 18848[T] = JCM 17042[T]); (13) *R. flavefaciens* strain C94[T] (ATCC 19208[T])

Properties	1	2	3	4	5	6	7	8	9	10	11	12	13
Catalase	+	+	+	+	−	+	−	−	+	−	+	−	−
Aesculin hydrolysis	+	+	−	+	−	+	V	−	−	−	+	+	−
Gelatine hydrolysis	+	−	+	+	−	+	+	−	Na	−	w	−	−
Acid production from													
Arabinose	+	−	−	−	+	+	−	−	−	−	−	−	−
Cellobiose	+	−	−	−	−	−	−	−	+	−	+	+	w
Erythritol	−	Na	−	−	Na	−	Na	−	Na	−	−	−	Na
Fructose	+	Na	+	+	−	+	+	+	−	−	w	−	−
Galactose	+	Na	Na	+	Na	+	+	+	−	−	−	−	−
Glucose	+	+	+	+	+	+	+	+	+	w	+	−	−
Lactose	+	+	+	+	+	−	+	−	−	−	+	−	−
Maltose	+	+	+	+	+	+	+	−	−	−	+	−	−
Mannitol	+	−	+	−	−	−	W	+	−	−	−	−	−
Mannose	+	−	w	−	−	−	W	−	+	−	w	−	−
Melibiose	+	Na	−	−	−	w	+	−	−	−	+	−	−
Raffinose	−	+	−	−	+	+	+	−	−	−	+	−	−
Rhamnose	−	−	−	−	+	+	Na	−	−	−	−	−	−
Ribose	−	Na	−	−	−	+	−	+	−	−	−	−	−
Saccharose	+	−	−	+	−	+	+	+	+	−	+	−	−
Salicin	+	−	−	w	−	+	W	−	−	−	−	−	−
Sorbitol	−	+	v	−	−	−	W	+	Na	−	−	−	Na
Starch	−	Na	−	−	Na	+	−	−	−	+	−	−	−
Trehalose	+	−	−	−	−	−	−	−	−	−	−	−	−
Xylose	+	−	−	−	+	+	+	−	−	−	w	−	−
Major end product of carbohydrate metabolism	A Ih	L A	F A L S	L A F	Na	F A L	L A B	A	A L S E	A F L P E	S A F	A S	A S F B L
G+C content (%)	42.4	43.4	45	42	45.7	43	40	47.6	44.2	39.1	43	53	43.2
Source	Human feces	Human feces	Human feces	Human feces	Human feces	Human feces	Human feces	Human feces	Rumen of cattle	Human feces	Human feces	Human feces	Human feces

A acetic acid, *F* formic acid, *L* lactic acid, *S* succinic acid, *E* ethanol, *P* pyruvic acid, *B* butyric acid, *Ih* isohexanoic acid, + positive reaction, − negative reaction, *Na* not available, *w* weakly reaction, *v* variable

D-trehalose, gelatine, L-arabinose and salicin. Reactions for D-raffinose, D-melezitose, D-sorbitol, glycerol, L-rhamnose, L-tryptophan and urea were found to be negative.

Strain AT7[T] was found to be susceptible to vancomycin (2 µg/mL), imipenem (0.047 µg/mL), ceftriaxone (0.75 µg/mL), rifampicin (0.002 µg/mL), benzyl penicillin (0.094 µg/mL), amoxicillin (0.094 µg/mL), cefotaxime (2 µg/mL), metronidazole (0.19 µg/mL), minocycline (0.0125 µg/mL), teicoplanin (0.016 µg/mL), erythromycin (0.025 µg/mL) and daptomycin (1 µg/mL). However, the strain was found to be resistant to amikacin (> 256 µg/mL). The minimum inhibitory concentration for each antimicrobial used is in parenthesis.

Total cellular fatty acid composition analysis of strain AT7[T] revealed that the most abundant fatty acids were $C_{16:0}$ (54%) and $C_{18:1n9}$ (30%). Minor amounts of other fatty acids ($C_{18:0}$, $C_{14:0}$, $C_{18:1n7}$, $C_{18:1n6}$, $C_{15:0}$, $C_{16:1n7}$, $C_{12:0}$, $C_{17:0}$, anteiso-$C_{15:0}$ and iso-$C_{15:0}$) were detected. The results of fatty acid analysis are summarised in Table 2.

Analysis of metabolic end products revealed that strain AT7[T] produces (after 72 h) acetic acid (17.1 ± 0.5 mM), isohexanoic acid (6 ± 0.2 mM), isobutanoic acid (2.3 ± 0.1 mM), butanoic acid (1.3 ± 0.1 mM), isopentanoic acid (1.3 ± 0.1 mM) and propanoic acid (0.7 ± 0.1 mM), but also small quantities (< 0.5 mM) of pentanoic and hexanoic acid.

Genomic analysis

Genome properties

The draft genome of strain AT7[T] has been deposited in EMBL-EBI under accession number FAVJ00000000 and is 3,069,882 bp long with 42.4% G+C content (Fig. 2). It is composed of five scaffolds and eight contigs. Among the 2925 predicted genes, 2867 are protein-coding genes and 58 are RNA genes (two 5S rRNA genes, one 16S rRNA gene, three 23S rRNA genes and fifty-two tRNA genes). A total of 2191 genes (76.4%) were assigned a putative function by COGs or NR blast. A total of 108 genes were identified as ORFans (4%). Using ARG-ANNOT (Gupta et al. 2014), no resistance genes were found, however, three genes (0.1%) were identified as PKS or NRPS (Conway and Boddy 2013). Using PHAST and RAST,

Table 2 Cellular fatty acid profiles of strain AT7[T] compared with those of closely related species; (1) Strain AT7[T]; (2) *Ruminococcus faecis* strain Eg2[T] (= KCTC 5757[T] = JCM 15917[T]); (3) *Ruminococcus gauvreauii* strain CCRI-16110[T] (= NML 060141[T] = CCUG 54292[T] = JCM 14987[T]) (4) *Ruminococcus champanellensis* strain 18P13[T] (= DSM 18848[T] = JCM 17042[T])

Fatty acids	1	2	3	4
anteiso-C15:0	< 1	ND	ND	**19.6**
anteiso-C17:0	0	ND	ND	2.8
C12:0	< 1	2.3	ND	ND
C13:1n12/C11:1 2-OH	0	1.9	ND	ND
C14:0	2.0 ± 0.2	**10**	**16.9**	ND
C15:0	< 1	ND	ND	ND
C15:2/C15:1n7	0	2.0	ND	ND
C16:0	**54.0 ± 4.2**	27.7	19.9	ND
C16:1n7	< 1	ND	ND	ND
C16:1n9	0	2.5	ND	ND
C17:0	< 1	ND	ND	0,4
C17:1n9/C17:2	0	2.7	ND	ND
C18: 1n11	0	ND	ND	ND
C18:0	9.0 ± 1.2	2.9	ND	0.7
C18:1c11/t9/t6	0	6.7	ND	ND
C18:1n11	0	ND	ND	ND
C18:1n6	2.0 ± 0.1	ND	ND	ND
C18:1n7	2.0 ± 1.2	ND	ND	ND
C18:1n9	**30.0 ± 2.3**	3.1	8.4	ND
C18:2n9, 12	0	3.3	ND	ND
iso-C13:03-OH	0	ND	ND	0.2
iso-C15:0	< 1	ND	ND	**26.6**
iso-C16:0	0	ND	ND	8.8
iso-C17:0	0	ND	ND	0.4

Date for 2–4 are taken from (Domingo et al. 2008; Kim et al. 2011; Chassard et al. 2012). *ND* not detected. Data were not available for *Ruminococcus torques* strain ATCC 27756[T], *Ruminococcus lactaris* strain ATCC 29176[T], *Clostridium glycyrrhizinilyticum* strain ZM35[T], *Coprococcus comes* strain ATCC 27758[T], *Ruminococcus gnavus* strain ATCC 29149[T], *Ruminococcus albus* strain 7[T], *Ruminococcus bromii* strain ATCC[T]; *Ruminococcus callidus* strain ATCC 27760[T] and *Ruminococcus flavefaciens* strain C94[T]

Bold values indicate major cellular fatty acids of the strains

1136 genes (40%) were found to be associated with mobilome elements. The remaining 483 genes (17%) were annotated as hypothetical proteins.
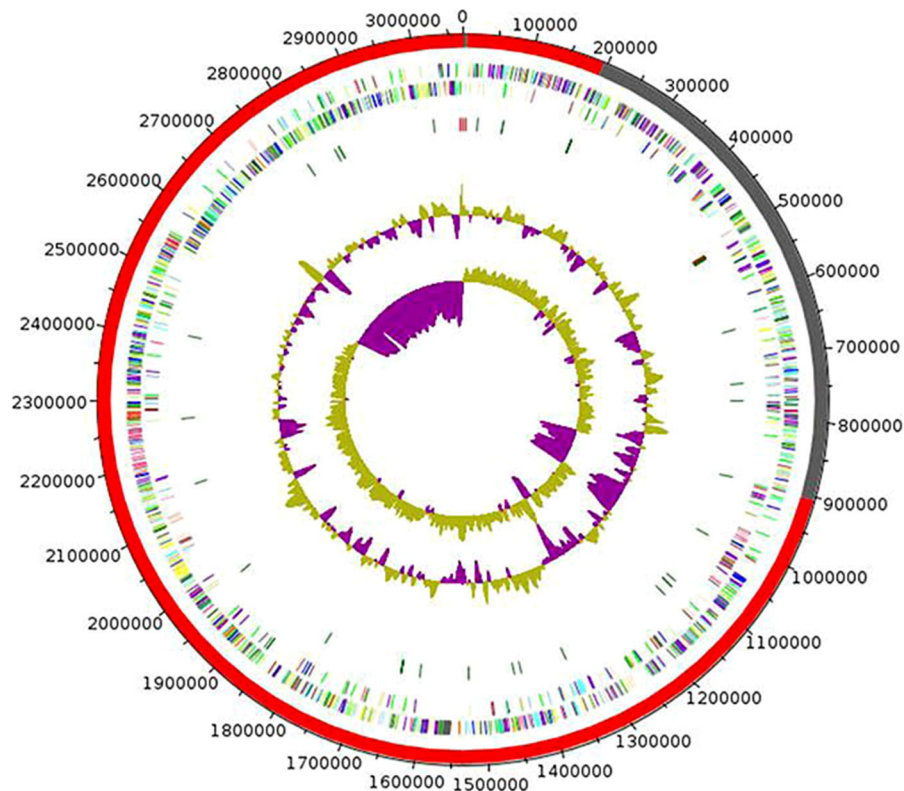
Fig. 2 Graphical circular map of the genome of strain AT7[T]. From outside to the centre: Contigs (red/grey), COG category of genes on the forward strand (three circles), genes on forward strand (blue circle), genes on the reverse strand (red circle), COG category on the reverse strand (three circles), GC content

16S gene-based phylogenetic analysis

16S rRNA gene sequence similarity values lower than 98.7% or 95%, have been used to assign strain to novel species or genera, respectively (Stackebrant and Ebers 2006; Kim et al. 2014; Yarza et al. 2014). The 16S gene sequence of strain AT7[T] exhibited a 95.2, 95.6, 95.6 and 95.9% nucleotide sequence similarity with *C. glycyrrhizinilyticum* strain ZM35[T] (= JCM 13368[T-] = DSM 17593[T]), *R. lactaris* strain ATCC 29176[T] (= VPI X6-29[T]), *R. faecis* strain Eg2[T] (= KCTC 5757[T] = JCM15917[T]) and *R. torques* strain JCM 6553[T] (= ATCC 27756[T] = VPI B2-51[T]), the closely related species with validly published names according to the phylogenetic analysis. The 16S rRNA gene sequence similarity values of strain AT7[T] and other members of the genus *Ruminococcus* are displayed in Table 3. Supplementary figure 3 (Fig. S3) shows a 16S rRNA gene tree for all *Ruminococcus* type strains plus type strains of type species and other representative species of genera in the families *Lachnospiraceae* and *Ruminococcaceae.* The 16S rRNA gene sequence of strain AT7[T] has been deposited in EMBL-EBI under accession number LN881607.

Genome comparison

The draft genome sequence of strain AT7[T] (3.07 Mb) is smaller than those of *Co. comes*, *R. faecis*, *R. flavefaciens*, *R. gnavus*, *R. gauvreauii* and *R. albus* (3.24, 3.26, 3.44, 3.62, 3.73 and 3.84 Mb respectively), larger than those of *R. bromii*, *R. champanellensis*, *R. lactaris* and *R. torques* (2.28, 2.54, 2.73 and 2.74 Mb respectively) but similar to that of *R. callidus* (3.09 Mb). Its G+C content (42.4%) is similar to that of *Co. comes* (42.5), lower than those of *R. gnavus*, *R. callidus*, *R. faecis*, *R. flavefaciens*, *R. lactaris*, *C. glycirrhizinilyticum*, *R. albus* and *R. champanellensis*

Table 3 16S rRNA gene sequence similarity values of strain AT7[T] obtained from comparisons with closely related species

RRNA sequences From	Strain AT7	M. faecis	M. lactaris	M. torques	M. glycyrrhizinilyticus	M. gnavus	Co. comes	R. gauvreauii	R. albus	R. bromii	R. callidus	R. champanellensis	R. flavefaciens
Similarity of 16S rRNA gene sequences													
Strain AT7[T] (LN881607)													
M. faecis strain Eg2[T] (FJ611794)	96												
M. lactaris strain ATCC 29176[T] (L76602)	96	96											
M. torques strain VPI B2-51[T] (L76604)	95	96	95										
M. glycyrrhizinilyticus strain ZM35[T] (AB233029)	95	96	94	95									
M. gnavus strain ATCC 29149[T] (X94967)	92	95	94	94	95								
Co. comes strain VPI C1-38[T] (EF031542)	94	95	94	94	96	94							
R. gauvreauii strain CCRI-16110[T] (EF529620)	91	93	92	92	92	93	93						
R. albus strain 7[T] (L76598)	85	86	86	86	83	84	83	84					
R. bromii strain ATCC 27255[T] (L76600)	82	83	82	82	82	93	82	83	89				
R. callidus strain ATCC 27760[T] (L76596)	84	84	84	85	85	84	84	84	90	89			
R. champanellensis strain 18P13[T] (AJ515913)	83	83	85	85	83	84	84	84	92	89	95		
R. flavefaciens strain C94[T] (L76603)	84	83	83	86	84	84	82	83	91	89	93	94	
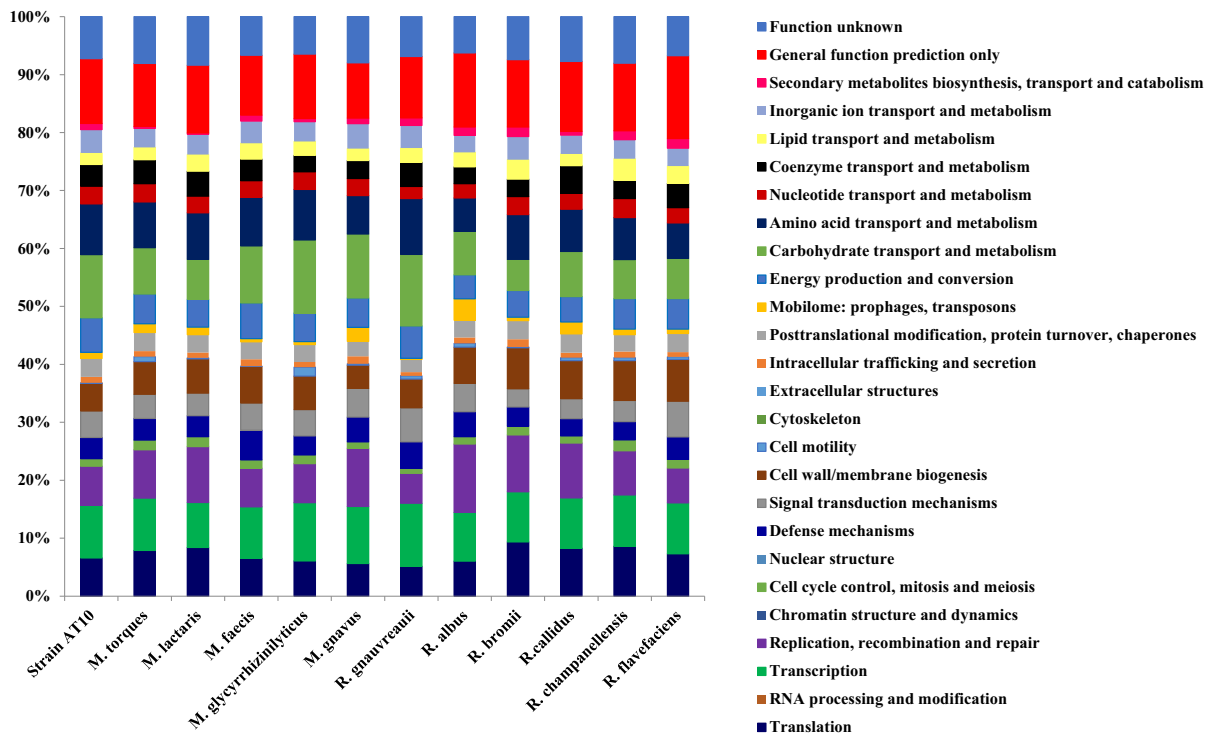
Fig. 3 Distribution of functional classes of predicted genes according to the clusters of orthologous groups of proteins of strain AT7[T] compared with closely related species

(43, 43, 43, 43.4, 44, 45, 45. 3, 45.8 and 53% respectively), but higher than those of *R. bromii*, *R. gauvreauii* and *R. torques* (40, 40 and 42% respectively). Its gene content (2925) is lower than those of *R. faecis*, *R. albus*, *C. glycirrhizinilyticum*, *Co. comes* and *R. gnavus* (3220; 3335; 3359, 3529 and 3744 respectively), but higher than those of *R. flavefaciens*, *R. gauvreauii*, *R. champanellensis*, *R. torques*, *R. lactaris*, and *R. bromii* (1807; 2110; 2371; 2491; 2486 and 2852 respectively). Even so, the distribution of genes into COG was similar among all compared genomes (Fig. 3 and Table 4). AGIOS values (Table 5) among compared species, except for strain AT7[T], ranged from 59% between *R. torques* and *R. champanellensis* to 75.9% between *R. lactaris* and *R. faecis*. When strain AT7[T] was compared to other species, this value ranged from 59.2% with *R. champanellensis* to 72.7% with *R. torques*. The dDDH values of strain AT7[T] ranged from 17.7% with *R. gauvreauii* to 29.2% with *R. callidus* and are shown in Table 6. The average amino acid identity values between strain AT7[T] and closely related species

ranged from 60.98% between *Co. comes* and strain AT7 to 73.49% between *R. faecis* and *R. lactaris*. However, these values were lower when strain AT7[T] and the group of closely related types strains were compared with *R. flavefaciens* and the species of the genus *Ruminococcus* sensu stricto as shown in Table 7.

Phylogenetic tree based on 271 concatenated orthologous genes from genomes of the 28 closest species

As *Ruminococcus* species can be separated into two different clusters belonging to two different taxonomic families (*Lachnospiraceae* and *Ruminococcaceae*), we decided to apply one of the best performing current taxonomic approaches based on genomic analysis using shared orthologous genes among closely related species (Fig. 4). Strain AT7[T] was found to cluster with *R. faecis*, *R. lactaris*, *R. torques*, *R. gnavus* and *C. glycyrrhizinilyticum*, forming a homogeneous cluster within the family

Table 4 Number of genes associated with the 25 general COG functional categories of strain AT7[T] compared to those of its closest species; (1) Strain AT7[T]; (2) *M. faecis* strain Eg2[T](= KCTC 5757[T] = JCM 15917[T]); (3) *M. lactaris* strain ATCC 29176[T](= VPI X6-29[T]); (4) *M. torques* strain ATCC 27756[T] (= VPI B2-51[T]); (5) *M. glycyrrhizinilyticus* strain ZM35[T] (= JCM 13368[T] = DSM 17593[T]); (6) *M. gnavus* strain ATCC 29149[T] (= VPI C7-9[T]); (7) *Co. comes* strain ATCC 27758[T] (= VPI C1-38[T]); (8) *R. gauvreauii* strain CCRI-16110[T] (= NML 060141[T] = CCUG 54292[T] = JCM 14987[T]); (9) *R. albus* strain 7[T] (= ATCC 27210[T] = DSM 20455[T] = JCM 14654[T]); (10) *R. bromii* strain V.P.I. 6883[T] (= ATCC 27255[T]); (11) *R. callidus* strain ATCC 27760[T] (= VPI S7-31[T]); (12) *R. champanellensis* strain 18P13[T] (= DSM 18848[T] = JCM 17042[T]); (13) *R. flavefaciens* strain C94[T] (ATCC 19208[T])

Genes	1	2	3	4	5	6	7	8	9	10	11	12	13
Translation	143	174	149	140	150	144	162	153	145	148	145	134	151
RNA processing and modification	0	0	0	0	0	0	0	0	0	0	0	0	0
Transcription	198	238	137	160	248	252	216	323	202	137	153	138	181
Replication, recombination and repair	146	178	171	148	166	257	206	154	283	155	166	119	124
Chromatin structure and dynamics	0	0	0	0	0	0	0	0	0	0	0	0	0
Cell cycle control, mitosis and meiosis	28	39	30	30	37	28	37	25	31	24	22	29	31
Nuclear structure	0	0	0	0	0	0	0	0	0	0	0	0	0
Defence mechanisms	80	137	64	66	81	110	98	137	103	52	52	49	80
Signal transduction mechanisms	99	125	69	74	112	127	133	174	118	50	61	57	127
Cell wall/membrane biogenesis	104	169	105	100	142	102	140	148	150	112	115	107	149
Cell motility	2	1	3	15	38	6	3	16	15	2	9	8	9
Cytoskeleton	0	0	0	0	0	0	0	0	0	0	0	0	0
Extracellular structures	0	0	0	0	0	0	0	0	0	0	0	0	0
Intracellular trafficking and secretion	22	32	16	17	22	33	25	19	25	21	15	17	17
Posttranslational modification, protein turnover, chaperones	69	81	55	57	74	66	77	65	71	51	57	45	66
Mobilome: prophages, transposons	23	14	23	27	13	62	14	7	89	9	36	15	16
Energy production and conversion	129	165	85	91	121	130	128	165	99	73	77	82	108
Carbohydrate transport and metabolism	237	263	122	142	311	282	212	367	181	85	137	105	144
Amino acid transport and metabolism	190	224	142	140	215	170	229	287	138	122	128	113	126
Nucleotide transport and metabolism	66	77	51	56	75	75	61	62	59	49	48	51	54
Coenzyme transport and metabolism	82	100	76	73	70	80	86	124	70	48	84	49	87
Lipid transport and metabolism	43	73	51	38	59	52	55	73	60	53	36	58	61
Inorganic ion transport and metabolism	88	102	62	58	84	111	74	116	70	63	57	51	64
Secondary metabolites biosynthesis, transport and catabolism	23	28	4	6	14	24	16	39	35	26	11	24	34
General function prediction only	243	276	207	193	274	244	235	315	307	184	212	182	295
Function unknown	157	178	148	143	159	204	169	204	150	117	136	125	139

COGs Clusters of Orthologous Groups database

Table 5 Pairwise comparison of strain AT7^T with closely related species using the AGIOS parameter; (1) Strain AT7^T; (2) *M. faecis* strain Eg2^T (= KCTC 5757^T = JCM 15917^T); (3) *M. lactaris* strain ATCC 29176^T (= VPI X6-29^T); (4) *M. torques* strain ATCC 27756^T (= VPI B2-51^T); (5) *M. glycyrrhizinilyticus* strain ZM35^T (= JCM 13368^T = DSM 17593^T); (6) *M. gnavus* strain ATCC 29149^T (= VPI C7-9^T); (7) *Co. comes* strain ATCC 27758^T = VPI C1-38^T; (8) *R. gauvreauii* strain CCRI-16110^T (= NML 060141^T = CCUG 54292^T = JCM 14987^T); (9) *R. albus* strain 7^T (= ATCC 27210^T = DSM 20455^T = JCM 14654^T); (10) *R. bromii* strain V.P.I. 6883^T (= ATCC 27255^T); (11) *R. callidus* strain ATCC 27760^T (= VPI S7-31^T); (12) *R. champanellensis* strain 18P13^T (= DSM 18848^T = JCM 17042^T); (13) *R. flavefaciens* strain C94^T(ATCC 19208^T)

Species	1	2	3	4	5	6	7	8	9	10	11	12	13
Strain AT7	**2869**	1002	1122	1177	1185	1256	987	1060	690	646	646	661	703
M. faecis	71.07	**3921**	1018	914	945	1017	925	912	613	579	609	564	609
M. lactaris	72.32	75.92	**2479**	1118	1055	1142	1000	1024	724	673	701	669	719
M. torques	72.57	72.04	73.10	**2489**	1077	1174	901	996	675	661	666	638	698
M. glycyrrhizinilyticus	71.73	70.58	71.86	71.88	**3359**	1184	910	1004	661	635	636	618	656
M. gnavus	72.70	71.45	72.71	71.96	72.68	**3760**	989	1092	710	663	693	642	710
Co. comes	69.10	71.82	71.07	68.85	69.69	70.17	**3529**	936	629	575	618	571	619
R. gauvreauii	65.90	65.20	66.72	66.16	66.53	66.88	66.32	**3790**	749	696	706	703	764
R. albus	60.41	60.49	61.22	60.76	60.15	60.95	60.86	60.45	**4051**	724	841	883	948
R. bromii	60.67	61.16	61.25	61.40	60.33	61.06	61.16	60.17	62.61	**2485**	715	729	723
R. callidus	59.86	60.38	61.27	60.17	61.01	60.08	61.05	61.09	63.97	61.36	**2847**	886	941
R. champanellensis	59.23	58.76	60.05	58.99	60.34	60.23	59.71	60.54	63.99	60.47	68.44	**2356**	935
R. flavefaciens	60.30	60.86	61.20	60.77	60.02	60.87	60.77	60.47	66.96	63.39	65.43	65.32	**3089**

Upper right, numbers of orthologous proteins shared between genomes; lower left, average percentage similarity of nucleotides corresponding to orthologous proteins shared between genomes and in bold, number of proteins for each species genome

Table 6 Pairwise comparison of strain AT7[T] with closely related species using the dDDH parameter; (1) Strain AT7[T]; (2) *M. faecis* strain Eg2[T](= KCTC 5757[T] = JCM 15917[T]); (3) *M. lactaris* strain ATCC 29176[T] (= VPI X6-29[T]); (4) *M. torques* strain ATCC 27756[T] (= VPI B2-51[T]); (5) *M. glycyrrhizinilyticus* strain ZM35[T] (= JCM 13368[T] = DSM 17593[T]); (6) *M. gnavus* strain ATCC 29149[T] (= VPI C7-9[T]); (7) *Co. comes* strain VPI C1-38[T] (= ATCC 27758[T]); (8) *R. gauvreauii* strain CCRI-16110[T] (= NML 060141[T-] = CCUG 54292[T] = JCM 14987[T]); (9) *R. albus* strain 7[T] (= ATCC 27210[T] = DSM 20455[T] = JC14654[T]); (10) *R. bromii* strain V.P.I. 6883[T] (= ATCC 27255[T]); (11) *R. callidus* strain ATCC 27760[T] (= VPI S7-31[T]); (12) *R. champanellensis* strain 18P13[T] (= DSM 18848[T] = JCM 17042[T]); (13) *R. flavefaciens* strain C94[T](= ATCC 19208[T])

Species	1	2	3	4	5	6	7	8	9	10	11	12	13
1		20.6% ± 2.3	19.4% ± 2.3	22.3% ± 2.3	18.9% ± 2.3	19.3% ± 2.3	24.1% ± 2.4	17.7% ± 2.2	26.7% ± 2.4	20.5% ± 2.3	29.2% ± 2.4	27.6% ± 2.4	27.1% ± 2.4
2			24.3% ± 2.3	23.2% ± 2.3	23.1% ± 2.4	25.1% ± 2.4	35.8% ± 2.5	24.4% ± 2.4	22.4% ± 2.3	16.7% ± 2.2	39.5% ± 2.5	20% ± 2.3	15.1% ± 2.1
3				24.6% ± 2.3	24.2% ± 2.4	21.3% ± 2.3	27.3% ± 2.5	21.9% ± 2.3	26.6% ± 2.4	19.5% ± 2.3	29.5% ± 2.4	23.5% ± 2.3	24.8% ± 2.4
4					24.5% ± 2.4	26.5% ± 2.4	27.2% ± 2.5	21.7% ± 2.3	25.8% ± 2.4	22.7% ± 2.4	38% ± 2.5	21.8% ± 2.3	26.5% ± 2.4
5						22.5% ± 2.4	24.2% ± 2.4	18.3% ± 2.3	28.4% ± 2.5	23.1% ± 2.4	23.6% ± 2.4	30.4% ± 2.5	40.6% ± 2.5
6							23.1% ± 2.3	19.6% ± 2.3	22.6% ± 2.4	21.7% ± 2.3	22.3% ± 2.3	26.8% ± 2.4	24.7% ± 2.4
7								23.9% ± 2.4	25.7% ± 2.4	21.8% ± 2.3	39.9% ± 2.5	28.8% ± 2.4	22.4% ± 2.4
8									18.3% ± 2.2	22.6% ± 2.4	19% ± 2.3	25.8% ± 2.4	18.8% ± 2.3
9										24.6% ± 2.4	24.4% ± 2.4	24.7% ± 2.4	18.8% ± 2.3
10											29.7% ± 2.4	19.3% ± 2.3	15.9% ± 2.2
11												20.4% ± 2.3	21.3% ± 2.3
12													17.7% ± 2.2
13													

Confidence intervals indicate inherent uncertainty in estimating DDH values from intergenomic distances based on models derived from empirical test data sets. These results are consistent with the 16S rRNA and phylogenomic analyses as well as the GGDC results: DDH, DNA-DNA hybridization and Genome-to-Genome Distance Calculator. HSP: high-scoring segment pairs

Table 7 The average amino acid identity values of strain AT7[T] compared with those of its phylogenetically close neighbours; (1) Strain AT7; (2) *M. faecis* strain Eg2[T] (= KCTC 5757[T] = JCM 15917[T]); (3) *M. lactaris* strain ATCC 29176[T] (= VPI X6-29[T]); (4) *M. torques* strain ATCC 27756[T] (= VPI B2-51[T]); (5) *M. glycyrrhizinilyticus* strain ZM35[T] (= JCM 13368[T] = DSM 17593[T]); (6) *M. gnavus* strain ATCC 29149[T] (= VPI C7-9[T]); (7) *Co. comes* strain ATCC 27758[T] (= VPI C1-38[T]), (8) *R. gauvreauii* strain CCRI-16110[T] (= NML 060141[T] = CCUG 54292[T] = JCM 14987[T]); (9) *R. albus* strain 7[T] (= ATCC 27210[T] = DSM 20455[T] = JCM 14654[T]); (10) *R. bromii* strain V.P.I. 6883[T] (= ATCC 27255[T]); (11) *R. callidus* strain ATCC 27760[T] (= VPI S7-31[T]); (12) *R. champanellensis* strain 18P13[T] (= DSM 18848[T] = JCM 17042[T]); (13) *R. flavefaciens* strain C94[T](ATCC 19208[T])

	1	2 (%)	3 (%)	4 (%)	5 (%)	6 (%)	7 (%)	8 (%)	9 (%)	10 (%)	11 (%)	12 (%)	13 (%)
Strain AT7		65.8	66.8	69.2	67.9	68.2	60.9	53.8	43.3	44.0	43.3	43.5	43.7
M. faecis			73.4	67.0	64.9	65.0	67.1	53.9	44.0	44.6	46.1	44.1	43.7
M. lactaris				68.6	65.4	65.7	64.4	54.7	44.3	45.2	45.4	44.2	44.3
M. torques					67.3	66.0	60.6	54.7	44.0	44.6	44.3	44.0	44.3
M. glycyrrhizinilyticus						67.3	60.9	54.7	43.5	44.1	44.7	43.8	43.5
M. gnavus							60.3	54.3	43.3	43.9	44.2	43.1	43.3
Co. comes								54.2	44.0	44.5	44.9	43.8	43.8
R. gauvreauii									42.8	43.7	43.3	43.6	43.5
R. albus										46.2	50.0	50.9	53.6
R. bromii											47.0	47.5	46.9
R. callidus												55.4	54.7
R. champanellensis													54.7
R. flavefaciens													
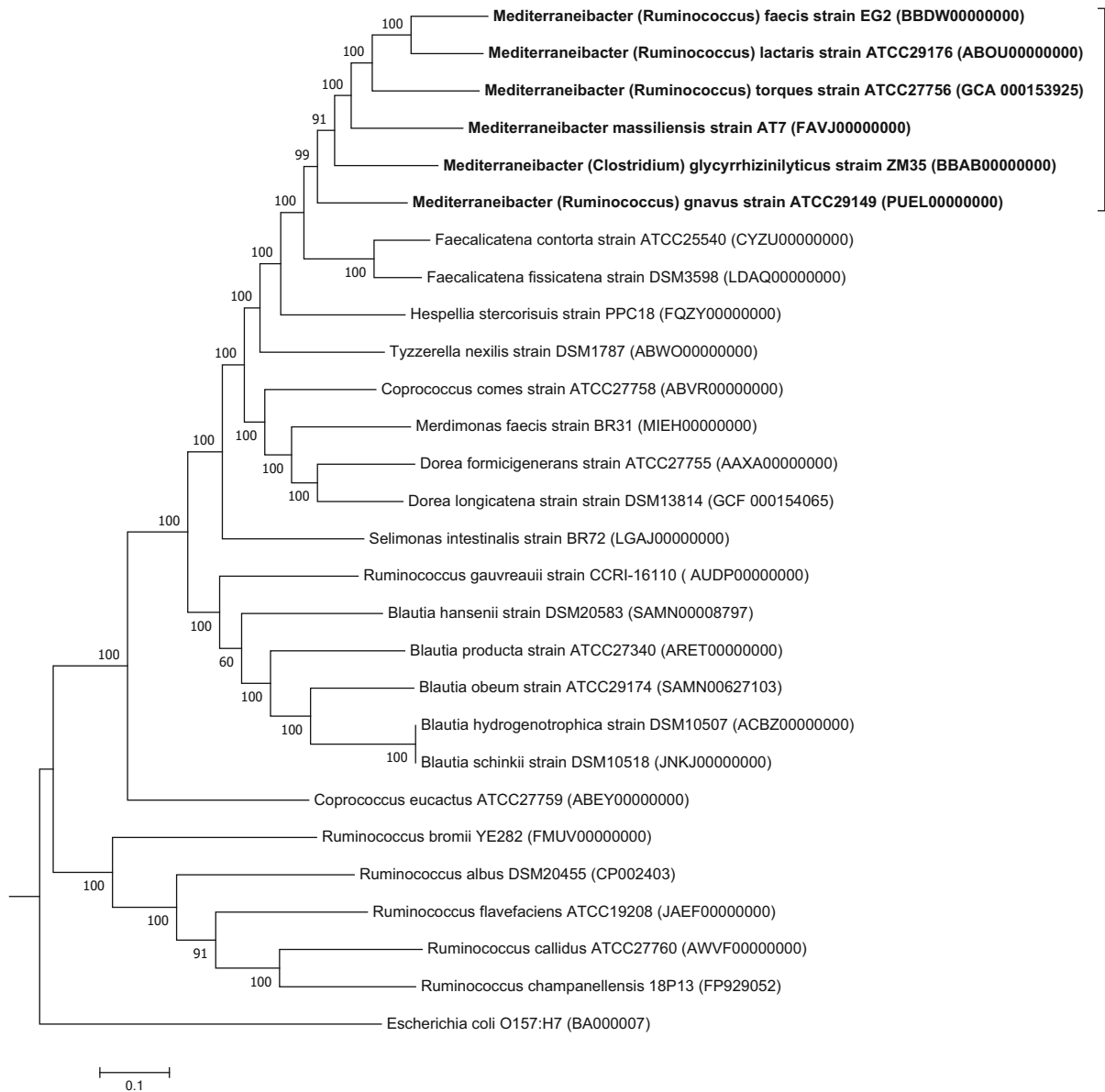
Fig. 4 Phylogenetic tree based on the 271 concatenated orthologous genes from the genomes of 28 related species. All 28 genomes were downloaded from NCBI (www.ncbi.nlm.nih. gov). For orthologous detection, we applied Proteinortho with default values (Lechner et al. 2011). All orthologous genes were aligned using Muscle (Edgar 2004) then concatenated. Phylogenetic reconstruction was performed using maximum likelihood method with the Kimura 2 parameter model and bootstrap value of 100

Lachnospiraceae. *R. gauvreauii* was also recovered as part of the family *Lachnospiraceae* but was not consistently related to the newly identified cluster (Fig. 4). In contrast, *R. bromii*, *R. albus*, *R. champanellensis*, *R. callidus* and *R. flavefaciens*, the type species of the genus *Ruminococcus*, formed a distinct

cluster. Based on these observations, 16S gene similarities (Table 3), number of shared orthologous proteins (Table 5), average of genomic identity of orthologous gene sequences (AGIOS—Table 5), and average amino acid identity (AAI—Table 7), we propose a new genus, *Mediterraneibacter*, to include

a new species, *Mediterraneibacter massiliensis*, represented by the type strain AT7[T] and to clarify the taxonomy of *Ruminococcus* species by reclassification of most of those species that do not cluster with the type species of the genus *Ruminococcus* in phylogenetic analyses. The phenotypic, chemotaxonomic, 16S similarities and genomic comparisons are shown in Tables 1, 2, 3, 4, 5, 6 and 7.

The 16S gene similarity between strain AT7[T] and *R. gnavus* (92%) was lower than the usual threshold of 94% for delineating genera. However, recent findings suggest that using only the 16S rRNA gene similarity is not adequate and that genomic analysis based on shared orthologous genes is much more robust (Fox et al. 1992; Coenye et al. 2005; Konstantinidis and Tiedje 2005; Varghese et al. 2015). Indeed, the phylogenetic tree based on 271 concatenated shared orthologous genes (Fig. 4), the number of shared proteins (Table 5), AGIOS (Table 5) and AAI (Table 7) all confirm that *R. gnavus* should be included in the new genus.

The sequence of the 16S ribosomal RNA gene alone does not allow satisfactory discrimination of the species in the *Lachnospiraceae* family. This is illustrated by the very low bootstrap values (Figure S3). These values are all below 70% for nodes between species of the new genus (accordingly not shown in Fig. S3). In the phylogenetic tree based on 271 shared orthologous genes (Fig. 4), the bootstrap values of the nodes between the species of the new genus are between 91 and 100% and the bootstrap of the node that differentiates the new *Mediterraneibacter* genus and the closely related genus *Faecalicatena* is 100%. This means that the creation of the new genus is based on very robust results (concatenated phylogenetic tree based on 271 shared orthologous genes) whereas the analysis based on the 16S ribosomal gene alone was associated with a very high risk of phylogenetic error.

Based on these findings, we propose to reclassify these four *Ruminococcus* species, namely *R. faecis*, *R. lactaris*, *R. torques* and *R. gnavus* and *C. glycyrrhizinilyticum* within the new genus *Mediterraneibacter* as *Mediterraneibacter faecis* comb. nov., *Mediterraneibacter lactaris* comb. nov., *Mediterraneibacter torques* comb. nov., *Mediterraneibacter gnavus* comb. nov. and *Mediterraneibacter glycyrrhizinilyticus* comb. nov. In addition, we observed that *R. gauvreauii* should probably be

reclassified in the *Blautia* genus but further analyses specifically focusing on this genus are necessary.

Discussion and conclusion

Strain AT7[T] was considered to represent a new species of the new genus *Mediterraneibacter* based on its MALDI-TOF spectrum (Fig. 1), which could not be identified on our database that contains more than 8000 spectra, 16S rRNA similarity level and genomic characteristics. Comparison of this bacterial species with other closely related species (Table 1) showed that strain AT7[T] can be differentiated by its metabolism of mannitol, mannose, salicin and trehalose. The dDDH (Table 6) values are very low when compared to closely related species, using threshold set at 70% according to Meier-Kolthoff et al. (2013a). The genomic comparisons (AGIOS and dDDH) reported in Tables 5 and 6 confirm that the similarities between strain AT7 and closely related species are in accordance with the proposition of a new species. Phenotypic differences, together with phylogenetic and genomic findings, allow us to propose strain AT7[T] (= CSUR P2086[T] = DSM 100837[T)]) as the type strain of *Mediterraneibacter massiliensis* gen. nov., sp. nov.

This new bacterium is potentially important for human health because it has been isolated from a morbidly obesity patient. It is currently known that some species of the family *Lachnospiraceae* family, namely *R. gnavus*, *Blautia obeum* and *Coprococcus catus* are strongly associated with weight gain, both in humans and in experimental models (Sepp et al. 2013; Petriz et al. 2014; Ziętak et al. 2016). More recently, *R. gnavus* was associated with adiposity in a microbiome-wide association study (MWAS) (Beaumont 2016). *R. gnavus* was also associated with obesity in another large-scale metagenomic study (Le Chatelier et al. 2013). This is particularly interesting because, to our knowledge, *R. gnavus* is one of the rare bacteria consistently associated with obesity and/or adiposity. However, based on our comprehensive phylogenetic analyses, *R. gnavus* should be classified in the genus *Mediterraneibacter*. Correcting this classification of this species is important as inaccurate nomenclature could lead researchers to draw erroneous conclusions about the role of the members of the genus *Ruminococcus* sensu stricto with regard to weight and adiposity regulation. Accordingly, the reclassification

of *R. gnavus* will help prevent confusion and will help studies analysing relationships between obesity and the gut microbiota.

In addition, we investigated the presence of 16S rRNA from strain AT7^T in the high throughput DNA and RNA sequence read archive (SRA) using an online open resource (Lagkouvardos et al. 2016). We found metagenomic sequences with a similarity greater than 97% with strain AT7^T in several gut metagenomes (human, bovine, chicken, mouse, rat, pig, primate and insect), skin (mouse, human) metagenomes, human oral metagenome, human lung metagenome, vaginal metagenome, food metagenome, as well as in environmental samples (wastewater, groundwater, seawater, marine sediment, bioreactor, hydrothermal vent, sludge, soil and insect). Metagenomic sequences corresponding to strain AT7^T were found in 7.9% (10844/135936) of all metagenomes and 30.7% (6191/20156) of the human gut metagenomes present in this database. Accordingly, the bacterium described here is found in the human mature anaerobic gut microbiota (HMAGM) (Million et al. 2017), consistent with its isolation from the stool sample of a 37-year-old French woman living in Marseille, who suffered from morbid obesity.

The Digital Protologue TaxoNumbers (http://imedea.uib-csic.es/dprotologue/index.php) of *M. massiliensis* gen. nov., sp. nov., *M. faecis* comb. nov., *M. lactaris* comb. nov., *M. torques* comb. nov., *M. gnavus* comb. nov. and *M. glycyrrhizinilyticus* comb. nov. are GA00061/TA00494, TA00495, TA00496, TA00497, TA00498 and TA00499, respectively.

Description of *Mediterraneibacter* gen. nov.

Mediterraneibacter (Me.di.ter.ra.ne.i.bac'ter. L. neut. n. *mediterraneum mare*, the Mediterranean sea; N.L. masc. n. *bacter* a rod; N.L. masc. n. *Mediterraneibacter* a rod from the Mediterranean Sea).

Gram-stain positive, asporogenous, non-motile, coccoid or coccobacillary-shaped, catalase positive and obligately anaerobic. The major end products of carbohydrate metabolism are acetic acid, formic acid and lactic acid. The DNA G+C content of the ranges from 42 to 45 mol %. The type species of the genus is *Mediterraneibacter massiliensis*, which was isolated from human faeces.

Description of *Mediterraneibacter massiliensis* sp. nov.

Mediterraneibacter massiliensis (mas.si.li.en'sis. L. masc. adj. *massiliensis*, of Massilia, the Latin name for Marseille).

In addition to the characteristics in the genus description, cells are coccobacillary-shaped, with a width ranging from 0.2 to 0.4 μm and a length ranging from 1 to 1.4 μm. Colonies are translucent with a diameter of 0.5–1 mm on 5% sheep blood Columbia agar. Oxidase negative. Optimum growth temperature is 37 °C under anaerobic conditions and pH tolerance ranges from 6.5 to 8.5. The major fatty acids are $C_{16:0}$ and $C_{18:1n9}$. The major end product of carbohydrate metabolism also include isohexanoic acid and isobutanoic acid. The draft genome of the type strain is 3,069,882 bp long with a DNA G+C content of 42.4%.

The type strain AT7^T has been deposited in the CSUR and DSM collections under numbers CSUR P2086 and DSM 100837, respectively. The type strain was isolated from the stool sample of a 37-year-old obese French woman. The draft genome and 16S rRNA sequences of the type strain have been deposited in EMBL-EBI under accession numbers FAVJ00000000 and LN881607, respectively.

Description of *Mediterraneibacter faecis* comb. nov.

Mediterraneibacter faecis (fae'cis. L. gen. n. *faecis*, of faeces, referring to its faecal origin).

Basonym: *Ruminococcus faecis* Kim et al. 2011.

The description of *Mediterraneibacter faecis* is the same as that given for *Ruminococcus faecis* (Kim et al. 2011). The type strain is Eg2^T (= KCTC 5757^T = JCM 15917^T).

Description of *Mediterraneibacter lactaris* comb. nov.

Mediterraneibacter lactaris (lac.ta'ris. L. masc. adj. *lactaris* milk-drinking [referring to its rapid fermentation of lactose and curding of milk]).

Basonym: *Ruminococcus lactaris* (Moore et al. 1976) Approved Lists 1980.

The description of *Mediterraneibacter lactaris* is the same as given for *Ruminococcus lactaris* (Moore et al. 1976). The type strain is ATCC 29176^T (= VPI X6-29^T).

Description of *Mediterraneibacter torques* comb. nov.

Mediterraneibacter torques (tor'ques. L. n. *torques* twisted necklace [referring to appearance of the chains from broth cultures]).

Basonym: *Ruminococcus torques* (Holdeman and Moore 1974) Approved Lists 1980.

The description of *Mediterraneibacter torques* is the same as given for *Ruminococcus torques* (Holdeman and Moore 1974). The type strain is ATCC 27756^T (= VPI B2-51^T).

Description of *Mediterraneibacter gnavus* comb. nov.

Mediterraneibacter gnavus (gna'vus. L. masc. adj. *gnavus* busy, active [referring to the active fermentative ability of this species]).

Basonym: *Ruminococcus gnavus* (Moore et al. 1976) Approved Lists 1980.

The description of *Mediterraneibacter gnavus* is the same as given for *Ruminococcus gnavus* (Moore et al. 1976). The type strain is ATCC 29149 (= VPI C7-9).

Description of *Mediterraneibacter glycyrrhizinilyticus* comb. nov.

Mediterraneibacter glycyrrhizinilyticus (gly.cy.rrhi. zi.ni.ly'ti.cus. N.L. neut. n. *glycyrrhizinum* glycyrrhizin [a sugar from the roots of Glycyrrhiza species], N.L. masc. adj. *lyticus* dissolving, able to dissolve, N.L. masc. adj. *glycyrrhizinilyticus* glycyrrhizin dissolving).

Basonym: *Clostridium glycyrrhizinilyticum* Sakuma et al. 2006.

The description of *Mediterraneibacter glycyrrhizinilyticus* is the same as given for *Clostridium glycyrrhizinilyticum* (Sakuma et al. 2006). The type strain is strain ZM35^T (= JCM 13368^T = DSM 17593^T).

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

References

Auch AF, von Jan M, Klenk H-P, Göker M (2010) Digital DNA–DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci 2:117–134

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol J Comput Mol Cell Biol 19:455–477

Beaumont M (2016) Genetic and environmental factors affecting the human gut microbiom in obesity. Student thesis. Doctoral thesis, Doctor of Philosophy, King's College

Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783–795

Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J (2009) DNAPlotter: circular and linear interactive genome visualization. Bioinform Oxf Engl 25:119–120

Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA (2012) Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinform Oxf Engl 28:464–469

Chassard C, Delmas E, Robert C, Lawson PA, Bernalier-Donadille A (2012) *Ruminococcus champanellensis* sp. nov., a cellulose-degrading bacterium from human gut microbiota. Int J Syst Evol Microbiol 62:138–143

Citron DM, Ostovari MI, Karlsson A, Goldstein EJ (1991) Evaluation of the E test for susceptibility testing of anaerobic bacteria. J Clin Microbiol 29:2197–2203

Coenye T, Gevers D, Van de Peer Y, Vandamme P, Swings J (2005) Towards a prokaryotic genomic taxonomy. FEMS Microbiol Rev 29:147–167

Conway KR, Boddy CN (2013) ClusterMine360: a database of microbial PKS/NRPS biosynthesis. Nucleic Acids Res 41:D402–D407

Dione N, Sankar SA, Lagier J-C, Khelaifia S, Michele C, Armstrong N, Richez M, Abrahão J, Raoult D, Fournier P-E (2016) Genome sequence and description of *Anaerosalibacter massiliensis* sp. nov. New Microbes New Infect 10:66–76

Domingo M-C, Huletsky A, Boissinot M, Bernard KA, Picard FJ, Bergeron MG (2008) *Ruminococcus gauvreauii* sp. nov., a glycopeptide-resistant species isolated from a human faecal specimen. Int J Syst Evol Microbiol 58:1393–1397

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Fournier P-E, Lagier J-C, Dubourg G, Raoult D (2015) From culturomics to taxonomogenomics: a need to change the taxonomy of prokaryotes in clinical microbiology. Anaerobe 36:73–78

Fox GE, Wisotzkey JD, Jurtshuk P (1992) How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int J Syst Bacteriol 42:166–170

Gouret P, Thompson JD, Pontarotti P (2009) PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinform 10:298

Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, Rolain J-M (2014) ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 58:212–220

Holdeman LV, Moore WEC (1974) New genus, *Coprococcus*, twelve new species, and emended descriptions of four previously described species of bacteria from human feces. Int J Syst Evol Microbiol 24:260–277

Hungate RE (1957) Microorganisms in the rumen of cattle fed a constant ration. Can J Microbiol 3:289–311

Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform 11:119

Kent WJ (2002) BLAT–the BLAST-like alignment tool. Genome Res 12:656–664

Kim M-S, Roh SW, Bae J-W (2011) *Ruminococcus faecis* sp. nov., isolated from human faeces. J Microbiol Seoul Korea 49:487–491

Kim M, Oh H-S, Park S-C, Chun J (2014) Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol 64:346–351

Kong L-C, Tap J, Aron-Wisnewsky J, Pelloux V, Basdevant A, Bouillot J-L, Zucker J-D, Dore J, Clement K (2013) Gut microbiota after gastric bypass in human obesity: increased richness and associations of bacterial genera with adipose tissue genes. Am J Clin Nutr 98:16–24

Konstantinidis KT, Tiedje JM (2005) Towards a genome-based taxonomy for prokaryotes. J Bacteriol 187:6258–6264

Krogh A, Larsson B, Von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874

Lagier J-C, Armougom F, Million M, Hugon P, Pagnier I, Robert C, Bittar F, Fournous G, Gimenez G, Maraninchi M, Trape J-F, Koonin EV, La Scola B, Raoult D (2012) Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 18:1185–1193

Lagier J-C, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P, Caputo A, Cadoret F, Traore SI, Seck EH, Dubourg G, Durand G, Mourembou G, Guilhot E, Togo A, Bellali S, Bachar D, Cassir N, Bittar F, Delerce J, Mailhe M, Ricaboni D, Bilen M, Dangui Nieko NPM, Dia Badiane NM, Valles C, Mouelhi D, Diop K, Million M, Musso D, Abrahão J, Azhar EI, Bibi F, Yasir M, Diallo A, Sokhna C, Djossou F, Vitton V, Robert C, Rolain JM, La Scola B, Fournier P-E, Levasseur A, Raoult D (2016) Culture of previously uncultured members of the human gut microbiota by culturomics. Nat Microbiol 1:16203

Lagkouvardos I, Joseph D, Kapfhammer M, Giritli S, Horn M, Haller D, Clavel T (2016) IMNGS: a comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. Sci Rep 6:33721

Lawson PA, Finegold SM (2015) Reclassification of *Ruminococcus obeum* as *Blautia obeum* comb. nov. Int J Syst Evol Microbiol 65:789–793

Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto J-M, Kennedy S, Leonard P, Li J, Burgdorf K, Grarup N, Jørgensen T, Brandslund I, Nielsen HB, Juncker AS, Bertalan M, Levenez F, Pons N, Rasmussen S, Sunagawa S, Tap J, Tims S, Zoetendal EG, Brunak S, Clément K, Doré J, Kleerebezem M, Kristiansen K, Renault P, Sicheritz-Ponten T, de Vos WM, Zucker J-D, Raes J, Hansen T, Bork P, Wang J, Ehrlich SD, Pedersen O, Guedon E, Delorme C, Layec S, Khaci G, van de Guchte M, Vandemeulebrouck G, Jamet A, Dervyn R, Sanchez N, Maguin E, Haimet F, Winogradski Y, Cultrone A, Leclerc M, Juste C, Blottière H, Pelletier E, LePaslier D, Artiguenave F, Bruls T, Weissenbach J, Turner K, Parkhill J, Antolin M, Manichanh C, Casellas F, Boruel N, Varela E, Torrejon A, Guarner F, Denariaz G, Derrien M, van Hylckama Vlieg J E T, Veiga P, Oozeer R, Knol J, Rescigno M, Brechot C, M'Rini C, Mérieux A, Yamada T (2013) Richness of human gut microbiome correlates with metabolic markers. Nature 500:541–546

Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ (2011) Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinform 12:124

Liu JR (2002) Emended description of the genus *Trichococcus*, description of *Trichococcus collinsii* sp. nov., and reclassification of *Lactosphaera pasteurii* as *Trichococcus pasteurii* comb. nov. and of *Ruminococcus palustris* as *Trichococcus palustris* comb. nov. in the low-G+C Gram-positive bacteria. Int J Syst Evol Microbiol 52:1113–1126

Liu C, Finegold SM, Song Y, Lawson PA (2008) Reclassification of *Clostridium coccoides*, *Ruminococcus hansenii*, *Ruminococcus hydrogenotrophicus*, *Ruminococcus luti*, *Ruminococcus productus* and *Ruminococcus schinkii* as *Blautia coccoides* gen. nov., comb. nov., *Blautia hansenii* comb. nov., *Blautia hydrogenotrophica* comb. nov., *Blautia luti* comb. nov., *Blautia producta* comb. nov., *Blautia schinkii* comb. nov. and description of *Blautia wexlerae* sp. nov., isolated from human faeces. Int J Syst Evol Microbiol 58:1896–1902

Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964

Matta J, Zins M, Feral-Pierssens AL, Carette C, Ozguler A, Golberg M, Czernichow S (2016) Prévalence du surpoids, de l'obésité et des facteurs de risque cardio-métaboliques dans la cohorte Constances. Bull Epidémiol Hebd 35–36:640–646

Matuschek E, Brown DFJ, Kahlmeter G (2014) Development of the EUCAST disk diffusion antimicrobial susceptibility testing method and its implementation in routine microbiology laboratories. Clin Microbiol Infect 20:O255–O266

Meier-Kolthoff JP, Göker M, Spröer C, Klenk H-P (2013a) When should a DDH experiment be mandatory in microbial taxonomy? Arch Microbiol 195:413–418

Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M (2013b) Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinform 14:60

Meier-Kolthoff JP, Klenk H-P, Göker M (2014) Taxonomic use of DNA G+C content and DNA–DNA hybridization in the genomic age. Int J Syst Evol Microbiol 64:352–356

Million M, Diallo A, Raoult D (2017) Gut microbiota and malnutrition. Microb Pathog 106:127–138

Moore WEC, Cato EP, Holdeman LV (1972) *Ruminococcus bromii* sp. n. and emendation of the description of *Ruminococcus Sijpestein*. Int J Syst Bacteriol 22:78–80

Moore ERB, Johnson JL, Holdeman LV (1976) Emendation of *Bacteroidaceae* and *Butyrivibrio* and descriptions of *Desulfomonas* gen. nov. and ten new species in the genera *Desulfomonas*, *Butyrivibrio*, *Eubacterium*, *Clostridium*, and *Ruminococcus*. Int J Syst Evol Microbiol 26:238–252

Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, Mullany EC, Biryukov S, Abbafati C, Abera SF, Abraham JP, Abu-Rmeileh NME, Achoki T, AlBuhairan FS, Alemu ZA, Alfonso R, Ali MK, Ali R, Guzman NA, Ammar W, Anwari P, Banerjee A, Barquera S, Basu S, Bennett DA, Bhutta Z, Blore J, Cabral N, Nonato IC, Chang J-C, Chowdhury R, Courville KJ, Criqui MH, Cundiff DK, Dabhadkar KC, Dandona L, Davis A, Dayama A, Dharmaratne SD, Ding EL, Durrani AM, Esteghamati A, Farzadfar F, Fay DFJ, Feigin VL, Flaxman A, Forouzanfar MH, Goto A, Green MA, Gupta R, Hafezi-Nejad N, Hankey GJ, Harewood HC, Havmoeller R, Hay S, Hernandez L, Husseini A, Idrisov BT, Ikeda N, Islami F, Jahangir E, Jassal SK, Jee SH, Jeffreys M, Jonas JB, Kabagambe EK, Khalifa SEAH, Kengne AP, Khader YS, Khang Y-H, Kim D, Kimokoti RW, Kinge JM, Kokubo Y, Kosen S, Kwan G, Lai T, Leinsalu M, Li Y, Liang X, Liu S, Logroscino G, Lotufo PA, Lu Y, Ma J, Mainoo NK, Mensah GA, Merriman TR, Mokdad AH, Moschandreas J,

Naghavi M, Naheed A, Nand D, Narayan KMV, Nelson EL, Neuhouser ML, Nisar MI, Ohkubo T, Oti SO, Pedroza A, Prabhakaran D, Roy N, Sampson U, Seo H, Sepanlou SG, Shibuya K, Shiri R, Shiue I, Singh GM, Singh JA, Skirbekk V, Stapelberg NJC, Sturua L, Sykes BL, Tobias M, Tran BX, Trasande L, Toyoshima H, van de Vijver S, Vasankari TJ, Veerman JL, Velasquez-Melendez G, Vlassov VV, Vollset SE, Vos T, Wang C, Wang X, Weiderpass E, Werdecker A, Wright JL, Yang YC, Yatsuya H, Yoon J, Yoon S-J, Zhao Y, Zhou M, Zhu S, Lopez AD, Murray CJL, Gakidou E (2014) Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 384:766–781

Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R (2014) The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res 42:D206–D214

Petriz BA, Castro AP, Almeida JA, Gomes CP, Fernandes GR, Kruger RH, Pereira RW, Franco OL (2014) Exercise induction of gut microbiota modifications in obese, non-obese and hypertensive rats. BMC Genom 15:511

Rainey FA (2010) Family VIII. *Ruminococcaceae* fam. nov. In: De Vos P, Garrity GM, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB (eds) Bergey's manual of systematic bacteriology, 2nd edn. Springer, New York

Rainey FA, Janssen PH (1995) Phylogenetic analysis by 16S ribosomal DNA sequence comparison reveals two unrelated groups of species within the genus *Ruminococcus*. FEMS Microbiol Lett 129:69–73

Ramasamy D, Mishra AK, Lagier J-C, Padhmanabhan R, Rossi M, Sentausa E, Raoult D, Fournier P-E (2014) A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol 64:384–391

Rodriguez-R LM, Konstantinidis KT (2014) Bypassing cultivation to identify bacterial species: culture-independent genomic approaches identify credibly distinct clusters, avoid cultivation bias, and provide true insights into microbial species. Microbe Mag 9:111–118

Rosselló-Mora R (2006) DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation. In: Molecular identification, systematics, and population structure of prokaryotes. Springer, pp 23–50

Sakuma K, Kitahara M, Kibe R, Sakamoto M, Benno Y (2006) *Clostridium glycyrrhizinilyticum* sp. nov., a glycyrrhizin-hydrolysing bacterium isolated from human faeces. Microbiol Immunol 50:481–485

Sasser M (2006) Bacterial identification by gas chromatographic analysis of fatty acids methyl esters (GC-FAME)

Seng P, Drancourt M, Gouriet F, La Scola B, Fournier P-E, Rolain JM, Raoult D (2009) Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 49:543–551

Sepp E, Lõivukene K, Julge K, Voor T, Mikelsaar M (2013) The association of gut microbiota with body weight and body

mass index in preschool children of Estonia. Microb Ecol Health Dis 24:19231

Sijpesteijn AK (1949) Cellulose-decomposing bacteria from the rumen of cattle. Antonie Van Leeuwenhoek 15:49–52

Stackebrant E, Ebers J (2006) Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 33:152–155

Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci USA 101:11030–11035

Togo AH, Durand G, Khelaifia S, Armstrong N, Robert C, Cadoret F, Di Pinto F, Delerce J, Levasseur A, Raoult D, Million M (2017) *Fournierella massiliensis*, gen. nov., sp. nov., a new human-associated member of the family *Ruminococcaceae*. Int J Syst Evol Microbiol 67:1393–1399

Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A (2015) Microbial species delineation using whole genome sequences. Nucleic Acids Res 43:6761–6771

Willems A, Collins MD (1995) Phylogenetic analysis of *Ruminococcus flavefaciens*, the type species of the genus *Ruminococcus*, does not support the reclassification of *Streptococcus hansenii* and *Peptostreptococcus productus* as ruminococci. Int J Syst Bacteriol 45:572–575

Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol 12:635–645

Zhang H, DiBaise JK, Zuccolo A, Kudrna D, Braidotti M, Yu Y, Parameswaran P, Crowell MD, Wing R, Rittmann BE, Krajmalnik-Brown R (2009) Human gut microbiota in obesity and after gastric bypass. Proc Natl Acad Sci USA 106:2365–2370

Zhao G, Nyman M, Jönsson JA (2006) Rapid determination of short-chain fatty acids in colonic contents and faeces of humans and rats by acidified water-extraction and direct-injection gas chromatography. Biomed Chromatogr BMC 20:674–682

Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS (2011) PHAST: a fast phage search tool. Nucleic Acids Res 39:347–352

Ziętak M, Kovatcheva-Datchary P, Markiewicz LH, Ståhlman M, Kozak LP, Bäckhed F (2016) Altered microbiota contributes to reduced diet-induced obesity upon cold exposure. Cell Metab 23:1216–1223

CrossMark

CORRECTION

Correction to: Description of *Mediterraneibacter massiliensis*, gen. nov., sp. nov., a new genus isolated from the gut microbiota of an obese patient and reclassification of *Ruminococcus faecis*, *Ruminococcus lactaris*, *Ruminococcus torques*, *Ruminococcus gnavus* and *Clostridium glycyrrhizinilyticum* as *Mediterraneibacter faecis* comb. nov., *Mediterraneibacter lactaris* comb. nov., *Mediterraneibacter torques* comb. nov., *Mediterraneibacter gnavus* comb. nov. and *Mediterraneibacter glycyrrhizinilyticus* comb. nov.

**Amadou Hamidou Togo · Awa Diop · Fadi Bittar · Marie Maraninchi ·
René Valero · Nicholas Armstrong · Grégory Dubourg · Noémie Labas ·
Magali Richez · Jeremy Delerce · Anthony Levasseur · Pierre-Edouard Fournier ·
Didier Raoult · Matthieu Million**

Correction to: Antonie van Leeuwenhoek
https://doi.org/10.1007/s10482-018-1104-y

Subsequent to the publication of the above article, it has been noticed that the designation of the type strain is not correct. The strain referred to throughout the article as strain AT7[T] should be designated as strain Marseille-P2086[T] (= CSUR P2086[T] = DSM 100837[T]). The corrected for protologue for the species *Mediterraneibacter massiliensis*, represented by strain Marseille-P2086[T] as type strain, is given below.

A. H. Togo · N. Armstrong · G. Dubourg ·
M. Richez · J. Delerce · A. Levasseur ·
D. Raoult · M. Million (✉)
Aix Marseille Univ, IRD, MEPHI, IHU-Méditerranée
Infection, 19-21 Boulevard Jean Moulin, 13005 Marseille,
France
e-mail: matthieumillion@gmail.com

A. Diop · F. Bittar · N. Labas · P.-E. Fournier
Aix Marseille Univ, IRD, VITROME, IHU-Méditerranée
Infection, 19-21 Boulevard Jean Moulin, 13005 Marseille,
France

M. Maraninchi · R. Valero
NORT "Nutrition, Obesity and Risk of Thrombosis",
INSERM1062, INRA1260, Aix Marseille Université,
13385 Marseille, France

R. Valero
CHU Hôpital de la Conception, Service Nutrition,
Maladies Métaboliques et Endocrinologie, APHM,
13385 Marseille, France

🙋 Springer

Description of *Mediterraneibacter massiliensis* sp. nov.

Mediterraneibacter massiliensis (mas.si.li.en'sis. L. masc. adj. *massiliensis*, of Massilia, the Latin name for Marseille).

In addition to the characteristics in the genus description, cells are coccobacillary-shaped, with a width ranging from 0.2 to 0.4 μm and a length ranging from 1 to 1.4 μm. Colonies are translucent with a diameter of 0.5–1 mm on 5% sheep blood Columbia agar. Oxidase negative. Optimum growth temperature is 37 °C under anaerobic conditions and pH tolerance ranges from 6.5 to 8.5. The major fatty acids are $C_{16:0}$ and $C_{18:1n9}$. The major end product of carbohydrate metabolism also includes isohexanoic acid and isobutanoic acid. The draft genome of the type strain is 3,069,882 bp long with a DNA G + C content of 42.4%.

The type strain Marseille-P2086^T has been deposited in the CSUR and DSMZ culture collections under numbers CSUR P2086 and DSM 100837, respectively. The type strain was isolated from the stool sample of a 37-year-old obese French woman. The draft genome and 16S rRNA sequences of the type strain have been deposited in EMBL-EBI under accession numbers FAVJ00000000 and LN881607, respectively.

Article IX:

Fournierella massiliensis **gen. nov., sp. nov., a new human associated member of the family *Ruminococcaceae*.**

Amadou Hamidou Togo, Guillaume Durand, Saber Khelaifia, Nicholas Armstrong, Catherine Robert, Frédéric Cadoret, Fabrizio Di Pinto, Jérémy Delerce, Anthony Levasseur, Didier Raoult and Matthieu Million

Fournierella massiliensis gen. nov., sp. nov., a new human-associated member of the family *Ruminococcaceae*

Amadou Hamidou Togo,[1] Guillaume Durand,[1] Saber Khelaifia,[1] Nicholas Armstrong,[1] Catherine Robert,[1] Frédéric Cadoret,[1] Fabrizio Di Pinto,[1] Jérémy Delerce,[1] Anthony Levasseur,[1] Didier Raoult[1,2] and Matthieu Million[1,*]

Abstract

An anaerobic bacterium, strain AT2^T, was isolated from the fresh stool sample of a healthy French man using the culturomics approach. The 16S rRNA gene sequence analysis showed that strain AT2^T had 95.2 % nucleotide sequence similarity with *Gemmiger formicilis* ATCC 27749^T, the phylogenetically closest species with standing in nomenclature. Cells are Gram-stain-negative, catalase- and oxidase-negative, obligately anaerobic, non-motile, non-spore-forming, rod-shaped, and the bacilli were mesothermophilic. The major fatty acids were $C_{16:0}$ (43.8 %) and $C_{18:1n9}$ (20 %). The DNA G+C content of the strain based on its genome sequence was 56.8 mol%. Based on the phenotypic, biochemical and phylogenetic analysis, we propose the creation of the genus *Fournierella* gen. nov., which contains strain AT2^T (=CSUR P2014^T=DSM 100451^T) as the type strain of the type species *Fournierella massiliensis* gen. nov., sp. nov.

Culturomics is a new approach for the characterization of living microbial diversity in any environmental or human sample [1]. With the development of new technologies such as high-throughput sequencing enabling public access to the complete genome sequences of many bacterial species, we proposed the inclusion of the complete genome sequence analysis in a polyphasic approach to describe new bacterial taxa [2]. This strategy, which we named taxono-genomics, combines phenotypic characteristics, notably the matrix assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS spectrum and genomic properties [3, 4].

During an exploratory study of fresh stool by culturomics [1], an isolate was obtained and a new genus was proposed to accommodate this strain as a member of the family *Ruminococcaceae* [5]. At the time of writing, the family *Ruminococcaceae* contained 17 genera, including *Acetanaerobacterium*, *Acetivibrio*, *Anaerobacterium*, *Anaerofilum*, *Anaerotruncus*, *Ercella*, *Ethanoligenens*, *Faecalibacterium*, *Fastidiosipila*, *Gemmiger*, *Hydrogenoanaerobacterium*, *Oscillibacter*, *Oscillospira*, *Papillibacter*, *Ruminococcus*, *Sporobacter* and *Subdoligranulum* (www.bacterio.net/ruminococcaceae.html). Among members of the family, *Faecalibacterium prausnitzii* is one of the most abundant bacteria of the human gut. It is an extremely oxygen-sensitive bacterium that is difficult to cultivate, even in

anaerobic conditions [6]. *F. prausnitzii* sustains growth in the presence of low partial pressure of oxygen, in presence of antioxidants [7], and showed mutualism with epithelial cells, possibly through mucin [8]. *F. prausnitzii* is one of the leading representatives of the human healthy mature anaerobic gut microbiota (HMAGM), suggesting the link between dietary antioxidants and maintenance of the HMAGM [9, 10]. It contributes to maintaining host–microbial homeostasis by secreting a microbial anti-inflammatory molecule that inhibits cellular NF-κB signalling and inflammation [11]. Changes in the abundance of *F. prausnitzii* have been linked to dysbiosis in several human disorders [12]. Here, we propose the main phenotypic, phylogenetic and genomic properties of strain AT2^T (=CSUR P2014^T=DSM 100451^T), that is close to but substantially differs from *Gemmiger formicilis*.

Strain AT2^T was isolated from a fresh stool sample collected from a healthy 28-year-old French man in January 2015. The stool sample was immediately stored at 4 °C after collection until being used for culture. The donor gave a written informed consent and the study was validated by the ethics committee of the Institut Federatif de Recherche IFR48 under agreement number 09–022. To isolate the novel strain, 1 g stool sample was injected in an anaerobic blood culture bottle (BACTEC Lytic/10 Anaerobic/F

139

Culture Vials) supplemented with 4 ml filter-sterilized rumen fluid and 5 % sheep blood, and then incubated at 37 °C. After incubation for 3 days, 100 µl culture suspension was collected, plated on 5 % sheep blood-enriched Columbia agar (BioMérieux) and incubated at 37 °C in an anaerobic atmosphere for 48 h. Emerging colonies were subcultured individually for purification using the same conditions, and identified by MALDI-TOF MS as described by Seng *et al.* [13]. Isolates were re-streaked three times and purity was confirmed by direct examination and MALDI-TOF MS analysis. Purity was confirmed when all 12 spots from 12 different colonies yielded 12 perfectly matching MALDI-TOF MS spectra. When the strain was not identified by MALDI-TOF MS, its 16S rRNA gene sequence was assessed as previously reported [14] using the fD1-rP2 primers, a GeneAmp PCR System 2720 thermal cycler (Applied Biosystems) and an ABI Prism 3130-XL capillary sequencer (Applied Biosciences).

For taxonomic assignment, the Chromas Pro 1.34 software (Technelysium Pty.) was used to correct sequences. Pairwise sequence similarities were calculated using the method recommended by Meier-Kolthoff [15] for the 16S rRNA gene sequence, available via the GGDC web server [16] and at (http://ggdc.dsmz.de/). Phylogenies were inferred by the GGDC web server using the DSMZ phylogenomics pipeline [17] adapted to single genes. A multiple sequence alignment was created with MUSCLE [18]. A maximum-likelihood (ML) tree was inferred from the alignment with RAxML [19]. Rapid bootstrapping in conjunction with the autoMRE boot stopping criterion [20] and subsequent search for the best tree was used. The sequences were checked for a compositional bias using the X^2 test as implemented in PAUP* [21].

Different growth temperatures (25, 28, 37, 45 and 55 °C) were tested. Growth of the strain was tested under anaerobic and microaerophilic conditions using GENbag Anaer and GENbag Microaer systems, respectively (BioMérieux), and under aerobic conditions, with or without 5 % CO_2. Transmission electron microscopy of the strain, using a Tecnai G20 (FEI company) at an operating voltage of 60 kV, was performed after negative staining. The pH for growth of the strain was tested at a range from pH 6 to 8.5. Tolerance to NaCl was tested using a range of 5–100 g l^{-1} NaCl on Schaedler agar with 5 % sheep blood (BioMérieux) in an anaerobic atmosphere. Gram staining was performed and observed using a Leica DM 2500 photonic microscope (Leica Microsystems) with a ×100 oil immersion lens. In addition to Gram staining, the KOH test was carried out to confirm the cell-wall type according to the procedures described elsewhere [22, 23]. Staining was also performed under anaerobic conditions as differences can occur with exposure to oxygen [24].

Motility of the bacterium was assessed using a Leica DM 1000 photonic microscope (Leica Microsystems) at ×100 magnification. A thermic shock at 80 °C for 20 min on fresh colonies of the strain was carried out in order to test sporulation. The viability of cells was checked by subculturing them on the same media before heating, while the motility of strain AT2^T was tested observing fresh colonies using a DM 1000 photonic microscope (Leica Microsystems) with a ×100 oil-immersion objective lens. Catalase (BioMérieux) activity was determined in 3 % hydrogen peroxide solution, and oxidase activity was assessed using an oxidase reagent (Becton Dickinson). Biochemical properties of the strain were investigated using API ZYM, 20A and 50CH strips (BioMérieux) according to the manufacturer's instructions. The antibiotic susceptibility of strain AT2^T was tested using the disc diffusion method [25]; the results are shown in the supplementary data. Cellular fatty acid methyl ester (FAME) analysis was performed by GC/MS. Two samples were prepared with approximately 5 mg bacterial biomass per tube harvested from several culture plates. FAMEs were prepared as described by Sasser [26]. GC/MS analyses were carried out as described before [27]. Briefly, FAMEs were separated using an Elite 5 MS column and monitored by mass spectrometry (Clarus 500 - SQ 8 S; Perkin Elmer). Spectral database search was performed using MS Search 2.0 operated with the Standard Reference Database 1A (NIST) and the FAMEs mass spectral database (Wiley).

For the analysis of short-chain fatty acid (SCFA) production, a Wilkins-Chalgren-Anaerobe broth supplemented with cysteine and DTT (reduced WCA broth described previously by Kläring *et al.* and Pfeiffer *et al.* [28, 29]) and prepared using strictly anaerobic techniques (100 % N_2) was used. Samples were collected at 24, 48 and 72 h after inoculation. Acetic, propanoic, butanoic, isobutanoic, pentanoic, hexanoic and heptanoic acids were purchased from Sigma Aldrich. A stock solution was prepared in water/methanol (50 %, v/v) at 0.1 M for each SCFA and then stored at −20 °C. Calibration standards were freshly prepared in water: 0.05, 0.5, 1, 5 and 10 mM. Culture samples, prepared in duplicate, were centrifuged 5 min at 13 000 r.p.m. and the supernatants were collected. All solutions were adjusted to pH 2–3 with HCl before injection. SCFAs were measured with a Clarus 500 chromatography system connected to a SQ8 S mass spectrometer (Perkin Elmer). Analysis was performed with an Elite FFAP column (30 m, 0.25 mm id, 0.25 µm film thickness) such as detailed previously [30]. Injection volume was 0.5 µl (split less, 200 °C). Helium was supplied at 1 ml min^{-1} as the carrier gas. Compounds were separated according to a linear temperature gradient from 100 to 200 °C at 8 °C min^{-1}. Selected ion recording mass spectrometry SCFA analysis by GC/MS was performed using the following masses: 43 m/z for isobutanoic acid; 60 m/z for acetic, butanoic, pentanoic, hexanoic and heptanoic acids; 74 m/z for propanoic acid. The transfer line and the electron impact source were set at 200 °C. Quadratic calibration curves were automatically fitted with an acceptable coefficient of determination above 0.999 and deviation below 20 % (Turbo mass 6.1, Perkin Elmer). SCFA quantities in samples were presented after subtraction of the quantities found in the blank samples.

Genomic DNA (gDNA) of the strain was sequenced using a MiSeq sequencer (Illumina) 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 [31]. The gDNA was quantified by a Qubit assay with the high sensitivity kit (Life technologies) to 72.2 ng μl^{-1}. The mate-pair library was prepared with 1.5 μg genomic DNA using the Nextera mate-pair Illumina guide. The gDNA sample was simultaneously fragmented and tagged with a mate-pair junction adapter. The fragmentation pattern was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies) with a DNA 7500 labchip. The DNA fragments ranged in size from 1.5 kb up to 11 kb with a maximum at 6.7 kb. No size selection was performed and 412 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with a maximum at 1033 bp on the Covaris device S2 in T6 tubes (Covaris). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies) and the final concentration library was measured at 24.1 nmol l^{-1}. The libraries were normalized at 4 nM and pooled. After a denaturation step and dilution, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and sequencing run were performed in a single 2×251 bp run.

MALDI-TOF MS failed to identify strain AT2^T at the genus and species levels. Therefore, its spectrum was added to our database to improve its content. For phylogenetic analysis, the input nucleotide matrix comprised 11 operational taxonomic units and 1572 characters, 405 of which were variable and 267 of which were parsimony-informative. The base-frequency check indicated no compositional bias (*P*=0.99, α=0.05). ML analysis under the GTR+Gamma model yielded a highest log likelihood of −5973.03, whereas the estimated alpha parameter was of 0.17. The ML bootstrapping converged after 100 replicates; the average support was of 87.75 %.

The 16S rRNA gene sequencing showed that the strain AT2^T exhibited 95.2 % nucleotide sequence similarity with *Gemmiger formicilis*, the phylogenetically closest species with standing in nomenclature [32]. The resulting phylogenetic tree highlighting the position of strain AT2^T with the phylogenetically closest species with a validly published name is shown in Fig. 1 (see also Figs S1 and S2, available in the online Supplementary Material), and strain AT2^T was thus classified in the family *Ruminococcaceae* [5]. Differences in MALDI-TOF MS spectra between strain AT2^T and other closely related species with available spectrum are presented in Fig. S3.

Colonies of strain AT2^T obtained on 5 % sheep blood-enriched Columbia agar (BioMérieux) were translucent with a diameter of 0.3 to 1 mm. Growth of the strain was observed in anaerobic and microaerophilic atmospheres at 28, 37 and 45 °C but optimal growth was observed in an anaerobic atmosphere at 37 °C after incubation for 48 h. No growth was obtained at 55 °C or in an aerobic atmosphere.

Cells of strain AT2^T were Gram-stain-negative (confirmed by the KOH test and under anaerobic conditions), rod-shaped, measured 0.5 μm in diameter and 2 μm in length (Fig. 2), were non-motile, non-spore-forming and without catalase and oxidase activities. The strain grew at a pH ranging from 6 to 8.5, with optimal growth at pH 7.0 to 7.3. No growth was observed on Schaedler agar enriched with 5 % sheep blood at 10 g l^{-1} NaCl. The main characteristics of strain AT2^T compared to the closest species are shown in Table 1. The classification and general features of strain AT2^T are summarized in Table S1. Analysis of the total cellular fatty acid composition demonstrated that the major fatty acid was the saturated acid $C_{16:0}$ (43.8 %) followed by the unsaturated acid $C_{18:1n9}$ (20 %). Values represent the GC area percentage from total identified fatty acid methyl esters only (aldehydes, dimethyl acetates and unidentified 'summed features' described previously were not included). Cellular fatty acid profiles of strain AT2^T compared with other closely related species are summarized in Table 2.

Strain AT2^T produced SCFAs after 24, 48 and 72 h of culture in reduced WCA broth. After 72 h, the production of acetic acid was predominant (>10 mM), higher than butanoic (6.0±0.3 mM), isobutanoic (2.4±0.1 mM), propanoic (0.6±0.1 mM), pentanoic (0.1±0.1 mM), isopentanoic and isohexanoic acids (the last two were not quantified). Hexanoic and heptanoic acids were not detected.

The draft genome of strain AT2^T was deposited in EMBL-EBI under accession number FAUK00000000 (Fig. S4), it is 3 829 842 bp long with a G+C content of 56.8 %. It is composed of 19 scaffolds (27 contigs). Of the 3632 predicted genes, 3553 were protein-coding genes, and 79 were RNAs (one 16S rRNA, four 23S rRNA, six 5S rRNA, 68 tRNAs). A total of 2514 genes (70.7 %) were assigned a putative function by COGs or NR BLAST comparison. A total of 298 genes (8.4 %) were identified as ORFans. Using ARG-ANNOT [33], two genes (0.06 %) associated with resistance were detected and seven genes (0.20 %) associated to polyketide synthase (PKS) or nonribosomal peptide (NRPS) [34] were discovered through genome analysis. Using PHAST and RAST, 1799 genes (50.6 %) were associated to mobilome elements. The remaining genes (616) were annotated as hypothetical proteins (Table S2). The distribution of genes into COGs functional categories is shown in Table S3. Considering closest species with available genome, the digital DNA–DNA hybridization (dDDH) values ranged from 23.4 % with *Subdoligranulum variabile* BI 114^T to 36.6 % with *F. prausnitzii* ATCC 27768^T (Table S4).

Compared with the closest phylogenetic species (Fig. 1), the phylogenetic distance between strain AT2^T and its closest neighbour was superior to the distance between *G. formicilis* X2-56^T and *S. variabile* BI 114^T, between *G. formicilis* X2-56^T and *F. prausnitzii* ATCC 27768^T and between *S. variabile* BI 114^T and *F. prausnitzii* ATCC 27768^T. Phylogenomics analysis was not possible given the unavailability of the closest species' genomes, but genomic comparisons (dDDH and AGIOS, reported in Tables S4 and S5, respectively)

141
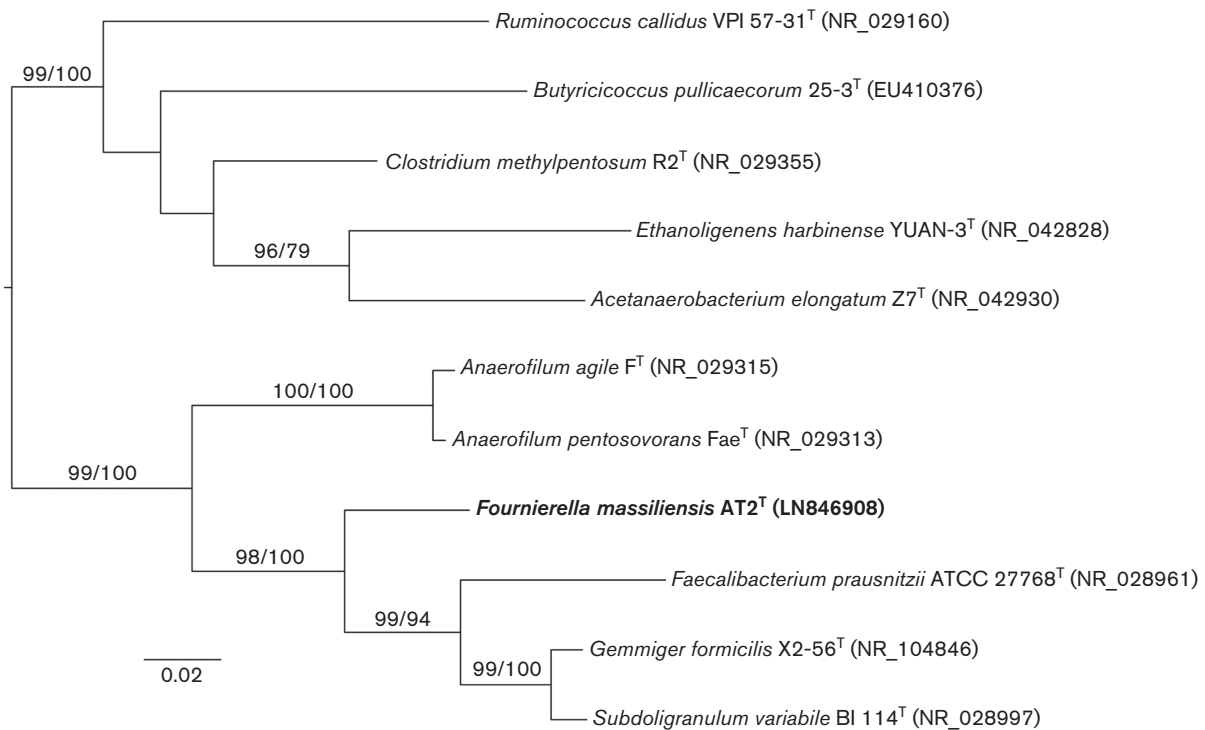
Fig. 1. ML tree inferred under the GTR+GAMMA model and rooted by midpoint-rooting. The branches are scaled in terms of the expected number of substitutions per site. Numbers above the branches are support values when larger than 60 % from ML (left) and MP (right) bootstrapping.

confirmed that the similarities between strain AT2[T] and the closest species are in accordance with the proposition of a new genus in the family *Ruminococcaceae*. Moreover, the G+C% difference exceeded 1 % (−2.2 % compared with *G. formicilis* X2-56[T], +4.6 % compared with *S. variabile* BI 114[T]). According to Qin *et al.* [35], a strain from a new genus will have less than 50 % pairwise percentage of conserved proteins with its closest phylogenetic neighbours. This percentage was 25.4 % (902/3553) with *F. prausnitzii* ATCC 27768[T] and 29.4 % (1045/3553) with *S. variabile* BI 114[T], confirming strain AT2[T] as a member of a new genus (Table S5). The genome of *G. formicilis* was not available but the 16S rRNA gene phylogenetic distance between strain AT2[T] and *G. formicilis* X2-56[T] was very similar to that of *F. prausnitzii* ATCC 27768[T] (Fig. 1). The 16S rRNA gene similarity values further support the proposal of a novel genus. Indeed, Yarza *et al.* [36] reported a median sequence identity of 96.4 % (95 % confidence interval 96.2 to 96.55) to distinguish two genera. This confirms our strain as a new genus (95.2 % 16S rRNA gene sequence similarity with *G. formicilis* X2-56[T], its closest phylogenetic neighbour). Kim *et al.* [37] also confirmed a taxonomic coherence between genomic and 16S rRNA gene sequence similarity for taxonomic demarcation of prokaryotes.

By comparison with reference strains of other closely related species (Table 1), strain AT2[T] differed in the combination
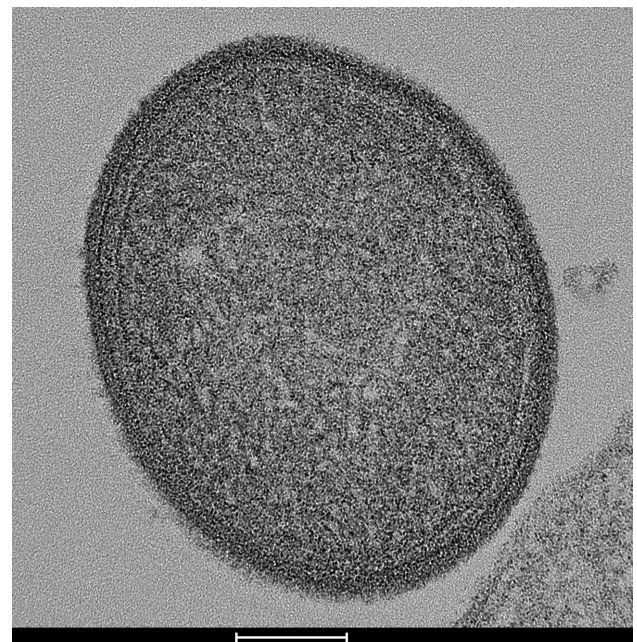
Fig. 2. Transmission electron micrograph of strain AT2[T], obtained using a Tecnai G20 (FEI company) at an operating voltage of 60kV. Bar, 100 nm.

142

of nitrate reductase activity (presence), use of L-arabinose (absence), and production of SCFA (acetic acid was the major SCFA produced, only a small amount of butyric acid was produced). Interestingly, strain AT2^T produced acetic acid while the reference strain of *F. prausnitzii* (ATCC 27768^T) consumed it. Moreover, MALDI-TOF analysis did not allow identifying previously known species. These phenotypic differences along with genomic and phylogenetic findings led us to propose that strain AT2^T (=CSUR P2014^T =DSM 100451^T) is the representative strain of a novel species of a new genus within the family *Ruminococcaceae* for which we propose the name *Fournierella massiliensis* gen. nov., sp. nov.

This bacterium was isolated from the faeces of a 28-year-old healthy French man living in Marseilles, France and may have a beneficial role in the gut through butyrate production. Butyrate is the preferred energy source for colonic epithelial cells and is thought to play an important role in maintaining colonic health in humans [38]. Moreover, the production of a significant amount of acetate promotes further the butyrate production in the gut since fifty percent of the butyrate-producing isolates are net acetate consumers during growth, probably because they employ the butyryl coenzyme A-acetyl coenzyme A1 transferase pathway for butyrate production [39].

DESCRIPTION OF *FOURNIERELLA* GEN. NOV.

Fournierella (Four.nier.el′la. N.L. fem. n. *Fournierella* named after the French clinical microbiologist Pierre-Edouard Fournier for his contribution to the taxono-genomic description of the bacteria).

Cells are Gram-negative-staining and the non-motile bacilli are 0.5 μm in diameter and 2 μm in length, and anaerobic. Optimal growth is observed at 37 °C and pH tolerance ranges from pH6–8.5. Cells do not produce catalase and oxidase.

The type species is *Fournierella massiliensis*.

DESCRIPTION OF *FOURNIERELLA MASSILIENSIS* SP. NOV.

Fournierella massiliensis (mas.si.li.en′sis. L. fem. adj. *massiliensis* of *Massilia*, the Latin name of Marseilles).

In addition to the characteristics given in the genus description, colonies grown on 5 % sheep blood-enriched Columbia agar are white with a diameter of 1 mm. Unable to produce indole. Using an API 50CH strip (BioMérieux), positive reactions are observed for glycerol, D-galactose, D-glucose, D-fructose, D-mannose, methyl α-D-glucopyranoside, aesculin ferric citrate, salicin, maltose, lactose, melibiose, sucrose, D-melezitose, rafinose, turanose and potassium 5-ketogluconate. Negative reactions are observed for erythritol, D-arabinose, L-arabinose, D-ribose, D-xylose, L-xylose, D-adonitol, methyl β-D-xylopyranoside, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl α-D-mannopyranoside, *N*-acetylglucosamine, amygdalin, arbutin, cellobiose, trehalose, inulin, starch, glycogen, xylitol, gentiobiose, D-xylose, D-tagatose, D-fucose, L-fucose, D-arabitol, L-arabitol, potassium gluconate and potassium 2-ketogluconate.

Using an API 20A strip (BioMérieux), positive reactions are observed for D-glucose, lactose, sucrose, D-mannose, aesculin ferric citrate, salicin, maltose, D-melezitose, glycerol and

Table 1. Differential characteristics of strain AT2^T compared with other cloesly related species

Strains: 1, AT2^T (data from this study); 2, *Gemmiger formicilis* X2-56^T (=ATCC 27749^T) [40]; 3, *Subdoligranulum variabile* BI 114^T (=DSM 15176^T) [41]; 4, *Faecalibacterium prausnitzii* ATCC 27768^T [6]; 5, *Anaerofilum pentosovorans* Fae^T (=DSM 7168^T) [42]; 6, *Anaerofilum agile* strain F^T (=DSM 4272^T) [42]. All strains were strict anaerobes. +, Positive; −, negative; NA, data not available; v, variable; w, weak.

	1	2	3	4	5	6
Cell diameter (μm)	0.5–2	0.3–1	0.6–2.5	0.5–0.9/2–14	0.2–0.6/3–6	0.2–0.6/3–6
Gram stain	−	v	−	−	v	v
Motility	−	−	−	−	v	+
Production of nitrate reductase	+	−	−	−	NA	NA
Utilization of:						
L-Arabinose	−	NA	NA	NA	+	+
D-Mannose	+	NA	+	NA	+	+
D-Mannitol	−	NA	−	NA	+	+
Maltose	+	+	+	W	+	+
Short chain fatty acid production						
Acetate	Major	Minor	Minor	Utilization	Major	Major
Formate	NA	Major	NA	Major	Major	Major
Lactate	NA	NA	Major	Major	Major	Major
Propionate	Minor	NA	NA	NA	−	−
Butyrate	Minor	Major	Major	Major	−	−
Isolation source	Human faeces	Human faeces	Human faeces	Human faeces	Industrial wastewater bioreactor	Sewage sludge

Table 2. Cellular fatty acid profile of strain AT2^T compared with other related species

Strains: 1, AT2^T (data from this study); 2, *Subdoligranulum variabile* BI 114^T[41]; 3, *Intestinimonas butyriciproducens* SRB-521–5-I^T[28]; 4, *Ethanoligenens harbinense* YUAN-3^T (=JCM 12961^T=CGMCC 1.5033^T); 5, *Hydrogenoanaerobacterium saccharovorans* SW512^T (=AS 1.5070^T=JCM 14861^T) [43]; 6, *Acetanaerobacterium elongatum* Z7^T (=JCM 12359^T=AS 1.5012^T) [44]. Values are % of total fatty acids. Only cellular fatty acid >1 % in strain AT2^T are included. NA, Not available.

Fatty acid	1	2	3	4	5	6
$C_{14:0}$	11.8	6.2	67.4	NA	15. 6	NA
$C_{15:0}$	1.0	<1	NA	NA	NA	NA
$C_{16:0}$	43.8	33.0	3.9	4.98	29.1	NA
$C_{18:0}$	10.0	11.6	3.7	NA	NA	NA
$C_{18:1n9}$	20.3	38.5	5.7	NA	NA	NA
$C_{18:2n6}$	1.2	NA	<1	NA	NA	NA

raffinose but no reaction is obtained for urease, D-mannitol, D-xylose, gelatin, D-cellulose, D-sorbitol, L-arabinose, L-rhamnose and trehalose. Using an API ZYM strip (Bio-Mérieux), exhibits esterase (C4), esterase lipase (C8), acid phosphatase, naphtol phosphohydrolase, α-galactosidase (melibase), α-glucosidase (maltase), β-glucosidase (cellulose), and *N*-acetyl-β-glucosaminidase (chitinase) activities. No alkaline phosphatase, lipase, leucine arylamidase, valine arylamidase, cystine arylamidase, trypsin, α-chymotrypsin, β-galactosidase, β-glucuronidase, α-mannosidase and α-fucosidase activities are observed. The major SCFA produced is acetic acid but butyric acid, isobutyric acid and propionic acid are also produced in small amounts. The predominant cellular fatty acids are the saturated acid $C_{16:0}$ and the unsaturated acid $C_{18:1n9}$.

The type strain is AT2^T (=CSUR P2014^T=DSM 100451^T), isolated from the faeces of a healthy 28-year-old French male. The DNA G+C content of the type strain is 56.8 % (genome sequence). The type strain is susceptible to amoxicillin, amoxicillin/clavulanic acid, cefalexin, ciprofloxacin, doxycycline, erythromycin, gentamicin, imipenem, metronidazole, nitrofurantoin, oxacillin, penicillin G, rifampicin, trimethoprim-sulfamethoxazole, tobramycin and vancomycin, but resistant to ceftriaxone.

References
1. **Lagier JC, Khelaifia S, Alou MT, Ndongo S, Dione N** *et al.* Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat Microbiol* 2016;1:16203.

2. **Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi M** *et al.* A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. *Int J Syst Evol Microbiol* 2014;64:384–391.

3. **Ramasamy D, Kokcha S, Lagier JC, Nguyen TT, Raoult D** *et al.* Genome sequence and description of *Aeromicrobium massiliense* sp. nov. *Stand Genomic Sci* 2012;7:246–257.

4. **Lagier JC, Elkarkouri K, Rivet R, Couderc C, Raoult D** *et al.* Non contiguous-finished genome sequence and description of *Senegalemassilia anaerobia* gen. nov., sp. nov. *Stand Genomic Sci* 2013;7: 343–356.

5. **Rainey FA.** Family VIII. *Ruminococcaceae* fam. nov. In: De Vos P, Garrity GM, Jones D, Krieg NR, Ludwig W *et al.* (editors). *Bergey's Manual of Systematic Bacteriology*, 2nd ed, vol. 3, (The Firmicutes). Dordrecht, Heidelberg, London, New York: Springer; 2009. pp. 1016.

6. **Duncan SH, Hold GL, Harmsen HJ, Stewart CS, Flint HJ.** Growth requirements and fermentation products of *Fusobacterium prausnitzii*, and a proposal to reclassify it as *Faecalibacterium prausnitzii* gen. nov., comb. nov. *Int J Syst Evol Microbiol* 2002;52:2141–2146.

7. **Khan MT, Duncan SH, Stams AJ, van Dijl JM, Flint HJ** *et al.* The gut anaerobe *Faecalibacterium prausnitzii* uses an extracellular electron shuttle to grow at oxic-anoxic interphases. *Isme J* 2012; 6:1578–1585.

8. **Sadaghian Sadabad M, von Martels JZ, Khan MT, Blokzijl T, Paglia G** *et al.* A simple coculture system shows mutualism between anaerobic *Faecalibacteria* and epithelial Caco-2 cells. *Sci Rep* 2015;5:17906.

9. **Million M, Diallo A, Raoult D.** Gut microbiota and malnutrition. *Microb Pathog* 2016:30212–30216.

10. **Million M, Tidjani Alou M, Khelaifia S, Bachar D, Lagier JC** *et al.* Increased gut redox and depletion of anaerobic and methanogenic prokaryotes in severe acute malnutrition. *Sci Rep* 2016;6:26051.

11. **Hornef MW, Pabst O.** Real friends: *Faecalibacterium prausnitzii* supports mucosal immune homeostasis. *Gut* 2016;65:365–367.

12. **Miquel S, Martín R, Rossi O, Bermúdez-Humarán LG, Chatel JM** *et al.* *Faecalibacterium prausnitzii* and human intestinal health. *Curr Opin Microbiol* 2013;16:255–261.

13. **Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE** *et al.* Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin Infect Dis* 2009;49:543–551.

14. **Drancourt M, Bollet C, Carlioz A, Martelin R, Gayral JP** *et al.* 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. *J Clin Microbiol* 2000;38:3623–3630.

15. **Meier-Kolthoff JP, Göker M, Spröer C, Klenk HP.** When should a DDH experiment be mandatory in microbial taxonomy? *Arch Microbiol* 2013;195:413–418.

16. **Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M.** Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 2013;14:60.

17. **Meier-Kolthoff JP, Klenk HP, Göker M.** Taxonomic use of DNA G +C content and DNA-DNA hybridization in the genomic age. *Int J Syst Evol Microbiol* 2014;64:352–356.

18. **Edgar RC.** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792–1797.

19. **Stamatakis A.** RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30: 1312–1313.

20. **Pattengale ND, Alipour M, Bininda-Emonds OR, Moret BM, Stamatakis A.** How many bootstrap replicates are necessary? *J Comput Biol* 2010;17:337–354.

21. **Swofford DL.** *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4.0 B10. Sunderland: Sinauer Associates; 2002.

22. Halebian S, Harris B, Finegold SM, Rolfe RD. Rapid method that aids in distinguishing Gram-positive from Gram-negative anaerobic bacteria. *J Clin Microbiol* 1981;13:444–448.

23. Powers EM. Efficacy of the Ryu nonstaining KOH technique for rapidly determining Gram reactions of food-borne and waterborne bacteria and yeasts. *Appl Environ Microbiol* 1995;61:3756–3758.

24. Johnson MJ, Thatcher E, Cox ME. Techniques for controlling variability in Gram staining of obligate anaerobes. *J Clin Microbiol* 1995;33:755–758.

25. Le Page S, van Belkum A, Fulchiron C, Huguet R, Raoult D *et al.* Evaluation of the PREVI® isola automated seeder system compared to reference manual inoculation for antibiotic susceptibility testing by the disk diffusion method. *Eur J Clin Microbiol Infect Dis* 2015;34:1859–1869.

26. Sasser M. *Bacterial Identification by Gas Chromatographic Analysis of Fatty Acids Methyl Esters (GC-FAME)*, Technical Note 101. Newark, DE: MIDI Inc; 2006.

27. Dione N, Sankar SA, Lagier JC, Khelaifia S, Michele C *et al.* Genome sequence and description of *Anaerosalibacter massiliensis* sp. nov. *New Microbes New Infect* 2016;10:66–76.

28. Kläring K, Hanske L, Bui N, Charrier C, Blaut M *et al.* *Intestinimonas butyriciproducens* gen. nov., sp. nov., a butyrate-producing bacterium from the mouse intestine. *Int J Syst Evol Microbiol* 2013;63:4606–4612.

29. Pfeiffer N, Desmarchelier C, Blaut M, Daniel H, Haller D *et al.* *Acetatifactor muris* gen. nov., sp. nov., a novel bacterium isolated from the intestine of an obese mouse. *Arch Microbiol* 2012;194:901–907.

30. Zhao G, Nyman M, Jönsson JA. Rapid determination of short-chain fatty acids in colonic contents and faeces of humans and rats by acidified water-extraction and direct-injection gas chromatography. *Biomed Chromatogr* 2006;20:674–682.

31. Illumina. 2012. Illumina data processing of nextera® mate pair reads on illumina sequencing platforms. 2012. www.illumina.com/documents/products/technotes/technote_nextera_matepair_data_processing.pdf. [accessed 05 January 2017].

32. Nomenclatural changes validly published in the 2002 issues of the IJSEM. www.bacterio.net/-changes-2002.html.

33. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M *et al.* ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother* 2014;58:212–220.

34. Conway KR, Boddy CN. ClusterMine360: a database of microbial PKS/NRPS biosynthesis. *Nucleic Acids Res* 2013;41:D402–D407.

35. Qin QL, Xie BB, Zhang XY, Chen XL, Zhou BC *et al.* A proposed genus boundary for the prokaryotes based on genomic insights. *J Bacteriol* 2014;196:2210–2215.

36. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W *et al.* Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* 2014; 12:635–645.

37. Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol* 2014;64:346–351.

38. Barcenilla A, Pryde SE, Martin JC, Duncan SH, Stewart CS *et al.* Phylogenetic relationships of butyrate-producing bacteria from the human gut. *Appl Environ Microbiol* 2000;66:1654–1661.

39. Duncan SH, Barcenilla A, Stewart CS, Pryde SE, Flint HJ. Acetate utilization and butyryl coenzyme a (CoA):acetate-CoA transferase in butyrate-producing bacteria from the human large intestine. *Appl Environ Microbiol* 2002;68:5186–5190.

40. Gossling J, Moore WEC. *Gemmiger formicilis*, n.gen., n.sp., an anaerobic budding bacterium from intestines. *Int J Syst Bacteriol* 1975;25:202–207.

41. Holmstrøm K, Collins MD, Møller T, Falsen E, Lawson PA. *Subdoligranulum variabile* gen. nov., sp. nov. from human feces. *Anaerobe* 2004;10:197–203.

42. Zellner G, Stackebrandt E, Nagel D, Messner P, Weiss N *et al.* *Anaerofilum pentosovorans* gen. nov., sp. nov., and *Anaerofilum agile* sp. nov., two new, strictly anaerobic, mesophilic, acidogenic bacteria from anaerobic bioreactors. *Int J Syst Bacteriol* 1996;46: 871–875.

43. Song L, Dong X. *Hydrogenoanaerobacterium saccharovorans* gen. nov., sp. nov., isolated from H2-producing UASB granules. *Int J Syst Evol Microbiol* 2009;59:295–299.

44. Chen S, Dong X. *Acetanaerobacterium elongatum* gen. nov., sp. nov., from paper mill waste water. *Int J Syst Evol Microbiol* 2004; 54:2257–2262.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.

Article X:

Anaerotruncus massiliensis **sp. nov., a succinate-producing bacterium isolated from human stool from an obese patient after bariatric surgery.**

Amadou Hamidou Togo, Awa Diop, Grégory Dubourg, Saber Khelaifia, Magali Richez, Nicholas Armstrong, Marie Maraninchi, Pierre-Edouard Fournier, Didier Raoult and Matthieu Million.

Anaerotruncus massiliensis sp. nov., a succinate-producing bacterium isolated from human stool from an obese patient after bariatric surgery

A. H. Togo[1,2], A. Diop[1,3], G. Dubourg[1,2], S. Khelaifia[1,2], M. Richez[1,2], N. Armstrong[1,2], M. Maraninchi[4], P.-E. Fournier[3], D. Raoult[1,2] and M. Million[1,2]

1) IHU-Méditerranée Infection, 2) Aix Marseille Univ, IRD, APHM, MEPHI, Marseille, 3) Aix Marseille Université, IRD, VITROME, and 4) Aix-Marseille Université, NORT 'Nutrition, Obesity and Risk of Thrombosis,' INSERM1062, INRA1260, Marseille, France

Abstract

A new bacterium, strain AT3[T], was isolated by microbial culturomics from a faecal sample from a Frenchman after bariatric surgery. The isolate exhibited 96.6% 16S ribosomal RNA gene nucleotide sequence similarity with *Anaerotruncus colihominis* strain WAL 14565[T] = CCUG 45055[T] = CIP 107754[T]. Phenotypic and genomic characteristics showed that the new strain represents a novel species, for which the name *Anaerotruncus massiliensis* sp. nov. is proposed. The type strain is strain AT3[T] = CSUR P2007[T] = DSM 100567[T].
© 2019 The Authors. Published by Elsevier Ltd.

Corresponding author: M. Million, Aix-Marseille Université, IRD198, APHM, MEPHI, IHU-Méditerranée Infection, 19-21 Boulevard Jean Moulin, 13385, Marseille cedex 05, France.
E-mail: matthieumillion@gmail.com

Introduction

Obesity is a major public health problem [1,2]. It increases the risk of metabolic diseases such as type 2 diabetes and cardiovascular diseases such as high blood pressure [3]. The global obesity rate has been steadily increasing since 1980 [4]. The influence of the gut microbiota in human health and disease has been revealed in the recent years. Recently obesity has been associated with gut microbiota dysbiosis [5–7].

The various treatment regimens proposed to treat obesity are mainly dietary measures as well as medical and surgical treatments. Bariatric surgery has been documented as one of the most effective treatments for obesity and has been associated with increased microbial diversity [8,9]. We studied stool samples from obese patients before and after bariatric surgery using the microbial culturomics approach [10]. During this study, a novel Gram-negative anaerobic bacterium was isolated from a stool sample collected after bariatric surgery from an obese man.

Here we characterize a novel bacterial species, strain AT3, belonging to *Anaerotruncus* genus [11], together with the genome sequencing assembly, annotation and comparison. At the time we described this new species, the genus *Anaerotruncus* only contained one species, the type species *Anaerotruncus colihominis* strain WAL 14565[T] = CCUG 45055[T] = CIP 107754[T] [11].

Methods

The stool sample was collected from a 47-year-old obese Frenchman with a body mass index of 35.3 kg/m^2 in November 2011. Written informed consent was obtained from the patient at the nutrition, metabolic disease and endocrinology service at La Timone Hospital, Marseille, France. The study and the assent procedure were approved by the local ethics committee of IFR 48 (approval 09-022, 2010). Stool samples were stored at −80°C after collection until use.

Growth of strain AT3 was performed in May 2015. Sterile stool extract was preincubated in a blood culture bottle

enriched with rumen fluid and sheep's blood, then seeded on 5% sheep's blood Columbia agar (bioMérieux, Marcy l'Etoile, France) as described elsewhere [12]. Strain purification and identification were performed as previously described [12,13].

The 16S ribosomal RNA (rRNA) gene sequencing of the strain AT3 was performed as previously reported [14] using the fD1-rP2 primers, a GeneAmp PCR System 2720 thermal cycler (Applied Biosystems; Thermo Fisher Scientific, Waltham, MA, USA) and an ABI Prism 3130-XL capillary sequencer (Applied Biosciences, Saint Aubin, France). For taxonomic assignment, CodonCode Aligner (CodonCode, Centerville, MA, USA) software was used to correct sequences, and BLASTn searches were performed by the National Center for Biotechnology Information (NCBI) web server (http://blast.ncbi.nlm.nih.gov.gate1.inist.fr/Blast.cgi) for neighbour search. Sequences were aligned using CLUSTAL W [15] and phylogenetic inferences were obtained using the maximum likelihood method [16] by MEGA 7 software [17].

The colonies' morphology and pigmentation were observed after cultivation of the strain on a 5% sheep's blood Columbia agar plate (bioMérieux) at 37°C in anaerobic condition for 48 hours. Phenotypic characteristic analysis of the strain was performed as previously described [12,18]. The growth of the strain for temperature ranged from 25 to 55°C when assessed on sheep's blood–enriched Columbia agar (bioMérieux) under different atmospheric conditions. The pH growth of strain AT3 was assessed at various pH range from 6 to 8.5 under anaerobic condition at 37°C. The saltiness of the strain using various concentrations (5, 10, 50, 75 and 100 g/L) of NaCl on Schaedler agar with 5% sheep's blood (bioMérieux) under an anaerobic atmosphere at 37°C was also tested.

Biochemical assays using API Gallery strips—API ZYM, API 20A, API 50 CH and rapid ID 32 A (bioMérieux)—were performed according the manufacturer's instruction.

Cellular fatty acid methyl ester and carbohydrate metabolism end product of strain AT3 were assessed as previously reported [12]. Antimicrobial susceptibility of the strain was also tested according the European Committee on Antimicrobial Susceptibility Testing (EUCAST) recommendation [19,20] using 12 antibiotics.

The antimicrobial agent susceptibility of the strain was tested according the EUCAST recommendations [19,20] using 12 antibiotics, including Etest strips for amikacin (0.016–256 μg/mL), vancomycin (0.016–256 μg/mL), imipenem (0.002–32 μg/mL), ceftriaxone (0.016–256 μg/mL), rifampicin (0.002–32 μg/mL), benzyl penicillin (0.002–32 μg/mL), amoxicillin (0.016–256 μg/mL), cefotaxime (0.002–32 μg/mL), metronidazole (0.016–256 μg/mL), minocycline (0.016–256 μg/mL), teicoplanin (0.016–256 μg/mL), erythromycin (0.016–256 μg/mL) and daptomycin (0.016–256 μg/mL) (bioMérieux).

Genomic DNA of strain AT3 was sequenced on MiSeq Technology (Illumina, San Diego, CA, USA) with the mate-pair strategy and assembly deployed as previously described [18,21]. The genome annotation of strain AT3 was performed as previously described [12] and compared to the genomes of his relatives species. For the comparison, genomes of closest species were automatically retrieved from the 16S rRNA tree and thus, *Anaerotruncus colihominis* strain WAL 14565^T (= CCUG 45055^T = CIP 107754^T) (ABGD00000000) [11], *Hydrogenoanaerobacterium saccharovorans* strain SW512^T (= AS 1.5070^T = JCM 14861^T) (FOCG00000000) [22], *Ruthenibacterium lactatiformans* strain 585−1^T (= DSM 100348^T = VKM B-2901^T) (JXXK00000000) [23], *Acutalibacter muris* strain KB18^T (= DSM 26090^T = KCTC 15540^T) (CP021422) [24], *Acetanaerobacterium elongatum* strain Z7^T (= JCM 12359^T = AS 1.5012^T) (FNID00000000) [25], *Ruminococcus flavefaciens* strain C94^T (= ATCC 19208^T) (JAEF00000000) [26] and *Ruminococcus champanellensis* strain 18P13^T (= DSM 18848^T = JCM 17042^T) (FP929052) [27] were selected. For each selected genome, complete genome sequence, proteome genome sequence and Orfeome genome sequence were retrieved from NCBI's FTP. All proteomes were analysed with proteinOrtho [28]. Then for a couple of genomes, a similarity score was computed. This score was the mean value of nucleotide similarity of average genomic identity of orthologous gene sequences between the two genomes studied [29]. An annotation of the entire proteome was performed to define the distribution of functional classes of predicted genes according to the clusters of orthologous groups of proteins (using the same method as for the genome annotation). The genome of strain AT3 was locally aligned 2 by 2 using the BLAT algorithm [30,31] against each the selected genomes previously cited. Digital DNA-DNA hybridization values were estimated from a generalized model [32]. Average amino acid identity was also calculated on the basis of the overall similarity between two genomic data sets of the online proteins (http://enve-omics.ce.gatech.edu/aai/index) [33,34].

Results

The spectrum generated from clean isolate spots was unable to match those in Bruker's and our own database, so the strain's 16S RNA gene was sequenced. BLASTn searches performed using the 16S rRNA gene sequences showed that strain AT3 exhibited 96.6% sequence similarity with *A. colihominis* strain WAL 14565^T = CCUG 45055^T = CIP 107754^T [11], classified in the *Ruminococcaceae* family created by Rainey [35]. A maximum likelihood phylogenetic tree based on 16S rRNA gene sequences showed relationships between strain AT3 and

some related taxa (Fig. 1). The numbers of nodes correspond to the percentages of bootstrap values obtained by repeating the analysis to 1000 times to generate a consensus tree.

This new isolate was a Gram-negative−staining, nonmotile and rod-shaped bacterium. The strain was catalase positive but oxidase negative. The growth of the strain occurred between 28 to 45°C, but optimal growth was observed at 37°C after 72 hours' incubation in anaerobic atmosphere on Columbia agar with 5% sheep's blood (bioMérieux). No growth was observed in aerobic atmosphere. No growth of this bacterium was observed using 10 g/L of NaCl concentration on Schaedler agar (bioMérieux). The pH for growth ranged from 6 to 8.5, but the optimal pH was 7.5. The colonies were opalescent, with a mean diameter of 0.5 to 1.5 mm. Haemolysis was not observed on Columbia agar plate (bioMérieux). Cells had a diameter ranging from 0.3 to 2.9 μm by electron microscopy (Fig. 2).

Using the API 50 CH strip, positive reactions were observed with glycerol, erythritol, adonitol, methyl-βD-xylopyranoside, galactose, glucose, mannose, sorbitol, esculin, trehalose, inulin, glycogen, gentiobiose, tagatose and potassium 5−cetogluconate and starch, whereas arabinose, ribose, xylose, fructose, sorbose, rhamnose, dulcitol, inositol, mannitol, methyl-αD-mannopyranoside, methyl-αD-glucopyranoside, *N*-acetylglucosamine, amygdaline, arbutine, salicin, cellobiose, maltose, lactose, melibiose, saccharose, melezitose, raffinose, sylitol, turanose, lyxose, fucose, arabitol and potassium gluconate were negative. Using the API 20A strip, glucose, maltose, mannose, trehalose and rhamnose are fermented, while mannitol, lactose, saccharose, salicin, xylose, arabinose,

glycerol, cellobiose, melezitose, raffinose, sorbitol and rhamnose are not. Aesculin and gelatine are hydrolyzed while urease and indole are not produced. Using the rapid ID 32 A strip, positive reactions were observed with alkaline phosphatase, arginine arylamidase, leucyl glycine arylamidase, phenylalanine arylamidase, leucine arylamidase, pyroglutamic acid arylamidase, tyrosine arylamidase, alanine arylamidase, glycine arylamidase, histidine arylamidase, glutamyl glutamic acid arylamidase, serine arylamidase and arginine dihydrolase. Mannose and raffinose are not fermented. Urease is not produced, while and indole are produced and nitrate is reduced. Using the API ZYM strip, positive reactions were observed with alkaline phosphatase, esterase (C-4), esterase lipase (C-8), acid phosphatase, naphthol-AS-BI-phosphohydrolase and α-glucosidase, while
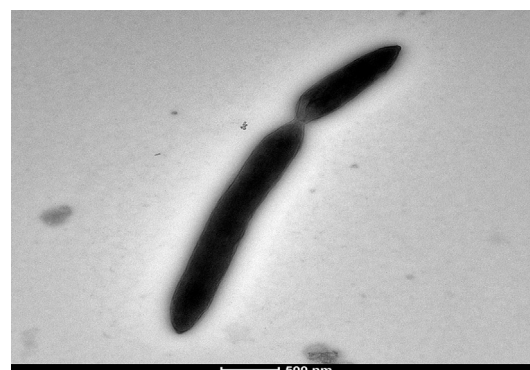
FIG. 2. Transmission electron microscopy of strain AT3 using a Tecnai G20 (FEI) at operating voltage of 60 kV. Scale bar, 500 μm.

FIG. 1. 16S rRNA-based maximum likelihood (ML) phylogenetic tree highlighting position of strain AT3 with closely related species. Corresponding GenBank accession numbers for 16S rRNA genes of each strain are indicated. Sequences were aligned using CLUSTAL W with default parameters and phylogenies were inferred by MEGA 7 software. Branches are scaled in terms of expected number of substitutions per site. Numbers above branches are support values when higher than 75% from ML bootstrapping. Scale bar represents a nucleotide sequence divergence of 2%.
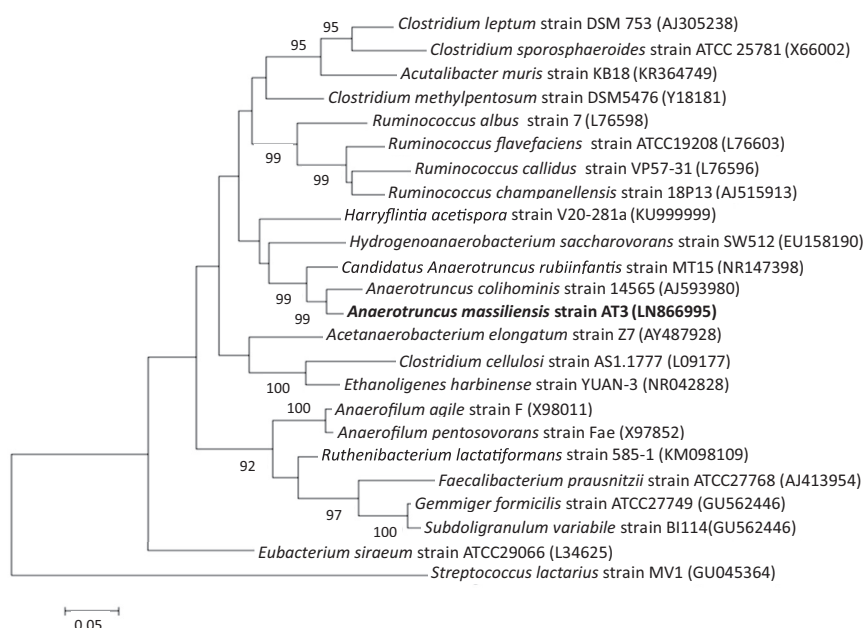
TABLE 1. Differential characteristics of strain AT3 compared to its closest species

Properties	1	2	3	4	5	6	7	8
Cells size (µm)	0.3/2.9	0.5/2−5	0.3−0.4/2.0−14.5	1.6/0.4	—	0.2−0.4/4.0−8.0	0.8/0.9	0.9/1.3
Gram stain	—	+	_	_	—	+	+	+
Motility	—	—	_	_	—	+	—	—
Spore forming	—	—	—	_	+	—	—	—
Aesculin hydrolysis	+	—	_	+	+	+	—	+
Gelatine hydrolysis	+	—	_	—	w	+	—	—
Urease	—	—	—	—	—	—	NA	—
Indole	+	+	—	—	—	+	—	—
Catalase	+	—	—	—	—	—	—	—
Acid from:								
Arabinose	—	—	+	_	—	+	+	—
Cellobiose	—	+	—	—	—	+	+	+
Glucose	+	+	+	+	—	+	—	—
Lactose	—	—	+	—	—	—	+	—
Maltose	v	+	NA	v	—	+	—	—
Mannose	v	+	+	—	—	—	—	—
Rafinose	—	—	+	—	—	+	—	—
Sucrose	—	—	+	v	NA	+	—	—
Salicin	—	—	_	v	—	+	—	—
Starch	+	—	—	+	—	—	NA	—
Trehalose	+	—	+	_	—	—	—	—
Alkaline phosphatase	+	—	NA	—	+	w	NA	+
Major end product of carbohydrate metabolism	A, B, Ib, Ip, Ih	A, B	A, E	L, S, A	NA	A, E	A, F, E	A, S
G+C content (%)	63.7	54	41.9	49.6	54.6	50.4	39−44	53.05
Isolated from	Human faeces	Human faeces	UASB reactor	Human faeces	Human faeces	Human faeces	Human faeces	Human faeces

+, positive result; −, negative result; v, variable result; w, weakly positive result; NA, data not available; UASB, upflow anaerobic sludge blanket; A, acetic acid; B, butyric acid; ib, isobutyric acid; Ip, isopentanoic acid; Ih, isohexanoic acid; L, lactic acid; E, ethanol; F, formic; S, succinic.
1, Strain AT3; 2, *Anaerotruncus colihominis*; 3, *Hydrogenoanaerobacterium saccharovorans*; 4, *Ruthenibacterium lactatiformans*; 5, *Acutalibacter muris*; 6, *Acetanaerobacterium elongatum*; 7, *Ruminococcus flavefaciens*, 8, *Ruminococcus champanellensis*. Data from [Lawson *et al.* 2004], [Song *et al.* 2004], [Shkoporov *et al.* 2016], [Lagkouvardos *et al.* 2016], [Chen *et al.* 2004] and [Chassard *et al.* 2012].

lipase, leucine arylamidase, valine arylamidase, cystine arylamidase, trypsin, α-chymotrypsin, α-galactosidase, β-galactosidase, β-glucuronidase, β-glucosidase, *N*-acetyl-β-glucosaminidase, α-mannosidase, α-mannosidase and α-fructosidase were negative.

The major cellular fatty of the new strain were the anteiso-$C_{15:0}$ (29%), $C_{16:0}$ (20.6%) and $C_{14:0}$ (19.5%). The different characteristics of strain AT3^{T} compared to its closet species are listed in Table 1. The composition of cellular fatty acids for this isolate is shown in Table 2. The end products of carbohydrate metabolism of the strain AT3 after 72 hours' incubation in Wilkins Chalgren anaerobe broth (MilliporeSigma, St Louis, MO, USA) were acetic acid (12.3 ± 0.5 mM), butanoic acid (7.6 ± 2.0 mM), isobutanoic acid (1.4 ± 0.1 mM), isopentanoic acid (1.0 ± 0.2 mM) and isohexanoic acids (1.0 ± 0.1 mM).

Antibiotic susceptibility tested according the EUCAST recommendation gave the following MIC results: >256, 0.125, 32, >256, 0.006, 0.016, 0.19, 0.016, 1.5, 0.125, 256 and 0.5 µg/mL, respectively, for amikacin, vancomycin, imipenem, ceftriaxone, rifampicin, benzyl penicillin, amoxicillin, metronidazole, minocycline, teicoplanin, erythromycin and daptomycin.

The draft genome of strain AT3 is 3 145 950 bp long with 63.6% G+C content (Fig. 3). It is composed of seven scaffolds (eight contigs). Within the 2927 predicted genes, 2868 were protein-coding genes and 59 were RNAs genes (four 5S rRNA genes, two 16S rRNA genes, two 23S rRNA genes and 51 tRNA genes). A total of 2106 genes (74.4%) were assigned to a putative function by COGs or by nonredundant blast. There

were 398 genes (13.9%) associated with peptide signal proteins. ARG-ANNOT (Antibiotic Resistance Gene−Annotation) analysis found two genes (0.07%) associated with antimicrobial agent resistance [36] and four genes (0.14%) associated with polyketide synthase or nonribosomal peptide synthetases [37]. Using the online RAST Server version 2.0 tool, 20 genes were found to be associated with antibiotic resistance (four genes for fluoroquinolone resistance, four for tetracycline resistance and 12 for multidrug resistance efflux pumps), nine genes associated with invasion and intracellular resistance, five to copper homeostasis, nine to cobalt−zinc−cadmium resistance, two to mercuric reductase, one to mercury resistance operon, one to copper homeostasis:copper tolerance and one to cadmium resistance. No gene associated with virulence, toxins or bacteriocins has been found [38]. There were 193 genes (6.7%) identified as ORFans. The remaining 481 (16.8%) were

TABLE 2. Cellular fatty acid composition of strain AT3

Fatty acid	1	2	3	4	5	6
Anteiso-$C_{15:0}$	29.0	NA	<1	9.8	6.04	19.56
$C_{16:0}$	20.6	29.1	6.6	22.1	NA	NA
$C_{14:0}$	19.5	15.6	2.2	NR	8.03	NA
$C_{18:1n9}$	5.6	NA	31.9	12.2	NR	NA
Iso-$C_{14:0}$	5.6	NA	NA	NR	32.11	NA
Iso-$C_{15:0}$	5.5	11.6	NA	16.8	42.83	26.57

NA, no data available. Values are percentages of total fatty acids.
1, Strain AT3; 2, *Hydrogenoanaerobacterium saccharovorans*; 3, *Ruthenibacterium lactatiformans*; 4, *Acutalibacter muris*; 5, *Acetanaerobacterium elongatum*; 6, *Ruminococcus champanellensis*. Data from [Song *et al.* 2004], [Shkoporov *et al.* 2016], [Lagkouvardos *et al.* 2016], [Chen *et al.* 2004] and [Chassard *et al.* 2012].
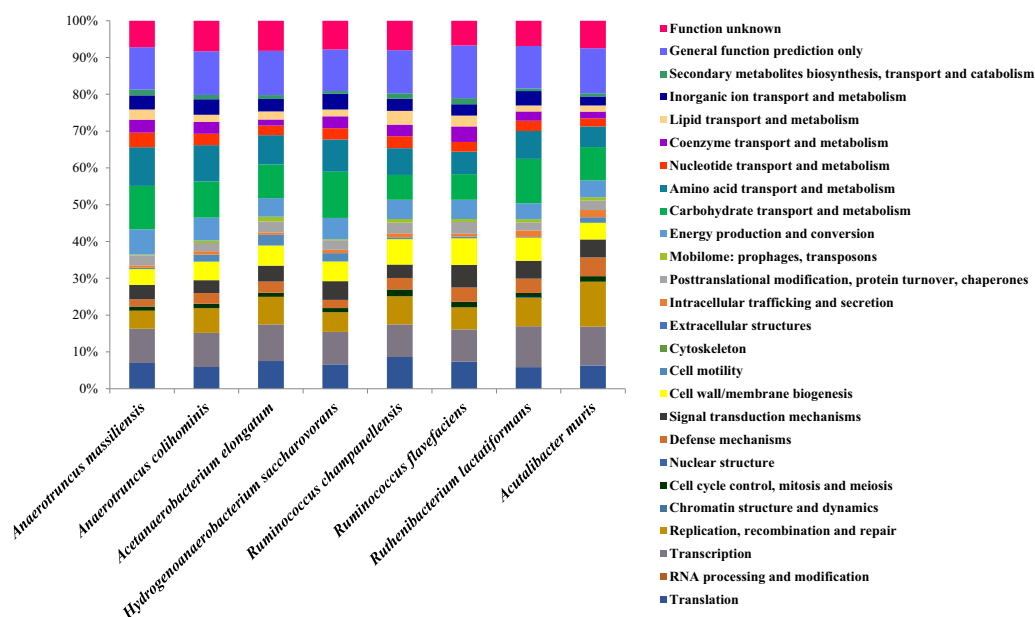
FIG. 3. Graphical circular map of genome of strain AT3. From outside to centre: Genes on forward strand coloured by COGs categories (only genes assigned to COGs), genes on reverse strand coloured by COGs categories (only gene assigned to COGs), RNA genes (tRNAs green, rRNAs red), GC content and GC skew. COGs, Clusters of Orthologous Groups database; rRNA, ribosomal RNA; tRNA, transfer RNA.

annotated as hypothetical proteins. The distribution of genes into COGs functional categories of strain AT3 with its nearest neighbours' species is presented in Table 3 and Fig. 4. The numbers of orthologous proteins shared between genomes is shown in Table 4. The genome of the strain was deposited in EMBL-EBI (European Molecular Biology Laboratory–European

Bioinformatics Institute) under accession number OEJM0000000. The digital DNA-DNA hybridization values of strain AT3 compared to its closely related species ranged from 21.5% with *A. muris* to 33.7% with *R. champanellensis* and are shown in Table 5. The average amino acid identity between the strain AT3 and closely related species was evaluated (Table 6).

TABLE 3. Number of genes associated with 26 general COGs functional categories of strain AT3ᵀ

Code	1	2	3	4	5	6	7	8	Description
J	151	141	144	149	153	148	151	134	Translation
A	0	0	0	0	0	0	0	0	RNA processing and modification
K	201	222	191	282	256	195	181	138	Transcription
L	106	161	116	201	294	147	124	119	Replication, recombination and repair
B	0	0	0	1	0	0	0	0	Chromatin structure and dynamics
D	24	29	25	28	36	22	31	29	Cell cycle control, mitosis and meiosis
Y	0	0	0	0	0	0	0	0	Nuclear structure
v	42	69	48	98	123	60	80	49	Defence mechanisms
T	85	84	111	124	119	85	127	57	Signal transduction mechanisms
M	93	121	116	160	109	107	149	107	Cell wall/membrane biogenesis
N	8	45	49	5	37	57	9	8	Cell motility
Z	0	0	0	0	0	0	0	0	Cytoskeleton
w	0	0	0	0	0	0	0	0	Extracellular structures
U	14	23	20	44	47	13	17	17	Intracellular trafficking and secretion
O	58	53	56	56	62	59	66	45	Posttranslational modification, protein turnover, chaperones
X	5	16	6	22	22	25	16	15	Mobilome: prophages, transposons
C	148	149	126	109	111	99	108	82	Energy production and conversion
G	255	236	272	307	219	179	144	105	Carbohydrate transport and metabolism
E	227	236	191	193	135	157	126	113	Amino acid transport and metabolism
F	85	74	66	71	53	51	54	51	Nucleotide transport and metabolism
H	77	78	71	64	45	33	87	49	Coenzyme transport and metabolism
I	59	46	40	41	40	43	61	58	Lipid transport and metabolism
P	83	101	93	99	59	66	64	51	Inorganic ion transport and metabolism
Q	33	28	16	19	21	21	34	24	Secondary metabolites biosynthesis, transport and catabolism
R	249	284	245	292	296	236	295	182	General function prediction only
S	156	199	170	175	181	161	139	125	Function unknown

COGs, Clusters of Orthologous Groups database.
1, Strain AT3; 2, *Anaerotruncus colihominis*; 3, *Hydrogenoanaerobacterium saccharovorans*; 4, *Ruthenibacterium lactatiformans*; 5, *Acutalibacter muris*; 6, *Acetanaerobacterium elongatum*; 7, *Ruminococcus flavefaciens*; 8, *Ruminococcus champanellensis*.
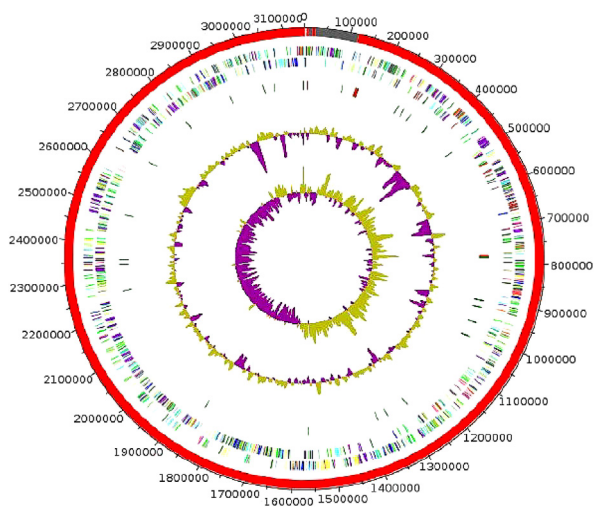
FIG. 4. Graphical circular map of chromosome. From outside to centre: Genes on forward strand coloured by COGs categories (only genes assigned to COGs), genes on reverse strand coloured by COGs categories (only gene assigned to COGs), RNA genes (tRNAs green, rRNAs red), GC content and GC skew. COGs, Clusters of Orthologous Groups database; rRNA, ribosomal RNA; tRNA, transfer RNA.

TABLE 4. Numbers of orthologous proteins shared between genomes (upper right), average percentage similarity of nucleotides corresponding to orthologous protein shared between genomes (lower left) and numbers of proteins per genome (bold)

	1	2	3	4	5	6	7	8
1	**2868**	1240	1158	965	843	978	813	808
2	73.7%	**3656**	1183	6433	3823	987	806	823
3	61.3%	63.0%	**2932**	954	855	1028	814	818
4	65.7%	64.3%	60.7%	**3908**	818	847	847	758
5	64.5%	62.5%	59.8%	63.8%	**3823**	831	756	764
6	64.2%	63.9%	64.8%	63.3%	62.5%	**2746**	815	816
7	62.4%	60.8%	62.4%	59.9%	60.5%	62.3%	**3089**	843
8	64.4%	63.5%	62.1%	63.8%	63.5%	63.7%	66.0%	**2356**

1, Strain AT3; 2, *Anaerotruncus colihominis*; 3, *Hydrogenoanaerobacterium saccharovorans*; 4, *Ruthenibacterium lactatiformans*; 5, *Acutalibacter muris*; 6, *Acetanaerobacterium elongatum*; 7, *Ruminococcus flavefaciens*; 8, *Ruminococcus champanellensis*.

These values ranged from 47% with *R. flavefaciens* to 68.5% with *A. colihominis*.

Discussion

When we compared the strain AT3 with the closest relative species differed by the catalase production, we found that his genome G+C content (63.7 vs. 41.9 to 54.6 for *A. colihominis*, *H. saccharovorans*, *R. lactatiformans*, *A. muris*, *A. elongatum*, *R. flavefaciens* and *R. champanellensis*, its closest relative species). The strain AT3 strain is phylogenetically close to *A. colihominis* but differs from it in Gram staining; *A. colihominis* is Gram staining positive while the new strain is Gram staining negative. The anteiso-$C_{15:0}$ (29%) is one of the major cellular fatty acid for the new strain, but it is absent from *A. colihominis*.

On the basis of the phenotypic, chemotaxonomic, genomic and phylogenetic characteristics, a novel bacterium isolated from the stool sample of a morbidly obese Frenchman after bariatric surgery, termed *Anaerotruncus massiliensis* sp. nov., is proposed. The type strain is strain AT3[T] = CSURP P2007[T] = DSM 100567[T]. A preliminary report of main characteristics was previously published as a new species announcement [39], but the present study reports the full characterization of this putative new species.

Description of *Anaerotruncus massiliensis* sp. nov.

Anaerotruncus massiliensis (mas.si.li.en'sis, L. fem. adj., *massiliensis*, 'of Massilia,' the Latin name of Marseille) is a Gram-negative-staining, catalase positive, oxidase negative, nonmotile non−spore-forming and anaerobic rod 0.3 μm wide and 2.9 μm long. Sheep's blood Columbia agar−grown colonies were opalescent with a mean diameter 0.5 to 1.5 mm after 72 hours'

TABLE 5. Pairwise comparison of strain AT3 with other species using GGDC, formula 2 (DDH estimates based on identities/HSP length)[a]

	1	2	3	4	5	6	7	8
1	100	26.0% ± 2.4	26.4% ± 2.4	27.1% ± 2.4	21.5% ± 2.4	22.1% ± 2.4	22.6% ± 2.4	33.7% ± 2.4
2		100	27.3% ± 2.4	24.8% ± 2.4	21.1% ± 2.4	20.6% ± 2.4	36.0% ± 2.5	35.9% ± 2.4
3			100	32.6% ± 2.45	34.3% ± 2.4	19.4% ± 2.4	28.0% ± 2.5	35.5% ± 2.5
4				100	19.1% ± 2.3	27.1% ± 2.4	27.7% ± 2.5	34.1% ± 2.5
5					100	22.0% ± 2.4	34.5% ± 2.5	22.0% ± 2.4
6						100	34.1% ± 2.5	33.7% ± 2.5
7							100	21.1% ± 2.4
8								100

DDH, DNA-DNA hybridization; GGDC, Genome-to-Genome Distance Calculator; HSP, high-scoring segment pairs; rRNA, ribosomal RNA.
1, Strain AT3; 2, *Anaerotruncus colihominis*; 3, *Hydrogenoanaerobacterium saccharovorans*; 4, *Ruthenibacterium lactatiformans*; 5, *Acutalibacter muris*; 6, *Acetanaerobacterium elongatum*; 7, *Ruminococcus flavefaciens*; 8, *Ruminococcus champanellensis*.
[a]Confidence intervals indicate inherent uncertainty in estimating DDH values from intergenomic distances based on models derived from empirical test data sets. These results are consistent with 16S rRNA and phylogenomic analyses as well as GGDC results.

TABLE 6. Average amino acid identity values of strain AT3 compared to its phylogenetically close neighbours

	1	2	3	4	5	6	7	8
1	100	68.5% ± 18.5	54.0% ± 16.4	49.9% ± 15.6	48.1% ± 15.2	51.8% ± 16.5	47.0% ± 15.5	48.7% ± 15.6
2		100	52.9% ± 16.4	50.1% ± 16.7	46.9% ± 15.3	50.3% ± 16.2	45.9% ± 15.5	48.1% ± 15.5
3			100	48.7% ± 15.6	46.4% ± 15.4	53.2% ± 16.6	47.3% ± 16.1	48.8% ± 16.1
4				100	47.2% ± 15.3	47.4% ± 15.6	44.2% ± 15.1	46.7% ± 14.8
5					100	46.4% ± 15.1	45.2% ± 15.0	46.4% ± 14.9
6						100	47.3% ± 15.9	48.8% ± 15.8
7							100	54.7% ± 17.9
8								100

1, Strain AT3; 2, *Anaerotruncus colihominis*; 3, *Hydrogenoanaerobacterium saccharovorans*; 4, *Ruthenibacterium lactatiformans*; 5, *Acutalibacter muris*; 6, *Acetanaerobacterium elongatum*; 7, *Ruminococcus flavefaciens*; 8, *Ruminococcus champanellensis*.

incubation in anaerobic atmosphere. Its optimum growth temperature was 37°C, pH tolerance ranges from pH 6 to 8.5 and maximal saltness for growth is 5 g/L of NaCl concentration. Using API strips, glycerol, erythritol, adonithol, methyl-βD-xylopyranoside, galactose, glucose, mannose, sorbitol, esculin, trehalose, inulin, glycogen, gentiobiose, tagatose and potassium 5-cetogluconate were positive. Using 50CH, glucose, maltose, mannose, trehalose, rhamnose were positive and aesculin and gelatin hydrolase was observed with API 20A strip; alkaline phosphatase, esterase (C-4), esterase lipase (C-8), acid phosphatase, naphthol-AS-BI-phosphohydrolase and α-glucosidase were positive with API ZYM strip; indole production, nitrate reduction, alkaline phosphatase, arginine arylamidase, leucyl glycine arylamidase, phenylalanine arylamidase, leucine arylamidase, pyroglutamic acid arylamidase, tyrosine arylamidase, alanine arylamidase, glycine arylamidase, histidine arylamidase, glutamyl glutamic acid arylamidase and serine arylamidase were positive with rapid ID 32 A strip. Its major cellular fatty acids are; anteiso-$C_{15:0}$, $C_{16:0}$ and $C_{14:0}$. The genome of this strain is 3 145 950 bp long with 63.6% G+C content. The type strain, $AT3^T$ (= $CSURP\ P2007^T$ = $DSM\ 100567^T$) was isolated from the stool sample of a French morbidly obese man after bariatric surgery. The 16S rRNA gene sequence and whole-genome shotgun sequence of strain AT3 were deposited in GenBank with accession numbers LN866995 and OEJM00000000, respectively.

Conflict of Interest

None declared.

Acknowledgements

References

[1] Cecil JE, Tavendale R, Watt P, Hetherington MM, Palmer CNA. An obesity-associated FTO gene variant and increased energy intake in children. N Engl J Med 2008;359(24):2558—66.

[2] Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science 2007;316(5826):889—94.

[3] Fruh SM. Obesity: Risk factors, complications, and strategies for sustainable long-term weight management. J Am Assoc Nurse Pract 2017 Oct;29(S1):S3—14.

[4] Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980—2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014;384(9945):766—81.

[5] Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. Nature 2006;444(7122):1022—3.

[6] Million M, Maraninchi M, Henry M, Armougom F, Richet H, Carrieri P, et al. Obesity-associated gut microbiota is enriched in *Lactobacillus reuteri* and depleted in *Bifidobacterium animalis* and *Methanobrevibacter smithii*. Int J Obes 2012;36(6):817—25.

[7] Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 2006;444(7122):1027—31.

[8] Kong L-C, Tap J, Aron-Wisnewsky J, Pelloux V, Basdevant A, Bouillot J-L, et al. Gut microbiota after gastric bypass in human obesity: increased richness and associations of bacterial genera with adipose tissue genes. Am J Clin Nutr 2013;98(1):16—24.

[9] Zhang H, DiBaise JK, Zuccolo A, Kudrna D, Braidotti M, Yu Y, et al. Human gut microbiota in obesity and after gastric bypass. Proc Natl Acad Sci U S A 2009;106(7):2365—70.

[10] Lagier J-C, Hugon P, Khelaifia S, Fournier P-E, La Scola B, Raoult D. The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. Clin Microbiol Rev 2015 Jan;28(1):237—64.

[11] Lawson PA, Song Y, Liu C, Molitoris DR, Vaisanen M-L, Collins MD, et al. *Anaerotruncus colihominis* gen. nov., sp. nov., from human faeces. Int J Syst Evol Microbiol 2004;54(Pt 2):413–7.

[12] Togo AH, Durand G, Khelaifia S, Armstrong N, Robert C, Cadoret F, et al. *Fournierella massiliensis*, gen. nov., sp. nov., a new human-associated member of the family *Ruminococcaceae*. Int J Syst Evol Microbiol 2017;67:1393–9.

[13] Seng P, Drancourt M, Gouriet F, La Scola B, Fournier P-E, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 2009;49(4):543–51.

[14] Drancourt M, Bollet C, Carlioz A, Martelin R, Gayral JP, Raoult D. 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. J Clin Microbiol 2000 Oct;38(10):3623–30.

[15] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994 Nov 11;22(22):4673–80.

[16] Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 1981;17(6):368–76.

[17] Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 2016;33(7):1870–4.

[18] Diop K, Diop A, Khelaifia S, Robert C, Pinto FD, Delerce J, et al. Characterization of a novel Gram-stain-positive anaerobic coccus isolated from the female genital tract: Genome sequence and description of Murdochiella vaginalis sp. nov. Microbiology Open 2018;7(3):e00570.

[19] Citron DM, Ostovari MI, Karlsson A, Goldstein EJ. Evaluation of the E test for susceptibility testing of anaerobic bacteria. J Clin Microbiol 1991;29(10):2197–203.

[20] Matuschek E, Brown DFJ, Kahlmeter G. Development of the EUCAST disk diffusion antimicrobial susceptibility testing method and its implementation in routine microbiology laboratories. Clin Microbiol Infect 2014;20(4):O255–66.

[21] Togo AH, Khelaifia S, Lagier J-C, Caputo A, Robert C, Fournier P-E, et al. Noncontiguous finished genome sequence and description of Paenibacillus ihumii sp. nov. strain AT5. New Microbes New Infect 2016;10:142–50.

[22] Song L, Dong X. *Hydrogenoanaerobacterium saccharovorans* gen. nov., sp. nov., isolated from H2-producing UASB granules. Int J Syst Evol Microbiol 2009 Feb 1 [cited 2018 Nov 9];59(2):295–9.

[23] Shkoporov AN, Chaplin AV, Shcherbakova VA, Suzina NE, Kafarskaia LI, Bozhenko VK, et al. *Ruthenibacterium lactatiformans* gen. nov., sp. nov., an anaerobic, lactate-producing member of the family *Ruminococcaceae* isolated from human faeces. Int J Syst Evol Microbiol 2016 Aug 1;66(8):3041–9.

[24] Lagkouvardos I, Pukall R, Abt B, Foesel BU, Meier-Kolthoff JP, Kumar N, et al. The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. Nat Microbiol 2016 Oct;1(10).

[25] Chen S. *Acetanaerobacterium elongatum* gen. nov., sp. nov., from paper mill waste water. Int J Syst Evol Microbiol 2004 Nov 1;54(6):2257–62.

[26] Sijpesteijn AK. Cellulose-decomposing bacteria from the rumen of cattle. Antonie Van Leeuwenhoek 1949;15(1):49–52.

[27] Chassard C, Delmas E, Robert C, Lawson PA, Bernalier-Donadille A. *Ruminococcus champanellensis* sp. nov., a cellulose-degrading bacterium from human gut microbiota. Int J Syst Evol Microbiol 2012 Jan;62(Pt 1): 138–43.

[28] Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics 2011;12:124.

[29] Ramasamy D, Mishra AK, Lagier J-C, Padhmanabhan R, Rossi M, Sentausa E, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol 2014;64(2):384–91.

[30] Auch AF, von Jan M, Klenk H-P, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci 2010;2(1):117–34.

[31] Kent WJ. BLAT–the BLAST-like alignment tool. Genome Res 2002 Apr;12(4):656–64.

[32] Meier-Kolthoff JP, Göker M, Spröer C, Klenk H-P. When should a DDH experiment be mandatory in microbial taxonomy? Arch Microbiol 2013 Jun;195(6):413–8.

[33] Konstantinidis KT, Tiedje JM. Towards a genome-based taxonomy for prokaryotes. J Bacteriol 2005 Sep 15;187(18):6258–64.

[34] Rodriguez-R LM, Konstantinidis KT. Bypassing cultivation to identify bacterial species: culture-independent genomic approaches identify credibly distinct clusters, avoid cultivation bias, and provide true insights into microbial species. Microbe Mag 2014 Mar 1;9(3):111–8.

[35] Rainey FA. Family VIII. *Ruminococcaceae* fam. nov. In: De Vos P, Garrity GM, Jones D, Krieg NR, Ludwig W, Rainey FA, et al., editors. Bergey's Manual of Systematic Bacteriology. 2nd edVol. 3. New York: Springer; 2010. p. 1016–43.

[36] Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 2014;58(1):212–20.

[37] Conway KR, Boddy CN. ClusterMine360: a database of microbial PKS/NRPS biosynthesis. Nucleic Acids Res 2013;41(Database issue): D402–7.

[38] Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 2014;42(Database issue):D206–14.

[39] Togo AH, Valero R, Delerce J, Raould D, Million M. "*Anaerotruncus massiliensis*," a new species identified from human stool from an obese patient after bariatric surgery. New Microbes New Infect 2016 Nov;14:56–7.

Article XI:

Butyricimonas phoceensis **sp. nov., a new anaerobic species isolated from the human gut microbiota of a French morbidly obese patient.**

Amadou Hamidou Togo, Awa Diop, Grégory Dubourg, T. T. Nguyen, Claudia Andrieu, Aurélia Caputo, Carine Couderc, Pierre-Edouard Fournier, Marie Maraninchi, René Valero, Didier Raoult and Matthieu Million

Butyricimonas phoceensis sp. nov., a new anaerobic species isolated from the human gut microbiota of a French morbidly obese patient

A. H. Togo[1], A. Diop[1], G. Dubourg[1], T. T. Nguyen[1], C. Andrieu[1], A. Caputo[1], C. Couderc[1], P.-E. Fournier[1], M. Maraninchi[2,3], R. Valero[2,3], D. Raoult[1,4] and M. Million[1]

1) *Aix Marseille Université, URMITE, Institut Hospitalier Universitaire Méditerranée-Infection, UM63, CNRS7278, IRD198, INSERM1095,* 2) *Aix Marseille Université, NORT "Nutrition, Obesity and Risk of Thrombosis", INSERM1062, INRA1260,* 3) *APHM, CHU Hôpital de la Conception, Service Nutrition, Maladies Métaboliques et Endocrinologie, F-13385 Marseille, France and* 4) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Butyricimonas phoceensis strain AT9 (= CSUR 2478 = DSM 100838) was isolated from a stool sample from a morbidly obese French patient living in Marseille using the culturomics approach. The genome of this Gram-negative-staining, anaerobic and non−spore forming rod bacillus is 4 736 949 bp long and contains 3947 protein-coding genes. Genomic analysis identified 173 genes as ORFans (4.5%) and 1650 orthologous proteins (42%) not shared with the closest phylogenetic species, *Butyricimonas virosa*. Its major fatty acid was the branched acid iso-C15:0 (62.3%).

© 2016 The Author(s). Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases.

Corresponding author: M. Million, Aix Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, INSERM 1095, Marseille, France
E-mail: matthieumillion@gmail.com

Introduction

Butyricimonas phoceensis strain AT9 (= CSUR P2478 = DSM 100838 was isolated from the faeces of a 57-year-old French woman living in Marseille with class III morbid obesity (body mass index (BMI) 55.8 kg/m^2). This isolate is part of an exploratory study of the gut flora from obese patients before and after bariatric surgery. Bariatric surgery is the most effective treatment for morbid obesity for sustainable weight loss and leads to an enrichment of the gut flora [1]. The goal of our study was to compare microbial diversity of the gut flora in obese patients before and after bariatric surgery by culturomics. The aim of culturomics is to exhaustively explore the microbial ecosystem of gut flora by using different culture

conditions followed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) identification [2].

The conventional approaches used in the delineation of bacterial species are 16S rRNA sequence identity and phylogeny [3], genomic (G+C content) diversity and DNA-DNA hybridization (DDH) [4,5]. However, these approaches present some difficulties, mainly as a result of their cutoff values, which change according to species or genera [6]. The accession of new technology tools, such as high-throughput sequencing, has made available nucleotide sequence libraries for many bacterial species [7]. We recently suggested incorporating genomic data in a polyphasic taxonogenomics approach to describe new bacteria. This procedure considers phenotypic characteristics, genomic analysis and the MALDI-TOF MS spectrum comparison [8,9].

Here we propose a classification and a set of characteristics for *Butyricimonas phoceensis* strain AT9, together with the description of complete genome sequencing, annotation and comparison as new species belonging to the genus *Butyricimonas*. The genus *Butyricimonas* was established in 2009 by

Sakamoto and encompasses four described species (*B. faecihominis, B. synergistica, B. paravirosa* and *B. virosa*). They were isolated from rat or human faeces and belong to the family *Porphyromonadaceae* [10,11]. The family *Porphyromonadaceae* contains 11 genera: *Porphyromonas* (type genus)*, Barnesiella, Butyricimonas, Dysgonomonas, Macellibacteroides, Odoribacter, Paludibacter, Parabacteroides, Petrimonas, Proteiniphilum* and *Tannerella* [12]. *Butyricimonas virosa* bacteraemia has been described in patients with colon cancer [13,14] and in patients with posttraumatic chronic bone and joint infection [14].

Materials and Methods

Sample collection

A stool sample was collected from a 57-year-old obese French woman (BMI 55.8 kg/m^2; 150 kg, 1.64 m tall) in June 2012. Written informed consent was obtained from the patient at the nutrition, metabolic disease and endocrinology service at La Timone Hospital (Marseille, France). The study and assent procedure were approved by the local ethics committee (IFR 48, no. 09-022, 2010). The stool sample was stored at −80°C after collection.

Isolation and identification of strain

Strain isolation was performed in May 2015. Stool extract was preincubated in blood culture bottles enriched with lamb rumen juice and sheep's blood in anaerobic atmosphere as described elsewhere [2]. The culture was followed closely for 30 days. At different time points (days 1, 3, 7, 10, 15, 21 and 30), a seeding of the preincubated product was performed on sheep's blood–enriched Columbia agar (bioMérieux, Marcy l'Etoile, France) during 48 hours of incubation in an anaerobic atmosphere at 37°C. Colonies that emerged were cultivated in the same isolated conditions.

The colonies were then identified by MALDI-TOF MS as previously described [15]. Briefly, one isolated bacterial colony was picked up with a pipette tip from a culture agar plate and spread as a thin smear on a MTP 384 MALDI-TOF MS target plate (Bruker Daltonics, Leipzig, Germany). Measurement and identification were performed as previously described [16]. When a bacterium was unidentifiable, 16S rRNA gene amplification and sequencing were performed.

The 16S rRNA PCR coupled with sequencing were performed using GeneAmp PCR System 2720 thermal cyclers (Applied Biosystems, Bedford, MA, USA) and ABI Prism 3130xl Genetic Analyzer capillary sequencer (Applied Biosystems) respectively [17]. Chromas Pro 1.34 software (Technelysium, Tewantin, Australia) was used to correct sequences, and BLASTn searches were performed at the National Center for Biotechnology Information (NCBI) website (http://blast.ncbi.nlm.nih.gov.gate1.inist.fr/Blast.cgi).

Phylogenetic analysis

A custom Python script was used to automatically retrieve all species from the same family of the new species and download 16S sequences from NCBI by parsing NCBI results and NCBI taxonomy page. The scripts also remove species that are not found on the List of Prokaryotic Names With Standing in Nomenclature (LPSN) website (http://www.bacterio.net/). The script retains the most appropriate 16S sequence (the longest sequence with the smallest number of degenerate nucleotides) whilst also retaining one sequence from another genus as an outside group. It then aligns and trims the extremities of the sequences. Sequences were aligned using Muscle v3.8.31 with default parameters, and phylogenetic inferences were obtained using neighbour-joining method with 500 bootstrap replicates within MEGA6 software.

Phenotypic and biochemical characterization

Growth conditions. Different growth temperatures (28, 37, 45 and 55 °C) were tested on sheep's blood–enriched Columbia agar (bioMérieux). Growth of this strain was tested under anaerobic conditions using the GENbag anaer system (bioMérieux), microaerophilic conditions using the GENbag microaer system (bioMérieux) and under aerobic conditions with or without 5% CO_2. The tolerance to salt of this strain over a range salt concentrations (0–100 g/L) on Schaedler agar with 5% sheep's blood (bioMérieux) under anaerobic atmosphere was performed.

Microscopy. A heat shock at 80°C for 20 minutes was performed for the sporulation test. A fresh colony was observed between blades and slats using a photonic microscope Leica DM 1000 (Leica Microsystems, Nanterre, France) at 40× to assess the motility of the bacteria. Gram staining was performed and observed using a photonic microscope Leica DM 2500 with a 100× oil-immersion objective lens. Transmission electron microscopy using a Tecnai G20 device (FEI Company, Limeil-Brevannes, France) at an operating voltage of 60 kV was performed to observe strain AT9 after negative colouration.

Biochemical assays. Biochemical assays were performed using API Gallery systems (API ZYM, API 20A and API 50CH) according to the manufacturer's instructions (bioMérieux). Detection of catalase (bioMérieux) and oxidase (Becton Dickinson, Le Pont de Claix, France) was also performed according to the manufacturer's instructions.

Antibiotic susceptibility. The antibiotic susceptibility of the strain was tested using a disk diffusion method [18] for 21 antibiotics

including the following: amoxicillin 25 μg/mL, amoxicillin–clavulanic acid 30 μg/mL, ceftriaxone 30 μg, ciprofloxacin 5 μg, clindamycin (DA15), colistin (CT50), Dalacin 15 μg/mL, doripenem 10 μg/mL, doxycycline 30 IU, erythromycin 15 IU, fosfomycin 10 μg, gentamicin 500 μg, gentamicin 15 μg, imipenem 10 μg/mL, metronidazole 4 μg/mL, oxacillin 5 μg, penicillin G 10 IU, rifampicin 30 μg, sulfamethoxazole 23.75 μg, trimethoprim 1.25 μg, teicoplanin (TEC30) and vancomycin 30 μg (i2a, Montpellier, France). The 1200 scan was used for the interpretation of results (Interscience, Saint-Nom-La-Bretèche, France).

Fatty acid analysis. Fresh colonies from a plate of Columbia agar with 5% sheep's blood were collected after 48 hours' incubation at 37°C for fatty acid analysis. Cellular fatty acid analysis was performed by gas chromatography/mass spectrometry (GC/MS). Two samples were prepared with approximately 100 mg of bacterial biomass each collected from a culture plate. Cellular fatty acid methyl esters were prepared as described by Sasser [19]. GC/MS analyses were carried out as previously described [20]. Briefly, fatty acid methyl esters were separated using an Elite 5-MS column and monitored by a Clarus 500 gas chromatograph equipped with a SQ8S MS detector (PerkinElmer, Courtaboeuf, France). Fatty acid methyl esters were identified by using the spectral database search using MS Search 2.0 operated with the Standard Reference Database 1A (National Institute of Standards and Technology, Gaithersburg, MD, USA) and the FAMEs mass spectral database (Wiley, Chichester, UK).

Genome sequencing and assembly

Genomic DNA (gDNA) of strain AT9 was sequenced using MiSeq Technology (Illumina, San Diego, CA, USA) with the mate-pair strategy. The gDNA was barcoded so it could be mixed with 11 other projects with the Nextera Mate Pair sample prep kit (Illumina). The gDNA was quantified by a Qubit assay with the high-sensitivity kit (Thermo Fisher Scientific Life Sciences, Waltham, MA, USA) to 325 ng/μL. The mate-pair library was prepared with 1.5 μg of genomic DNA using the Nextera mate pair Illumina guide. The genomic DNA sample was simultaneously fragmented and tagged with a mate-pair junction adapter. The pattern of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) with a DNA 7500 lab chip. The DNA fragments ranged in size from 1.5 to 11 kb with an optimal size at 4.8 kb. No size selection was performed, and 600 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal at 966 bp on the Covaris S2 device in T6 tubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies), and the final concentration library was measured at 24.3 nmol/L. The libraries were normalized at 2 nM and pooled. After a denaturation step and dilution at 15 pM, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and sequencing run were performed in a single 39-hour run at a 2 × 151 bp read length. Total information of 8.9 Gb was obtained from a 1009K/mm² cluster density, with a cluster passing quality control filters of 91.5% (17 486 000 passing filter-paired reads). Within this run, the index representation for strain AT9 was determined to be 8.38%. The 1 465 998 paired reads were trimmed then assembled in six scaffolds using Spades software [21].

Genome annotation and comparison

Open reading frames (ORFs) were predicted using Prodigal [22] with default parameters. Nevertheless, the predicted ORFs were excluded if they spanned a sequencing gap region (contains N). The predicted bacterial protein sequences were searched against the GenBank and Clusters of Orthologous Groups (COGs) databases using BLASTP (*E* value 1e-03 coverage). If no hit was found, it searched against the nr (nonredundant) database using BLASTP with an *E* value of 1e-03, coverage 70% and identity 30%. If the sequence length was smaller than 80 amino acids, we used an *E* value of 1e-05. The tRNAs and rRNAs were predicted using the tRNA Scan-SE and RNAmmer tools respectively [23,24]. Phobius was used to foresee the signal peptides and number of transmembrane helices respectively [25]. Mobile genetic elements were foretold using PHAST and RAST [26,27]. ORFans were identified if none of the BLASTP runs provided positive results (*E* value was lower than 1e-03 for an alignment length greater than 80 amino acids. If alignment lengths were smaller than 80 amino acids, we used an *E* value of 1e-05). Artemis and DNA Plotter were used for data management and visualization of genomic features respectively [28,29]. Genomes were automatically retrieved from the 16S rRNA tree using Xegen software (PhyloPattern) [30]. For each selected genome, complete genome sequence, proteome genome sequence and Orfeome genome sequence were retrieved from the NCBI FTP site. All proteomes were analysed with proteinOrtho [31]. Then for each couple of genomes, a similarity score was computed. This score is the mean value of nucleotide similarity between all couple of orthologous genes between the two genomes studied (average genomic identity of orthologous gene sequences (AGIOS)) [7]. For the genomic comparison of strain AT9, we used *Butyricimonas virosa* (type) strain JCM15149T (Genbank project number: JAEW00000000), *Odoribacter laneus* strain YIT12061 (ADMC00000000), *Bacteroides plebeius* strain DSM17135
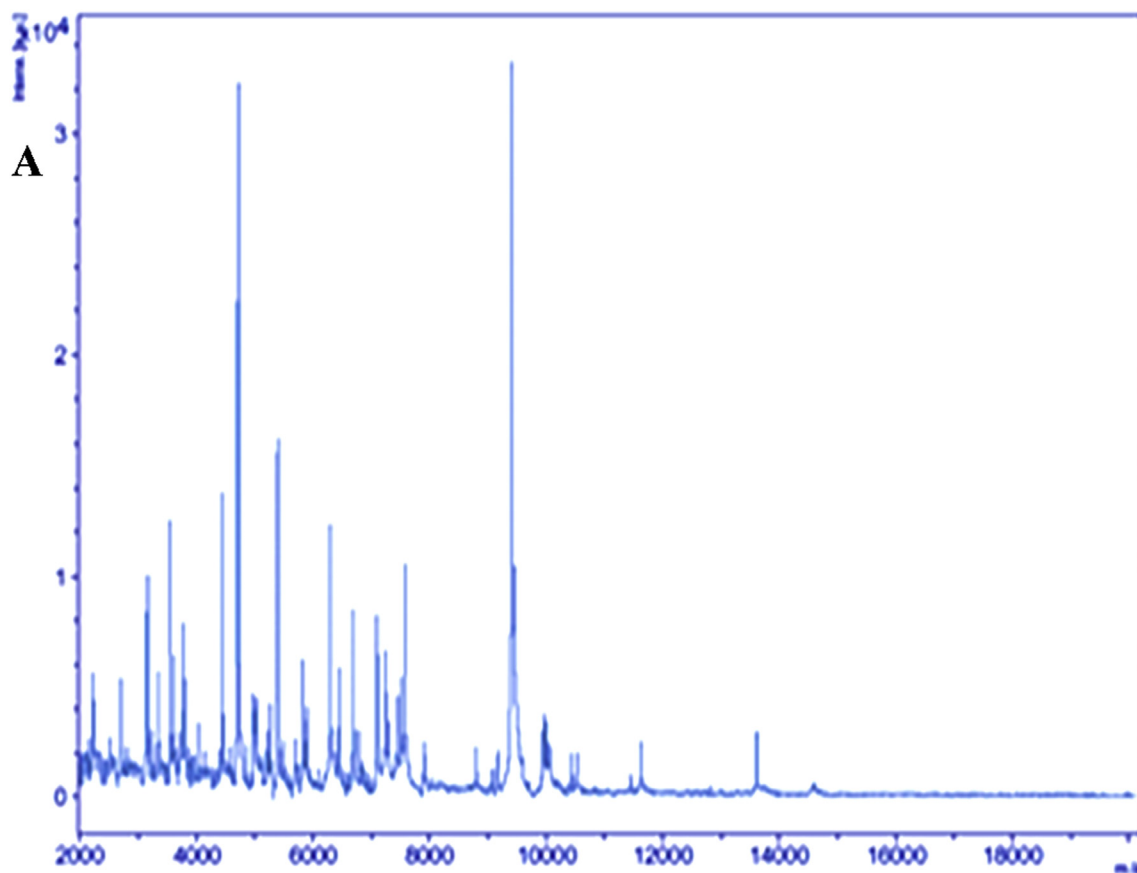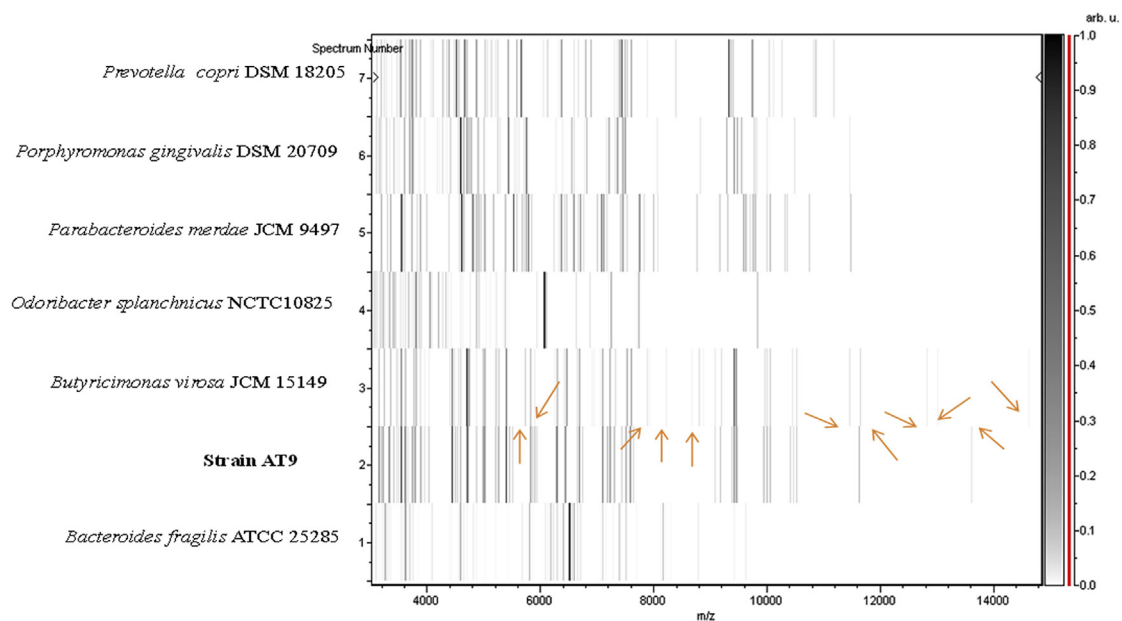
FIG. 1. MALDI-TOF MS analysis of *Butyricimonas phoceensis* strain AT9. (a) Reference mass spectrum from strain AT9. (b) Gel view comparing strain AT9 to other close species. Gel view displays raw spectra of loaded spectrum files arranged in pseudo−gel-like look. The *x*-axis records m/z value. Left *y*-axis displays running spectrum number originating from subsequent spectra loading. Peak intensity is expressed by greyscale scheme code. Colour bar and right *y*-axis indicate relation between colour peak is displayed with and peak intensity in arbitrary units. Displayed species are indicated at left. Arrows indicated discordant peaks between strain AT9 and its closest phylogenetic neighbour, *Butyricimonas virosa*.

(ABQC00000000), *Paraprevotella clara* strain YIT11840 (AFFY00000000), *Parabacteroides merdae* ATCC43184 (AAXE00000000), *Porphyromonas catoniae* ATCC 51270 (JDFF00000000) and *Odoribacter splanchnicus* strain DSM20712 (CP002544). An annotation of the entire proteome was performed to define the distribution of functional classes of predicted genes according to the clusters of orthologous groups of proteins (using the same method as for the genome annotation). The genome of strain AT9 was locally aligned pairwise using the BLAT algorithm [28,29] against each of the selected genomes previously cited, and DDH values were estimated from a generalized model [32]. Annotation and comparison processes were performed in the multiagent software system DAGOBAH [33], which includes Figenix [34] libraries that provide pipeline analysis.

Results

Phylogenetic analysis

The spectrum generated from clean strain AT9 spots did not match with those identified from the Bruker database even when two strains of *Butyricimonas virosa*, including the type strain

(JCM15149T), were included in the database (Fig. 1a). The phylogenetic analysis, performed using 16S rRNA gene sequences, showed that our strain AT9 exhibited 98.3, 97.8, 97.5 and 94.2% similarity with *Butyricimonas virosa* JCM 15149T, *Butyricimonas faecihominis* JCM 18676T, *Butyricimonas paravirosa* JCM 18677T and *Butyricimonas synergistica* JCM 15148T respectively [9,10] (Table 1). However, this percentage remains lower than the 98.7% 16S rRNA gene sequence threshold recommended by Kim *et al.* [35] to delineate a new species. The neighbour-joining phylogenetic tree (Fig. 2), based on 16S rRNA gene sequences, shows the relationships between strain AT9 and some related taxa. The 16S rRNA sequence of strain AT9 was deposited in European Molecular Biology Laboratory–European Bioinformatics Institute (EMBL-EBI) under accession number LN881597. A gel view was performed in order to see the spectra differences of strain AT9 with other related bacteria. Eleven discordant peaks were found when we compared strain AT9 and the *B. virosa* JCM15149T profile (Fig. 1b).

Phenotypic and biochemical characterization

The growth of strain AT9 occurred between 28 to 37°C, but optimal growth was observed at 37°C after 48 hours' incubation in anaerobic atmosphere. It is an anaerobic bacillus, but it

TABLE 1. Percentage 16S rRNA gene similarity within *Butyricimonas* genus

	B. faecihominis JCM 18676T	B. paravirosa JCM 18677T	B. synergistica JCM 15148T	B. virosa JCM 15149T	B. phoceensis strain AT9
B. faecihominis JCM 18676T	100	97.30	94.07	96.84	97.77
B. paravirosa JCM 18677T		100	94.75	96.84	97.51
B. synergistica JCM 15148T			100	94.22	94.20
B. virosa JCM 15149T				100	98.38
B. phoceensis strain AT9					100
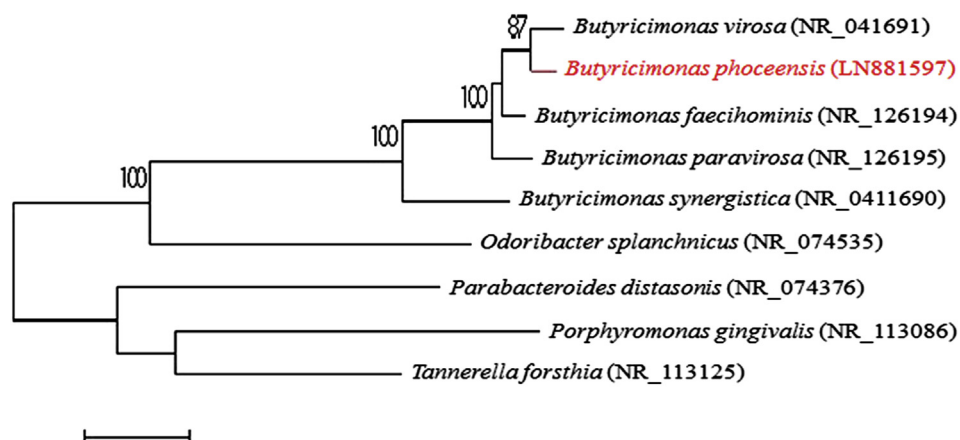
FIG. 2. Phylogenetic tree based on 16S rRNA highlighting position of *Butyricimonas phoceensis* strain AT9 relative to other close species. Corresponding GenBank accession numbers for 16S rRNA genes are indicated at right of strains in tree. Sequences were aligned using Muscle v3.8.31 with default parameters, and phylogenetic inferences were obtained using neighbour-joining method with 500 bootstrap replicates within MEGA6 software. Scale bar represents 2% nucleotide sequence divergence.

can also grow in microaerophilic atmospheres at 28°C. The colonies were ~1 to 2 mm in diameter and opalescent on 5% sheep's blood−enriched Columbia agar. Growth of this isolate was observed using 5 g of salt on Schaedler agar with 5% sheep's blood but not with 10 g/L of salt. This bacterium is not able to form spores. It is a Gram-negative stain (Fig. 3a); it is a motile rod-shaped bacterium that is catalase positive and oxidase negative. Cell diameter ranges 0.5 to 1.5 μm, with a mean diameter of 1 μm by electron microscopy (Fig. 3b). Table 2 summarizes the classification and main features of strain AT9.

Using the API ZYM strip, we observed that strain AT9 possesses alkaline phosphatase, esterase (C4), esterase lipase (C8), naphthol-AS-BI-phosphohydrolase, phosphatase acid and N-acetyl-β-glucosaminidase activities; there were no activities for the other enzymes tested. Using API 20A strip, positive reactions were obtained for indole, D-glucose, D-lactose, glycerol and D-mannose. Using the API 50 CH strip, positive reactions were observed only with esculin ferric citrate and potassium 2-ketogluconate. The differences of characteristics compared to other representatives of the genus *Butyricimonas* are detailed in Table 3.
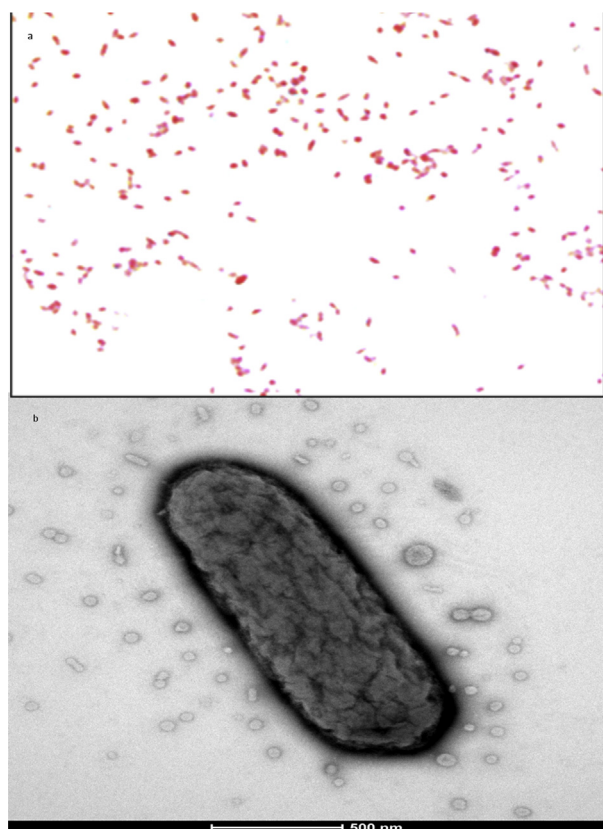
FIG. 3. Phenotypic features of *Butyricimonas phoceensis* strain AT9. (a) Gram stain. (b) Transmission electron microscopy using Tecnai G20 (FEI Company) at operating voltage of 60 kV. Scale bar = 500 nm.

TABLE 2. Classification and general features of *Butyricimonas phoceensis* strain AT9

Property	Term
Current classification	Domain: *Bacteria*
	Phylum: *Bacteroidetes*
	Class: *Bacteroidia*
	Order: *Bacteroidales*
	Family: *Porphyromonadaceae*
	Genus: *Butyricimonas*
	Species: *B. phoceensis*
	Type strain: AT9
Gram stain	Negative
Cell shape	Rod
Motility	Motile
Sporulation	Non−spore forming
Temperature range	Mesophile
Optimum temperature	37°C
Oxygen requirement	Anaerobic
Carbon source	Unknown
Energy source	Unknown
Habitat	Human gut
Biotic relationship	Free living
Pathogenicity	Unknown
Isolation	Human faeces

Of the 21 antibiotics tested, strain AT9 was susceptible to gentamicin 500 μg, vancomycin, doxycycline, trimethoprim−sulfamethoxazole, rifampicin, amoxicillin 25 μg/mL, metronidazole 4 μg/mL, amoxicillin−clavulanic acid 30 μg/mL, imipenem 10 μg/mL, penicillin G, teicoplanin and doripenem 10 μg/mL and was resistant to erythromycin, oxacillin, gentamicin 15 μg, colistin, ceftriaxone, ciprofloxacin, clindamycin, dalacin 15 μg/mL and fosfomycin. Analysis of the total cellular fatty acid composition demonstrated that the major fatty acid detected was the branched iso-C15:0 acid (62.3%). Hydroxy and cyclo fatty acids were also detected (Table 4).

Genome properties

The draft genome of strain AT9 (Fig. 4) (accession no. FBYB00000000) is 4 736 949 bp long with 42.51% G+C content (Table 5). It is composed of six scaffolds comprising seven contigs. Of the 4007 predicted genes, 3947 were protein-coding genes and 60 were RNAs (four genes 5S rRNA, one 16S rRNA, one 23S rRNA and 54 tRNA). A total of 2386 genes (60.45%) were assigned as putative functions (by COGs or by NR BLAST), 178 genes (4.51%) were identified as ORFans and ten genes were associated with polyketide synthase or non-ribosomal peptide synthetase [36]. Using ARG-ANNOT [37], three genes associated with resistance were found, including *TetQ, TetX* (which confers resistance to tetracycline) and *ErmF* (which confers resistance to erythromycin). This could represent the *in silico/in vitro* discordance for antibiotic resistance prediction, as strain AT9 was resistant to erythromycin but susceptible to doxycycline. The remaining 1316 genes (33.34%) were annotated as hypothetical proteins. Genome statistics are provided in Table 5. Table 6 lists the distribution of genes into COGs functional categories of strain AT9.

TABLE 3. Differential characteristics of strain *Butyricimonas phoceensis* strain AT9 with *Butyricimonas* species

Property	Strain AT9	B. virosa	B. faecihominis	B. paravirosa	B. synergistica
Cell diameter width/length (μm)	0.5/1.75	0.6–0.8/2.5–5	0.7–1/3–5	0.8–1/2–12.4	0.5–1/3–6
Oxygen requirement	–	–	–	–	–
Gram stain	–	–	–	–	–
Motility	+	–	–	–	–
Spore formation	–	–	–	–	–
Production of:					
Catalase	+	+	+	+	–
Oxidase	–	–	–	–	–
Urease	–	–	–	–	–
Indole	+	+	+	+	+
β-Galactosidase	+	+	+	+	+
N-acetyl-glucosaminidase	+	+	+	+	+
Utilization of:					
L-Arabinose	–	–	+	–	–
D-Mannose	+	–	+	+	+
D-Mannitol	–	–	–	–	–
D-Glucose	+	+	+	+	+
D-Maltose	–	–	+	–	+
Isolation source	Human faeces	Rat faeces	Human faeces	Rat faeces	Human faeces
DNA G+C content (mol%)	42.5	46.5	45.2	44.9	46.4

TABLE 4. Cellular fatty acid profiles of strain *Butyricimonas phoceensis* strain AT9 compared to other closely related *Butyricimonas* species

Fatty acid	Strain AT9	B. faecihominis JCM 18676T	B. paravirosa JCM 18677T	B. synergistica JCM 15148T	B. virosa JCM 15149T
C4:0	TR	NA	NA	NA	NA
C12: 0	NA	TR	TR	NA	NA
C14:0	TR	TR	1.8	NA	1.3
C15: 0	TR	TR	NA	NA	NA
C16:0	3.7	2.8	3.2	2.4	2.1
C18:0	TR	TR	TR	1.0	TR
iso-C5: 0	2.9	NA	NA	NA	NA
iso-C11: 0	NA	TR	TR	NA	NA
iso-C13: 0	NA	1.0	1.0	NA	TR
iso-C15: 0	62.3	64.6	57.6	61.8	68.6
anteiso-C15: 0	1.2	1.8	1.7	2.0	1.5
iso-C17: 0	NA	1.0	TR	NA	TR
C14:0 3-OH	TR	NA	NA	NA	NA
C16: 0 3-OH	4.8	1.7	6.3	1.6	5.2
C17: 0 3-OH	9.0	NA	NA	NA	NA
iso-C15: 0 3-OH	NA	TR	1.8	1.6	1.7
iso-C17: 0 3-OH	NA	5.3	10.6	14.9	10.4
C18: 2n6	2.9	NA	NA	NA	NA
C18:1n5	2.1	NA	NA	NA	NA
C16:1n7	TR	NA	NA	NA	NA
iso-C17: 0	TR	1.0	TR	NA	TR
iso-C15:1n5	TR	NA	NA	NA	NA
C18: 1ω9c	NA	8.3	9.5	12.6	6.0
C18: 2ω6, 9c	NA	1.4	1.5	2.3	1.2
C9, 10-methylene C16:0	7.0	NA	NA	NA	NA

Number are percentages. NA, not available; TR, trace amounts <1%.

Genome comparison

The draft genome (4.74 Mb) sequence of strain AT9 is smaller than those of *Butyricimonas synergistica* (4.77 Mb), but larger than those of *Butyricimonas virosa, Porphyromonas catoniae, Bacteroides plebeius, Paraprevotella clara, Odoribacter laneus, Parabacteroides merdae* and *Odoribacter splanchnicus* (4.72, 2.04, 3.27, 3.65, 4.43, 3.77 and 4.39 MB respectively).

The G+C content of strain AT9 (42.5%) is smaller than those of *Butyricimonas virosa, Odoribacter splanchnicus, Bacteroides plebeius, Parabacteroides merdae, Paraprevotella clara, Butyricimonas synergistica* and *Porphyromonas catoniae* (46.5, 43.4, 44.3, 44.8, 45.3, 48.1, 46.4 and 51.0% respectively) but larger than those of *Odoribacter laneus* (40.55). Fig. 5 shows that the distribution of genes into COGs categories was similar in all genomes compared. In addition, strain AT9 shared 2297, 1535, 742, 1720, 999, 1173, 2108 and 960 orthologous genes with *B. virosa, O. laneus, P. catoniae, O. splanchnicus, B. plebeius, P. merdae, B. synergistica* and *P. clara* respectively (Table 6). Accordingly, strain AT9 has 1650 (42%) of 3947 orthologous proteins not shared with its closest phylogenetic neighbour, *B. virosa*. The AGIOS values ranged from 53.3 to 76.2% among the compared closest species except strain AT9. When strain AT9 was compared to other close species, the AGIOS values ranged from 53.5% with *P. catoniae* to 97.7% with *B. virosa* (Table 7).
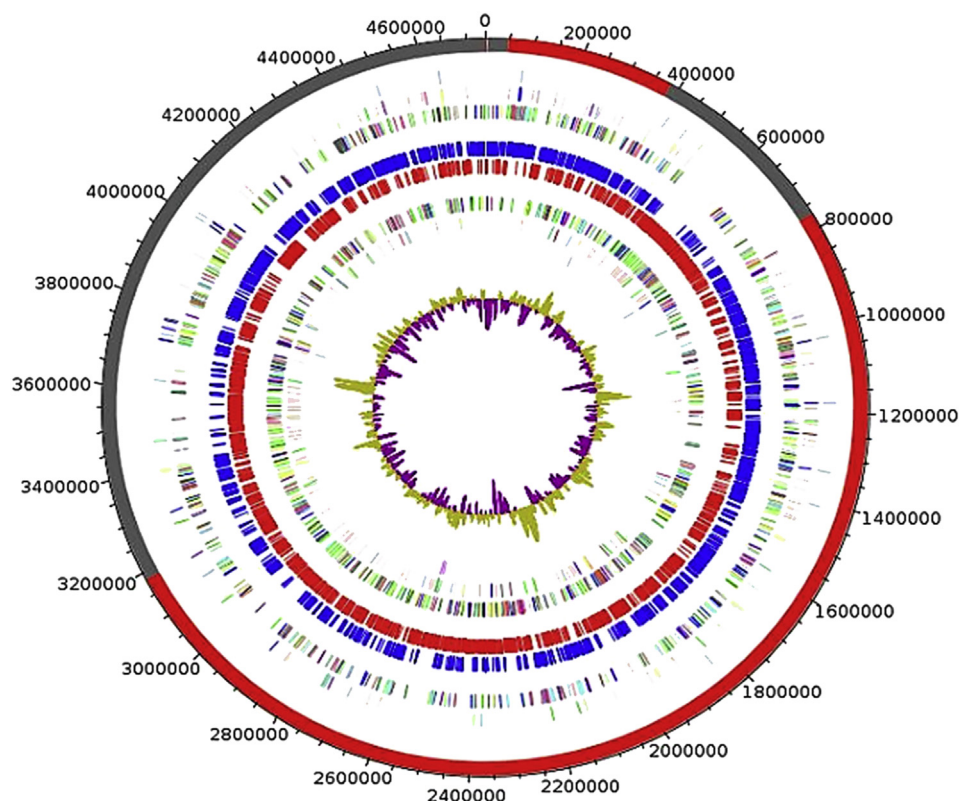
FIG. 4. Graphical circular map of genome of *Butyricimonas phoceensis* strain AT9. From outside to centre: contigs (red/grey), COGs category of genes on forward strand (three circles), genes on forward strand (blue circle), genes on reverse strand (red circle), COGs category on reverse strand (three circles), GC content.

The DDH value was 80.2% ± 2.7 with *B. virosa*, 17.7% ± 2.2 with *O. laneus*, 21.4% ± 2.3 with *B. plebeius*, 20.2% ± 2.3 with *P. clara*, 19.1% ± 2.2 with *P. merdae*, 18.3% ± 2.2 with *P. catoniae* and 17.3% ± 2.2 with *O. splanchnicus* (Table 8).

TABLE 5. Nucleotide content and gene count levels of the genome of *Butyricimonas phoceensis* strain AT9

	Genome (total)	
Attribute	Value	% of total
Size (bp)	4 736 949	100
G+C content (bp)	2 013 756	42.51
Coding region (bp)	4 330 163	91.40
Total genes	4007	100
RNA genes	60	1.50
Protein-coding genes	3947	98.50
Genes with function prediction	2386	60.45
Genes assigned to COGs	1880	47.63
Genes with peptide signals	1185	30.02
Gene associated to PKS or NRPS	10	0.25
Genes associated to ORFan	178	4.51
Genes associated to mobilome	1109	28.10
Genes associated to toxin/antitoxin	70	1.8
Genes associated to resistance genes	3	0.076
Genes with paralogues (*E* value 1e-10)	1449	36.71
Genes with paralogues (*E* value 1e-25)	1098	27.82
Gene associated to hypothetical protein	1316	33.34
Genes larger than 5000 nucleotides	5	0

COGs, Clusters of Orthologous Groups database; PKS, polyketide synthase; NRPS, nonribosomal peptide synthase.

Discussion

Strain AT9 is part of an exploratory culturomics study of the gut flora from obese patients before and after bariatric surgery. The aim of culturomics is to exhaustively explore the microbial ecosystem of gut flora by using different culture conditions followed by MALDI-TOF MS identification [2]. The phylogenetic analysis, performed using 16S rRNA sequences, showed that strain AT9 exhibited 98.3% similarity with *Butyricimonas virosa*. However, this percentage remains lower than the 98.7% 16S rRNA gene sequence threshold recommended to delineate a new species [3,38].

The genus *Butyricimonas* was established in 2009 by Sakamoto and includes four described species [9–11]. All the species of the genus *Butyricimonas* are anaerobic. These bacteria are isolated in human or rat faeces. To evaluate the genomic similarity with other closest species, we determined two parameters: DDH [39] and AGIOS [7]. Although the values of DDH (80.2%) and AGIOS (97.7%) were very high between strain AT9 and *Butyricimonas virosa* (type strain JCM15149T), we found several discrepancies justifying the description of a new species, including motility, D-mannose utilization (absent in

TABLE 6. Number of genes associated with the 25 general COGs functional categories of *Butyricimonas phoceensis* strain AT9

Code	Value	% value	Description
J	193	4.89	Translation
A	0	0	RNA processing and modification
K	192	4.87	Transcription
L	111	2.81	Replication, recombination and repair
B	0	0	Chromatin structure and dynamics
D	23	0.58	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	85	2.15	Defence mechanisms
T	174	4.41	Signal transduction mechanisms
M	200	5.06	Cell wall/membrane biogenesis
N	20	0.51	Cell motility
Z	4	0.10	Cytoskeleton
W	3	0.07	Extracellular structures
U	28	0.71	Intracellular trafficking and secretion
O	91	2.30	Posttranslational modification, protein turnover, chaperones
X	32	0.81	Mobilome: prophages, transposons
C	122	3.09	Energy production and conversion
G	92	2.33	Carbohydrate transport and metabolism
E	120	3.04	Amino acid transport and metabolism
F	60	1.52	Nucleotide transport and metabolism
H	99	2.51	Coenzyme transport and metabolism
I	69	1.75	Lipid transport and metabolism
P	199	5.04	Inorganic ion transport and metabolism
Q	26	0.66	Secondary metabolites biosynthesis, transport and catabolism
R	150	3.80	General function prediction only
S	67	1.69	Function unknown
—	2067	52.36	Not in COGs

COGs, Clusters of Orthologous Groups database.

B. virosa but present in *B. paravirosa, B. synergistica* and *B. faecihominis*), MALDI-TOF MS spectrum (11 different peaks), different GC% (42.5 vs. 46.5% for *B. virosa*), high proportion of orthologous proteins not shared between the two species (1650/3947 (42%)) and different COGs repartition ((D) cell cycle control 110 vs. 124, (P) transport of inorganic ions 32 vs. 20 for strain AT9 and *B. virosa* respectively).

Conclusion

On the basis of phenotypic, chemotaxonomic, phylogenetic and genomic information, a novel species belonging to the genus *Butyricimonas* is proposed with the name *Butyricimonas phoceensis* sp. nov. The type strain is AT9. This bacterium was isolated from the faeces of a 57-year-old obese French woman living in Marseille after bariatric surgery. The isolation of this new species demonstrates that microbial culturomics extends the repertoire of human gut anaerobes, which are of critical importance to decipher the links among gut microbiota, health and disease, including obesity.
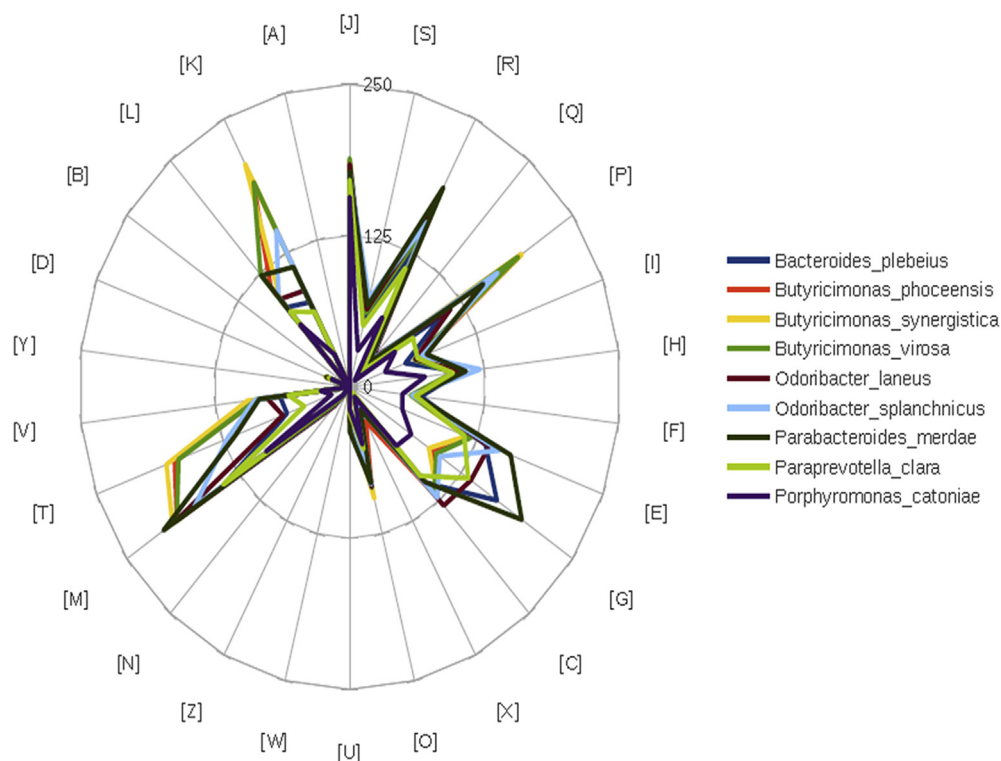
FIG. 5. Distribution of functional classes of predicted genes according to clusters of orthologous groups of proteins from *Butyricimonas phoceensis* strain AT9.

TABLE 7. Numbers of orthologous proteins shared between genomes (upper right), average percentage similarity of nucleotides corresponding to orthologous protein shared between genomes (lower left) and numbers of proteins per genome (bold)

	Odoribacter laneus	Butyricimonas phoceensis strain AT9	Porphyromonas catoniae	Odoribacter splanchnicus	Bacteroides plebeius	Parabacteroides merdae	Butyricimonas virosa	Butyricimonas synergistica	Paraprevotella clara
O. laneus	**3103**	1535	745	1472	1005	1187	1519	1480	964
B. phoceensis strain AT9	57.52	**3947**	742	1720	999	1173	2297	2108	960
P. catoniae	53.37	53.50	**1597**	737	726	826	729	725	746
O. splanchnicus	59.11	68.17	53.65	**3497**	977	1149	1702	1604	960
B. plebeius	55.52	62.17	55.11	62.84	**2643**	1175	986	963	1059
P. merdae	55.38	63.08	55.30	63.65	66.34	**4384**	1154	1130	1123
B. virosa	57.47	97.79	53.41	68.18	62.13	62.87	**3934**	2086	950
B. synergistica	57.22	76.18	53.75	68.24	62.15	62.93	76.24	**3874**	926
P. clara	54.31	61.92	54.84	62.34	68.02	65.40	61.87	62.09	**2847**

TABLE 8. Pairwise comparison of *Butyricimonas phoceensis* strain AT9 with other species using GGDC, formula 2 (DDH estimates based on identities/HSP length)[a]

	Strain AT9	Odoribacter laneus	Bacteroides plebeius	Butyricimonas virosa	Paraprevotella clara	Parabacteroides merdae	Porphyromonas catoniae	Odoribacter splanchnicus
Strain AT9	100% ± 0	17.7% ± 2.2	21.4% ± 2.3	80.2% ± 2.7	20.2% ± 2.3	19.1% ± 2.3	18.3% ± 2.3	17.3% ± 2.2
O. laneus		100% ± 0	19% ± 2.3	18.2% ± 2.3	20.5% ± 2.3	18.9% ± 2.3	19.6% ± 2.3	18.2% ± 2.3
B. plebeius			100% ± 0	19.9% ± 2.3	20.3% ± 2.3	21.5% ± 2.3	17.6% ± 2.2	18.4% ± 2.3
B. virosa				100% ± 0	20.3% ± 2.3	19.4% ± 2.3	19.0% ± 2.3	17.4% ± 2.2
P. clara					100% ± 0	18.9% ± 2.3	17.8% ± 2.2	17.7% ± 2;2
P. merdae						100% ± 0	17.6% ± 2.2	21.5% ± 2.3
P. catoniae							100% ± 0	18% ± 2.2
O. splanchnicus								100% ± 0

DDH, DNA-DNA hybridization; GGDC, Genome-to-Genome Distance Calculator; HSP, high-scoring segment pairs.
[a]Confidence intervals indicate inherent uncertainly in estimating DDH values from intergenomic distances based on models derived from empirical test data sets (which are always limited in size). These results are in accordance with 16S rRNA and phylogenomic analyses as well as GGDC results.

Taxonomic and nomenclatural proposals

Description of strain AT9 sp. nov.. *Butyricimonas phoceensis* (pho.ce.en.sis, N.L. gen. n. *phoceensis*, based on the acronym of the Phocean city where the type strain was isolated). Cells are Gram-negative-staining, non−spore forming, motile, rod-shaped bacilli, with a size of 0.5 to 1.5 μm in diameter. Colonies are opalescent with a diameter of 1 to 2 mm on 5% sheep's blood−enriched Columbia agar. The strain is oxidase negative and catalase positive. It has an optimum growth temperature of 37°C and is anaerobic, but it is able to grow in microaerophilic condition at 28°C. Using API Gallery systems, positive reactions were observed for alkaline phosphatase, esterase (C4), esterase lipase (C8), naphthol-AS-BI-phosphohydrolase, phosphatase acid, *N*-acetyl-β-glucosaminidase, indole, D-glucose, D-lactose, glycerol and D-mannose, esculin ferric citrate and potassium 2-ketogluconate. Cells are susceptible to gentamicin 500 μg, vancomycin, doxycycline, trimethoprim−sulfamethoxazole, rifampicin, penicillin G and teicoplanin. The major fatty acid detected was iso-C15:0. The length of the genome is 4 736 949 bp with 42.51% G+C content. The 16S rRNA gene sequence and whole-genome shotgun sequence of *B. phoceensis* strain AT9 were deposited in EMBL-EBI under accession numbers LN881597 and FBYB00000000, respectively. The type strain AT9 (= CSUR P2478 = DSM 100838) was isolated from the stool sample of a French obese woman. The habitat of this microorganism is the human digestive gut.

Acknowledgements

Conflict of Interest

None declared.

References

[1] Zhang H, DiBaise JK, Zuccolo A, Kudrna D, Braidotti M, Yu Y, et al. Human gut microbiota in obesity and after gastric bypass. Proc Natl Acad Sci U S A 2009;106:2365−70.

[2] Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 2012;18:1185−93.

[3] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152–5.

[4] Garrity GM, Trüper HG, Whitman WB, Grimont PAD, Nesme X, Frederiksen W, et al. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Microbiol 2002;52:1043–7.

[5] Rosselló-Mora R. DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation. In: Molecular identification, systematics, and population structure of prokaryotes. New York: Springer; 2006. p. 23–50.

[6] Welker M, Moore ERB. Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. Syst Appl Microbiol 2011;34:2–11.

[7] Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi M, Sentausa E, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol 2014;64(Pt 2):384–91.

[8] Kokcha S, Mishra AK, Lagier JC, Million M, Leroy Q, Raoult D, et al. Non-contiguous-finished genome sequence and description of *Bacillus timonensis* sp. nov. Stand Genomic Sci 2012;6:346–55.

[9] Mishra AK, Lagier JC, Nguyen TT, Raoult D, Fournier PE. Non-contiguous-finished genome sequence and description of *Peptoniphilus senegalensis* sp. nov. Stand Genomic Sci 2013;7:370–81.

[10] Sakamoto M, Takagaki A, Matsumoto K, Kato Y, Goto K, Benno Y. *Butyricimonas synergistica* gen. nov., sp. nov. and *Butyricimonas virosa* sp. nov., butyric acid–producing bacteria in the family '*Porphyromonadaceae*' isolated from rat faeces. Int J Syst Evol Microbiol 2009;59(Pt 7):1748–53.

[11] Sakamoto M, Tanaka Y, Benno Y, Ohkuma M. *Butyricimonas faeciho-minis* sp. nov. and *Butyricimonas paravirosa* sp. nov., isolated from human faeces, and emended description of the genus *Butyricimonas*. Int J Syst Evol Microbiol 2014;64(Pt 9):2992–7.

[12] Sakamoto M. The Family *Porphyromonadaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. The Prokaryotes—other major lineages of bacteria and the Archaea. Berlin: Springer; 2014. p. 811–24.

[13] Ulger Toprak N, Bozan T, Birkan Y, Isbir S, Soyletir G. *Butyricimonas virosa*: the first clinical case of bacteraemia. New Microbes New Infect 2015;4:7–8.

[14] Ferry T, Laurent F, Ragois P, Chidiac C, Lyon BJI Study Group. Post-traumatic chronic bone and joint infection caused by *Butyricimonas* spp., and treated with high doses of ertapenem administered subcutaneously in a 30-year-old obese man. BMJ Case Rep 2015;2015:212359.

[15] Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 2009;49:543–51.

[16] Hugon P, Ramasamy D, Lagier JC, Rivet R, Couderc C, Raoult D, et al. Non-contiguous-finished genome sequence and description of *Alistipes obesi* sp. nov. Stand Genomic Sci 2013;7:427–39.

[17] Nkamga VD, Huynh HTT, Aboudharam G, Ruimy R, Drancourt M. Diversity of human-associated *Methanobrevibacter smithii* isolates revealed by multispacer sequence typing. Curr Microbiol 2015;70:810–5.

[18] Le Page S, van Belkum A, Fulchiron C, Huguet R, Raoult D, Rolain JM. Evaluation of the PREVI® Isola automated seeder system compared to reference manual inoculation for antibiotic susceptibility testing by the disk diffusion method. Eur J Clin Microbiol Infect Dis 2015;34:1859–69.

[19] Sasser M. Bacterial identification by gas chromatographic analysis of fatty acids methyl esters (GC-FAME). Technical note 101. Newark, DE: MIDI; 2006.

[20] Dione N, Sankar SA, Lagier JC, Khelaifia S, Michele C, Armstrong N, et al. Genome sequence and description of *Anaerosalibacter massiliensis* sp. nov. New Microbes New Infect 2016;10:66–76.

[21] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol J Comput Mol Cell Biol 2012;19:455–77.

[22] Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:119.

[23] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955–64.

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

[25] Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol 2004;338:1027–36.

[26] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res 2011;39(Web Server issue):W347–52.

[27] Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 2014;42(Database issue):D206–14.

[28] Kent WJ. BLAT—the BLAST-like alignment tool. Genome Res 2002;12:656–64.

[29] Auch AF, von Jan M, Klenk HP, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci 2010;2:117–34.

[30] Gouret P, Thompson JD, Pontarotti P. PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinformatics 2009;10:298.

[31] Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics 2011;12:124.

[32] Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 2013;14:60.

[33] Gouret P, Paganini J, Dainat J, Louati D, Darbo E, et al. Integration of evolutionary biology concepts for functional annotation and automation of complex research in evolution: the multi-agent software system DAGOBAH. In: Springer-Verlag, editor. Evolutionary biology—concepts, biodiversity, macroevolution and genome evolution. Amsterdam: Springer-Verlag; 2011. p. 71–87.

[34] Gouret P, Vitiello V, Balandraud N, Gilles A, Pontarotti P, Danchin EG. FIGENIX: intelligent automation of genomic annotation: expertise integration in a new software platform. BMC Bioinformatics 2005;6:198.

[35] Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol 2014;64(Pt 2):346–51.

[36] Conway KR, Boddy CN. ClusterMine360: a database of microbial PKS/NRPS biosynthesis. Nucleic Acids Res 2013;41(Database issue):D402–7.

[37] Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 2014;58:212–20.

[38] Konstantinidis KT, Stackebrandt E. Defining taxonomic ranks. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E, editors. The Prokaryotes. New York: Springer; 2006.

[39] Auch AF, Klenk HP, Göker M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. Stand Genomic Sci 2010;2:142–8.

Article XII:

Noncontiguous finished genome sequence and description of *Paenibacillus ihumii* sp. nov. strain AT5.

Amadou Hamidou Togo, Saber Khelaifia, Jean-Christophe Lagier, Aurelia Caputo, Catherine Robert, Pierre-Edouard Fournier, Marie Maraninchi, René Valero, Didier Raoult and Matthieu Million.

Noncontiguous finished genome sequence and description of *Paenibacillus ihumii* sp. nov. strain AT5

A. H. Togo[1], S. Khelaifia[1], J.-C. Lagier[1], A. Caputo[1], C. Robert[1], P.-E. Fournier[1], M. Maraninchi[2], R. Valero[2,3], D. Raoult[1,4] and M. Million[1]

1) *Aix Marseille Université, URMITE, Institut Hospitalier Universitaire Méditerranée-Infection, UM63, CNRS7278, IRD198, INSERM1095,* 2) *Aix Marseille Université, NORT "Nutrition, Obesity and Risk of Thrombosis", INSERM1062, INRA1260,* 3) *APHM, CHU Hôpital de la Conception, Service Nutrition, Maladies Métaboliques et Endocrinologie, Marseille, France* and 4) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Paenibacillus ihumii sp. nov. strain AT5 (= CSUR 1981 = DSM 100664) is the type strain of *P. ihumii*. This bacterium was isolated from a stool sample from a morbidly obese French patient using the culturomics approach. The genome of this Gram-negative, facultative anaerobic, motile and spore-forming bacillus is 5 924 686 bp long. Genomic analysis identified 253 (5%) of 3812 genes as ORFans and at least 2599 (50.03%) of 5194 orthologous proteins not shared with the closest phylogenetic species.

New Microbes and New Infections © 2016 The Authors. Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases.

Corresponding author: M. Million, Aix Marseille Université, URMITE, Institut Hospitalier Universitaire Méditerranée-Infection, UM63, CNRS7278, IRD198, INSERM1095, Marseille, France
E-mail: matthieumillion@gmail.com

Introduction

Paenibacillus ihumii strain AT5 (= CSUR P1981 = DSM 100664) is the type strain of *P. ihumii* sp. nov. This isolate is part of an exploratory study of the gut flora from obese patients before and after bariatric surgery using a microbial culturomics approach, the aim of which is to exhaustively explore the microbial ecosystem of gut flora by using different culture conditions [1]. This bacterium was isolated from a stool sample collected before bariatric surgery from a 33-year-old Frenchwoman living in Marseille with morbid obesity.

The conventional parameters used in the delineation of bacterial species include 16S rRNA sequence identity and phylogeny [2], genomic (G + C content) diversity and DNA-DNA hybridization (DDH) [3,4]. However, these methods present some shortfalls, mainly due to their cutoff values, which vary according to species or genera [5]. The advent of new technology tools such as high-throughput sequencing has enabled us to access descriptions of many bacterial species in the public nucleotide sequence library [6]. Recently we proposed including genomic data in a polyphasic approach to describe new bacterial taxa (taxonogenomics). This strategy considers phenotypic characteristics, genomic analysis and the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) spectrum and comparison [7,8]. These characteristics support the circumscription and description of the species *P. ihumii* as a novel bacterium. Here we provide a brief classification and set of characteristics for *P. ihumii* sp. nov. strain AT5 alongside the description of the complete genome sequencing and annotation.

Materials and Methods

Sample collection

A stool sample was collected from a 33-year-old obese Frenchwoman (body mass index 38.6 kg/m²; 100 kg, 1.61 m

tall) in November 2011. Written consent was obtained from the patient at the Nutrition, Metabolic Disease and Endocrinology service at Timone Hospital, Marseille, France. The study and consent procedures were approved by the local IFR 48 ethics committee under consent 09-022, 2010. The stool sample was stored at −80°C after collection and studied using microbial culturomics as previously reported [1].

Isolation and identification of strain

Growth of *P. ihumii* strain AT5 was performed in May 2015. The sterile stool extract was preincubated in blood culture bottles enriched with rumen fluid and sheep's blood as described elsewhere [1]. The culture was monitored for 30 days. On various days (days 1, 3, 7, 10, 15, 21 and 30), a seeding of the preincubated product was done on sheep's blood−enriched Columbia agar (bioMérieux, Marcy l'Etoile, France) and incubated for 24 hours in an aerobic atmosphere at 37°C. The colonies that emerged were cultivated under the same conditions for isolation. They were then identified by MALDI-TOF as described elsewhere [9]. In short, one isolated bacterial colony was picked up with a pipette tip from a culture agar plate and spread as a thin smear on an MTP 384 MALDI-TOF target plate (Bruker Daltonics, Leipzig, Germany). Each smear was overlaid with 2 μL of matrix solution (saturated solution of α-cyano-4-hydroxycinnamic acid in 50% acetonitrile, 2.5% trifluoroacetic acid) and allowed to dry for 5 minutes. Measurements were performed with a Microflex spectrometer (Bruker). Spectra were recorded in the positive linear mode for the mass range from 2000 to 20 000 Da (parameter settings: ion source 1 (ISI), 20 kV; IS2, 18.5 kV; lens, 7 kV). A spectrum was obtained after 675 shots with variable laser power. The time of acquisition was between 30 seconds and 1 minute per spot. Identification was carried out as previously reported [10].

The 16S rRNA PCR, coupled with sequencing, was performed using GeneAmp PCR System 2720 thermal cyclers (Applied Biosystems, Foster City, CA, USA) and ABI Prism 3130xl Genetic Analyser capillary sequencer (Applied Biosystems) respectively [11]. Chromas Pro 1.34 software (Technelysium, Tewantin, Australia) was used to correct sequences, and BLASTn searches were performed in National Center for Biotechnology Information (NCBI; http://blast.ncbi.nlm.nih.gov.gate1.inist.fr/Blast.cgi). Sequences were then aligned using Clustal W, and phylogenetic inferences were obtained using the neighbour-joining method with MEGA 6 (Molecular Evolutionary Genetics Analysis version 6) software. The numbers of nodes correspond to the percentages of bootstrap values obtained by repeating the analysis 1000 times in order to generate a consensus tree.

Growth conditions

Different growth temperatures (25, 28, 37, 45 and 55°C) were tested on sheep's blood−enriched Columbia agar (bioMérieux). Growth of this strain was tested in an anaerobic atmosphere using the GENbag anaer system (bioMérieux), in a microaerophilic atmosphere using GENbag microaer system (bioMérieux) and in an aerobic atmosphere with or without 5% CO_2. The saltiness of this species was tested using 5% NaCl on Schaedler agar with 5% sheep's blood (bioMérieux) in an aerobic atmosphere.

Biochemical, sporulation and motility assays

Biochemical assays were performed using API Gallery systems: API ZYM (bioMérieux), API 20NE (bioMérieux) and API50 CH (bioMérieux). Detection of catalase (bioMérieux) and oxidase (Becton Dickinson, Franklin Lakes, NJ, USA) was also conducted. A thermal shock at 100°C for half an hour was carried out in order to test sporulation. A fresh colony was observed between blades and slats using a Leica DM 1000 photonic microscope (Leica Microsystems, Wetzlar, Germany) at 40× to assess the motility of the bacteria.

Microscopy

Transmission electron microscopy using a Tecnai G20 device (FEI Company, Limeil-Brevannes, France) at an operating voltage of 60 kV was performed to observe the *P. ihumii* strain AT5 after negative colouration. Gram staining was performed and observed using a Leica DM 2500 photonic microscope with a 100× oil-immersion objective lens.

Antibiotic susceptibility

Antibiotic susceptibility of the strain was tested using 18 antibiotics, including amoxicillin 25 μg, amoxicillin 20 μg/clavulanic acid 10 μg, cefalexin 30 μg, ceftriaxone 30 μg, ciprofloxacin 5 μg, doxycycline 30 IU, erythromycin 15 IU, nitrofurantoin 300 μg, gentamicin 500 μg, gentamicin 15 μg, imipenem 10 μg, metronidazole 4 μg, oxacillin 5 μg, penicillin G 10 IU, rifampicin 30 μg, trimethoprim 1.25 μg/sulfamethoxazole 23.75 μg, tobramycin 10 μg and vancomycin 30 μg (i2a, Montpellier, France). The Scan 1200 was used to interpret the results (Interscience, Saint-Nom-La-Bretèche, France).

Genome sequencing and assembly

P. ihumii genomic DNA (gDNA) was sequenced on the MiSeq Technology (Illumina, San Diego, CA, USA) using 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 gDNA was quantified by a Qubit assay using the high sensitivity kit (Thermo Fisher Scientific, Waltham, MA) to 145 ng/μL. The mate pair library was prepared with 1 μg of gDNA using the Nextera mate pair Illumina guide. The gDNA

sample was simultaneously fragmented and tagged with a mate pair junction adapter. The pattern of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) using a DNA 7500 lab chip. DNA fragments were ranged in size from 1.5 to 11 kb, with an optimal size of 3.987 kb. No size selection was performed, and only 334 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared into small fragments with an optimal size of 1051 bp on the Covaris device S2 in T6 tubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity BioAnalyzer LabChip (Agilent), and the final concentration library was measured at 2.90 nmol/L. The libraries were normalized at 2 nM and pooled. Following a denaturation step and dilution at 15 pM, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and a sequencing run were performed in a single 2 × 301 bp run. In total, 7.3 Gb of information was obtained from a 511K/mm^2 cluster density, with a cluster passing quality control filters of 97.0% (12 079 000 passing filter paired reads). Within this run, the index representation for *P. ihumii* was determined to 10.02%. The 1 210 259 paired reads were trimmed and assembled to 12 scaffolds using the SPAdes software [12].

Genome annotation and comparison

Open reading frames (ORFs) were predicted using Prodigal [13] with default parameters. Nevertheless, the predicted ORFs were excluded if they spanned a sequencing gap region (containing N). The predicted bacterial protein sequences were searched against the GenBank database and the Clusters of Orthologous Groups (COGs) database using BLASTP (*E* value 1e-03, coverage 0.7 and 30% identity). If no hit was found, it searched against the NR database using BLASTP with an *E* value of 1e-03, coverage 0.7 and 30% identity. If the sequence length was smaller than 80 amino acids, we used an *E* value of 1e-05. The tRNAs and rRNAs were predicted using the tRNA Scan-SE and RNAmmer tools respectively [14,15]. SignalP and TMHMM were used to foresee the signal peptides and the number of transmembrane helices respectively [16,17]. Mobile genetic elements were predicted using PHAST and RAST [18,19]. ORFans were identified if their BLASTP *E* value was lower than 1e-03 for an alignment length greater than 80 amino acids. If alignment lengths were smaller than 80 amino acids, we used an *E* value of 1e-05. Artemis and DNA Plotter were used for data management and visualization of genomic features respectively [20,21]. Genomes were automatically retrieved from the 16S rRNA tree using Xegen software (PhyloPattern) [22]. For each selected genome, complete genome sequence, proteome genome sequence and Orfeome genome sequence were retrieved from the FTP site of National Center for

Biotechnology Information (NCBI). All proteomes were analysed using proteinOrtho [23]. A similarity score was then computed for each pair of genomes. This score is the mean value of nucleotide similarity between all orthologous pairs in the two genomes studied (average genomic identity of orthologous gene sequences, AGIOS) [6]. For the genomic comparison, we used *P. ihumii* strain AT5 (CYXK00000000), *P. fonticola* strain DSM21315 (ARMT00000000), *P. peoriae* strain KCTC 3763 (CP011512), *P. stellifer* strain DSM 14472 (CP009286), *P. terrae* strain HPL-003 (CP003107) and *P. borealis* strain DSM 13188 (CP009285). An annotation of the entire proteome was performed to define the distribution of functional classes of predicted genes according to the clusters of orthologous groups of proteins (using the same method as for the genome annotation). *P. ihumii* genome was locally aligned two by two by using the BLAT algorithm [24,25] against each of the selected genomes previously cited, and DDH values were estimated from a generalized model. The DDH threshold is less than 70%

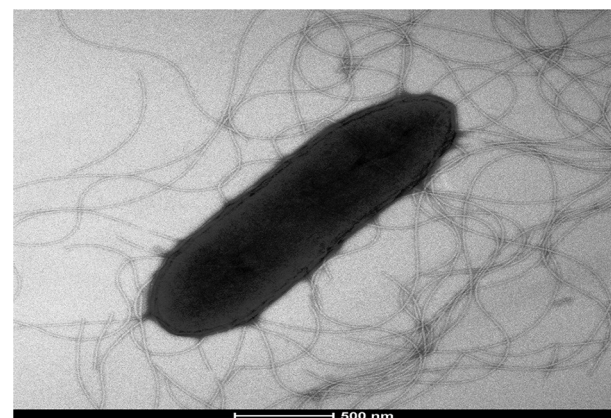**FIG. 1.** Gram staining of *Paenibacillus ihumii* strain AT5.

FIG. 2. Transmission electron microscopy of *Paenibacillus ihumii* strain AT5 using Tecnai G20 (FEI Company) at operating voltage of 60 kV. Scale bar = 500 nm.

Property	Term
Current classification	Domain: *Bacteria*
	Phylum: *Firmicutes*
	Class: *Bacilli*
	Order: *Bacillales*
	Family: *Paenibacillaceae*
	Genus: *Paenibacillus*
	Species: *P. ihumii*
	Type strain: AT5
Gram stain	Negative
Cell shape	Rod
Motility	Motile
Sporulation	Sporulating
Temperature range	Mesophile
Optimum temperature	37°C
Oxygen requirement	Facultative anaerobic
Carbon source	Unknown
Energy source	Unknown
Habitat	Human gut
Biotic relationship	Free living
Pathogenicity	Unknown
Isolation	Human faeces

for a species to be considered as new species [26]. Annotation and comparison processes were performed using the Multi-Agent software system DAGOBAH [27], including Figenix [28] libraries that provide pipeline analysis.

Results

Phenotypic and biochemical characterization

The *P. ihumii* strain AT5 is a Gram-negative motile rod which is catalase and oxidase negative (Fig. 1). The growth of the strain occurred between 28 to 55°C, but optimal growth was observed at 37°C after 24 hours of incubation in an aerobic atmosphere. The colonies were approximately 1 to 2 mm in diameter and grey on 5% sheep's blood−enriched Columbia agar. Cells had a diameter ranging from 0.50 to 1.75 μm, with a mean diameter of 1 μm measured using electron microscopy (Fig. 2). No growth of this bacterium was observed using 5% NaCl on Schaedler agar with 5% sheep's blood. This bacterium is a facultative anaerobe bacillus but can also grow in a microaerophilic atmosphere. The strain was able to form spores. Table 1 summarizes the classification and main features of *P. ihumii*.

Of all the 18 antibiotics tested, the *P. ihumii* strain AT5 was susceptible to all of them except metronidazole and tobramycin. Using an API ZYM strip, we observed that *P. ihumii* possesses alkaline phosphatase, esterase (C4), esterase lipase (C8), naphthol-AS-BI-phosphohydrolase, α-galactosidase (melibase), β-galactosidase (hydrolase), α-glucosidase (maltase) and β-glucosidase (cellulose) activities. However, there are no activities of lipase (C14), leucin arylamidase, valine arylamidase, cystine arylamidase, trypsin, α-chymotrypsin, phosphatase acid, β-glucuronidase, N-acetyl-β-glucosaminidase, α-mannosidase and α-fucosidase. Using an API 20 NE strip, positive reactions were obtained for nitrate reduction, β-galactosidase, arabinose, β-glucosidase, mannose, mannitol, N-acetyl-glucosamine and maltose. Negative reactions were obtained for L-tryptophan, D-glucose, L-arginine, urea, gelatine, potassium gluconate, capric acid, adipic acid, malic acid, trisodium citrate and phenyl acetic acid. Using an

TABLE 2. **Differential characteristics of *Paenibacillus ihumii* strain AT5, *Paenibacillus borealis*, *Paenibacillus fonticola*, *Paenibacillus peoriae*, *Paenibacillus stellifer* and *Paenibacillus terrae***

| Property	*P. ihumii*	*P. borealis*	*P. fonticola*	*P. peoriae*	*P. stellifer*	*P. terrae*
Cell diameter width/length (μm)	0.5/1.75	0.7−1/3−5	0.8−1/2−12.4	0.5−1/3−6	0.6−0.8/2.5−5	1.3−1.8/4−7
Oxygen requirement	+/−	+/−	+/−	+/−	+/−	+/−
Gram stain	−	−	v	+	+	v
Growth with NaCl 5%	−	−	−	−	−	−
Motility	+	+	+	+	+	+
Spore formation	+	+	+	+	+	+
Production of:						
Catalase	−	+	+	+	+	+
Oxidase	−	−	−	−	−	−
Nitrate reductase	+	−	−	v	−	+
Urease	−	NA	+	−	NA	−
β-Galactosidase	+	NA	+	NA	NA	+
N-acetyl-glucosamine	+	+	−	v	−	+
Utilization of:						
L-Arabinose	+	+	+	+	+	+
D-Ribose	+	+	−	+	−	+
D-Mannose	+	+	−	+	+	+
D-Mannitol	+	NA	−	+	−	+
D-Glucose	+	+	−	+	−	+
D-Fructose	+	+	−	+	+	+
D-Maltose	+	+	−	+	+	+
D-Lactose	+	+	−	+	NA	+
D-Xylose	+	+	−	+	+	+
Habitat	Human gut	Spruce forest humus	Warm springs	Soil/rotting vegetation	Food-packing board	Soil
Genome size	5.92	8.16	6.30	5.77	5.66	6.08
DNA G + C content (mol%)	50.2	51.39	47.68	46.44	53.54	46.77

+, positive result; −, negative result; +/−, facultative anaerobic; v, variable result; NA, data not available.

API 50 CH strip, we demonstrated that *P. ihumii* is able to ferment L-arabinose, D-ribose, D-xylose, methyl-βD-xylopranoside, D-galactose, D-glucose, D-fructose, D-mannose, D-mannitol, methyl-αD-glucopyranoside, N-acetylglucosamine, amygdalin, arbutin, esculin ferric citrate, salicin, D-cellobiose, D-maltose, D-lactose, D-melibiose, D-saccharose, D-trehalose, D-raffinose, starch, glycogen, gentiobiose and D-lyxose. No fermentation were recorded for glycerol, erythritol, D-arabinose, L-xylose, D-adonitol, L-sorbose, L-rhamnose, dulcitol, inositol, D-sorbitol, methyl-α-D-mannopyranoside, inulin, D-melezitose, xylitol, D-turanose, D-tagatose, D-fucose, L-fucose, D-arabitol, L-arabitol, potassium gluconate, potassium 2-ketogluconate and potassium 5-ketogluconate. Table 2 presents a comparison of the different characteristics with other representatives of the *Paenibacillus* genus.

Phylogenetic analysis

No match could be found from the spectrum generated from clean *P. ihumii* strain AT5 spots and those in the Bruker database (Fig. 3). The phylogenetic analysis, performed using 16S rRNA sequences, showed that *P. ihumii* sp. nov. strain AT5 exhibited 98.2% identity with *Paenibacillus lentus* [29], classified in the *Paenibacillaceae* family created by Ash in 1993 [30]. However, this percentage remains lower than the 98.7% 16S rRNA gene sequence threshold recommended by Stackebrandt and Ebers [2] to delineate a new species. This enables us to say that the *P. ihumii* strain AT5 is a new species within the *Paenibacillaceae* family (Table 1). A neighbour-joining phylogenetic tree (Fig. 4) based on 16S rRNA gene sequences shows the relationships between *P. ihumii* and some related taxa. The
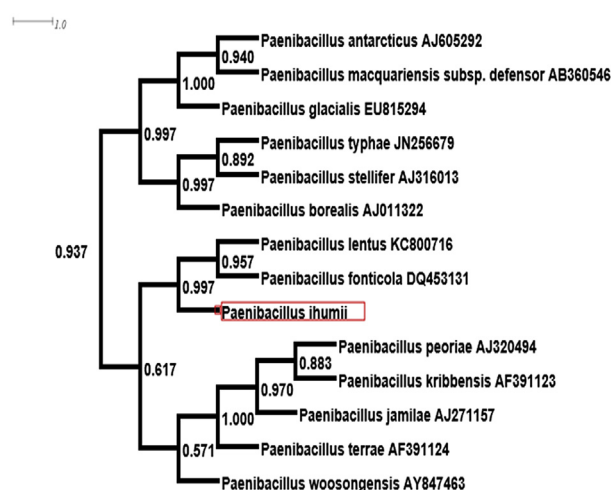
FIG. 4. Phylogenetic tree highlighting position of *Paenibacillus ihumii* strain AT5 relative to other close species. Sequences are recovered using nucleotide blast against 16S rRNA Database of Silva 'All-Species Living Tree' project (LTPs119). Sequences were aligned using muscle and phylogenetic inferences obtained using approximately maximum likelihood method within Fast Tree software. Numbers at nodes are support local values computed using Shimodaira-Hasegawa test. Corresponding GenBank accession numbers for 16S rRNA genes are indicated at right of strains in tree.

P. ihumii 16S rRNA sequence was deposited in European Molecular Biology Laboratory−European Bioinformatics Institute (EMBL-EBI) under accession number LN881615. A gel view was performed in order to observe spectra differences between *P. ihumii* and other close bacteria (Fig. 5).
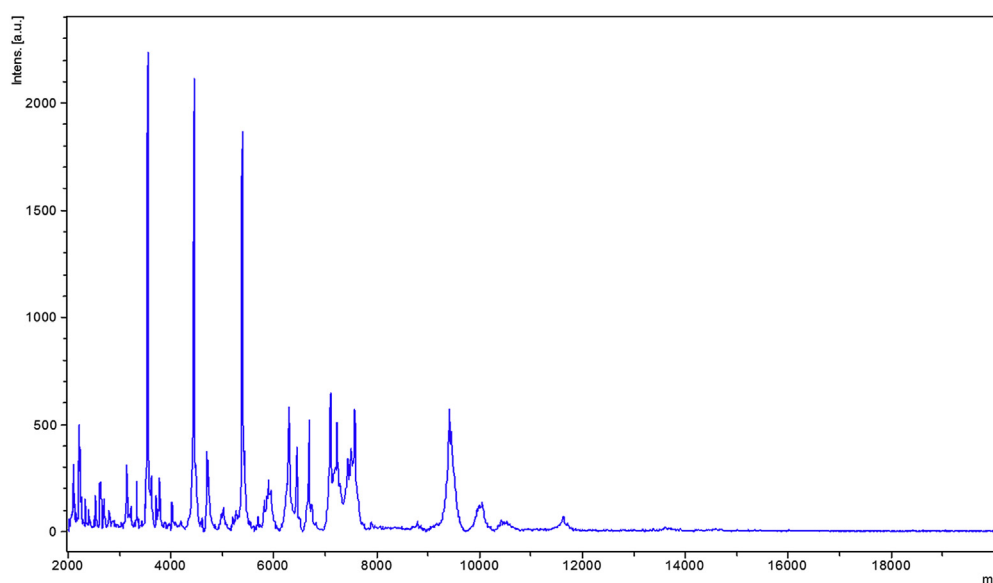
FIG. 3. MALDI-TOF reference mass spectrum from *Paenibacillus ihumii* strain AT5. MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.
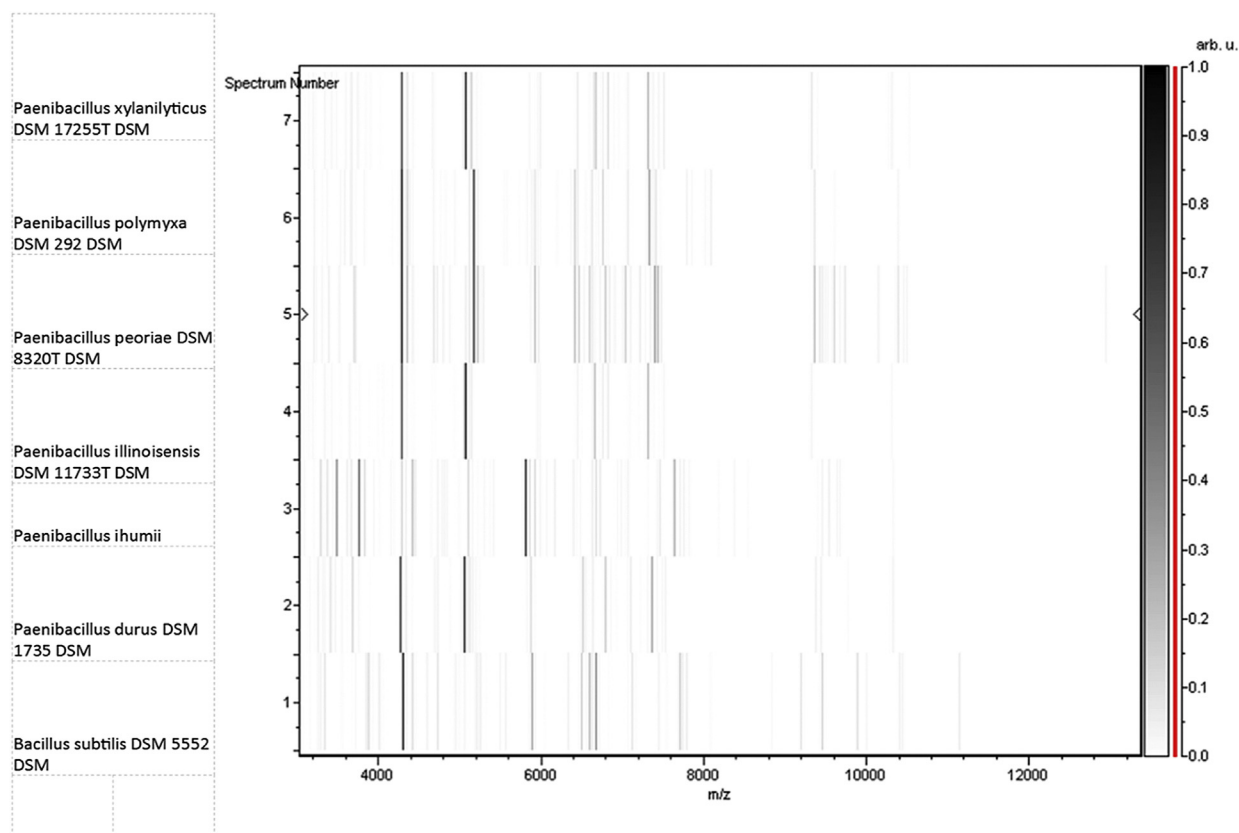
FIG. 5. Gel view comparing *Paenibacillus ihumii* strain AT5 to other species. Gel view displays raw spectra of loaded spectrum files arranged in pseudo-gel-like look. *x*-axis records *m/z* value. Left *y*-axis displays running spectrum number originating from subsequent spectra loading. Peak intensity is expressed by greyscale scheme code. The right *y*-axis indicates the peak intensity according to the colour of this peak, in arbitrary units. Displayed species are indicated on left.

Genome properties

The genome of *P. ihumii* strain AT5 (Fig. 6) is 5 924 686 bp long with 50.20% G + C content (Table 3). It consists of 12 scaffolds (composed of 16 contigs). Of the 5274 predicted genes, 5194 were protein-coding genes and 80 were RNAs (six 5S rRNA, five 16S rRNA, two 23S rRNA, 67 tRNA). A total of 3812 genes (73.39%) were assigned as having a putative function (by cogs or by NR blast), and 253 genes (4.87%) were identified as ORFans. The remaining genes (896 genes, 17.25%) were annotated as hypothetical proteins. Using ARG-ANNOT [31], no resistance genes were found. Nevertheless, 15 genes associated to polyketide synthase or nonribosomal peptide synthetase [32] were discovered through genome analysis and implicated in the production of secondary metabolites. The distribution of genes into COGs functional categories is presented in Table 4.

Genome comparison

P. borealis, *P. fonticola*, *P. peoriae*, *P. stellifer* and *P. terrae* are species closely related to *P. ihumii* with available genomes (Table 1) and were consequently chosen for this comparative analysis. The G + C content of *P. ihumii* is smaller than that of
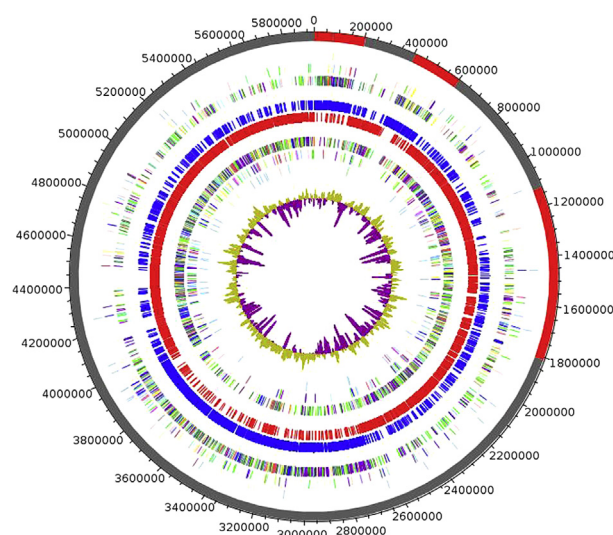
FIG. 6. Circular graphical map of genome. From outside to centre: Contigs (red/grey), COGs category of genes on forward strand (three circles), genes on forward strand (blue circle), genes on reverse strand (red circle), COGs category on reverse strand (three circles), G + C content. COGs, Clusters of Orthologous Groups database.

TABLE 3. Nucleotide content and gene count levels of the chromosome

	Genome (total)	
Attribute	Value	% of total
Size (bp)	5 924 686	100
G+C content (bp)	2 973 782	50.20
Coding region (bp)	5 065 111	85.49
Extrachromosomal elements	0	0
Total genes	5274	100
RNA genes	80	1.51
Protein-coding genes	5194	98.48
Genes with function prediction	3812	73.39
Genes assigned to COGs	3881	74.72
Genes with peptide signals	743	14.30
Genes with transmembrane helices	1327	25.54

COGs, Clusters of Orthologous Groups database.

P. borealis and *P. stellifer* (50.20, 51.39 and 53.54% respectively) but larger than that of *P. fonticola*, *P. peoriae* and *P. terrae* (47.68, 46.44 and 46.77% respectively). The gene content of *P. ihumii* is smaller than that of *P. fonticola*, *P. borealis* and *P. terrae* (5194, 5645, 6213 and 5525 respectively) but larger than that of *P. peoriae* and *P. stellifer* (5122 and 4464 respectively). Fig. 7 shows that the distribution of genes into COGs categories was similar across all compared genomes. In addition, *P. ihumii* shared 2595, 2141, 1998, 2217 and 2414 orthologous genes with *P. fonticola, P. peoriae, P. stellifer, P. terrae* and *P. borealis* respectively (Table 5). The AGIOS values ranged from 57.59 to 87.31% among compared *Paenibacillus* species with the exception of *P. ihumii*. When *P. ihumii* was compared to other *Paenibacillus* species, the AGIOS value ranged from 57.88% for *P. fonticola* to 67.99% for *P. peoriae* (Table 5). DDH was 28% ± 2.43 for *P. fonticola*, 17.5% ± 2.23 for *P. peoriae,* 18.2% ± 2.26 for *P. stellifer,* 17.6% ± 2.23 for *P. terrae* and 17.6% ± 2.24 for *P. borealis* (Table 6). These data confirm *P. ihumii* as a unique species.

Conclusion

On the basis of phenotypic, genomic and phylogenetic analyses, we formally propose the creation of *P. ihumii* sp. nov., which contains the strain AT5. This bacterium was isolated from a stool sample of a 33-year-old morbidly obese Frenchwoman living in Marseille.

Taxonomic and Nomenclatural Proposals

Description of *P. ihumii* strain AT5 sp. nov.

Paenibacillus ihumii (i.hum.i'i. N.L. gen. n. *ihumii,* based on the acronym IHUMI, the Institut Hospitalo-Universitaire Méditerranée-Infection in Marseille, France, where the type strain was isolated).

Cells are Gram-negative, spore-forming, motile, rod-shaped bacilli with a size of 0.5−1.75 μm. Colonies are grey, with a diameter of 1−2 mm on 5% sheep's blood−enriched Columbia agar. The strain is catalase and oxidase negative. It has an optimum growth temperature of 37°C and is a facultative anaerobe, able to grow in a microaerophilic atmosphere.

Using API Gallery systems, positive reactions were observed for alkaline phosphatase, esterase (C4), esterase lipase (C8), naphthol-AS-BI-phosphohydrolase, α-galactosidase, β-galactosidase, α-glucosidase, β-glucosidase, nitrate reduction, β-galactosidase, mannose, mannitol, N-acetylglucosamine, arabinose, maltose, L-arabinose, D-ribose, D-xylose, methyl-βD-xylopranoside, D-galactose, D-glucose, D-fructose, D-mannose, D-mannitol, methyl-αD-glucopyranoside, N-acetylglucosamine, amygdalin, arbutin, esculin ferric citrate, salicin, D-cellobiose, D-maltose, D-lactose, D-melibiose, D-saccharose, D-trehalose, D-raffinose, starch, glycogen, gentiobiose and D-lyxose.

Cells are susceptible to amoxicillin, amoxicillin/clavulanic acid, cefalexin, ceftriaxone, ciprofloxacin, doxycycline, erythromycin, nitrofurantoin, gentamicin, imipenem, oxacillin, penicillin G, rifampicin, trimethoprim/sulfamethoxazole and vancomycin, and are resistant to metronidazole and tobramycin.

The length of the genome is 5 924 686 bp with 50% G + C content. The 16S rRNA gene sequence and whole-genome shotgun sequence of the *P. ihumii* strain AT5 were deposited

TABLE 4. Number of genes associated with 26 general COGs functional categories

Code	Value	% of total	Description
J	271	5.21	Translation
A	0	0	RNA processing and modification
K	462	8.89	Transcription
L	133	2.56	Replication, recombination and repair
B	1	0.01	Chromatin structure and dynamics
D	64	1.23	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	165	3.17	Defence mechanisms
T	281	5.41	Signal transduction mechanisms
M	214	4.12	Cell wall/membrane biogenesis
N	81	1.55	Cell motility
Z	9	0.17	Cytoskeleton
W	21	0.40	Extracellular structures
U	49	0.94	Intracellular and trafficking secretion
O	159	3.06	Post-translational modification, protein turnover, chaperones
C	150	2.88	Energy production and conversion
G	623	11.99	Carbohydrate transport and metabolism
E	269	5.17	Amino acid transport and metabolism
F	110	2.11	Nucleotide transport and metabolism
H	211	4.06	Coenzyme transport and metabolism
I	125	2.40	Lipid transport and metabolism
P	212	4.08	Inorganic ion transport and metabolism
Q	95	1.82	Secondary metabolites biosynthesis, transport and catabolism
R	399	7.68	General function prediction only
S	233	4.48	Function unknown
—	1313	25.27	Not in COGs

COGs, Clusters of Orthologous Groups database.
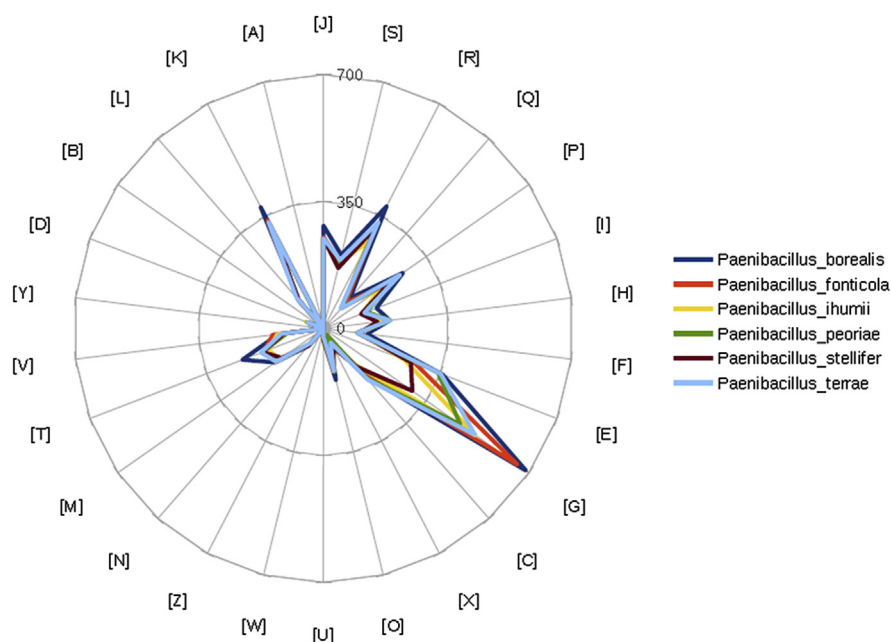
FIG. 7. Distribution of functional classes of predicted genes according to clusters of orthologous groups of proteins.

TABLE 5. Numbers of orthologous proteins shared between genomes (upper right), average percentage similarity of nucleotides corresponding to orthologous protein shared between genomes (lower left) and numbers of proteins per genome (bold)

	Paenibacillus ihumii	Paenibacillus fonticola	Paenibacillus peoriae	Paenibacillus stellifer	Paenibacillus terrae	Paenibacillus borealis
P. ihumii	**5194**	2595	2141	2328	2217	2414
P. fonticola	57.88	**5645**	2249	2067	2328	2517
P. peoriae	67.99	67.76	**5122**	2084	2792	2448
P. stellifer	58.11	57.59	57.65	**4464**	2141	2345
P. terrae	67.96	67.73	87.31	57.69	**5525**	2549
P. borealis	58.25	57.88	58.20	72.26	58.19	**6213**

TABLE 6. Pairwise comparison of Paenibacillus ihumii with other species using GGDC, formula 2 (DDH estimates based on identities/HSP length)[a]

	Paenibacillus ihumii	Paenibacillus fonticola	Paenibacillus peoriae	Paenibacillus stellifer	Paenibacillus terrae	Paenibacillus borealis
P. ihumii	100%	28% ± 2.43	17.5% ± 2.23	18.2% ± 2.26	17.6% ± 2.23	17.6% ± 2.24
P. fonticola		100%	16.8% ± 2.21	17.1% ± 2.22	17.2% ± 2.22	17% ± 2.22
P. peoriae			100%	18.2% ± 2.26	33.7% ± 2.47	18.1% ± 2.75
P. stellifer				100%	18.3% ± 2.26	19.7% ± 2.3
P. terrae					100%	18.1% ± 2.25
P. borealis						100%

[a]Confidence intervals indicate inherent uncertainty in estimating DDH values from intergenomic distances based on models derived from empirical test data sets (which are always limited in size). These results are in accordance with the 16S rRNA and phylogenomic analyses as well as the GGDC results. DDH, DNA-DNA hybridization; GGDC, Genome-to-Genome Distance Calculator; HSP, high-scoring segment pairs.

in EMBL-EBI under accession numbers LN881615 and CYXK00000000 respectively.

The AT5 type strain (= CSUR P1981 = DSM 100664) was isolated from a stool sample from an obese Frenchwoman. The habitat of this microorganism is the human digestive tract.

Acknowledgements

Conflict of Interest

None declared.

References

[1] Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 2012;18:1185–93.

[2] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152–5.

[3] Stackebrandt E, Frederiksen W, Garrity GM, Grimont PA, Kämpfer P, Maiden MC, et al. Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Microbiol 2002;52:1043–7.

[4] Rosselló-Mora R. DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation. In: Stackebrandt E, editor. Molecular identification, systematics, and population structure of prokaryotes. Berlin: Springer Verlag; 2006. p. 23–50.

[5] Welker M, Moore ERB. Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. Syst Appl Microbiol 2011;34:2–11.

[6] Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi M, Sentausa E, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol 2014;64(pt 2):384–91.

[7] Kokcha S, Mishra AK, Lagier JC, Million M, Leroy Q, Raoult D, et al. Non contiguous-finished genome sequence and description of *Bacillus timonensis* sp. nov. Stand Genomic Sci 2012;6:346–55.

[8] Mishra AK, Lagier JC, Nguyen TT, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Peptoniphilus senegalensis* sp. nov. Stand Genomic Sci 2013;7:370–81.

[9] Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 2009;49:543–51.

[10] Hugon P, Ramasamy D, Lagier JC, Rivet R, Couderc C, Raoult D, et al. Non contiguous-finished genome sequence and description of *Alistipes obesi* sp. nov. Stand Genomic Sci 2013;7:427–39.

[11] Nkamga VD, Huynh HTT, Aboudharam G, Ruimy R, Drancourt M. Diversity of human-associated *Methanobrevibacter smithii* isolates revealed by multispacer sequence typing. Curr Microbiol 2015;70:810–5.

[12] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 2012;19:455–77.

[13] Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:119.

[14] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955–64.

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

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

[17] Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001;305:567–80.

[18] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res 2011;39(Web Server issue): W347–52.

[19] Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 2014;42(Database issue):D206–14.

[20] Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 2012;28:464–9.

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

[22] Gouret P, Thompson JD, Pontarotti P. PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinformatics 2009;10:298.

[23] Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics 2011;12:124.

[24] Kent WJ. BLAT—the BLAST-like alignment tool. Genome Res 2002;12:656–64.

[25] Auch AF, von Jan M, Klenk HP, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci 2010;2:117–34.

[26] Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 2013;14:60.

[27] Gouret P, Paganini J, Dainat J, Louati D, Darbo E, Pontarotti P, et al. Integration of evolutionary biology concepts for functional annotation and automation of complex research in evolution: the multi-agent software system DAGOBAH. In: Pontarotti P, editor. Evolutionary biology—concepts, biodiversity, macroevolution and genome evolution. Berlin: Springer Verlag; 2011. p. 71–87.

[28] Gouret P, Vitiello V, Balandraud N, Gilles A, Pontarotti P, Danchin EG. FIGENIX: intelligent automation of genomic annotation: expertise integration in a new software platform. BMC Bioinformatics 2005;6:198.

[29] Li YF, Calley JN, Ebert PJ, Helmes EB. *Paenibacillus lentus* sp. nov., a β-mannanolytic bacterium isolated from mixed soil samples in a selective enrichment using guar gum as the sole carbon source. Int J Syst Evol Microbiol 2014;64(pt 4):1166–72.

[30] Ash C, Priest FG, Collins MD. Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Proposal for the creation of a new genus *Paenibacillus*. Antonie Van Leeuwenhoek 1993–1994;64:253–60.

[31] Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 2014;58:212–20.

[32] Conway KR, Boddy CN. ClusterMine360: a database of microbial PKS/ NRPS biosynthesis. Nucleic Acids Res 2013;41(Database issue): D402–7.

Article XIII:

Draft genome and description of *Eisenbergiella massiliensis* strain AT11[T]: A new species isolated from human feces after bariatric surgery.

Amadou H. Togo, Awa Diop, Matthieu Million, Marie Maraninchi, Jean-Christophe Lagier,

Catherine Robert, Fabrizio Di Pinto, Didier Raoult, Pierre-Edouard Fournier and Fadi Bittar

CrossMark

Draft Genome and Description of *Eisenbergiella massiliensis* Strain AT11[T]: A New Species Isolated from Human Feces After Bariatric Surgery

Amadou H. Togo[1] · Awa Diop[2] · Matthieu Million[1] · Marie Maraninchi[3] · Jean-Christophe Lagier[1] · Catherine Robert[2] · Fabrizio Di Pinto[1] · Didier Raoult[1] · Pierre-Edouard Fournier[2] · Fadi Bittar[1,4]

Abstract

A novel strain of a Gram-stain negative, non-motile, non-spore forming rod-shaped, obligate anaerobic bacterium, designated AT11[T], was isolated from a stool sample of a morbidly obese woman living in Marseille, France. This bacterium was characterized using biochemical, chemotaxonomic, and phylogenetic methods. The 16S rRNA gene sequence analysis showed that strain AT11[T] had a 97.8% nucleotide sequence similarity with *Eisenbergiella tayi* strain B086562[T], the closest species with standing in nomenclature. The major cellular fatty acids of the novel isolate were $C_{16:0}$ followed by saturated or unsaturated C_{18} fatty acids ($C_{18:1}n9$, $C_{18:1}n5$ and $C_{18:0}$). The draft genome of strain AT11[T] is 7,114,554 bp long with 48% G+C content. 6176 genes were predicted, including 6114 protein-coding genes and 62 were RNAs (with 2 5S rRNA genes, two 16S rRNA genes, two 23S rRNA genes, and 56 tRNA genes). The digital DNA–DNA hybridization (dDDH) relatedness between the new isolate and *E. tayi* strain B086562[T] was 23.1% ± 2.2. Based on the phenotypic, chemotaxonomic, genomic, and phylogenetic characteristics, *Eisenbergiella massiliensis* sp. nov., is proposed. The type strain is AT11[T] (= DSM 100838[T] = CSUR P2478[T]).

Introduction

The number of people suffering from obesity has increased in recent decades [25]. It has been well established that the gut microbiota contributes to the development of human metabolic disorders such as obesity [18, 24]. Bariatric surgery is the most effective treatment for morbid obesity. It

induces a sustainable weight loss, improves complications related to obesity, and increases the diversity of the gut flora [14, 34].

We conducted a study comparing the gut microbiota from obese patients before and after bariatric surgery using a new microbial high-throughput culture approach known as culturomics [16]. This new approach makes it possible to isolate and describe the living microbial diversity of any environmental and clinical sample. Using culturomics, we isolated a new anaerobic bacterium, strain AT11[T], from a stool sample harvested following bariatric surgery. The discovery of this bacterium has been previously reported as a new species announcement without a thorough description [31].

Herein, strain AT11[T] was analyzed by a polyphasic approach in order to describe it as a new bacterial taxon. This combines phenotypic characteristics, the matrix laser desorption ionization-time of flight mass spectrometry (MALDI-TO MS) spectrum, and genomic properties known as taxono-genomics [27].

Here, we propose a classification and a set of phenotypic, chemical, and chemotaxonomic characteristics of a new bacterial species: strain AT11[T], which belongs to

✉ Fadi Bittar
fadi.bittar@univ-amu.fr

[1] Aix Marseille Univ, IRD, APHM, MEPHI, IHU-Méditerranée Infection, Marseille, France

[2] Aix Marseille Univ, IRD, APHM, SSA, VITROME, IHU-Méditerranée Infection, Marseille, France

[3] Aix Marseille Univ, NORT "Nutrition, Obesity and Risk of Thrombosis", INSERM1062, INRA1260, 13385 Marseille, France

[4] IHU-Méditerranée Infection, 19-21 Bd Jean Moulin, 13005 Marseille, France

Springer

the genus *Eisenbergiella* [1], together with the description of the complete genome sequencing, annotation, and genomic comparison. To date, this genus includes only one species *Eisenbergiella tayi*, the type strain B086562^T (= LMG 27400^T = DSM 26961^T = ATCC BAA-2558^T) as reported in List of Prokaryotic Names with Standing in Nomenclature (http://www.bacterio.net/ruminococcus.html).

Materials and Methods

Ethics and Sample Collection

Once informed consent had been obtained, stool samples were collected before and after surgery. These samples were obtained from a 56-year-old obese French woman following bariatric surgery on April 27, 2011. All samples were stored at − 80 °C before culturing. The study and the assent procedure were approved by the local ethics committee of IFR 48, under ascent number 09-022, 2010.

Isolation and Identification of the Strain

Strain AT11^T was first grown on July 22, 2015. One gram of stool was pre-incubated in BD BACTEC™ Lytic/10 Anaerobic/F Culture Vials media culture bottles (Becton, Dickenson and Company, Le Pont de Claix, France) enriched with 4 ml of filtered rumen juice and 4 ml of sheep blood. The pre-incubated product was cultured on 5% sheep blood-enriched Columbia agar (bioMérieux, Marcy l'Etoile, France) as described elsewhere [32]. This strain was isolated 21 days after pre-incubation. The resulting colonies were then identified using MALDI-TOF mass spectrometry (Bruker Daltonics, Leipzig, Germany) as previously described [29]. When the spectra of a bacterium are not identified by MALDI-TOF MS screening, 16S rRNA gene amplification and sequencing is performed.

Phylogenetic Analysis

The 16S rRNA gene amplification PCR and sequencing were performed using GeneAmp PCR System 2720 thermal cyclers (Applied Bio systems, Bedford, MA, USA) and ABI Prism 3130xl Genetic Analyzer capillary sequencer (Applied Bio systems), respectively, as described by Drancourt et al. [6]. The CodonCode Aligner was used to correct sequences and BLASTn searches were performed on the NCBI (National Centre for Biotechnology Information) web server at http://blast.ncbi.nlm.nih.gov.gate1.inist.fr/Blast

.cgi for the taxonomic assignation. Pairwise sequence similarities were calculated using the method recommended by Meier-Kolthoff et al. [23] and as described previously [33]. Sequences were aligned using ClustalW with default parameters and phylogenies were inferred using the GGDC web server available at http://ggdc.dsmz.de/ using the DSMZ phylogenomics pipeline.

Phenotypic, Biochemical, and Chemotaxonomic Characterization

Different growth temperatures (room temperature, 28, 37, 45, and 55 °C) were tested on sheep blood-enriched Columbia agar (bioMérieux) under anaerobic conditions using GENbag anaer system (bioMérieux), microaerophilic conditions using GENbag microaer system (bioMérieux), and aerobic conditions, with or without 5% CO_2.

Phenotypic and biochemical characteristics were performed as described elsewhere [32]. In addition to the three API gallery systems (API® ZYM, API® 20A, and API® 50 CH) usually used in our laboratory, API® Rapid ID 32A gallery system was added and the tests were done according to the manufacturer's instructions (bioMérieux).

E test strips for Amikacin 0.016–256 µg/ml, Vancomycin 0.016–256 µg/ml, Imipenem 0.002–32 µg/ml, Ceftriaxone 0.016–256 µg/ml, Rifampicin 0.002–32 µg/ml, Benzyl penicillin 0.002–32 µg/ml, Amoxicillin 0.016–256 µg/ml, Minocycline 0.016–256 µg/ml, Teicoplanin 0.016–256 µg/ml, Erythromycin 0.016–256 µg/ml, and Daptomycin 0.016–256 µg/ml (bioMérieux) were used for the antimicrobial agent susceptibility of strain AT11^T as recommended by EUCAST [4, 22]. Breakpoint tables for the interpretation of MICs and inhibition zone diameters, version 7.1, 2017, were used to interpret the results: these are available at http://www.eucast.org.

Cellular fatty acid methyl ester (FAME) analysis of this was then performed using gas chromatography/mass spectrometry (GC/MS) as described by Dione et al. [5].

Genome Sequencing and Assembling

The genomic DNA of strain AT11^T was sequenced and assembled as described in previous studies [33]. It was quantified by a Qubit assay using the high sensitivity kit (Life Technologies, Carlsbad, CA, USA) to 107.7 ng/µl and mechanically sheared with a circular shear to small fragments with an optimal length of 1401 bp using the Covaris device S2 in T6 tubes (Covaris, Woburn, MA, USA). A High Sensitivity Bioanalyzer LabChip (Agilent Technologies Inc, Santa Clara, CA, USA) was used to visualize the library profile and the final concentration library was measured at 34.4 nmol/l. The libraries were then normalized

and pooled at 2 nM. After a denaturation step and dilution at 15 pM, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and a sequencing run were performed in a single 2×251-bp run. A total of 5.6 Gb of information was obtained from the 589 K/mm^2 cluster density with a cluster passing quality control filters of 96.1% (11,444,000 passing filter paired reads). Within this run, the index representation for strain AT11[T] was determined to 6.46%. The 697,439 paired reads were trimmed and assembled.

Genome Annotation and Comparison

Open reading frames (ORFs) were predicted using Prodigal [10] with default parameters but the predicted ORFs were excluded if they were spanning a sequencing gap region (contain N). The predicted bacterial protein sequences were searched again the Clusters of Orthologous Groups (COG) using BLASTP with an E value of 1e^{-03}, a coverage of 0.7, and a percent identity of 30%. If no hit was found, a search was conducted against the Nucleotide Redundant (NR) database using the same parameters. If the length of sequence was smaller than 80 amino acids, a 1e^{-05} E value was used. The tRNAScanSE tool [21] was used to find tRNA genes, while ribosomal RNAs were found by using RNAmmer [15]. Lipoprotein signal peptides and the number of transmembrane helices was predicted using Phobius [11]. ORFans were identified if all the BLASTP performed gave no positive results with an E value smaller than 1e^{-03} for ORFs with a sequence size larger than 80 amino acid or an E value smaller than 1e^{-05} for ORFs with a sequence length smaller than 80 amino acids. Paralog genes were defined by blasting each protein gene against all protein genes of this genome. For pseudogenes, the first step was

to define the closed species genomes used for comparison analysis. Then, the potential missing genes in the genomes of interest were identified. All processes of annotation and comparison were performed using the Multi-Agent Software System DAGOBAH [8] that includes Figenix [9]. Genomic similarity was evaluated via digital DNA–DNA hybridization (dDDH) using the Genome to Genome Distance Calculator (GGDC) 2.1-DSM web service (http://ggdc.dsmz.de/ggdc.php).

The species used for genomic comparison were retrieved from the 16S rDNA gene tree. The following strains were selected: *Blautia producta* strain ATCC 27340[T] (ARET00000000) [7, 19, 28], *Eisenbergiella tayi* strain B086562[T] (MCGH00000000) [1, 2], *Anaerostipes hadrus* strain DSM 3319[T] (AMEY00000000) [13], *Parasporobacterium paucivorans* strain DSM 15970[T] (GCA900167085) [3], *Eubacterium ruminantium* strain ATCC 17233[T] (GCA900167085) [3], *Clostridium bolteae* strain WAL 16351[T] (AGYH00000000) [30], and *Clostridium clostridioforme* strain ATCC 25,537 (GCA900113155) [12]. For each selected strain, the complete genome sequence was retrieved from the FTP of NCBI (National Center for Biotechnology Information). The proteome was analyzed using proteinOrtho [17]. For each couple of genomes, a similarity score was then computed.

Accession Numbers

The 16S rRNA gene sequence and whole-genome shotgun sequence of strain AT11[T] were deposited in EMBL-EBI under accession numbers LN881600 and OEZA00000000, respectively. The Digital Protologue database TaxonNumber for strain AT11[T] is TA00401.

Fig. 1 Phylogenetic tree based on 16S rRNA sequence comparison highlighting the position of strain AT11[T] against other most closely related type strains. The scale bar represents a 2% nucleotide sequence divergence

Results and Discussion

Phylogenetic Analysis

The spectrum generated from strain AT11^T spots did not match those of Bruker and our in-house database (Supplementary Fig. 1) available at http://www.mediterranee-infection.com/article.php?laref=933&titre=c-d-e. This new strain exhibited 97.76% nucleotide sequence similarity with *Eisenbergiella tayi*, the closest species with standing in nomenclature according to the 16S rDNA sequence analysis. Figure 1 presents the neighbor-joining phylogenetic tree (Fig. 1) based on 16S rRNA gene sequences and shows the relationships between strain AT11^T and some related taxa. This sequence of the strain was deposited in EMBL-EBI under accession number LN881600.

Phenotypic and Biochemical Characterization

Strain AT11^T is strictly anaerobic, its growth temperature was between 28 and 45 °C, and optimal growth was observed at 37 °C. Colonies appeared light gray in color and exhibited an irregular form with a diameter between 0.5 and 1.5 mm after 72 h of culture on Columbia agar with 5% sheep blood (bioMérieux). No growth was observed above 5 g/l (10–100 g/l) salt on Schaedler agar with 5% sheep blood (bioMérieux). Cells were Gram-negative, non-motile, non-spore-forming, catalase positive, and rod shaped, measuring 1–3 μm in length and 0.4–0.5 μm wide using electron microscopy (Supplementary Fig. 2). The negativity of Gram staining was confirmed by the positive KOH test, but the strain had a positive Gram structure in electron microscopy. The characteristics of strain AT11^T, according to API® gallery systems (50 CH, 20A, Zym and Rapid ID 32A), along with those of the closest species, *Eisenbergiella tayi* strain B086562^T, are listed in Supplementary Table 1 and the differences between these two species are presented in Table 1.

Hexadecanoic acid was the most abundant fatty acid (63%), followed by saturated and unsaturated C$_{18}$ fatty acids representing approximately (33%) of total relative abundance. The fatty acid profiles of strain AT11^T and the closest strain *E. tayi* B086562^T are shown in Table 2.

Antimicrobial agent susceptibility was tested according to the EUCAST recommendations leading to the following MIC results: 32, 0.5, 0.125, 1.6, 0.064, 0.38, 0.5, 0.5, and 0.125 μg/ml, respectively, for Amikacin, Vancomycin, Imipenem, Ceftriaxone, Rifampicin, Benzyl penicillin, Amoxicillin, Minocycline, and Teicoplanin.

Table 1 Differential characteristic of strain AT11^T with *Eisenbergiella tayi* B086562^T

Properties	*Eisenbergiella massiliensis* AT11^T	*Eisenbergiella tayi* B086562^{Ta}
Indole production	V	−
Arabinose	+	−
Arbutin	+	−
Cellulose	+	−
Dulcitol	+	−
Gelatin	+	−
Glucose	+	−
Lactose	+	−
Maltose	V	−
Mannitol	V	−
Mannose	+	−
Raffinose	+	−
Rhamnose	+	−
Saccharose	+	−
Salicin	+	−
Sorbose	+	−
Tagatose	+	−
Trehalose	+	−
Trypsin	+	−
Xylose	V	−
Potassium 5-cetogluconate	+	−
Acid phosphatase	+	−
Alkaline phosphatase	V	+
Arginine hydrolase	+	−
Esterase	+	−
Esterase lipase	+	−
Naphthol-AS-BI-phosphohydrolase	+	−
Tyrosine arylamidase	−	+
α-Arabinosidase	−	+
α-Fructosidase	+	−
β-Glucuronidase	+	−
Isolated from	Human feces	Blood

+ Positive, − negative

v Variable

[a] Data for *E. tayi* were obtained from Amir et al. [25]

Genome Properties

The genome deposited in EMBL-EBI under accession number OEZA00000000 (Fig. 2) is 7,114,554 bp long with 48% GC content. It is composed of 19 contigs consisting of 17 scaffolds. Of the 6176 predicted genes, 6114 were protein-coding genes and 62 were RNAs (two 5S rRNA genes, two 16S rRNA genes, two 23S rRNA genes, 56 tRNA genes). A total of 4321 genes (70.67%) were assigned a putative

Table 2 Cellular fatty acid composition (%) of strain AT11^T compared to its closest neighbor *Eisenbergiella tayi* strain B086562^T

Fatty acid	Name	Strain AT11	*E. tayi*[a]
$C_{16:0}$	Hexadecanoic acid	62.7	45.4
$C_{18:1n9}$	9-Octadecenoic acid	10.3	14.8
$C_{18:1n5}$	13-Octadecenoic acid	9.2	ND
$C_{18:0}$	Octadecanoic acid	7.8	12.8
$C_{18:2n6}$	9,12-Octadecadienoic acid	4.4	1.3
$C_{18:1n7}$	11-Octadecenoic acid	1.6	3.2
$C_{17:0}$	Heptadecanoic acid	1.5	ND
$C_{15:0}$	Pentadecanoic acid	<1	<1
$C_{14:0}$	Tetradecanoic acid	<1	6.3
$C_{16:1n7}$	9-Hexadecenoic acid	<1	ND
$C_{20:4n6}$	5,8,11,14-Eicosatetraenoic acid	<1	ND
9,10-Methylene-$C_{16:0}$	2-Hexyl-cyclopropaneoctanoic acid	<1	ND
Iso-$C_{16:0}$	14-Methyl-pentadecanoic acid	<1	ND
Iso-$C_{15:0}$	13-Methyl-tetradecanoic acid	<1	ND
C16: 0 2-OH	2-Hydroxyhexadecanoic acid	ND	1.6
$C_{17:2}$	Heptadecadienoic acid	ND	11.6
$C_{13:1\ cis\ 12}$	12-Tridecanoic acid	ND	2.1
Anteiso-$C_{15:0}$	12-Methyl-tetradecanoic acid	ND	<1

ND Not detected

[a]Data for *E. tayi* were obtained from Amir et al. [25]

function by COGs or by NR BLAST. 130 genes were identified as ORFans (2.13%). The remaining 1525 genes (24.94%) were annotated as hypothetical proteins. Two genes associated with Vancomycin (Vancomycin B-type resistance protein, VanW) resistance and 20 genes associated with beta-lactamase resistance were found using the RAST web server [26]. The remaining 1525 genes (24.94%) were annotated as hypothetical proteins.

Genome Comparison

The draft genome sequence of strain AT11^T (7.11 MB) is larger in size than those of *C. bolteae, B. producta, C. clostridioforme, Eubacterium ruminantium*, and *A. hadrus* (6.38, 6.09, 5.46, 2.84, and 2.77 MB, respectively) but almost equal to that of *E. tayi* (7.15). Its G+C (48%) content is lower than that of *C. clostridioforme* and *C. bolteae* (49 and 49.6%, respectively), but higher than that of *E. tayi, B. producta, Eubacterium ruminantium*, and *A. hadrus* (46.3, 45.7, 37.2, and 37.2, respectively). Its gene content (6114) is higher than that of *C. bolteae, B. producta, C. clostridioforme, A. hadrus*, and *Eubacterium ruminantium* (5892, 5666, 5376, 2716, and 2533, respectively) but lower than that of *E. tayi* at 6156. The distribution of genes into COG categories was not entirely similar in all compared genomes (Fig. 3). The average genomic identity of orthologous gene sequences (AGIOS)

values ranged from 61.7% between *C. bolteae* and *Eubacterium ruminantium* to 90.8% between *C. bolteae* and *C. clostridioforme* among compared species without strain AT11^T (Supplementary Table 2). When strain AT11^T was included in the comparison, these values ranged from 62.2% with *Eubacterium ruminantium* to 78.4% with *E. tayi* (Supplementary Table 2). The dDDH values for strain AT11^T ranged from 19.5% with *A. hadrus* to 34.4% with *C. clostridioforme* (Supplementary Table 3) with a probability of error of ±2%. These values are very low and below the cutoff of 70%, thus also confirming that this strain is a new species.

Based on the phenotypic, chemotaxonomic, genomic, and phylogenetic characteristics, a novel bacterium isolated from the stool sample of a morbidly obese French woman, under the name *Eisenbergiella massiliensis* sp. nov., is proposed. The type strain is AT11^T = DSM 100838^T = CSUR P2478^T.

Description of *Eisenbergiella massiliensis* sp. nov

Eisenbergiella massiliensis (*mas.si.li.en'sis*. L. fem. adj. massiliensis, of Massilia, the Latin name for Marseille). It is a strictly anaerobic bacterium which grows at a mesothermal temperature of 37 °C. The colonies grown on Columbia agar with 5% enriched sheep blood are light gray, non-hemolytic, and irregular with a diameter of 0.5 mm.
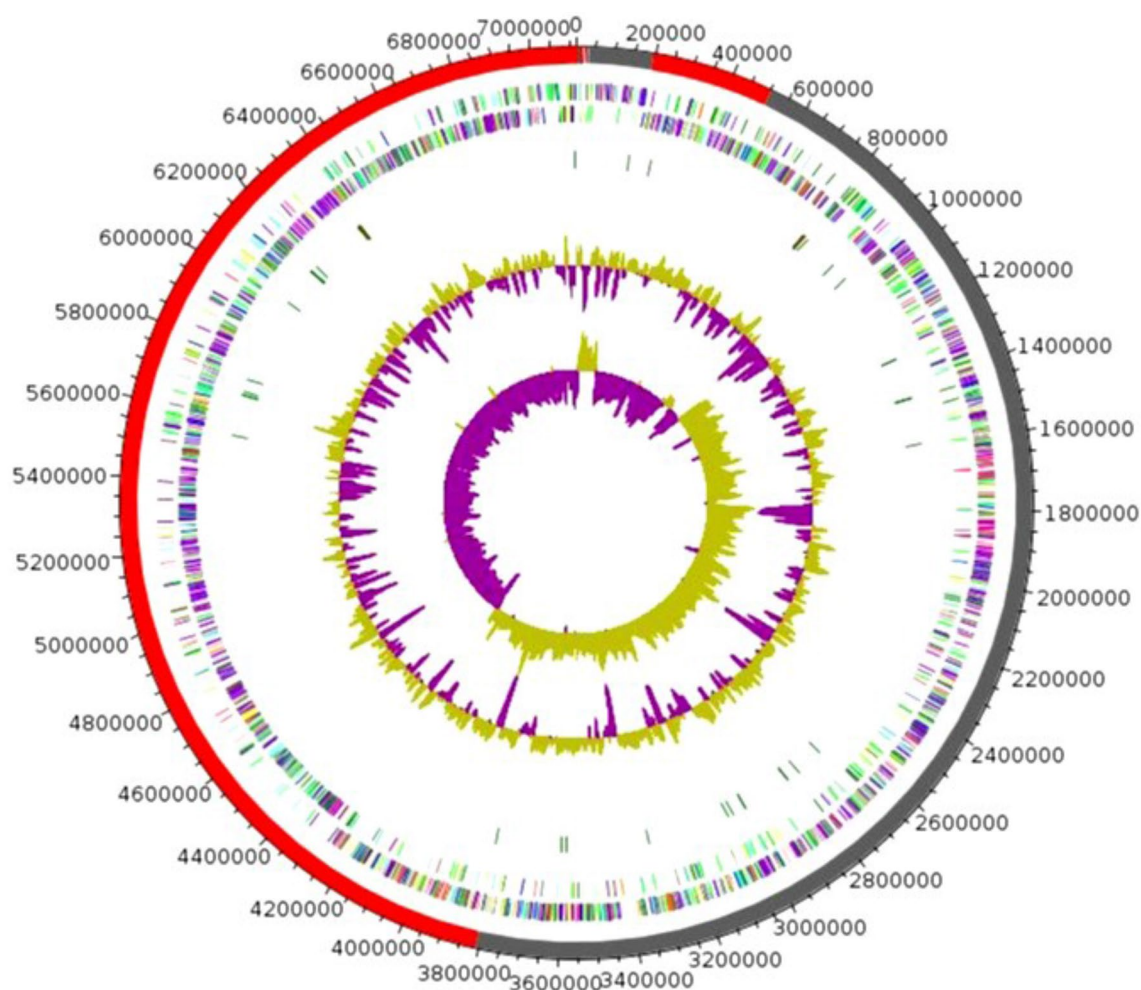
Fig. 2 Graphical circular map of the genome of strain AT11^T: From outside to the center: Contigs (red/gray), Clusters of Orthologous Groups (COGs) category of genes on the forward strand (three circles), genes on forward strand (blue circle), genes on the reverse strand (red circle), COGs category on the reverse strand (three circles), G+C content. (Color figure online)

Cells exhibit a negative Gram-stain, are non-spore-forming, non-motile, catalase positive, and rod-shaped bacilli, with a size of 0.4/2 µm. Using the API Gallery systems (API® ZYM API® 50CH API® 20A and API® rapid ID 32A) in anaerobic condition, positives reactions were observed for acid phosphatase, alkaline phosphatase, esterase, esterase lipase, naphthol-AS-BI-phosphohydrolase, N-acetyl-β-glucosaminidase, α-arabinosidase, α-fucosidase, α-galactosidase, β-galactosidase, α-glucosidase, β-glucosidase, β-glucuronidase, 6-phosphate-β-galactosidase, arbutin, D-cellobiose, D-glucose, D-lactose, D-lyxose, D-maltose, D-mannose, D-raffinose, D-saccharose, D-tagatose, D-trehalose, dulcitol, D-xylose, L-arabinose, L-rhamnose, L-sorbose, potassium 5-cetogluconate, and salicin. Urease and indole are not produced, gelatin was not liquefied and nitrate was not reduced, although esculin was hydrolyzed. The major cellular fatty acids detected were $C_{16:0}$ (62.7%) and $C_{18:1n9}$ (10.3%). Its genome, consisting of one chromosome, is 7,114,554 bp in length with 48% of G+C content. The type strain AT11^T=CSUR P2478^T=DSM 100838^T was isolated from the stool sample of a French morbidly obese woman following bariatric surgery.
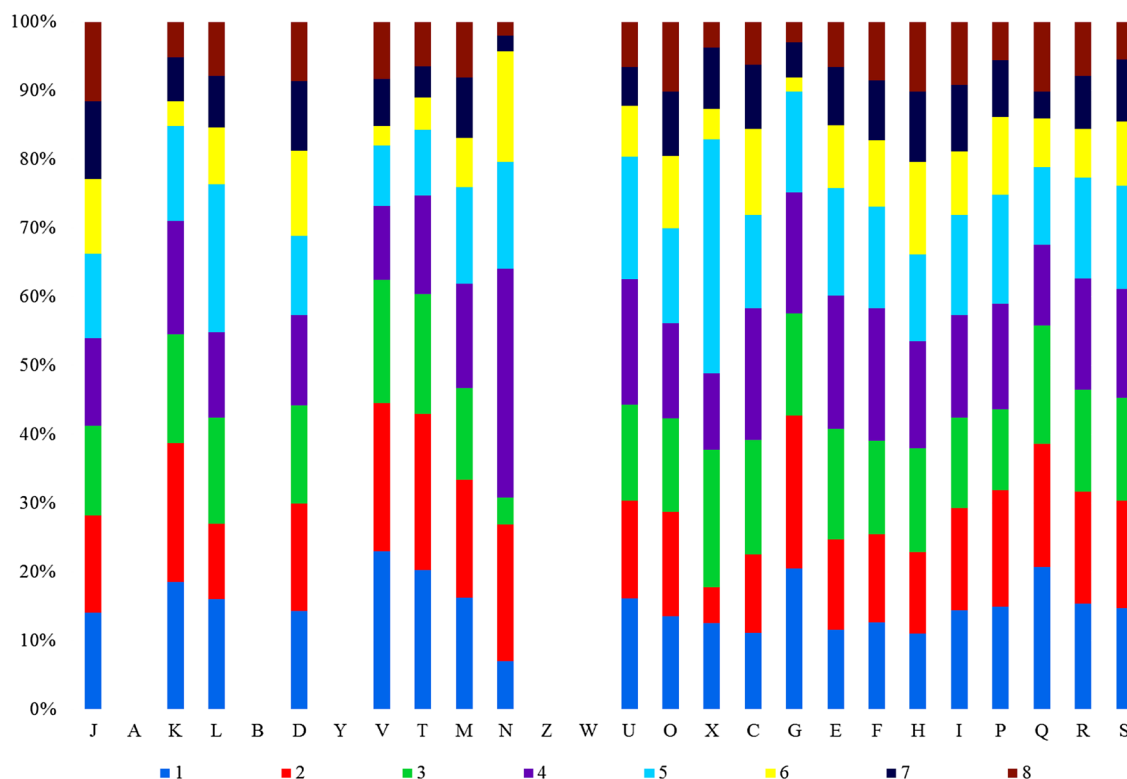
Fig. 3 Distribution of functional classes of predicted genes according to the Clusters of Orthologous Groups (COGs) of strain AT11^T with its closest species: 1, Strain AT11^T; 2, *Eisenbergiella tayi* strain DSM 26961^T; 3, *Blautia producta* strain ATCC 27340^T; 4, *Clostridium bolteae* strain WAL 16351^T; 5, *Clostridium clostridioforme* strain ATCC 25537^T; 6, *Parasporobacterium paucivorans* strain DSM 15970^T; 7, *Anaerostipes hadrus* strain ATCC 29173^T; and 8, *Eubacterium ruminantium* strain ATCC 17233^T. *Superscript T* Type strain, *ATCC* American Type Culture Collection, *DSM* Deutsche Sammlung von Mikroorganismen, *WAL* Wadsworth Anaerobe Laboratory

Compliance with Ethical Standards

Conflict of interest The author declares that they have no conflicts of interest.

References

1. Amir I, Bouvet P, Legeay C et al (2014) *Eisenbergiella tayi* gen. nov., sp. nov., isolated from human blood. Int J Syst Evol Microbiol 64:907–914

2. Bernard K, Burdz T, Wiebe D et al (2017) Characterization of isolates of *Eisenbergiella tayi*, a strictly anaerobic gram-stain variable bacillus recovered from human clinical materials in Canada. Anaerobe 44:128–132

3. Bryant MP (1959) Bacterial species of the rumen. Bacteriol Rev 23:125–153

4. Citron DM, Ostovari MI, Karlsson A, Goldstein EJ (1991) Evaluation of the E test for susceptibility testing of anaerobic bacteria. J Clin Microbiol 29:2197–2203

5. Dione N, Sankar SA, Lagier J-C et al (2016) Genome sequence and description of *Anaerosalibacter massiliensis* sp. nov. New Microbes New Infect 10:66–76

6. Drancourt M, Bollet C, Carlioz A et al (2000) 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. J Clin Microbiol 38:3623–3630

7. Ezaki T, Li N, Hashimoto Y et al (1994) 16S ribosomal DNA sequences of anaerobic cocci and proposal of *Ruminococcus hansenii* comb. nov. and *Ruminococcus productus* comb. nov. Int J Syst Bacteriol 44:130–136

8. Gouret P, Paganini J, Dainat J et al (2011) Integration of evolutionary biology concepts for functional annotation and automation of complex research in evolution: the multi-agent software system DAGOBAH. In: Pontarotti P (ed) Evolutionary biology—concepts, biodiversity, macroevolution and genome evolution. Springer, Berlin Heidelberg, pp 71–87

9. Gouret P, Vitiello V, Balandraud N et al (2005) FIGENIX: intelligent automation of genomic annotation: expertise integration in a new software platform. BMC Bioinform 6:198

10. Hyatt D, Chen G-L, Locascio PF et al (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform 11:119

11. Käll L, Krogh A, Sonnhammer ELL (2004) A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338:1027–1036

12. Kaneuchi C, Watanabe K, Terada A et al (1976) Taxonomic Study of *Bacteroides clostridiiformis* subsp. *clostridiiformis* (Burri and Ankersmit) Holdeman and Moore and of Related Organisms: Proposal of *Clostridium clostridiiformis* (Burri and Ankersmit) comb.

nov. and *Clostridium symbiosum* (Stevens) comb. nov. Int J Syst Bacteriol 26:195–204

13. Kant R, Rasinkangas P, Satokari R et al (2015) Genome sequence of the butyrate producing Anaerobic bacterium *Anaerostipes hadrus* PEL 85. Genome Announc 3:e00224-15

14. Kong L-C, Tap J, Aron-Wisnewsky J et al (2013) Gut microbiota after gastric bypass in human obesity: increased richness and associations of bacterial genera with adipose tissue genes. Am J Clin Nutr 98:16–24

15. Lagesen K, Hallin P, Rødland EA et al (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108

16. Lagier J-C, Hugon P, Khelaifia S et al (2015) The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. Clin Microbiol Rev 28:237–264

17. Lechner M, Findeiss S, Steiner L et al (2011) Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinform 12:124

18. Ley RE, Turnbaugh PJ, Klein S, Gordon JI (2006) Microbial ecology: human gut microbes associated with obesity. Nature 444:1022–1023

19. Liu C, Finegold SM, Song Y, Lawson PA (2008) Reclassification of *Clostridium coccoides, Ruminococcus hansenii. Ruminococcus hydrogenotrophicus, Ruminococcus luti, Ruminococcus productus* and *Ruminococcus schinkii* as *Blautia coccoides* gen. nov., comb. nov., *Blautia hansenii* comb. nov., *Blautia hydrogenotrophica* comb. nov., *Blautia luti* comb. nov., *Blautia producta* comb. nov., *Blautia schinkii* comb. nov. and description of *Blautia wexlerae* sp. nov., isolated from human faeces. Int J Syst Evol Microbiol 58:1896–1902

20. Lomans BP, Leijdekkers P, Wesselink J-J et al (2001) Obligate sulfide-dependent degradation of methoxylated aromatic compounds and formation of methanethiol and dimethyl sulfide by a freshwater sediment isolate, *Parasporobacterium paucivorans* gen. nov., sp. nov. Appl Environ Microbiol 67:4017–4023

21. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964

22. Matuschek E, Brown DFJ, Kahlmeter G (2014) Development of the EUCAST disk diffusion antimicrobial susceptibility testing method and its implementation in routine microbiology laboratories. Clin Microbiol Infect 20:O255–O266

23. Meier-Kolthoff JP, Göker M, Spröer C, Klenk H-P (2013) When should a DDH experiment be mandatory in microbial taxonomy? Arch Microbiol 195:413–418

24. Million M, Maraninchi M, Henry M et al (2012) Obesity-associated gut microbiota is enriched in *Lactobacillus reuteri* and depleted in *Bifidobacterium animalis* and *Methanobrevibacter smithii*. Int J Obes 36:817–825

25. Ng M, Fleming T, Robinson M et al (2014) Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. The Lancet 384:766–781

26. Overbeek R, Olson R, Pusch GD et al (2014) The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res 42:D206-214

27. Ramasamy D, Mishra AK, Lagier J-C et al (2014) A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol 64:384–391

28. Rettedal EA, Gumpert H, Sommer MOA (2014) Cultivation-based multiplex phenotyping of human gut microbiota allows targeted recovery of previously uncultured bacteria. Nat Commun 5:4714

29. Seng P, Drancourt M, Gouriet F et al (2009) Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 49:543–551

30. Song Y, Liu C, Molitoris DR et al (2003) *Clostridium bolteae* sp. nov., isolated from human sources. Syst Appl Microbiol 26:84–89

31. Togo AH, Khelaifia S, Bittar F et al (2016) '*Eisenbergiella massiliensis*', a new species isolated from human stool collected after bariatric surgery. New Microbes New Infect 13:15–16

32. Togo AH, Khelaifia S, Lagier J-C et al (2016) Noncontiguous finished genome sequence and description of *Paenibacillus ihumii* sp. nov. strain AT5. New Microbes New Infect 10:142–150

33. Togo AH, Durand G, Khelaifia S et al (2017) *Fournierella massiliensis*, gen. nov., sp. nov., a new human-associated member of the family *Ruminococcaceae*. Int J Syst Evol Microbiol 67:1393–1399

34. Zhang H, DiBaise JK, Zuccolo A et al (2009) Human gut microbiota in obesity and after gastric bypass. Proc Natl Acad Sci USA 106:2365–2370

Article XIV:

Hugonella massiliensis gen. nov., sp. nov., genome sequence, and description of a new strictly anaerobic bacterium isolated from the human gut.

Ziena Elsawi, **Amadou Hamidou Togo**, Mamadou Beye, Grégory Dubourg, Claudia Andrieu, Nicholas Armsrtong, Magali Richez, Fabrizio Di Pinto, Fadi Bittar, Noémie Labas, Pierre-Edouard Fournier, Didier Raoult and Saber Khelaifia

ORIGINAL RESEARCH

WILEY MicrobiologyOpen

Hugonella massiliensis gen. nov., sp. nov., genome sequence, and description of a new strictly anaerobic bacterium isolated from the human gut

Ziena Elsawi[1,*] | Amadou Hamidou Togo[1,*] | Mamadou Beye[1] | Grégory Dubourg[1] | Claudia Andrieu[1] | Nicholas Armsrtong[1] | Magali Richez[1] | Fabrizio di Pinto[1] | Fadi Bittar[1] | Noémie Labas[1] | Pierre-Edouard Fournier[1] | Didier Raoult[1,2] | Saber Khelaifia[1] (iD)

[1]Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, CNRS (UMR 7278), IRD (198), INSERM (U1095), AMU (UM63), Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine, Aix-Marseille Université, Marseille, France

[2]Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

Correspondence
Saber Khelaifia
URMITE, CNRS (UMR 7278), IRD (198), INSERM (U1095), AMU (UM63), Faculté de Médecine, Aix-Marseille Université, Marseille Cedex 5, France.
Email: khelaifia_saber@yahoo.fr

Funding information
This study was funded by the "Fondation Méditerranée Infection".

Abstract

The human gut is composed of a large diversity of microorganisms, which have been poorly described. Here, using culturomics, a new concept based on the variation in culture conditions and MALDI-TOF MS identification, we proceed to explore the microbial diversity of the complex ecosystem of the human gut. Using this approach, we isolated strain AT8[T] (=CSUR P2118 = DSM 101782) from stool specimens collected from a 51-year-old obese French woman. Strain AT8[T] is a strictly anaerobic, nonmotile, nonspore-forming gram-positive coccus that do not exhibit catalase and oxidase activities. 16S rDNA-based identification of strain AT8[T] demonstrated 92% gene sequence similarity with *Eggerthella lenta* DSM 2243, the phylogenetically closed validly named type species. Here, we present a set of features for the strain AT8[T] and the description of its complete genome sequence and annotation. The 2,091,845 bp long genome has a G+C content of 63.46% and encodes 1,849 predicted genes; 1,781 were protein-coding genes, and 68 were RNAs. On the basis of the characteristics reported here, we propose the creation of a new bacterial genus *Hugonella* gen. nov., belonging to the *Eggerthellaceae* family and including *Hugonella massiliensis* gen. nov., sp. nov., strain AT8[T] as the type strain.

KEYWORDS
culturomics, *Hugonella massiliensis* sp. nov., obesity, taxonogenomic

1 | INTRODUCTION

The human gut harbors a complex bacterial community known as microbiota. However, this ecosystem remains incompletely characterized and its diversity poorly described (Eckburg et al., 2005; Lozupone, Stombaugh, Gordon, Jansson, & Knight, 2012). Culturomics concepy

was recently proposed as a new alternative to explore this ecosystem and enriches the human microbiota repertoire. This method is based on the large variation in culture conditions and the use of rapid bacterial identification methods such as matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) and 16S rRNA gene amplification and sequencing of the colonies (Lagier

*These two authors contributed equally

et al., 2015). Traditionally, several parameters were used to identify and define a new bacterial species including 16S rRNA gene sequencing and phylogeny, genomic diversity of the G+C content, DNA-DNA hybridization (DDH) intensive phenotypic, and chemotaxonomic characterization (Ramasamy, Mishra, Lagier, Padhmanabhan, & Rossi, 2014; Welker & Moore, 2011). Nevertheless, some limits have been appeared notably because the cutoff values vary dramatically between species and genera (Rosselló-Móra, 2006). So in order to describe new bacterial species, we recently proposed a new method named taxonogenomics, which includes both genomic analysis and proteomic information obtained by MALDI-TOF analysis (Ramasamy et al., 2014). Using culturomics techniques (Lagier et al., 2015), we herein isolated strain AT8[T] from a stool specimen of a 51-year-old obese French woman (BMI 44.38 kg/m^2). Here, we present a classification and a set of characteristics of strain AT8[T] together with the description of its complete genome sequencing and annotation that allowed us to describe them as the first representative of a new bacterial genus classified into *Eggerthellaceae* family within the phylum *Actinobacteria*. The *Eggerthellaceae* family contains nine different genera, *Adlercreutzia*, *Asaccharobacter*, *Cryptobacterium*, *Denitrobacterium*, *Eggerthella*, *Enterorhabdus*, *Gordonibacter*, *Paraeggerthella*, and *Slackia* (Gupta et al. 2006). The species of this family are strictly anaerobic cocci and they do not form spores (Gupta et al. 2006).

2 | METHODS AND MATERIALS

2.1 | Ethics and sample collection

The stool sample was collected from a 51-year-old obese French woman (BMI 44.38 kg/m^2; weight 108 kg, 1.56 meters in height) in January 2012. Written consent was obtained from the patient at the Nutrition, Metabolic Disease and Endocrinology service, at La Timone Hospital, (Marseille, France). The study and the consent procedures were approved by the local IFR 48 ethics committee, under consent number 09-022, 2010. The stool sample was stored at −80°C after collection.

2.2 | Isolation of the strain

Strain AT8[T] was isolated in June 2015 by anaerobic culture. Approximately, 1 g of stool specimen was inoculated anaerobically in an anaerobic blood culture bottle supplemented with 5% (v/v) sheep blood and 5% (v/v) rumen fluid. pH was adjusted at 7.5 using KOH solution (10%) and the blood culture bottle was incubated at 37°C for 3 days. After 3 days incubation, subcultures were done on solid medium consisting of Columbia agar supplemented with 5% sheep blood and incubated anaerobically for 48 hr. All growing colonies were picked several times to obtain pure cultures.

2.3 | Strain identification by MALDI-TOF MS and 16S rRNA gene sequencing

The MALDI-TOF MS protein analysis consisted of picking an isolated colony and then depositing twelve distinct deposits on a MTP 96 MALDI-TOF target plate (Bruker Daltonics, Leipzig, Germany) to be analyzed. 2 µl of a matrix solution (saturated solution of α-cyano-4-hydroxycinnamic acid diluted in 50% acetonitrile and 2.5% of tri-fluoro-acetic acid) was added on each spot. Measurements and proteomic analysis of the isolate were carried out with a Microflex spectrometer (Bruker) as previously described (Seng et al., 2009). Protein spectra were imported into the MALDI BioTyper software (version 2.0, Bruker) and analyzed by standard pattern matching (with default parameter settings) against the spectra of the Bruker database (constantly incremented with our new spectra). A score >1.9 enabled the identification at the species level and a score < 1.7 did not enable any identification. Sequencing the 16S rRNA gene is needed to achieve the identification if the bacterium is not referenced in the database. The 16S rRNA gene amplification and sequencing were performed as previously described (Morel et al., 2015). For similarity level thresholds of 98.65% and 95%, a new species or a new genus was suggested, respectively, as proposed by Kim, Oh, Park, & Chun, (2014).

2.4 | Phylogenetic tree

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

2.5 | Growth conditions

Growth of the strain AT8[T] was tested under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems, respectively (bioMérieux, Marcy l'Etoile, France), and in aerobic conditions, with or without 5% CO_2. Different temperatures (25, 30, 37, 45°C) were tested to determine the optimal growth of the strain AT8[T]. Optimal salt concentration required for growth was determined by growing the strain at 0, 0.5, 1,and 1.5% of NaCl. The optimal pH for growth was determined by testing different pH: 5, 6, 6.5, 7, 7.5, 8, and 8.5.

2.6 | Morphological, biochemical, and antibiotic susceptibility tests

Sporulation assay was done by a thermic shock at 60°C for 20 min follow by a subculture on 5% sheep blood-enriched Columbia agar medium (bioMérieux). Using a DM1000 photonic microscope (Leica Microsystems, Nanterre, France) with a100X objective lens,

the motility of the strain from a fresh culture was observed. The colony's surface was observed on a 5% sheep blood agar culture medium after 24-hr' incubation at 37°C. In order to observe the cells morphology, they were fixed with 2.5% glutaraldehyde in 0.1mol/L cacodylate buffer for at least 1 hr at 4°C. A drop of the cell suspension was deposited for approximately 5 min on glow-discharged formvar carbon film on 400 mesh nickel grids (FCF400-Ni, EMS). The grids were dried on blotting paper and cells were negatively stained for 10 s with 1% ammonium molybdate solution in filtered water at RT. Electron micrographs were acquired with a Tecnai G[20] Cryo (FEI) transmission electron microscope operated at 200 keV. The gram coloration was performed using the color Gram 2 kit (bioMérieux) and observed using a DM1000 photonic microscope (Leica Microsystems).

For the biochemical characterization assays, available API ZYM and API 50 CH strips (bioMérieux) were performed according to the manufacturer's instructions. Cellular fatty acid methyl ester (FAME) analysis was performed by gaz chromatography/mass spectrometry (GC/MS). Two samples were prepared with approximately 20 mg of bacterial biomass per tube harvested from several culture plates. FAME were prepared as described by Sasser, (2006). GC/MS analyses were carried out as described before (Dione et al., 2016). Briefly, FAME were separated using an Elite 5-MS column and monitored by mass spectrometry (Clarus 500 - SQ 8 S, Perkin Elmer, Courtaboeuf, France). Spectral database search was performed using MS Search 2.0 operated with the Standard Reference Database 1A (NIST, Gaithersburg, USA) and the FAMEs mass spectral database (Wiley, Chichester, UK).

Susceptibility to antibiotics was tested using Müller Hinton agar medium (bioMérieux) according to EUCAST 2015 recommendations (http://www.eucast.org). The following antibiotics were tested: doxycycline, cefoxitin, ciprofloxacin, clindamycin, erythromycin, fosfomycin, linezolid, oxacillin, penicillin, pristinamycin, rifampicin, teicoplanin, trimethoprim-sulfamethoxazole, vancomycin, colistin, and metronidazole.

2.7 | DNA extraction and genome sequencing and assembly

Strain AT8[T] was cultured on ten petri dishes with 5% sheep blood Columbia agar. Genomic DNA (gDNA) of strain AT8[T] was extracted in two steps : a mechanical treatment was first performed by glass beads acid washed (G4649-500 g Sigma) using a FastPrep BIO 101 instrument (Qbiogene, Strasbourg, France) at maximum speed (6.5) for 3 × 30s. Then after a 2 hr lysozyme incubation at 37°C, DNA was extracted on the EZ1 biorobot (Qiagen) with EZ1 DNA tissues kit. The elution volume is 50 µl. gDNA was quantified by a Qubit assay with the high sensitivity kit (Life technologies, Carlsbad, CA, USA) to 29.1 ng/µl.

The gDNA of strain AT8[T] was sequenced on MiSeq Technology (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 gDNA using the Nextera mate

pair Illumina guide. The gDNA sample was simultaneously fragmented and tagged with a mate pair junction adapter. The pattern of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies Inc, Santa Clara, CA, USA) with a DNA 7500 labchip. The DNA fragments ranged in size from 1.5 kb up to 11 kb with an optimal size at 6.593 kb. No size selection was performed and 395 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal at 983 bp on the Covaris device S2 in T6 tubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies Inc, Santa Clara, CA, USA) and the final concentration library was measured at 23.68 nmol/L. The libraries were normalized at 3 nmol/L and pooled. After a denaturation step and dilution, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and sequencing run were performed in a single 2 × 251-bp run. Genome assembly was performed in pipelines that allow us to make assembly with different assembly's softwares, on trimmed or not trimmed data. Velvet (Zerbino & Birney, 2008), Spades (Bankevich et al., 2012), and Soap Denovo (Luo et al., 2012) were used on trimmed (MiSeq software and Trimmomatic (Bolger, Lohse, & Usadel, 2014) software) and on not trimmed data (only MiSeq software). For each of the six assemblies performed, GapCloser (Luo et al., 2012) was used to reduce gaps. Then contamination with Phage Phix was identified (blastn against Phage Phix174 DNA sequence) and eliminated. At the end, scaffolds with size under 800 bp were removed and scaffolds with depth value lower than 0.25 of the mean depth are removed (identified as possible contaminant). The best assembly is selected on different criterion (number of scaffolds, N50, number of N). For strain AT8[T], the best assembly was obtained with Spades and a coverage of 349. Default parameters were used, together with those specific parameters: –careful and kmer 77,99,127.

Total information of 10 Gb was obtained from a 690 K/mm^2 cluster density with a cluster passing quality control filters of 94.5% (16,542,000 passing filter paired reads). Within this run, the index representation for strain AT8[T] was determined to 7.36%. The 1,218,050 paired reads were trimmed then assembled in nine scaffolds.

2.8 | Genome annotation and comparison

The prediction of open reading frames (ORFs) was performed by Prodigal (http://prodigal.ornl.gov/) with default parameters but the predicted ORFs were excluded if they were spanning a sequencing gap region (containing N). The predicted bacterial protein sequences were searched against the clusters of orthologous groups (COG) (Galperin, Makarova, Wolf, & Koonin, 2015) using BLASTP (E-value 1e^{-03}, coverage 0.7, and identity percent 30%). If no hit was found, it was searched against the NR database using BLASTP with E-value of 1e^{-03} coverage 0.7 and identity percent of 30%, and if the sequence length was smaller than 80 amino acids, we used an E-value of 1e^{-05}. tRNA genes were found by the tRNAScanSE tool,

TABLE 1 Classification and general features of *Hugonella massiliensis* strain AT8[T] according to the MIGS recommendations (Field et al., 2008)

MIGS ID	Property	Term	Evidence code[a]
	Current classification	Domain: *Bacteria*	TAS (Woese, Kandler, & Wheelis, 1990)
		Phylum: *Actinobacteria*	TAS (Garrity & Holt, 2001)
		Class: *Coriobacteriia*	TAS (Skerman, McGowan, & Sneath, 1980)
		Order: *Eggerthellales*	TAS (Wade et al., 1999)
		Family: *Eggerthellaceae*	TAS (Moore, Cato, & Holdeman, 1971)
		Genus: *Hugonella*	IDA
		Species: *Hugonella massiliensis*	IDA
		Type strain: AT8[T]	IDA
	Gram stain	Positive	IDA
	Cell shape	Cocci	IDA
	Motility	Nonmotile	IDA
	Sporulation	nonspore forming	IDA
	Temperature range	Mesophile	IDA
	Optimum temperature	37°C	IDA
	pH	pH 6.5 to 8	
	Optimum pH	7	
MIGS-6.3	Salinity	0.5 to 1.5%	IDA
	Optimum salinity	0-0.5% NaCl	IDA
MIGS-22	Oxygen requirement	Strictly anaerobic	IDA
	Carbon source	Unknown	IDA
	Energy source	Unknown	IDA
MIGS-6	Habitat	Human gut	IDA
MIGS-15	Biotic relationship	Free living	IDA
	Pathogenicity	Unknown	NAS
	Biosafety level	Unknown	IDA
MIGS-14	Isolation	Human feces	IDA
MIGS-4	Geographic location	France	IDA
MIGS-5	Sample collection time	2012	IDA
MIGS-4.3	Depth	Surface	IDA
MIGS-4.4	Altitude	0 m above sea level	IDA

[a]Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Nontraceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from http://www.geneontology.org/GO.evidence.shtml of the Gene Ontology project (Ashburner et al., 2000). If the evidence is IDA, then the property was directly observed for a live isolate by one of the authors or an expert mentioned in the acknowledgments.

whereas ribosomal RNAs were found using RNAmmer (Lagesen et al., 2007; Lowe & Eddy, 1997). Using Phobius, the lipoprotein signal peptides and the number of transmembrane helices were predicted (Käll, Krogh, & Sonnhammer, 2004). ORFans were identified if all the BLASTP performed did not give positive results (*E*-value smaller than $1e^{-03}$ for ORFs with sequence size larger than 80 aa or *E*-value smaller than $1e^{-05}$ for ORFs with sequence length smaller 80 aa). The XEGEN software (Phylopattern) allowed us to automatically retrieve genomes from the 16S RNA tree (Gouret, Thompson, & Pontarotti, 2009). For each selected species, the complete genome sequence, proteome sequence, and Orfeome sequence were retrieved from the FTP of NCBI. The proteomes were analyzed with proteinOrtho (Lechner et al., 2011). Then for each couple of genomes, a similarity score (mean value of nucleotide similarity between all couple of orthologues between the two genomes studied) was computed by AGIOS software (Average Genomic Identity Of gene Sequences) (Ramasamy et al., 2014). An annotation of all proteome was done to determine the predicted genes functional classes' distribution according to the clusters of orthologous groups of proteins. The Multi-Agent software system DAGOBAH, which includes Figenix libraries for provide pipeline analysis and Phylopattern for tree manipulation, was used to perform the annotation and comparison processes (Gouret et al., 2011). Genome-to-Genome Distance Calculator (GGDC) analysis

TABLE 2 Differential characteristics of *Hugonella massiliensis* strain AT8[T], *Eggerthella lenta* (Kageyama, Benno, & Nakase, 1999), *Denitrobacterium detoxificans* (Anderson, Rasmussen, Jensen, & Allison, 2000), *Slackia exigua* (Kim et al., 2010), *Slackia heliotrinireducens* (Lanigan, 1976), *Gordonibacter pamelaeae* (Würdemann et al., 2009), *Adlercreutzia equolifaciens* (Maruo, Sakamoto, Ito, Toda, & Benno, 2008). na: Nonavailable data

Properties	Hugonella massiliensis	Eggerthella lenta	Denitrobacterium detoxificans	Slackia exigua	Slackia heliotrinireducens	Gordonibacter pamelaeae	Adlercreutzia equolifaciens
Cell diameter (μm)	0.8–1.2	0.4–0.8	0.5–1 × 1–1.5	0.5–1.0	0.8–1.2	0.6–1.01	0.6–1.5
Oxygen requirement	−	−	−	+	−	+	−
Gram stain	+	+	+	+	+	+	+
Salt requirement	−	na	na	−	+	−	na
Motility	−	−	−	−	−	+	−
Endospore formation	−	−	−	−	−	−	−
Indole	−	−	−	−	−	−	na
Production of							
Alkaline phosphatase	−	na	−	−	na	na	na
Catalase	−	−	−	−	−	+	na
Oxidase	−	−	−	−	−	−	na
Nitrate reductase	−	+	+	+	+	+	−
Urease	−	na	na	−	−	na	−
β-galactosidase	−	na	na	−	na	na	−
N-acetyl-glucosamine	+	na	na	na	na	na	na
Acid from							
L-Arabinose	+	−	na	−	−	−	+
Ribose	−	−	−	−	−	−	na
Mannose	+	−	na	−	−	−	−
Mannitol	−	−	−	na	na	−	+
Sucrose	−	na	na	na	−	na	na
D-glucose	−	−	−	−	−	−	−
D-fructose	−	−	−	−	na	−	na
D-maltose	−	−	na	na	na	−	na
D-lactose	−	na	na	na	na	na	na
Habitat	Human gut	Human gut	Ruminal microbes	Human gut	Human gut	Human gut	Human gut

TABLE 3 Cellular fatty acid composition (%) of _Hugonella massiliensis_ strain AT8[T]

Fatty acids	Name	Mean relative % [a]
C16:0	Hexadecanoic acid	35.0 ± 1.5
C14:0	Tetradecanoic acid	30.9 ± 0.5
C15:0 anteiso	12-methyl-tetradecanoic acid	7.3 ± 0.6
C18:1n9	9-Octadecenoic acid	6.5 ± 1.5
C15:0 iso	13-methyl-tetradecanoic acid	6.3 ± 0.8
C18:0	Octadecanoic acid	5.7 ± 1.7
C14:0 iso	12-methyl-Tridecanoic acid	3.4 ± 0.4
C15:0	Pentadecanoic acid	1.9 ± 0.1
C18:2n6	9,12-Octadecadienoic acid	1.5 ± 0.1
C18:1n7	11-Octadecenoic acid	1.5 ± 0.5
C17:0	Heptadecanoic acid	TR

[a]Mean peak area percentage; TR = trace amounts <1%.

was performed using the GGDC web server as previously reported (Meier-Kolthoff, Auch, Klenk, & Göker, 2013).

3 | RESULTS

3.1 | Phenotypic and biochemical features

Strain AT8[T] presents small, smooth, shiny, circular colonies with a diameter of 2–5 mm. Strain AT8[T] is gram-positive, nonmotile, and nonspore forming (Fig. S1). Cells are rod-shaped with a mean diameter of 0.8–1.2 μm (Fig. S2). Growth was obtained at temperatures ranging from 28 to 55°C and pH ranging from 6.5 to 8 with an optimum growth at 37°C and pH 7 after 48 hr incubation. The isolate did not require NaCl for growth; an optimal growth was observed at 0 or 0.5% of NaCl. No growth was observed at 1 and 1.5% of NaCl. Strain AT8[T] was strictly anaerobic and did not grow in aerobic or 5% CO_2
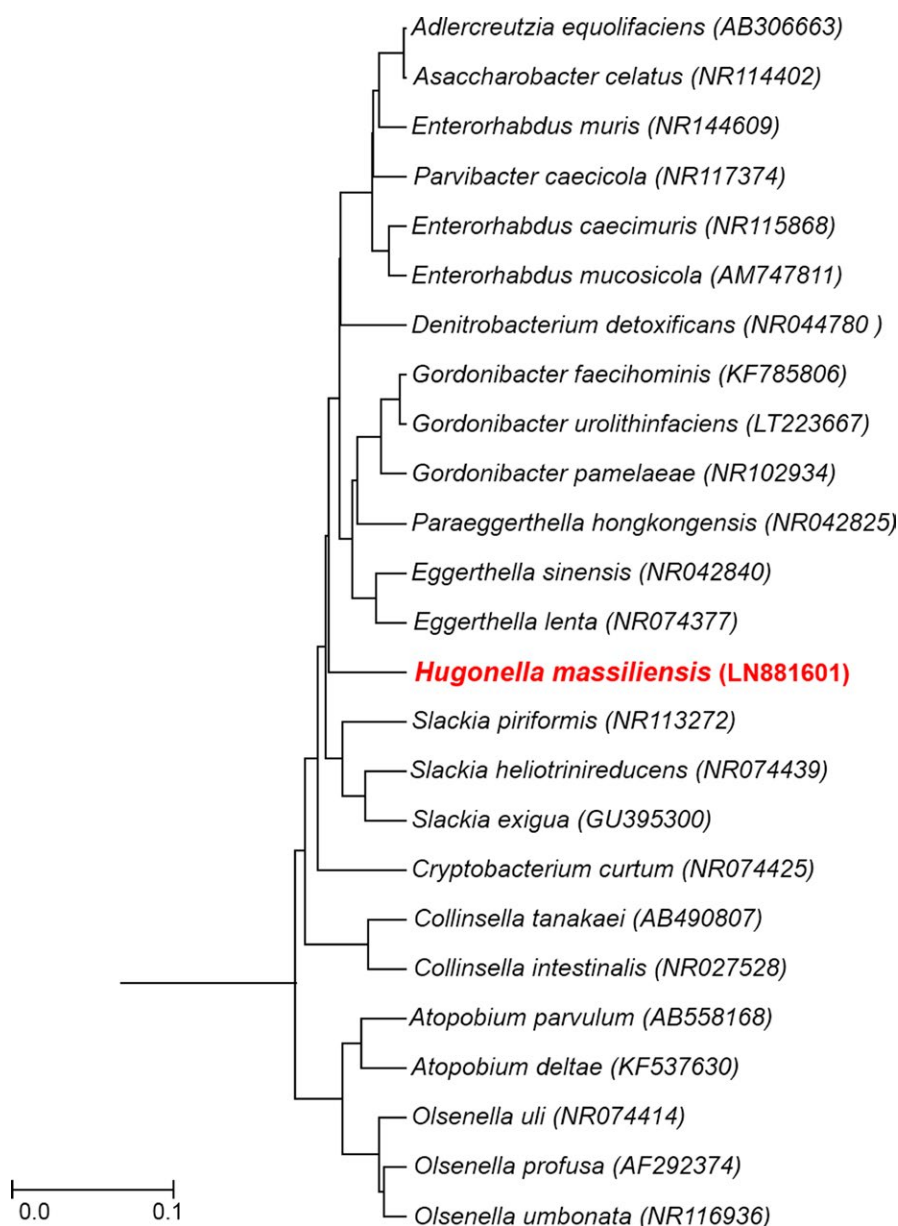
FIGURE 1 Phylogenetic tree highlighting the position of _Hugonella massiliensis_ strain AT8[T] relative to other close strains. Sequences were aligned using Muscle v3.8.31 with default parameters and phylogenetic inferences were obtained using neighbor-joining method with 500 bootstrap replicates, within MEGA6 software. The scale bar represents a 0.1% nucleotide sequence divergence

atmospheres. The principal characteristics of the strain and classification are present in Table 1.

Strain AT8^T has no catalase and oxidase activities. Using an API ZYM strip, positive reactions were observed for esterase (C4), esterase lipase (C8), naphthol-AS-BI-phosphohydrolase, lipase (C14), leucine arylamidase, valine arylamidase, trypsin, α-chymotrypsin, β-galactosidase, N-acetyl- β-glucosaminidase, α-galactosidase, β-glucuronidase, α-glucosidase, β-glucosidase, α-fucosidase, α-mannosidase, and negative reaction was observed for alkaline phosphatase. An API 50CH strip showed negative reactions for arbutin, salicin, D-maltose, D-fructose, D-sucrose, D-raffinose, glycerol, erythritol, D-ribose, D-xylose, L-xylose, D-adonitol, methyl-βD-xylopyranoside, D-glucose, D-galactose, D-lactose, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl-αD-mannopyranoside, methyl-αD-glucopyranoside, D-cellobiose, D-melibiose, D-trehalose, D-melezitose, starch, glycogen, xylitol, gentiobiose, D-turanose, D-lyxose, D-tagatose, D-fucose, L-fucose, D-arabitol, L-arabitol, and potassium gluconate; and positive reactions for D-mannose, amygdalin, N-acetyl-glucosamine, esculin ferric citrate, D-arabinose, inulin, potassium 2-ketogluconate, and potassium 5-ketogluconate. The phenotypic characteristics of Strain AT8^T were compared with the most closely related species (Table 2). The most abundant fatty acids were saturated structures: hexadecanoic acid C16:0 (35%) and tetradecanoic acid C14:0 (31%). Several branched fatty acids were also detected and only three unsaturated structures were described (Table 3).

Antimicrobial susceptibility testing demonstrates that the strain AT8^T was susceptible to cefoxitin, ciprofloxacin, clindamycin, fosfomycin, linezolid, oxacillin, penicillin, pristinamycin, rifampicin, teicoplanin, trimethoprim-sulfamethoxazole, vancomycin, and metronidazole, and resistant to doxycycline, erythromycin, and colistin.

3.2 | Strain identification and phylogenetic analysis

The spectra of strain AT8^T obtained by MALDI-TOF MS did not match with any strains in our database (Bruker, continuously incremented with our data) suggesting that our isolate could be a new isolate. We added the spectrum from strain AT8^T to URMITE database (Figure S3) (http://www.mediterranee-infection.com/article.php?laref=256&titre=urms-database). PCR-based identification of the16S rDNA of strain AT8^T (Accession number: LN881601) demonstrated 92% of 16S rRNA gene sequence similarity with the reference *Eggerthella lenta* (DSM 2243), the phylogenetically closest validated species (Figure 1). This value is under the threshold that allows the identification of a new genus, as established by Kim et al., (2014). Consequently, strain AT8^T is considered as the type strain of the first isolate of a new genus named *Hugonella* gen. nov., for which *Hugonella massiliensis* sp. nov strain AT8^T is the type strain. Finally, the gel view showed the mass spectra's differences with other closely related genera of *Eggerthellacae* family (Figure 2).

3.3 | Genome properties

The genome of *H. massiliensis* strain AT8^T (Accession number: FAUL00000000) contains 2,091,845 bp with 63.46% of G+C content
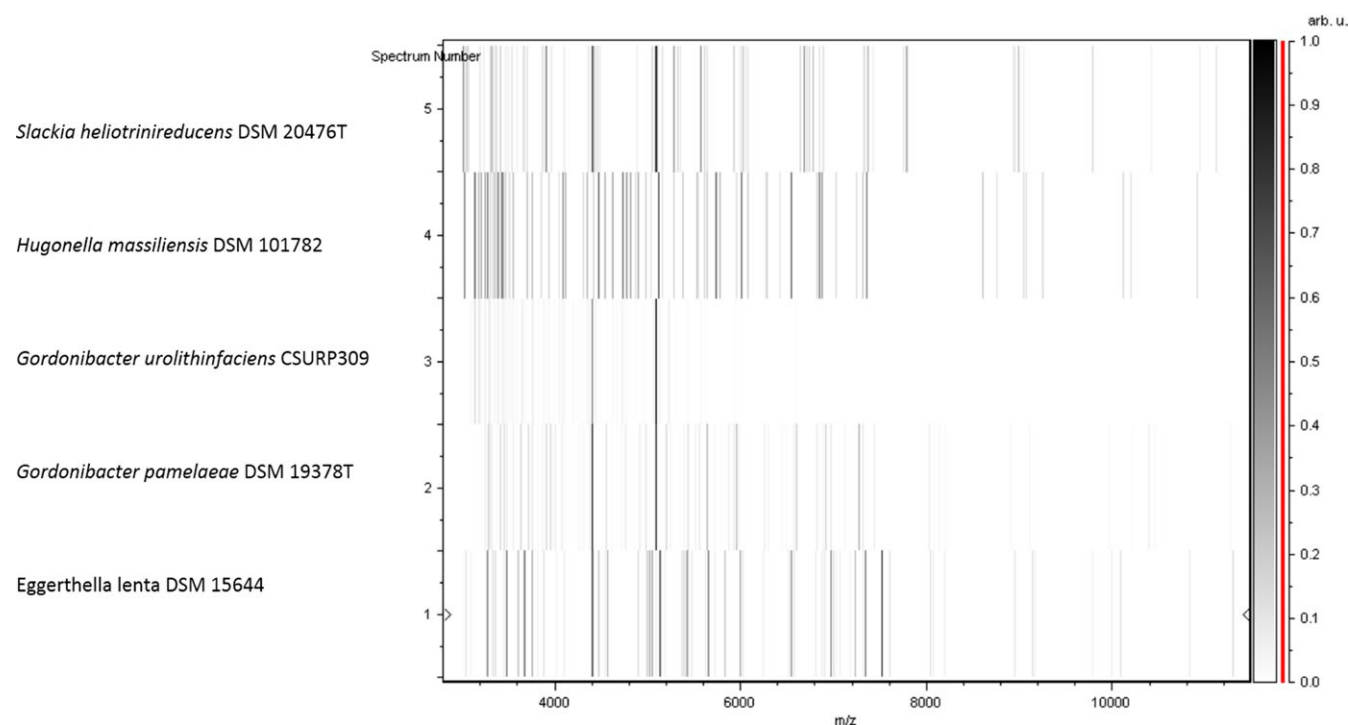
FIGURE 2 Gel view comparing *Hugonella massiliensis* strain AT8^T to other closely related species. The gel view displays the raw spectra of strain AT8^T of loaded spectrum files arranged in a pseudo-gel like look. The x-axis records the m/z value. The left y-axis displays the running spectrum number originating from subsequent spectra loading. The peak intensity is expressed by a gray scale scheme code. The right y-axis indicates the relation between the color of a peak and its intensity, in arbitrary units. Displayed species are indicated on the left

TABLE 4 The genome nucleotide content and gene count levels of *Hugonella massiliensis* strain AT8[T]

Attribute	Genome (total)	
	Value	% of total[a]
Size (bp)	2,091,845	100
G+C content (%)	1,326,454	63.46
Coding region (bp)	1,844,967	88.19
Total genes	1,849	100
RNA genes	68	3.69
Protein-coding genes	1,781	100
Genes with function prediction	1,438	80.74
Genes assigned to COGs	1,242	69.73
Genes with peptide signals	156	8.75
CRISPR repeats	0	0
ORFn genes	90	5.05
Genes associated with PKS or NRPS	8	0.44
No. of antibiotic-resistant genes	0	0

[a]The total is based on either the size of the genome in base pairs or the total number of protein-coding genes in the annotated genome.

rRNA, 50 genes are tRNA genes). A total of 1,438 genes (80.74%) were assigned as putative function (by cogs or by NR blast). Ninety genes (5.05%) were identified as ORFans. The remaining genes were annotated as hypothetical proteins (201 genes = 11.29%) (Table 4). Table 5 represents the gene's distribution into clusters of orthologous groups (COGs).

3.4 | Genome comparison

We made some comparisons with the closest annotated sequenced genomes currently available: *Slackia piriformis* ADMD00000000.1, *Slackia exigua* ACUX00000000.2, *Gordonibacter pamelaeae* FP929047.1, *Enterorhabdus caecimuris* ASSY00000000.1, *Slackia heliotrinireducens* CP001684.1, *Adlercreutzia equolifaciens* AP013105.1, *Eggerthella lenta* CP001726.1, *Cryptobacterium curtum* CP001682.1, and *Denitrobacterium detoxificans* (CP011402.1) (Table 6). The strain AT8[T] has a smaller draft genome sequence than those of *S. piriformis*, *S. exigua*, *A. equolifaciens*, *G. pamelaeae*, *E. caecimuris*, *S. heliotrinireducens*, *E. lenta*, and *D. detoxificans* (respectively, 2.09, 2.12, 2.10, 2.86, 3.61, 2.96, 3.17, 3.63 Mb and 2.45) but larger than those of *C. curtum* (1.62 Mb). The G+C content of *H. massiliensis* (63.46%) is larger than those of *S. piriformis*, *S. exigua*, *A. equolifaciens*, *C. curtum*, *S. Heliotrinireducens*, and
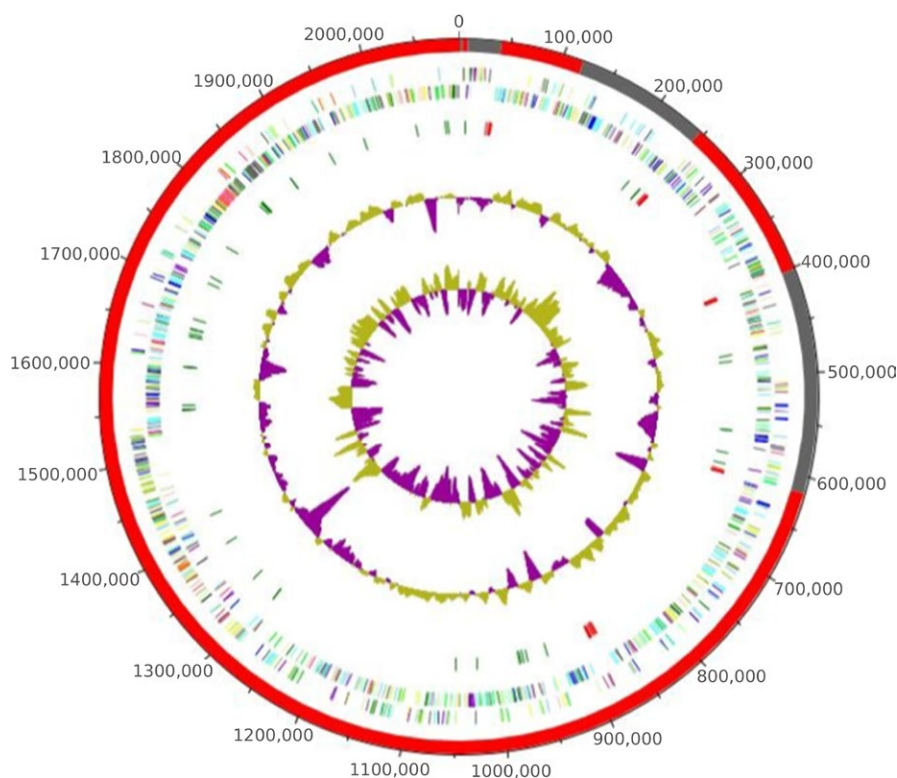
FIGURE 3 Graphical circular map of the chromosome. From outside to the center: Genes on the forward strand colored by Clusters of Orthologous Groups (COG) categories (only gene assigned to COG), genes on the reverse strand colored by COG categories (only gene assigned to COG), RNA genes (tRNAs green, rRNAs red), G+C content, and G+C skew

(Table 4, Figure 3) and is composed of nine scaffolds with 12 contigs. The draft genome was shown to encode 1,849 predicted genes, among which 1,781 were protein-coding genes, and 68 were RNAs (six genes are 5S rRNA, six genes are 16S rRNA, six genes are 23S

D. detoxificans (57.68, 62.15, 63.46, 50.91, 60.21%, and 59.5%, respectively), but smaller than those of *G. pamelaeae*, *E. caecimuris*, and *E. lenta* (65.48, 64.13, and 64.20%, respectively). The gene content of *H. massiliensis* is smaller than those of *S. piriformis*,

**TABLE 5** Number of genes associated with the 25 general COG functional categories

Code	Value	% of total	Description
[J]	154	8.646828	Translation
[A]	0	0	RNA processing and modification
[K]	72	4.0426726	Transcription
[L]	64	3.593487	Replication, recombination, and repair
[B]	0	0	Chromatin structure and dynamics
[D]	21	1.1791129	Cell cycle control, mitosis, and meiosis
[Y]	0	0	Nuclear structure
[V]	22	1.2352611	Defense mechanisms
[T]	49	2.7512634	Signal transduction mechanisms
[M]	56	3.1443012	Cell wall/membrane biogenesis
[N]	12	0.6737788	Cell motility
[Z]	0	0	Cytoskeleton
[W]	7	0.39303765	Extracellular structures
[U]	19	1.0668164	Intracellular trafficking and secretion
[O]	69	3.8742278	Posttranslational modification, protein turnover, chaperones
[X]	9	0.5053341	Mobilome: prophages, transposons
[C]	147	8.25379	Energy production and conversion
[G]	61	3.425042	Carbohydrate transport and metabolism
[E]	155	8.702975	Amino acid transport and metabolism
[F]	59	3.3127456	Nucleotide transport and metabolism
[H]	80	4.4918585	Coenzyme transport and metabolism
[I]	70	3.930376	Lipid transport and metabolism
[P]	74	4.154969	Inorganic ion transport and metabolism
[Q]	23	1.2914094	Secondary metabolites biosynthesis, transport, and catabolism
[R]	117	6.569343	General function prediction only
[S]	52	2.919708	Function unknown
_	539	30.263897	Not in COGs

[a]The total is based on the total number of protein-coding genes in the annotated genome.

TABLE 6 Genome comparison between *Hugonella massiliensis* strain AT8[T] and closely related species

Name of organisms	INSDC	Size (Mb)	G+C (%)	Total Genes
Hugonella massiliensis strain AT8[T]	FAUL00000000	2.091	63.46	1,781
Eggerthella lenta DSM2243	CP001726.1	3.63	64.2	3,070
Cryptobacterium curtum DSM15641	CP001682.1	1.62	50.91	1,357
Gordonibacter pamelaeae strain 7-10-1-b	FP929047.1	3.61	65.48	2,027
Adlercreutzia equolifaciens FJC-B9	AP013105.1	2.86	63.46	2,281
Slackia piriformis YIT_12062	ADMD00000000.1	2.12	57.68	1,799
Slackia exigua ATCC700122	ACUX00000000.2	2.09	62.15	2,029
Enterorhabdus caecimuris B7	ASSY00000000.1	2.96	64.13	2,455
Slackia heliotrinireducens DSM20476	CP001684.1	3.16	60.21	2,765
Denitrobacterium detoxificans DSM 21843	CP011402.1	2.45	59.5	1,989
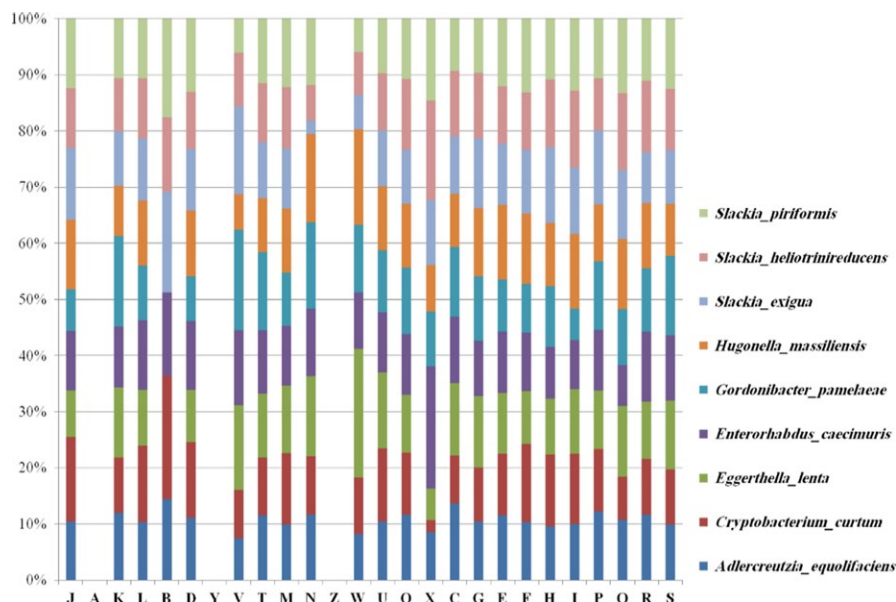
FIGURE 4 Distribution of functional classes of predicted genes according to the clusters of orthologous groups of proteins of *Hugonella massiliensis* strain AT8[T] and other close related species

S. exigua, *A. equolifaciens*, *G. pamelaeae*, *E. caecimuris*, *S. heliotrinireducens*, *E. lenta*, and *D. detoxificans* (1,781, 1,799, 2,029, 2,281, 2,027, 2,455, 2,765, 3,070, 1,989, respectively), but larger than those of *C. curtum* (1,357) (Table 6). The distribution of genes into COG categories is identical in all genomes compared (Figure 4). Calculation of the Average Genomic Identity of Orthologous gene Sequences (AGIOS) shows that *H. massiliensis* shared 800, 757, 854, 482, 796, 830, 922, and 844 orthologous genes with *S. piriformis*, *S. exigua*, *S. heliotrinireducens*, *G. pamelaeae*, *C. curtum*, *E. caecimuris*, *E. lenta*, and *A. equolifaciens*, respectively (Table 7). The AGIOS values ranged from 53.43% to 81.88% among species with standing in nomenclature, except *H. massiliensis*. When *H. massiliensis* was compared to the other closest species, the AGIOS values ranged from 57.10% with *S. exigua* to 71.59% with *G. pamelaeae* (Table 7). Among species with standing in nomenclature, dDDH values ranged from 19.4% to 29.4%. dDDH values between strain AT8[T] and compared species ranged from 19.8% with *A. equolifaciens* to 22% with *S. piriformis* and *C. curtum* (Table 8).

4 | DISCUSSION

Here, we used the culturomics approach to study the microbial diversity of the digestive tract of an obese subject in order to enrich the obese microbiota repertoire and establish a possible relationship between the biodiversity and morbid obesity in our patient. Here, we report the phenotypic characterization and genomic description of this new bacterial genus isolated from a stool specimen collected from a French female obese patient. Results reported here support the effectiveness of the culturomics procedure in the exploration of the bacterial diversity and the discovery of the remained unknown species to make their genomes available for eventual metagenomics study. In addition, culturing new bacteria permits the testing of

culture conditions and provides information on antibiotic susceptibility. It is noted that this work does not present a medical interest of this strain but only broadens the knowledge about the microbial diversity of the human gut and extends the repertoire of the new species.

5 | CONCLUSION

Based on phenotypic, phylogenetic, MALDI-TOF, and genomic analyses, we formally propose the creation of the genus *Hugonella* gen. nov., including *Hugonella massiliensis* gen. nov., sp. nov currently the only cultivated species. Indeed, *H. massiliensis* strain AT8[T] is only 92% 16S rRNA sequence similarity with *Eggerthella lenta* DSM2243 which gives it the status of a new genus. The strain has been isolated from a stool specimen of a morbidly obese French woman, as part of the culturomics study by anaerobic culture at 37°C. Several other bacterial species that remain undescribed were also isolated from different stool specimens using different culture conditions, suggesting that the human intestinal microbiota remains partially unknown and its diversity has yet to be fully explored.

5.1 | Description of *Hugonella* gen. nov

Hugonella (Hu.gon.el'la, ML. dim. suffix *tella*; M.L. fem. n. *Hugonella* named after the French bacteriologist Perrine Hugon). Belongs to the family *Eggerthellaceae* within the phylum *Actinobacteria*. Cells are gram-positive cocci, nonmotile, and nonspore forming. Mesophilic and do not require NaCl for growth. Cells do not produce catalase and oxidase. The most abundant fatty acids were saturated structures: hexadecanoic acid C16:0 (35%) and tetradecanoic acid C14:0 (31%). Several branched fatty acids were also detected and only three unsaturated structures were described. The DNA G+C content is approximately 63%. *Hugonella massiliensis* strain AT8[T] the type strain

TABLE 7 The number of orthologous protein shared between genomes (upper right), average percentage similarity of nucleotides corresponding to orthologous proteins shared between genomes (lower left), and numbers of proteins per genome (bold)

	Hugonella massiliensis	Slackia piriformis	Slackia exigua	Slackia heliotrinireducens	Gordonibacter pamelaeae	Cryptobacterium curtum	Enterorhabdus caecimuris	Eggerthella lenta	Adlercreutzia equolifaciens	Denitrobacterium detoxificans
Hugonella massiliensis	**1,781**	59.31	57.11	68.36	71.60	67.79	60.90	70.37	844	847
Slackia piriformis	800	**1,799**	798	870	528	713	892	983	911	805
Slackia exigua	757	61.71	**2,029**	825	437	736	761	844	789	786
Slackia heliotrinireducens	854	60.54	57.58	**2,765**	499	725	864	948	878	867
Gordonibacter pamelaeae	482	61.32	57.98	70.12	**2,027**	434	584	715	587	629
Cryptobacterium curtum	796	5743	53.43	63.75	65.51	**1,357**	730	834	745	785
Enterorhabdus caecimuris	830	59.30	62.29	59.98	63.70	56.45	**2,455**	1,076	1,143	894
Eggerthella lenta	922	60.55	57.88	69.33	81.88	64.70	62.93	**3,070**	1,108	940
Adlercreutzia equolifaciens	69.59	59.60	57.02	68.64	74.63	63.90	67.426	73.39	**2,281**	917
Denitrobacterium detoxificans	72 .5	67.69	68.26	68.56	71.11	68.49	68.71	70.08	68.74	**1,989**

TABLE 8 Pairwise comparison of *Hugonella massiliensis* with other species using GGDC, formula 2 (DDH estimates based on identities/HSP length)* upper right

	Hugonella massiliensis	Cryptobacterium curtum	Gordonibacter pamelaeae	Eggerthella lenta	Enterorhabdus caecimuris	Adlercreutzia equolifaciens	Slackia piriformis	Slackia exigua	Slackia heliotrinireducens	Denitrobacterium detoxificans
Hugonella massiliensis	**100 ± 00**	22 ± 2.35	20.5 ± 2.32	20.1 ± 2.31	20.6 ± 2.32	19.8 ± 2.3	22 ± 2.35	20.8 ± 2.33	20.3 ± 2.32	21.8 ± 2.35
Cryptobacterium curtum		**100 ± 00**	25.4 ± 2.41	22.4 ± 2.36	22.5 ± 2.36	24.1 ± 2.39	22.6 ± 2.36	23.6 ± 2.38	26.8 ± 2.42	21.1 ± 2.30
Gordonibacter pamelaeae			**100 ± 00**	29.4 ± 2.44	22.4 ± 2.36	22.4 ± 2.36	20.8 ± 2.33	20.3 ± 2.31	19.8 ± 2.3	19.7 ± 2.30
Eggerthella lenta				**100 ± 00**	21.8 ± 2.35	21.5 ± 2.34	20.9 ± 2.33	20.1 ± 2.31	19.4 ± 2.29	20.2 ± 2.40
Enterorhabdus caecimuris					**100 ± 00**	25.6 ± 2.41	22.3 ± 2.36	19.8 ± 2.3	19.5 ± 2.29	19.4 ± 2.30
Adlercreutzia equolifaciens						**100 ± 00**	21.3 ± 2.34	19.4 ± 2.29	19.9 ± 2.3	19.5 ± 2.40
Slackia piriformis							**100 ± 00**	21 ± 2.33	21.5 ± 2.34	20.2 ± 2.30
Slackia exigua								**100 ± 00**	20.7 ± 2.32	20.3 ± 2.30
Slackia heliotrinireducens									**100 ± 00**	20.5 ± 2.30
Denitrobacterium detoxificans										**100 ± 00**

Bold value: Presents the comparison between the strain and itself

was isolated from feces of a 51-year-old obese French woman (BMI 44.38 kg/m2).

5.2 | Description of *Hugonella massiliensis* sp. nov

Hugonella massiliensis (mas.si.li.en'sis, L. fem. adj., *massiliensis* of Massilia, the Roman name of Marseille, France, where the type strain was isolated). Strictly anaerobic gram-positive cocci, nonmotile, and nonspore forming. Cells diameter is of 0.8–1.2 μm. An optimal growth was observed at 37°C, pH 7, and NaCl is not required for growth. Colonies are smooth, shiny and measure 2–5 mm. No catalase and no oxidase activities were observed.

Using API strips, positive reactions were observed for esterase, esterase lipase, naphthol-AS-BI-phosphohydrolase, D-mannose, amygdalin, N-acetyl-glucosamine, esculin ferric citrate, D-arabinose, inulin, potassium 2-ketogluconate, potassium 5-ketogluconate, lipase (C14), leucine arylamidase, valine arylamidase, trypsin, α-chymotrypsin, β-galactosidase, N-acetyl- β-glucosaminidase, α-galactosidase, β-glucuronidase, α-glucosidase, β-glucosidase, α-fucosidase, and α-mannosidase. Strain AT8[T] was susceptible to cefoxitin, ciprofloxacin, clindamycin, fosfomycin, linezolid, oxacillin, penicillin, pristinamycin, rifampicin, teicoplanin, trimethoprim-sulfamethoxazole, vancomycin, and metronidazole. The most abundant fatty acids were hexadecanoic acid C16:0 (35%) and tetradecanoic acid C14:0 (31%).

The genome of *H. massiliensis* strain AT8[T] is 2,091,845 bp with 63.46% of G+C content. The 16S rRNA and genome sequences are deposited in EMBL-EBI under accession numbers LN881601 and FAUL00000000, respectively. MALDI-TOF Spectrum of strain AT8[T] is available in: (http://mediterraner-infection.com/article.php?laref=256&titre=urms-database). The habitat of the microorganism is the human gut. The type strain AT8[T] (=CSUR P2118 = DSM 101782) was isolated from the stool specimen of an obese French individual as part of a culturomics study.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Anderson, R. C., Rasmussen, M. A., Jensen, N. S., & Allison, M. J. (2000). *Denitrobacterium detoxificans* gen. nov., sp. nov., a ruminal bacterium that respires on nitrocompounds. *International Journal of Systematic and Evolutionary Microbiology*, 50(2), 633–638.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... Sherlock, G.. (2000). Gene ontology: Tool for the unification of biology. the gene ontology consortium. *Nature Genetics*, 25, 25–29.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., ... Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19(5), 455–477.

Bolger, A.M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120.

Dione, N., Sankar, S. A., Lagier, J. C., Khelaifia, S., Michele, C., Armstrong, N., ... Fournier, P. E. (2016). Genome sequence and description of *Anaerosalibacter massiliensis* sp. nov. *New Microbes and New Infections*, 10, 66–76.

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., ... Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science*, 308(5728), 1635–1638.

Field, D., Garrity, G., Gray, T., Morrison, N., Selengut, J., Sterk, P., ... Wipat, A. (2008). The minimum information about a genome sequence (MIGS) specification. *Nature Biotechnology*, 26(5), 541–547.

Galperin, M. Y., Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Research*, 43, D261–D269.

Garrity, G. M., & Holt, J. G. (2001). The Road Map to the Manual. In G. M. Garrity, D. R. Boone, & R. W. Castenholz (Eds.), *Bergey's Manual of Systematic Bacteriology*, Vol. 1, 2nd ed. (pp. 119–169). New York: Springer.

Gouret, P., Paganini, J., Dainat, J., Louati, D., Darbo, E., Pontarotti, P., & Levasseur, A. (2011). Integration of evolutionary biology concepts for functional annotation and automation of complex research in evolution: The multi-agent software system DAGOBAH. InP. Pontarotti, (Ed). *Evolutionary Biology- Concepts, Biodiversity, Macroevolution and Genome Evolution* (pp. 71–87). Berlin, Heidelberg: Springer-Verlag.

Gouret, P., Thompson, J. D., & Pontarotti, P. (2009). PhyloPattern: Regular expressions to identify complex patterns in phylogenetic trees. *BMC Bioinformatics*, 10, 298.

Gupta, R. S., Chen, W. J., Adeolu, M., Chai, Y. (2006). Molecular signatures for the class *Coriobacteriia* and its different clades; proposal for division of the class *Coriobacteriia* into the emended order *Coriobacteriales*, containing the emended family *Coriobacteriaceae* and *Atopobiaceae* fam. nov., and *Eggerthellales* ord. nov., containing the family *Eggerthellaceae* fam. nov., nov. IJSB (2006); 63: 3379–3397.

Kageyama, A., Benno, Y., & Nakase, T. (1999). Phylogenetic evidence for the transfer of *Eubacterium lentum* to the genus *Eggerthella* as *Eggerthella lenta* gen. nov., comb. nov. *International Journal of Systematic Bacteriology*, 49(4), 1725–1732.

Käll, L., Krogh, A., & Sonnhammer, E. L. (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology*, 338(5), 1027–1036.

Kim, M., Oh, H. S., Park, S. C., & Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology*, 64(2), 346–351.

Kim, K. S., Rowlinson, M. C., Bennion, R., Liu, C., Talan, D., Summanen, P., & Finegold, S. M. (2010). Characterization of *Slackia exigua* isolated from human wound infections, including abscesses of intestinal origin. *Journal of Clinical Microbiology*, 48(4), 1070–1075.

Lagesen, K., Hallin, P., Rodland, E. A., Staerfeldt, H. H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, 35(9), 3100–3108.

Lagier, J. C., Hugon, P., Khelaifia, S., Fournier, P. E., La Scola, B., & Raoult, D. (2015). The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. *Clinical Microbiology Reviews*, 28(1), 237–264.

Lanigan, G. W. (1976). *Peptococcus heliotrinreducans*, sp. nov., a cytochrome-producing anaerobe which metabolizes pyrrolizidine alkaloids. *Journal of General Microbiology*, 94(1), 1–10.

Lechner, M., Findeiss, S., Steiner, L., Marz, M., Stadler, P. F., & Prohaska, S. J. (2011). Proteinortho: Detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics*, *12*, 124.

Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, *25*(5), 955–964.

Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K., & Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature*, *489*(7415), 220–230.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., ... Wang, J. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience*, *1*, 18.

Maruo, T., Sakamoto, M., Ito, C., Toda, T., & Benno, Y. (2008). Adlercreutzia equolifaciens gen. nov., sp. nov., an equol-producing bacterium isolated from human faeces, and emended description of the genus I. *International Journal of Systematic and Evolutionary Microbiology*, *58*(5), 1221–1227.

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., & Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*, *14*, 60.

Moore, W. E. C., Cato, E. P., & Holdeman, L. V. (1971). *Eubacterium lentum* (Eggerth) Prevot 1938: Emendation of description and designation of the neotype strain. *International Journal of Systematic Bacteriology*, *21*, 307–310.

Morel, A. S., Dubourg, G., Prudent, E., Edouard, S., Gouriet, F., Casalta, J. P., ... Raoult, D. (2015). Complementarity between targeted real-time specific PCR and conventional broad-range 16S rDNA PCR in the syndrome-driven diagnosis of infectious diseases. *European Journal of Clinical Microbiology and Infectious Diseases*, *34*(3), 561–570.

Ramasamy, D., Mishra, A. K., Lagier, J. C., Padhmanabhan, R., Rossi, M., Sentausa, E., ... Fournier, P. E. (2014). A polyphasic strategy incorporating genomic data for the taxonomic description of new bacterial species. *International Journal of Systematic and Evolutionary Microbiology*, *64*(2), 384–391.

Rosselló-Móra, R. (2006). DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation. In E. Stackebrandt (Ed.), *Molecular identification, systematics, and population structure of prokaryotes*. Berlin Springer, p. 23–50.

Sasser, M. (2006). *Bacterial identification by gas chromatographic analysis of fatty acids methyl esters (GC-FAME) Technical note 101*. Newark, DE: MIDI Inc.

Seng, P., Drancourt, M., Gouriet, F., La Scola, B., Fournier, P. E., Rolain, J. M., & Raoult, D. (2009). Ongoing revolution in bacteriology: Routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clinical Infectious Diseases*, *49*(4), 543–551.

Skerman, V. B. D., McGowan, V., & Sneath, P. H. A. (1980). Approved lists of bacterial names. *International Journal of Systematic Bacteriology*, *30*, 225–420.

Wade, W. G., Downes, J., Dymock, D., Hiom, S. J., Weightman, A. J., Dewhirst, F. E., ... Coleman, B. (1999). The family Coriobacteriaceae: Reclassification of Eubacterium exiguum (Poco et al. 1996) and Peptostreptococcus heliotrinreducens (Lanigan 1976) as Slackia exigua gen. nov., comb. nov., and Slackia heliotrinireducens gen. nov., comb. nov., and Eubacterium lentum (Prevot 1938) as Eggerthella lenta gen. nov., comb. nov. *International Journal of Systematic Bacteriology*, *49*(2), 595–600.

Welker, M., & Moore, E. R. (2011). Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. *Systematic and Applied Microbiology*, *34*(1), 2–11.

Woese, C. R., Kandler, O, & Wheelis, M. L. (1990). Towards a natural system of organisms: Proposal for the domains archaea, bacteria, and eucarya. *Proceedings of the National Academy of Sciences USA*, *87*, 4576–4579.

Würdemann, D., Tindall, B. J., Pukall, R., Lünsdorf, H., Strömpl, C., Namuth, T., ... Oxley, A. P. (2009). *Gordonibacter pamelaeae* gen. nov., sp. nov., a new member of the *Coriobacteriaceae* isolated from a patient with Crohn's disease, and reclassification of *Eggerthella hongkongensis* Lau *et al.* 2006 as *Paraeggerthella hongkongensis* gen. nov., comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, *59*(6), 1405–1415.

Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, *18*(5), 821–829.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

Article XV:

Noncontiguous finished genome sequence and description of *Enterococcus massiliensis* sp. nov.

Stéphanie Le Page, Theresa Cimmino, **Amadou Togo**, Matthieu Million, C. Michelle, Saber Khelaifia, Jean-Christophe. Lagier, Didier Raoult and Jean-Marc Rolain

Noncontiguous finished genome sequence and description of *Enterococcus massiliensis* sp. nov.

S. Le Page, T. Cimmino, A. Togo, M. Million, C. Michelle, S. Khelaifia, J.-C. Lagier, D. Raoult and J.-M. Rolain

Aix-Marseille Université, URMITE UM 63 CNRS 7278 IRD 198 INSERM U1905, IHU Méditerranée Infection, Facultés de Médecine et de Pharmacie, Marseille, France

Abstract

Enterococcus massiliensis strain sp. nov. (= CSUR P1927 = DSM 100308) is a new species within the genus *Enterococcus*. This strain was first isolated from a fresh stool sample of a man during culturomics study of intestinal microflora. *Enterococcus massiliensis* is a Gram-positive cocci, facultative anaerobic and motile. *E. massiliensis* is negative for mannitol and positive for β-galactosidase, contrary to *E. gallinarum*. The complete genome sequence is 2 712 841 bp in length with a GC content of 39.6% and contains 2617 protein-coding genes and 70 RNA genes, including nine rRNA genes.

© 2016 The Author(s). Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases.

Corresponding author: J.-M. Rolain, Aix-Marseille Université, URMITE UM 63 CNRS 7278 IRD 198 INSERM U1905, IHU Méditerranée Infection, Facultés de Médecine et de Pharmacie, 27 boulevard Jean Moulin, 13385 Marseille cedex 05, France
E-mail: jean-marc.rolain@univ-amu.fr

Introduction

Enterococcus massiliensis sp. nov. strain AM1[T] (= CSUR P1927 = DSM 100308) belongs to the genus *Enterococcus*. This bacterium is a Gram-positive, facultative anaerobic, motile and unpigmented. It was isolated from a fresh stool sample of a human in Marseille as part of culturomics study [1]. Currently, a polyphasic approach that combines proteomic by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) analysis [2], genomic data with 16S rRNA sequence identity and phylogeny [3], genomic G+C content diversity and DNA-DNA hybridization (DDH) and phenotypic characterization is used to describe new bacterial species [4].

Enterococcus were classified in the genus of *Streptococcus* because of the presence of the D antigen until 1984 [5]. But after analysis of the genome of *Streptococcus faecalis* and *S. faecium*, these strains have been transferred to the genus *Enterococcus* [6]. Members of the genus *Enterococcus* are components of the intestinal flora of humans and animals. There are opportunistic pathogens with two principal strains: *Enterococcus faecalis* and *Enterococcus faecium*, responsible for nosocomial infections [7,8].

Here we present a summary classification and a set of features for *E. massiliensis* sp. nov. strain AM1[T] together with the description of the complete genome sequence and annotation. These characteristics support the circumscription of the species *E. massiliensis*.

Organism Information

A stool sample was collected in 2015 from a voluntary patient as a negative control and isolated on Columbia agar supplemented with 5% sheep's blood (bioMérieux, Marcy-l'Étoile, France) in aerobic and anaerobic condition using GasPak EZ Anaerobe Container System Sachets (Becton Dickinson (BD), San Diego, CA, USA) at 37°C. *Enterococcus massiliensis* was sequenced as part of a culturomics study aiming to isolate all bacterial species colonizing the human gut [9]. *Enterococcus massiliensis* strain AM1[T] (GenBank accession no. LN833866) exhibited a 97% 16S

rRNA nucleotide sequence similarity with *Enterococcus gallinarum* (JF915769), the phylogenetically closest validly published bacterial species (Fig. 1) after comparison with National Center for Biotechnology Information (NCBI) database. This value is lower than 98.7% 16S rRNA gene sequence similarity set as a threshold recommended by Stackebrandt and Ebers [3] to delineate a new species without carrying out DNA-DNA hybridization.

Growth occurred between 25°C and 37°C, but optimal growth was observed at 37°C, 24 hours after inoculation. Colonies were smooth and whitish, approximately 1 mm in diameter on 5% sheep's blood−enriched agar (bioMérieux). Growth of the strain was tested under anaerobic and microaerophilic conditions using GasPak EZ Anaerobe pouch (BD) and CampyGen Compact (Oxoid, Basingstoke, UK) systems, respectively, and in aerobic conditions, with or without 5% of CO_2. Growth was achieved under aerobic (with and without CO_2), microaerophilic and anaerobic conditions. Gram staining showed Gram-positive cocci without sporulation (Fig. 2A). A motility test was positive and realized with API M Medium (bioMérieux), a semisolid medium with an inoculation performed by swabbing one colony into the medium. After 24 hours of incubation, the growth of *E. massiliensis* was away from this stabbed line, characteristic of positive motility. Cells grown on agar exhibited a mean diameter of 0.5 μm and a mean length ranging from 1.1 to 1.3 μm (mean 1.2 μm), determined by negative staining transmission electron microscopy (Fig. 2B).

Differential phenotypic characteristics using API 50CH and API Zym system (bioMérieux) between *E. massiliensis* sp. nov. AM1[T] and other *Enterococcus* species [9] are presented in

Table 1. Antibiotic susceptibility testing was performed by the disk diffusion method on Müller-Hinton agar with blood (bioMérieux). *E. massiliensis* is susceptible to vancomycin, teicoplanin, linezolid, gentamicin, ciprofloxacin, doxycycline, rifampicin and pristinamycin and resistant or intermediate to penicillin G, oxacillin, ceftriaxone, cefoxitin, trimethoprim/sulfamethoxazole, fosfomycin, erythromycin and clindamycin.

Extended Features Descriptions

MALDI-TOF MS protein analysis was carried out as previously described [2] using a Microflex spectrometer (Bruker Daltonics, Leipzig, Germany). Twelve distinct deposits were done for strain AM1[T] from 12 isolated colonies. Twelve distinct deposits were done for strain AM1[T] from 12 isolated colonies. Spectra were imported into the MALDI BioTyper software, version 2.0 (Bruker), and analysed by standard pattern matching against 7765 bacterial spectra, including 92 spectra from 31 *Enterococcus* species, in the BioTyper database. Interpretation of scores was as follows: a score of ≥2 enabled the identification at the species level, a score of ≥1.7 but <2 enabled the identification at the genus level and a score of <1.7 did not enable any identification (scores established by the manufacturer, Bruker). For strain AM1[T], no significant MALDI-TOF MS score was obtained against the Bruker database, thus suggesting that our isolate was a new species. We incremented our database with the spectrum from strain AM1[T] (Fig. 3).
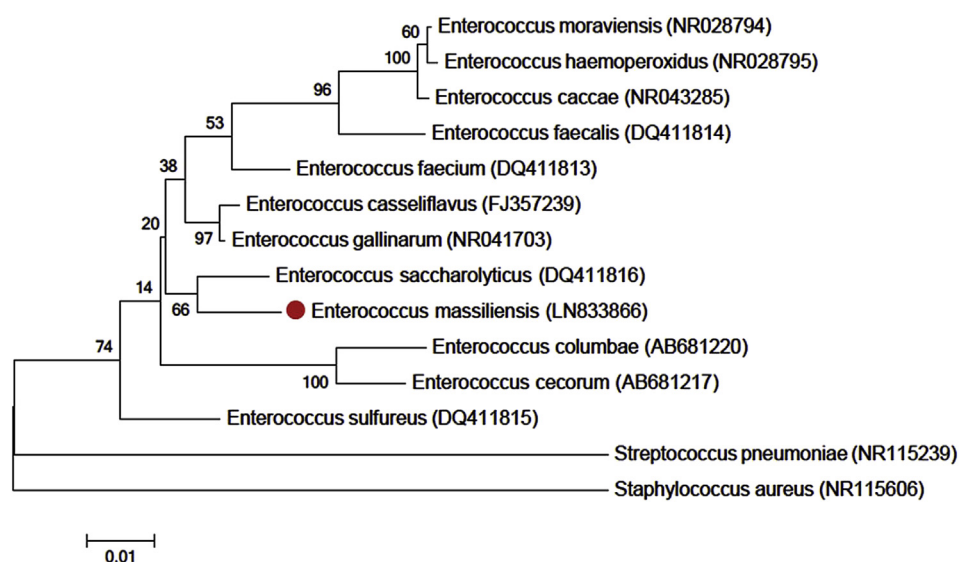
FIG. 1. Consensus phylogenetic tree highlighting position of *Enterococcus massiliensis* relative to other type strains within genus *Enterococcus* by 16S. GenBank accession numbers appear in brackets. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using maximum-likelihood method in MEGA6 software package. Numbers at nodes are percentages of bootstrap values from 1000 replicates that support nodes. *Streptococcus pneumoniae* and *Staphylococcus aureus* were used as outgroups. Scale bar = 1% nucleotide sequence divergence.
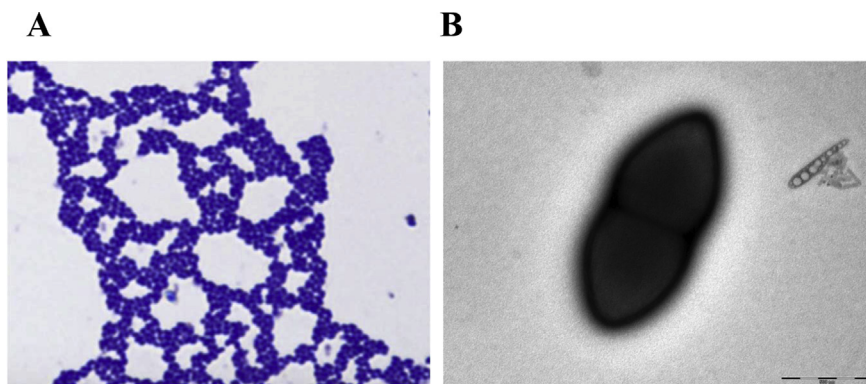
A   **B**

Genome Sequencing Information

Enterococcus massiliensis sp. nov. (GenBank accession no. CVRN00000000) is the 54 species described within *Enterococcus* genus.

After DNA extraction by the phenol−chloroform method, genomic DNA of *E. massiliensis* was sequenced on a MiSeq instrument (Illumina, San Diego, CA, USA) using paired end and mate pair strategies.

For genome annotation, open reading frames (ORFs) were predicted using Prodigal (http://prodigal.oml.gov/) with default parameters, but the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched for against the GenBank database (http://www.ncbi.nlm.nih.gov/genbank) and the Clusters of Orthologous Groups (COGs) databases using BLASTP. The tRNAScanSE tool [10] was used to find tRNA genes, whereas ribosomal RNAs were detected using RNAmmer [11] and BLASTn against the GenBank database.

The ARG-ANNOT database for acquired antibiotic resistance genes (ARGs) was used for a BLAST search using the Bio-Edit interface [12]. The assembled sequences were searched against the ARG database under moderately stringent conditions (e-value of 10^{-5}) for the *in silico* ARG prediction. *E. massiliensis* presents the *Lsa* gene, encoding a putative ABC protein *Lsa* with an identity to 72% with *Lsa* family ABC-F of *E. faecalis* in NCBI, which phenotypically confirms its resistance to clindamycin.

Analysis of presence of polyketide synthase (PKS) and non-ribosomal polyketide synthesis (NRPS) was performed by

TABLE 1. Differential characteristics of *Enterococcus massiliensis* sp. AM1, *E. faecalis*, *E. casseliflavus*, *E. gallinarum*, *E. haemoperoxidus*, *E. cecorum*, *E. sulfureus* and *E. caccae*

Property	*E. massiliensis*	*E. faecalis*	*E. casseliflavus*	*E. gallinarum*	*E. haemoperoxidus*	*E. cecorum*	*E. sulfureus*	*E. caccae*
Oxygen requirement	Facultative anaerobic	Facultative anaerobic	Facultative anaerobic	Facultative anaerobic	Facultative anaerobic	Facultative anaerobic	Facultative anaerobic	Facultative anaerobic
Gram stain	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive
Motility	Motile	−	Motile	Motile	−	−	−	−
Pigment	−	−	+	−	−	−	+	−
Production of:								
Alkaline phosphatase								
Catalase	−	−	−	−	−	−	−	−
Oxydase	−	−	−	−	−	−	−	−
β-Glucuronidase	−	−	−	−	NA	+	NA	NA
α-Galactosidase	−	−	−	−	NA	−	NA	NA
β-Galactosidase	+	−	−	−	NA	−	NA	NA
N-acetyl-glucosamine	+	−	+	+	+	−	+	+
Acid form:								
Mannitol	−	+	+	+	+	−	−	−
Sorbose	−	−	−	−	−	−	−	−
L-Arabinose	+	−	+	+	−	−	−	−
Sorbitol	−	+	v	−	−	−	−	−
D-Raffinose	+	−	+	+	−	+	+	−
Xylose	+	−	+	+	−	−	−	−
D-Trehalose	+	+	+	+	+	+	+	+
G+C content (%)	39.6	37.3	42.7	40.7	35.8	36.3	37.8	35.8
Habitat	Human stool	Intestine of mammals	Intestine of mammals	Intestine of mammals	Water	Commensal chicken	Plants	Human stool

+, positive result; −, negative result; v, variable result; NA, data not available.
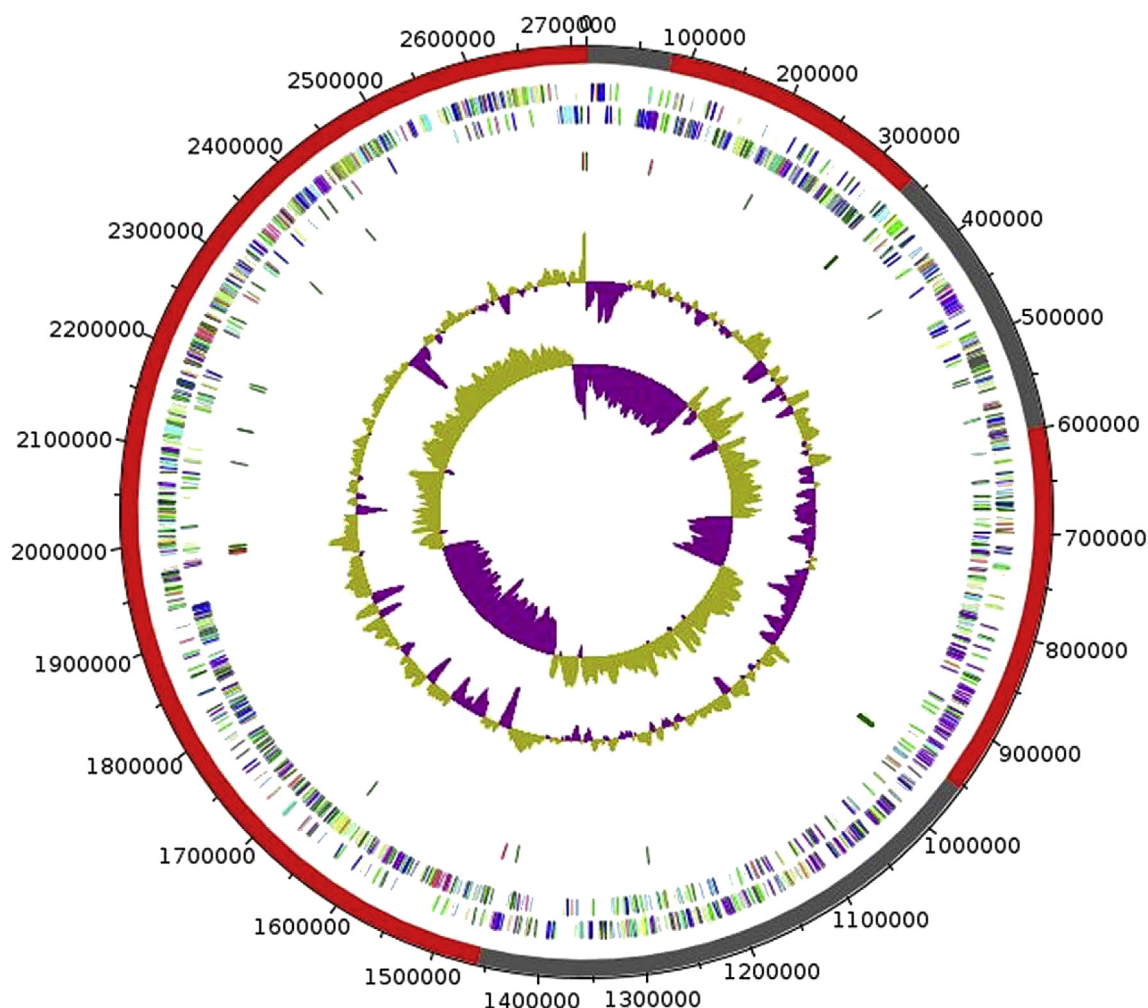
FIG. 3. Graphical circular map of chromosome. From outside to centre: genes on forward strand (coloured by COGs categories explain in Table 3), genes on reverse strand (coloured by COGs categories), RNA genes (tRNAs green, rRNAs red), G+C content, G+C skew.

discriminating the gene with a large size using a database realized in our laboratory; predicted proteins were compared against the nonredundant (nr) GenBank database using BLASTP and finally examined using antiSMASH [13]. Analysis of the genome revealed the absence of NRPKs and PKS. Lipoprotein signal peptides and the number of transmembrane helices were predicted using SignalP [14] and TMHMM [15], respectively. ORFans were identified if their BLASTP E value was lower than 10^{-3} for alignment length >80 amino acids.

We used the Genome-to-Genome Distance calculator (GGDC) web server (http://ggdc.dsmz.de) to estimate the overall similarity among the compared genomes and to replace the wet-lab DDH by a digital DDH [16,17]. GGDC 2.0 BLAST+ was chosen as alignment method, and the recommended formula 2 was taken into account to interpret the results.

We compared the genome of *E. massiliensis* with nine other genomes of *Enterococcus* strains. The genome is 2 712 841 bp long (one chromosome, no plasmid) with a GC content of 39.6%

(Table 2). The properties and statistics of the genome are summarized in Table 2. The draft genome of *E. massiliensis* is smaller than those of *E. moraviensis*, *E. haemoperoxidus*, *E. caccae*, *E. casseliflavus*, *E. gallinarum* and *E. faecalis* (3.60, 3.58, 3.56, 3.43,

TABLE 2. Nucleotide content and gene count levels of genome

Attribute	Value	% of total[a]
Genome size (bp)	2 712 841	100
DNA G+C content (bp)	1 075 567	39.6
DNA coding region (bp)	2 408 151	88.77
Total genes	2687	100
RNA genes	70	2.60
Protein-coding genes	2617	97.39
Genes with function prediction	1889	72.18
Genes assigned to COGs	1863	71.19
Genes with peptide signals	250	9.55
Genes with transmembrane helices	630	24.07

COGs, Clusters of Orthologous Groups database.
[a]Total is based on either size of genome in base pairs or total number of protein-coding genes in annotated genome.

TABLE 3. Number of genes associated with 25 general COGs functional categories

Color of COGs class	COGs class	Value	Percentage[a]	Description
	A	0	0	RNA processing and modification
	B	0	0	Chromatin structure and dynamics
	C	78	2.98	Energy production and conversion
	D	22	0.84	Cell cycle control, cell division, chromosome partitioning
	E	167	6.38	Amino acid transport and metabolism
	F	67	2.56	Nucleotide transport and metabolism
	G	248	9.48	Carbohydrate transport and metabolism
	H	46	1.76	Coenzyme transport and metabolism
	I	53	2.03	Lipid transport and metabolism
	J	154	5.88	Translation, ribosomal structure and biogenesis
	K	187	7.15	Transcription
	L	155	5.98	Replication, recombination and repair
	M	89	3.40	Cell wall/membrane/envelope biogenesis
	N	5	0.19	Cell motility
	O	54	2.06	Posttranslational modification, protein turnover, chaperones
	P	104	3.97	Inorganic ion transport and metabolism
	Q	20	0.76	Secondary metabolites biosynthesis, transport and catabolism
	R	260	9.94	General function prediction only
	S	190	7.26	Function unknown
	T	59	2.25	Signal transduction mechanisms
	U	24	0.91	Intracellular trafficking, secretion, and vesicular transport
	V	60	2.29	Defense mechanisms
	W	0	0	Extracellular structures
	Y	0	0	Nuclear structure
	Z	0	0	Cytoskeleton
	—	754	28.81	Not in COGs

COGs, Clusters of Orthologous Groups database.

[a]Total is based on total number of protein-coding genes in annotated genome.

3.16 and 2.96 Mb, respectively), but larger than those of *E. saccharolyticus*, *E. columbae*, *E. cecorum* and *E. sulfureus* (2.60, 2.58, 2.34 and 2.31, respectively). The G+C content of *E. massiliensis* is lower than those of *E. casseliflavus* and *E. gallinarum* (42.8 and 40.7) but greater than those of *E. moraviensis*, *E. haemoperoxidus*, *E. caccae*, *E. saccharolyticus*, *E. columbae*, *E. cecorum*, *E. sulfureus* and *E. faecalis* (39.6, 36.1, 35.7, 35.8, 36.9, 36.6, 36.4, 38.0 and 37.5, respectively). Of the 2687 predicted chromosomal genes, 2617 were protein-

coding genes and 70 were RNAs including 61 tRNAs and nine rRNAs (5S = 4, 23S = 2, 16S = 3). A total of 1889 genes (72.2%) were assigned to a putative function ([Fig. 3](), [Table 3]()). Seventy-one genes were identified as ORFans (2.71%), and the remaining genes were annotated as hypothetical proteins. The distribution of genes into COGs functional categories is presented in [Table 3]().

Conclusion and Perspectives

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Enterococcus massiliensis* sp. nov. AM1[T]. This strain was isolated in Marseille, France.

Taxonomic and Nomenclatural Proposals

Description of *Enterococcus massiliensis* sp. nov.

Enterococcus massiliensis (*massiliensis* because this strain was isolated in Massilia, the Latin name of Marseille, where the strain was sequenced).

Colonies were whitish and approximately 1 mm diameter on 5% sheep's blood−enriched agar. Cells are Gram-positive, non-haemolytic, facultative anaerobic with a mean length of 1.2 μm and a mean diameter of 0.6 μm. Growth occurred between 25°C to 37°C, but optimal growth was observed at 37°C. Alkaline phosphatase, esterase (C4), esterase lipase (C8), leucine arylamidase, acid phosphatase, β-galactosidase and N-acetyl-β-glucosaminidase activities were present. Esculin activity was also positive, but catalase, oxydase, β-galactosidase and N-acetyl-β-glucosaminidase were negative. Positive reaction were obtained for D-ribose, D-glucose, D-fructose, D-mannose and N-acetylglucosamine. *E. massiliensis* was susceptible to vancomycin, teicoplanin, linezolid, gentamicin, ciprofloxacin, doxycycline, rifampicin and pristinamycin, but resistant to trimethoprim/sulfamethoxazole, fosfomycin, erythromycin and clindamycin.

The G+C content of the genome is 39.6%. The 16S rRNA and genome sequences are deposited in GenBank under accession numbers LN833866 and [CVRN00000000](), respectively. The type strain AM1[T] (= CSUR P1927 = DSM 100308) was isolated from a fresh stool sample of a patient in Marseille, France.

Acknowledgements

Conflict of Interest

None declared.

References

[1] Lagier JC, Hugon P, Khelaifia S, Fournier PE, La Scola B, Raoult D. The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. Clin Microbiol Rev 2015;28:237−64.

[2] Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 2009;49:543−51.

[3] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152−5.

[4] Rosselo-Mora R. DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation. In: Stackebrandt E, editor. Molecular identification, systematics, and population structure of prokaryotes. Berlin: Springer; 2015. p. 23−50.

[5] Thiercelin E, Jouhaud L. Reproduction de l'enterocoque; taches centrales; granulations peripheriques et microblastes. Compt Rend Soc Biol 1903;55:686−8.

[6] Schleifer KH, Kilpper-Balz R. Transfer of *Streptococcus faecalis* and *Streptococcus faecium* to the genus *Enterococcus* nom. rev. as *Enterococcus faecalis* comb. nov. and *Enterococcus faecium* comb. nov. Int J Syst Bacteriol 1984;34:31−4.

[7] Bereket W, Hemalatha K, Getenet B, Wondwossen T, Solomon A, Zeynudin A, et al. Update on bacterial nosocomial infections. Eur Rev Med Pharmacol Sci 2012;16:1039−44.

[8] Miller WR, Munita JM, Arias CA. Mechanisms of antibiotic resistance in enterococci. Expert Rev Anti Infect Ther 2014;12:1221−36.

[9] Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 2012;18:1185−93.

[10] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955−64.

[11] Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 2007;35:3100−8.

[12] Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 2014;58:212−20.

[13] Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, et al. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res 2015;43:W237−43.

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

[15] Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001;305:567−80.

[16] Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence−based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 2016;21:14−60.

[17] Meier-Kolthoff JP, Klenk HP, Göker M. Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. Int J Syst Evol Microbiol 2014;64:352−6.

Article XVI:

Draft genome and description of *Negativicoccus massiliensis* strains Marseille-P2082[T], a new species isolated from the gut microbiota of an obese patient.

Amadou Hamidou Togo, Awa Diop, Matthieu Million, Saber Khelaifia, Noémie Labas, Frédéric Cadoret, Marie Maraninchi, Didier Raoult, Grégory Dubourg

1 **Draft genome and description of *Negativicoccus massiliensis* strains Marseille-P2082, a**

2 **new species isolated from the gut microbiota of an obese patient**

3 Amadou Hamidou Togo[1,2], Awa Diop[3], Matthieu Million[1,2], Saber Khelaifia[1,2], Marie

4 Maraninchi[4], Didier Raoult[1,2], Pierre-Edouard Fournier[2,3], Grégory Dubourg[1,2]*

5 1. Aix Marseille Univ, IRD198, AP-HM, MEPHI, Marseille, France.

6 2. IHU Méditerranée Infection, Marseille, France

7 3. Aix Marseille Univ, IRD, SSA, VITROME, IHU-Méditerranée Infection, Marseille,

8 France.

9 4. Aix Marseille Université, NORT "Nutrition, Obesity and Risk of Thrombosis",

10 INSERM1062, INRA1260, Marseille, France.

11

12 ***Corresponding author:** Gregory Dubourg, [1]Aix Marseille Univ, IRD198, APHM, MEPHI,

13 IHU-Méditerranée Infection, 19-21 Boulevard Jean Moulin, 13385, Marseille cedex 05,

14 France.

15 **Tel:** +33 413 732 401

16 **Fax:** +33 413 732 402

17 **email:** gregory.dubourg@univ-amu.fr

18 **Running Head:** *Negativicoccus massiliensis sp. nov.*

19 **Keywords:** *Culturomics*; *taxonomy*; *Negativicoccus massiliensis*; *human gut microbiota*;

20 *obesity*; *bariatric surgery*

Abstract

Strain Marseille-P2082, a new anaerobic, non-motile, non-spore forming, Gram-negative, coccoid bacterium was isolated from the feces of a 33 year-old obese French woman before bariatric surgery. The isolate exhibited 98.65% 16S rRNA gene nucleotide sequence similarity with *Negativicoccus succinicivorans* strain ADV 07/08/06-B-1388[T], its closest phylogenetic neighbour with standing in nomenclature. However, the dDDH relatedness between the new isolate and *N. succinicivorans* is 52.5 ± 2.7%. Strain Marseille-P2082 exhibited a genome of 1, 360, 589-bp-long with a 51.1% G+C content. Its major fatty acids were $C_{18:1n9}$, $C_{18:0}$ and $C_{16:0}$. Based on its phenotypic, genomic and phylogenetic characteristics, strain Marseille-P2082[T] (= CSURP2082 = DSM 100853) is proposed as type strain of the novel species *Negativicoccus massiliensis* sp. nov.

Introduction

Obesity is a major public health problem that increases the risk of several diseases such as metabolic diseases (type II diabetes) and cardiovascular diseases. According to Marie et al; the world rate of obesity has increased between 1980 and 2013 (Ng et al. 2014). Obesity results from an imbalance between inputs and energy expenditure. These process involves several factors such as genetic and environmental factors (Frayling et al. 2007; Cecil et al. 2008). More recently, obesity has been associated with an imbalance of the gut microbiota composition (Ley et al. 2006; Turnbaugh et al. 2006; Million et al. 2012).

Several methods of treating obesity have been proposed, mainly dietary and lifestyle measures, medical and surgical treatments. Bariatric surgery is one of the most effective treatments for obesity. It is also associated with an increase in microbial diversity (Zhang et al. 2009; Kong et al. 2013). For the purpose of assessing gut microbiota evolution from obese people and after bariatric surgery, we studied gut microbiota of the stool sample from obese subjects before and after bariatric surgery by culturomics (Lagier et al. 2015). During this study, we isolated a new anaerobic, Gram-negative coccus strain Maseille-P2082[T], from a stool sample of a 33 years-old French woman living in Marseille after bariatric surgery.

Here, we describe the phenotypic characteristics of this strain Maseille-P2082[T] together with the description of this complete genome sequencing and annotation.

Material and methods

Sample collection and isolation of strain Maseille-P2082[T]

The stool sample was collected in November 2011 from a 33 years-old obese French woman with a body mass index of 38.58 kg/m^2. Written and informed consent was obtained

57 from the patient at the Nutrition, Metabolic disease and Endocrinology service, in la Timone

58 Hospital, Marseille, France. The study and the assent procedure were approved by the local

59 ethics committee of IFR 48, under number 09 -022. The stool sample was stored at -80°C

60 after collection and studied by microbial culturomics approach.

61 The strain was isolated in anaerobic condition as described elsewhere (Togo et al. 2017).

62 **Phylogenetic analysis**

63 The 16S rRNA gene sequencing of the strain Maseille-P2082 was performed as

64 previously reported (Drancourt et al. 2000) using the fD1-rP2 primers, a GeneAmp PCR

65 System 2720 thermal cycler (Applied Bio systems, Bedford, MA, USA) and an ABI Prism

66 3130-XL capillary sequencer (Applied Biosciences, Saint Aubin, France).

67 For taxonomic assignment, CodonCode Aligner (101 Victoria Street Centerville, MA

68 02632, USA) was used to correct sequences and BLASTn searches was performed in NCBI

69 (National Centre for Biotechnology Information) web server at

70 http://blast.ncbi.nlm.nih.gov.gate1.inist.fr/Blast.cgi. Pairwise sequence similarities were

71 calculated using the method recommended by Meier-Kolthoff et al. (Meier-Kolthoff et al.

72 2013b) for the 16S rRNA gene available via the Genome to Genome Distance Calculator

73 (GGDC) web server (Meier-Kolthoff et al. 2013a, 2014) available at http://ggdc.dsmz.de/. A

74 multiple sequence alignment was created with MUSCLE (Edgar 2004). A Maximum

75 likelihood (ML) and a maximum parsimony (MP) trees were inferred from the alignment with

76 RAxML(Stamatakis 2014) and TNT (Goloboff et al. 2008), respectively. For ML, rapid

77 bootstrapping in conjunction with the autoMRE bootstopping criterion (Pattengale et al. 2010)

78 and subsequent search for the best tree was used; for MP, 1000 bootstrapping replicates were

79 used in conjunction with tree-bisection-and-reconnection branch swapping and ten random

80 sequence addition replicates. Reliability of the nodes was estimated as to the percentages of

81 bootstrap values obtained by repeating the analysis 1000 times to generate a consensus tree.

82   The sequences were checked for a compositional bias using the X^2 test as implemented in

83   PAUP* (Swofford 2002).

84   **Phenotypic and chemical characteristic analysis**

85   Colony morphology and pigmentation were observed after cultivation of the strain on

86   Columbia agar (bioMérieux, Marcy l'Etoile, France) at 37°C for 48 hours. Phenotypic

87   characteristics of the strain was performed as previously described (Togo et al. 2016a, 2017).

88   API strips; (API® ZYM, API® 20NE, and API®50 CH) were performed according to the

89   manufacturer's instruction (bioMérieux). The growth of strain Maseille-P2082 was attempted

90   at various pH values (6, 6.5, 7 and 8.5). The saltiness of the strain was also tested using

91   various NaCl concentrations (5, 10, 50, 75 and 100g/l) on schaedler agar enriched with 5%

92   Sheep Blood (bioMérieux) in anaerobic atmosphere at 37°C. Cellular fatty acid methyl ester

93   (FAME) analysis was performed by Gas Chromatography/Mass Spectrometry (GC/MS) as

94   described elsewhere (Togo et al. 2017).

95   **Genome sequencing and assembly**

96   Genomic DNA (gDNA) of strain Maseille-P2082 was sequenced using the MiSeq

97   Technology (Illumina Inc, San Diego, CA, USA) with the Mate Pair strategy and assembled

98   as previously described (Togo et al. 2016b).

99   **Genome annotation and comparison**

100   The genome annotation of strain Maseille-P2082 was performed as previously

101   described (Togo et al. 2017). For the comparison, genomes were automatically retrieved from

102   closely related species in the 16S rRNA tree, *Negativicoccus succinicivorans* strain ADV

103   07/08/06-B-1388[T] (Marchandin et al. 2010), *Veillonella tobetsuensis* strain JCM 17976

104   (Mashima et al. 2013), *Mitsuokella jalaludinii* strain M9 (Lan 2002), *Selenomonas artemidis*

105   strain ATCC 43528 (Moore et al. 1987), *Mitsuokella multacida* strain DSM 20544 (Mitsuoka

106   et al. 1974; Euzeby 1998) and *Selenomonas bovis* strain WG (Zhang and Dong 2009) are

107 selected. For each selected genome, complete genome sequence, proteome sequence and

108 Orfeome sequence were retrieved from the FTP of NCBI. All proteomes were analysed with

109 proteinOrtho (Lechner et al. 2011). Then for each couple of genomes, a similarity score was

110 computed. This score is the mean value of nucleotide similarity between all couple of

111 orthologous between the two genomes studied (Ramasamy et al. 2014). An annotation of the

112 entire proteome was performed to define the distribution of functional classes of predicted

113 genes according to the clusters of orthologous groups of proteins. The genome of strain

114 Maseille-P2082 was locally aligned 2-by-2 using BLSAT algorithm (Kent 2002; Auch et al.

115 2010) against each of the selected previously cited genomes and dDDH values were estimated

116 from a generalized model (Meier-Kolthoff et al. 2013a).

117 **Results and discussion**

118 **MALDI-TOF analysis**

119   The reference spectrum generated from isolated colonies was unable to match with

120 those of the Bruker database (Fig. 1). A gel view was performed in order to detect the spectral

121 differences of strain Maseille-P2082 and closet species (Fig. 2).

122 **Phylogenetic analysis**

123   The 16S rRNA phylogenetic analysis, showed that strain Maseille-P2082^T exhibited a

124 98. 65% sequence similarity with *N. succinicivorans* (Marchandin et al. 2010) classified in the

125 family *Veillonellaceae* created in 1971 (Rogosa 1971). Maximum likelihood phylogenetic tree

126 (Fig. 3) based on 16S rRNA gene sequences, showed that the strain Maseille-P2082 was most

127 closely related to *N. succinicivorans*. The 16S rRNA sequence of this strain was deposited in

128 EMBL-EBI under accession number LN881615.

129 **Phenotypic, biochemical and chemotaxonomic characterization**

130   This new isolate is a Gram-stain negative, non-motile, coccoid bacterium. The strain

131 has no catalase nor oxidase activity. Growth occurred between 28 and 55°C, but optimal

132   growth was observed at 37°C after 72 hours of incubation in anaerobic atmosphere on 5%

133   sheep blood Columbia agar (bioMérieux). No growth of the strain was observed in aerobic

134   atmosphere. No growth of this bacterium was observed using NaCl concentration of 10g/l or

135   more on 5% sheep blood schaedler agar (bioMérieux). Strain grew at pH value ranging from 6

136   to 8.5 but the optimal pH was 7.5. Colonies were circular, translucent and very small, with a

137   mean diameter less than 0.5 mm. No haemolysis was seen on Columbia agar (bioMérieux).

138   Cells had a diameter ranging from 0.2 to 0.5µm by electron microscopy (Fig. 4).

139   Using API 50 CH strip, positive fermentation reactions were observed with aesculin, D-

140   galactose, D-glucose, D-fructose, D-lactose, D-maltose, D-mannose, D-melezitose, D-ribose,

141   D-saccharose, D-tagatose, D-turanose, glycerol, N-acethylglucosamine, potassium 5-

142   cetogluconate and starch. Negative reactions were observed for adonitol, amygdaline,

143   arbutine, D-arabinose, D-arabitol, D-cellobiose, D-fucose, D-lyxose, D-mannitol, D-

144   melibiose, D-raffinose, D-sorbitol, D-trehalose, D-xylose, dulcitol, erythritol, gentiobiose,

145   glycogen, inositol, inulin, L-arabinose, L-arabitol, L-fucose, L-rhamnose, L-Sorbose, L-

146   xylose, methyl-αD-glucopyranoside, methyl-α-D-mannopyranoside, Methyl-βD-

147   Xylopyranoside, potassium gluconate and potassium 2-ceto gluconate, salicin, and xylitol.

148   Using API® Zym strip, positive reactions were observed for esterase, esterase lipase, valine

149   arylamidase, acid phosphatase, and naphtol-AS-BI-phosphohydrolase, but alkaline

150   phosphatase, lipase, leucine arylamidase, cystine arylamidase, trypsin, α-chymotrypsin, α-

151   galactosidase, β-galactosidase, β-glucuronidase, α-glucosidase, β-glucosidase, N-acetyl-β-

152   glucosaminidase, α-mannosidase and α-fructosidase activities were negative . Using API®

153   20NE strip, nitrate was not reduced to nitrite, indole was not produced, aesculin was not

154   hydrolysed, gelatine was not liquefied and 4-nitrophenyl-β-D-galactopyranoside was not

155   utilised while urease and arginine hydrolase activities were positive. D-glucose, L-arabinose,

156   D-mannose, D-mannitol, N-acetyl-glucosamine-, D-maltose, potassium gluconate, capric

157 acid, adipic acid, malic acid, trisodium citrate and phenyl acetic acid were negative. The strain

158 Maseille-P2082 differed from *N. succinicivorans* in its ability to ferment 16 sugars whereas *N.*

159 *succinicivorans* is assaccharolytic. The differential characteristics of strain Maseille-P2082[T]

160 compared with its closet neighbours are detailed in Table 1.

161   The most abundant cellular fatty analysis were $C_{18:1n9}$ (37%), $C_{18:0}$ (35%), $C_{16:0}$

162 (21.3%), $C_{18:2n6}$ (6.3%) and the weak amount of $C_{14:0}$ was less than 1%. The composition of

163 cellular fatty acid for this isolate is detailed in Table 2.

164 **Genome properties**

165   The draft genome from strain Maseille-P2082 was deposited in EMBL-EBI under

166 accession number LT700188. It is 1, 360, 589 bp long with a 51.1% G+C content (Fig. 5). It

167 is composed of 1 scaffold (composed of 1 contig). Of the 1,358 predicted genes; 1, 294 were

168 protein-coding genes and 64 were RNAs genes (4 rRNA (5S, 16S 23S, 52 tRNA genes). A

169 total of 995 genes (76.9 %) were assigned a putative function (by COGs or by NR blast), 23

170 genes were identified as ORFans (1. 8%), and 5 genes were associated with polyketide

171 synthase (PKS) or non-ribosomal peptide synthetase (NRPS) (Conway and Boddy 2013). The

172 remaining 967 genes (17.4%) were annotated as hypothetical proteins. Using ARG-ANNOT

173 (Gupta et al. 2014), no resistance gene was found. The distribution of genes into COGs

174 functional categories is presented in Table 3.

175 **Genome comparison**

176   The draft genome (1,361Mb) of the strain Maseille-P2082 is the smallest compared

177 genome (*N. succinicivorans*, *V. tobetsuensis*, *S. artemidis M. jalaludinii*, *M. multacida*, and *S.*

178 *bovis* (1.51, 2.16, 2.28, 2.41, 2.58 and 2.68 Mb, respectively). G+C content (51.1%) is greater

179 than those of *N. succinicivorans* and *V. tobetsuensis* (48.3 and 38.1 % respectively) but

180 smaller than those of *S. artemidis*, *S. bovis*, *M. jalaludinii* and *M. multacida* (57.3, 59.2, 58.1

181 and 58.1%, respectively. The gene content (1, 294) of strain Maseille-P2082[T] was the

182   smallest of those included in this comparison as the gene content of *N. succinicivorans*, *S.*

183   *artemidis*, *S. bovis*, *V. tobetsuensis*, *M. jalaludinii* and *M. multacida* was 1,679, 2,195, 2,510,

184   1,960, 2,183 and 2,558, respectively. However, the gene distribution into COGs categories

185   was similar among all compared genomes (Table 3 Fig. 6). The dDDH values ranged from

186   23% with *V. tobetsuensis* strain JCM 17976 to 52.5 % with *N. succinicivorans* strain ADV (

187   Table 4). The average amino acid identity (AAI) between the strain Maseille-P2082[T] and

188   closely related species was evaluated (Table 5). These value ranged from 47.18% with *S.*

189   *artemidis* to 94.96 with *N. succinicivorans*. The average nucleotide identity (ANI) between

190   strain Maseille-P2082 and *N. succinicivorans* was 93.5%, this was lower than the threshold

191   value (95-96%) recommended to delineate species (Konstantinidis and Tiedje 2005;

192   Rodriguez-R and Konstantinidis 2014).

Taxonomic and nomenclatural proposals

Description of *Negativicoccus massiliensis* sp. nov strain Maseille-P2082[T]

Negativicoccus massiliensis (*mas.si.li.en'sis*. L. fem. adj. *massiliensis*, of Massilia, the Latin name of Marseille, where *N. massiliensis* was first isolated). Cells are Gram-negative, non-motile non-spore forming and coccoid; with a mean diameter of 0.5µm. Columbia agar-grown colonies are circular, flat, translucent with a mean diameter smaller than 0.5 mm after 72 hour of incubation in anaerobic condition. Catalase and oxidase activities are absent. The optimum growth temperature is 37°C. Urease and arginine hydrolase activities are positive. The strain was not able to reduce nitrate to nitrite, indole was not produced, gelatin is not liquefied and aesculin was not hydrolysed. The length of the genome is 1,360589 bp with a 51.1% of G+C content. The 16S rRNA gene sequence and whole-genome shotgun sequence are deposited in EMBL-EBI under accession numbers LN876651 and LT700188, respectively.

The type strain Maseille-P2082[T] (=CSUR P2082, DSM 100853) was isolated from the stool sample of a French morbidly obese woman. The habitat of this microorganism is the human gut.

Conflict of interest

218   The authors declare no conflict of interest.

219   **References**

220   Auch AF, von Jan M, Klenk H-P, Göker M (2010) Digital DNA-DNA hybridization for

221   microbial species delineation by means of genome-to-genome sequence comparison.

222   Stand Genomic Sci 2:117–134. doi: 10.4056/sigs.531120

223   Cecil JE, Tavendale R, Watt P, et al (2008) An obesity-associated FTO gene variant and

224   increased energy intake in children. N Engl J Med 359:2558–2566. doi:

225   10.1056/NEJMoa0803839

226   Conway KR, Boddy CN (2013) ClusterMine360: a database of microbial PKS/NRPS

227   biosynthesis. Nucleic Acids Res 41:D402-407

228   Drancourt M, Bollet C, Carlioz A, et al (2000) 16S ribosomal DNA sequence analysis of a

229   large collection of environmental and clinical unidentifiable bacterial isolates. J Clin

230   Microbiol 38:3623–3630

231   Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high

232   throughput. Nucleic Acids Res 32:1792–1797. doi: 10.1093/nar/gkh340

233   Euzeby JP (1998) Taxonomic note: necessary correction of specific and subspecific epithets

234   according to Rules 12c and 13b of the International Code of Nomenclature of Bacteria

235   (1990 Revision). Int J Syst Bacteriol 48:1073–1075. doi: 10.1099/00207713-48-3-

236   1073

237   Frayling TM, Timpson NJ, Weedon MN, et al (2007) A common variant in the FTO gene is

238   associated with body mass index and predisposes to childhood and adult obesity.

239   Science 316:889–894. doi: 10.1126/science.1141634

240   Goloboff PA, Farris JS, Nixon KC (2008) TNT, a free program for phylogenetic analysis.

241   Cladistics 24:774–786. doi: 10.1111/j.1096-0031.2008.00217.x

242   Gupta SK, Padmanabhan BR, Diene SM, et al (2014) ARG-ANNOT, a new bioinformatic

243   tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents

244   Chemother 58:212–220

245   Kent WJ (2002) BLAT--the BLAST-like alignment tool. Genome Res 12:656–664

246   Kong L-C, Tap J, Aron-Wisnewsky J, et al (2013) Gut microbiota after gastric bypass in

247   human obesity: increased richness and associations of bacterial genera with adipose

248   tissue genes. Am J Clin Nutr 98:16–24. doi: 10.3945/ajcn.113.058743

249   Konstantinidis KT, Tiedje JM (2005) Towards a Genome-Based Taxonomy for Prokaryotes. J

250   Bacteriol 187:6258–6264. doi: 10.1128/JB.187.18.6258-6264.2005

251   Lagier J-C, Hugon P, Khelaifia S, et al (2015) The rebirth of culture in microbiology through

252   the example of culturomics to study human gut microbiota. Clin Microbiol Rev

253   28:237–264. doi: 10.1128/CMR.00014-14

254   Lan GQ (2002) Mitsuokella jalaludinii sp. nov., from the rumens of cattle in Malaysia. Int J

255   Syst Evol Microbiol 52:713–718. doi: 10.1099/ijs.0.01930-0

256   Lechner M, Findeiss S, Steiner L, et al (2011) Proteinortho: detection of (co-)orthologs in

257   large-scale analysis. BMC Bioinformatics 12:124. doi: 10.1186/1471-2105-12-124

258   Ley RE, Turnbaugh PJ, Klein S, Gordon JI (2006) Microbial ecology: human gut microbes

259   associated with obesity. Nature 444:1022–1023. doi: 10.1038/4441022a

260   Marchandin H, Teyssier C, Campos J, et al (2010) Negativicoccus succinicivorans gen. nov.,

261   sp. nov., isolated from human clinical samples, emended description of the family

262   Veillonellaceae and description of Negativicutes classis nov., Selenomonadales ord.

263   nov. and Acidaminococcaceae fam. nov. in the bacterial phylum Firmicutes. Int J Syst

264   Evol Microbiol 60:1271–1279

265   Mashima I, Kamaguchi A, Miyakawa H, Nakazawa F (2013) Veillonella tobetsuensis sp.

266   nov., an anaerobic, gram-negative coccus isolated from human tongue biofilms. Int J

267   Syst Evol Microbiol 63:1443–1449. doi: 10.1099/ijs.0.042515-0

268   Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M (2013a) Genome sequence-based species

269   delimitation with confidence intervals and improved distance functions. BMC

270   Bioinformatics 14:60. doi: 10.1186/1471-2105-14-60

271   Meier-Kolthoff JP, Göker M, Spröer C, Klenk H-P (2013b) When should a DDH experiment

272   be mandatory in microbial taxonomy? Arch Microbiol 195:413–418. doi:

273   10.1007/s00203-013-0888-4

274   Meier-Kolthoff JP, Klenk H-P, Göker M (2014) Taxonomic use of DNA G+C content and

275   DNA-DNA hybridization in the genomic age. Int J Syst Evol Microbiol 64:352–356.

276   doi: 10.1099/ijs.0.056994-0

277   Million M, Maraninchi M, Henry M, et al (2012) Obesity-associated gut microbiota is

278   enriched in Lactobacillus reuteri and depleted in Bifidobacterium animalis and

279   Methanobrevibacter smithii. Int J Obes 36:817–825. doi: 10.1038/ijo.2011.153

280   Mitsuoka T, Terada A, Watanabe K, Uchida K (1974) Bacteroides multiacidus, a New

281   Species from the Feces of Humans and Pigs. Int J Syst Bacteriol 24:35–41. doi:

282   10.1099/00207713-24-1-35

283   Moore LVH, Johnson JL, Moore WEC (1987) Selenomonas noxia sp. nov., Selenomonas

284   flueggei sp. nov., Selenomonas infelix sp. nov., Selenomonas dianae sp. nov., and

285   Selenomonas artemidis sp. nov., from the Human Gingival Crevice. Int J Syst

286   Bacteriol 37:271–280. doi: 10.1099/00207713-37-3-271

287   Ng M, Fleming T, Robinson M, et al (2014) Global, regional, and national prevalence of

288   overweight and obesity in children and adults during 1980–2013: a systematic analysis

289   for the Global Burden of Disease Study 2013. The Lancet 384:766–781. doi:

290   10.1016/S0140-6736(14)60460-8

291   Pattengale ND, Alipour M, Bininda-Emonds ORP, et al (2010) How many bootstrap

292   replicates are necessary? J Comput Biol J Comput Mol Cell Biol 17:337–354. doi:

293   10.1089/cmb.2009.0179

294   Ramasamy D, Mishra AK, Lagier J-C, et al (2014) A polyphasic strategy incorporating

295   genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol

296   Microbiol 64:384–391

297   Rodriguez-R LM, Konstantinidis KT (2014) Bypassing cultivation to identify bacterial

298   species: Culture-independent genomic approaches identify credibly distinct clusters,

299   avoid cultivation bias, and provide true insights into microbial species. Microbe Mag

300   9:111–118. doi: 10.1128/microbe.9.111.1

301   Rogosa M (1971) Transfer of Veillonella Prevot and Acidaminococcus Rogosa from

302   Neisseriaceae to Veillonellaceae fam. nov., and the Inclusion of Megasphaera Rogosa

303   in Veillonellaceae. Int J Syst Bacteriol 21:231–233. doi: 10.1099/00207713-21-3-231

304   Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of

305   large phylogenies. Bioinforma Oxf Engl 30:1312–1313. doi:

306   10.1093/bioinformatics/btu033

307   Swofford D (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).

308   Version 4.0b10. Sinauer Associates, Sunderland.

309   Togo AH, Diop A, Dubourg G, et al (2016a) Butyricimonas phoceensis sp. nov., a new

310   anaerobic species isolated from the human gut microbiota of a French morbidly obese

311   patient. New Microbes New Infect 14:38–48. doi: 10.1016/j.nmni.2016.07.010

312   Togo AH, Durand G, Khelaifia S, et al (2017) Fournierella massiliensis, gen. nov., sp. nov., a

313   new human-associated member of the family Ruminococcaceae. Int J Syst Evol

314   Microbiol 67:1393–1399. doi: 10.1099/ijsem.0.001826

315   Togo AH, Khelaifia S, Lagier J-C, et al (2016b) Noncontiguous finished genome sequence

316   and description of Paenibacillus ihumii sp. nov. strain AT5. New Microbes New Infect

317   10:142–150. doi: 10.1016/j.nmni.2016.01.013

318   Turnbaugh PJ, Ley RE, Mahowald MA, et al (2006) An obesity-associated gut microbiome

319   with increased capacity for energy harvest. Nature 444:1027–1031. doi:

320   10.1038/nature05414

321   Zhang H, DiBaise JK, Zuccolo A, et al (2009) Human gut microbiota in obesity and after

322   gastric bypass. Proc Natl Acad Sci U S A 106:2365–2370. doi:

323   10.1073/pnas.0812600106

324   Zhang K, Dong X (2009) Selenomonas bovis sp. nov., isolated from yak rumen contents. Int J

325   Syst Evol Microbiol 59:2080–2083. doi: 10.1099/ijs.0.007641-0

326

327 **Table 1:** Differential characteristics of strain Maseille-P2082 (1) compared with its closest species *N. succinicivorans* strain ADV (Data from

328 Marchandin et *al.*, 2010) (2); *V. tobetsuensis* strain JCM 17976 (Data from Mashima et *al.*, 2013) (3); *M. jalaludinii* strain M9 (Data from Lan et

329 *al.*, 2002) (4); *S. artemidis* strain ATCC 43528 (Data from Moore et *al.*, 1987) (5); *M. multacida* strain DSM 20544 (Data from Mitsuoka et *al.*,

330 1974) (6); *S. bovis* strain WG (Data from Zhang and Dong, 2009) (7)

Properties	1	2	3	4	5	6	7
Motility	–	–	–	–	+	–	+
Spore formation	–	–	–	–	–	–	–
Indole production	–	–	–	–	NR	–	–
Nitrate reductase	–	–	+	+	+	+	–
Acid production from:							
Aesculin	+	–	–	+	–	+	+
Arabinose	–	–	–	v	–	+	+
Cellobiose	–	–	–	+	–	+	+
Fructose	+	–	–	+	+	+	–
Galactose	+	–	–	+	–	+	–
Glucose	+	–	–	+	+	+	+

Glycerol	+	–	–	+	–	–	–
Inositol	–	–	–	+	–	+	–
Lactose	+	–	–	+	–	+	+
Maltose	+	–	–	+	+	+	–
Mannitol	+	–	–	–	+	+	–
Mannose	+	–	–	+	–	+	+
Melibiose	–	–	–	+	–	+	+
Raffinose	–	–	–	–	w	+	+
Rhamnose	–	–	–	–	–	+	–
Ribose	–	–	–	+	–	+	–
Salicin	–	–	–	+	–	+	+
Sorbitol	–	–	–	+	–	–	–
Sucrose	+	–	–	+	+	+	+
Trehalose	+	–	–	+	–	+	+
Xylose	–	–	–	v	–	+	–
Xylitol	–	–	–	+	–	+	–

Isolation source	Human feces	human clinical samples	Human tongue biofilm	Cattle rumen	Human gingival crevice	Human and Pig feces	Yak rumen
G+C content (mol %)	51.1	48.3	38.1	58.1	57.3	58.1	59.2

331 + = Positive; – = Negative; w = weakly positive; v = Variable; NR = Data Not Reported, in bold difference between strain Marseille-P2082 and

332 *N. succinicivorans* strain ADV

333 **Table 2:** Cellular fatty acid composition of strain Maseille-P2082

Fatty acids	Mean relative %[a]
$C_{18:1n9}$	37.2 ± 11.9
$C_{18:0}$	34.7 ± 14.8
$C_{16:0}$	21.3 ± 0.6
$C_{18:2\,n6}$	6.3 ± 2.8
$C_{14:0}$	<1

334 [a]Mean peak area percentage ± standard deviation.

Table 3: Number of genes associated with the 26 general COG functional categories of strain Maseille-P2082T

1) Strain Maseille-P2082; 2) *N. succinicivorans* ADV 07/08/06-B-1388[T]; 3) *V. tobetsuensis* strain JCM 17976; 4) *M. jalaludinii* strain M9; 5) *S. artemidis* strain ATCC 43528; 6) *M. multacida* strain DSM 20544; 7) *S. bovis* strain WG.

| Code	1	2	4	6	5	7	3	Description
J	163	169	178	177	170	181	178	Translation
A	0	0	0	0	0	0	0	RNA processing and modification
K	45	55	98	104	71	102	54	Transcription
L	64	76	70	72	68	89	68	Replication, recombination and repair
B	0	0	0	0	0	0	1	Chromatin structure and dynamics
D	27	31	31	33	31	34	28	Cell cycle control, mitosis and meiosis
Y	0	0	0	0	0	0	0	Nuclear structure
V	27	31	44	41	42	48	36	Defence mechanisms
T	31	40	54	58	62	80	32	Signal transduction mechanisms
M	70	82	105	110	100	120	99	Cell wall/membrane biogenesis
N	10	12	11	10	49	54	10	Cell motility
Z	0	0	0	0	0	0	0	Cytoskeleton

W	4	5	4	8	3	12	4	Extracellular structures
U	22	22	32	31	33	39	32	Intracellular trafficking and secretion
O	44	51	73	81	75	74	53	Posttranslational modification, protein turnover, chaperones
X	1	6	49	29	23	33	20	Mobilome: prophages, transposons
C	45	58	114	121	99	107	104	Energy production and conversion
G	36	47	171	182	107	135	64	Carbohydrate transport and metabolism
E	85	113	168	184	162	180	163	Amino acid transport and metabolism
F	52	62	73	79	53	67	57	Nucleotide transport and metabolism
H	78	88	116	116	122	121	116	Coenzyme transport and metabolism
I	37	39	43	45	44	50	34	Lipid transport and metabolism
P	62	71	86	85	116	71	103	Inorganic ion transport and metabolism
Q	16	19	20	23	17	26	28	Secondary metabolites biosynthesis, transport and catabolism
R	85	96	135	152	127	135	112	General function prediction only
S	51	61	71	77	68	74	81	Function unknown

338

Tableau 4: Pairwise comparison of strain Maseille-P2082 with other species using GGDC, formula 2 (DDH estimates based on identities / HSP length)*

1) Strain Maseille-P2082; 2) *N. succinicivorans* ADV 07/08/06-B-1388[T]; 3) *V. tobetsuensis* strain JCM 17976; 4) *M. jalaludinii* strain M9; 5) *S. artemidis* strain ATCC 43528; 6) *M. multacida* strain DSM 20544; 7) *S. bovis* strain WG

	1	2	3	4	5	6	7
1	100%	52.5% ± 2.7	23.0% ± 2.35	33.4% ± 2.45	32.8% ± 2.45	28.2% ± 2.45	23.3% ± 2.35
2		100%	24.2% ± 2.35	33.1% ± 2.45	17.5% ± 2.2	28.1% ± 2.45	24.8% ± 2.4
3			100%	32.1% ± 2.45	33.5% ± 2.25	30.9% ± 2.45	33.9% ± 2.45
4				100%	18.5% ± 2.3	36.6% ± 2.45	22.2% ± 2.35
5					100%	18.3% ± 2.25	19.1% ± 2.3
6						100%	22.3% ± 2.4
7							100%

Tableau 5: The average amino acid identity values of strain Maseille-P2082 compared with those of its phylogenetically close neighbours

1) Strain Maseille-P2082; 2) *N. succinicivorans* ADV 07/08/06-B-1388[T]; 3) *V. tobetsuensis* strain JCM 17976; 4) *M. jalaludinii* strain M9; 5) *S. artemidis* strain ATCC 43528; 6) *M. multacida* strain DSM 20544; 7) *S. bovis* strain WG

	1	2	3	4	5	6	7
Strain Maseille-P2082	100%	94.96%	50.88%	47.46%	47.18%	47.48%	47.36%
N. succinicivorans		100%	51.19%	47.13%	46.94%	47.25%	46.84%
V. tobetsuensis			100%	47.49%	46.92%	47.35%	46.88%
M. jalaludinii				100%	61.41%	91.06%	69.61%
S. artemidis					100%	61.76%	60.73%
M. multacida						100%	69.88%
S. bovis							100%
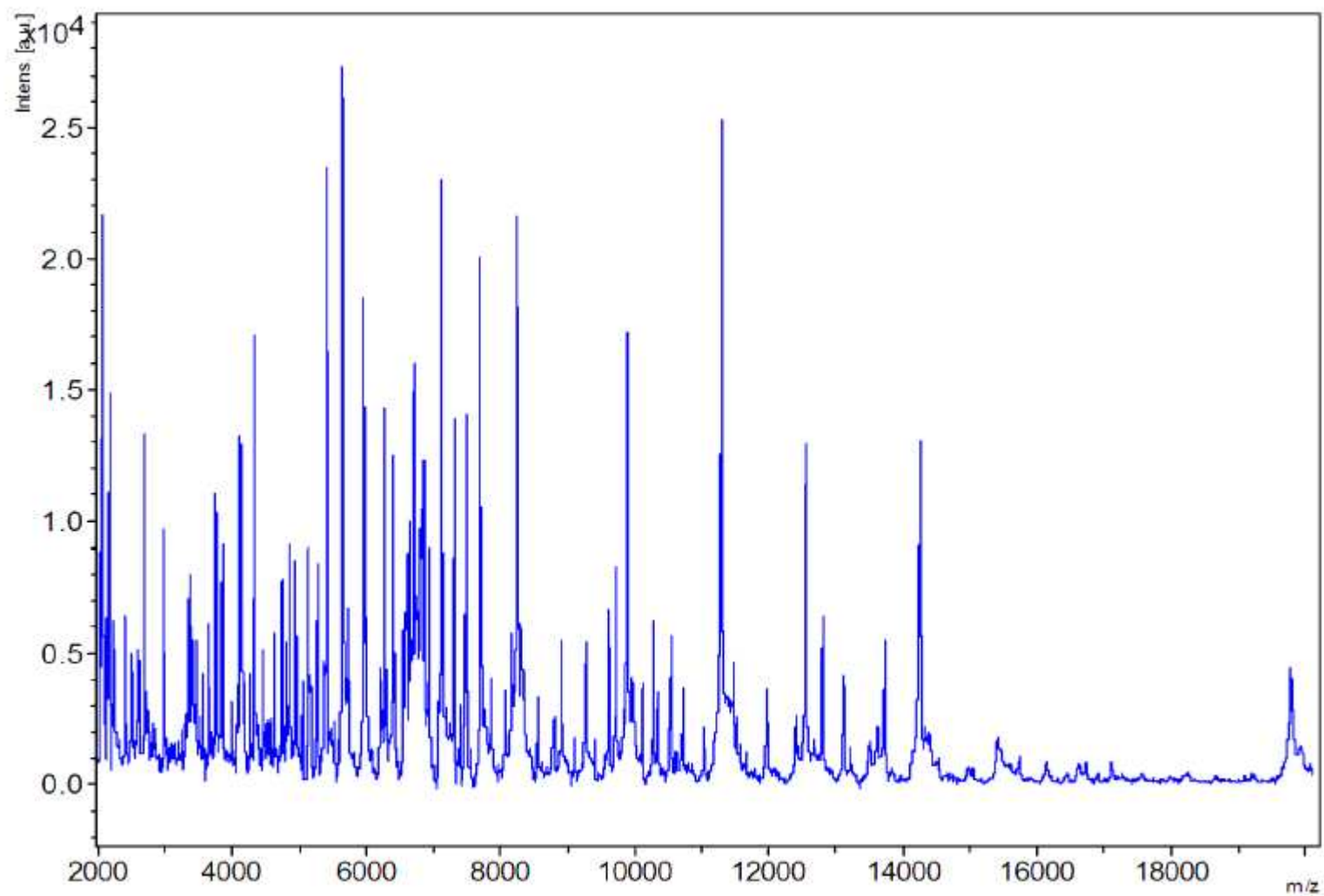
348 **Figure 1: Reference mass spectrum from strain Maseille-P2082^T:** Spectra from 12 individual colonies were compared and a reference

231

349   spectrum was generated.
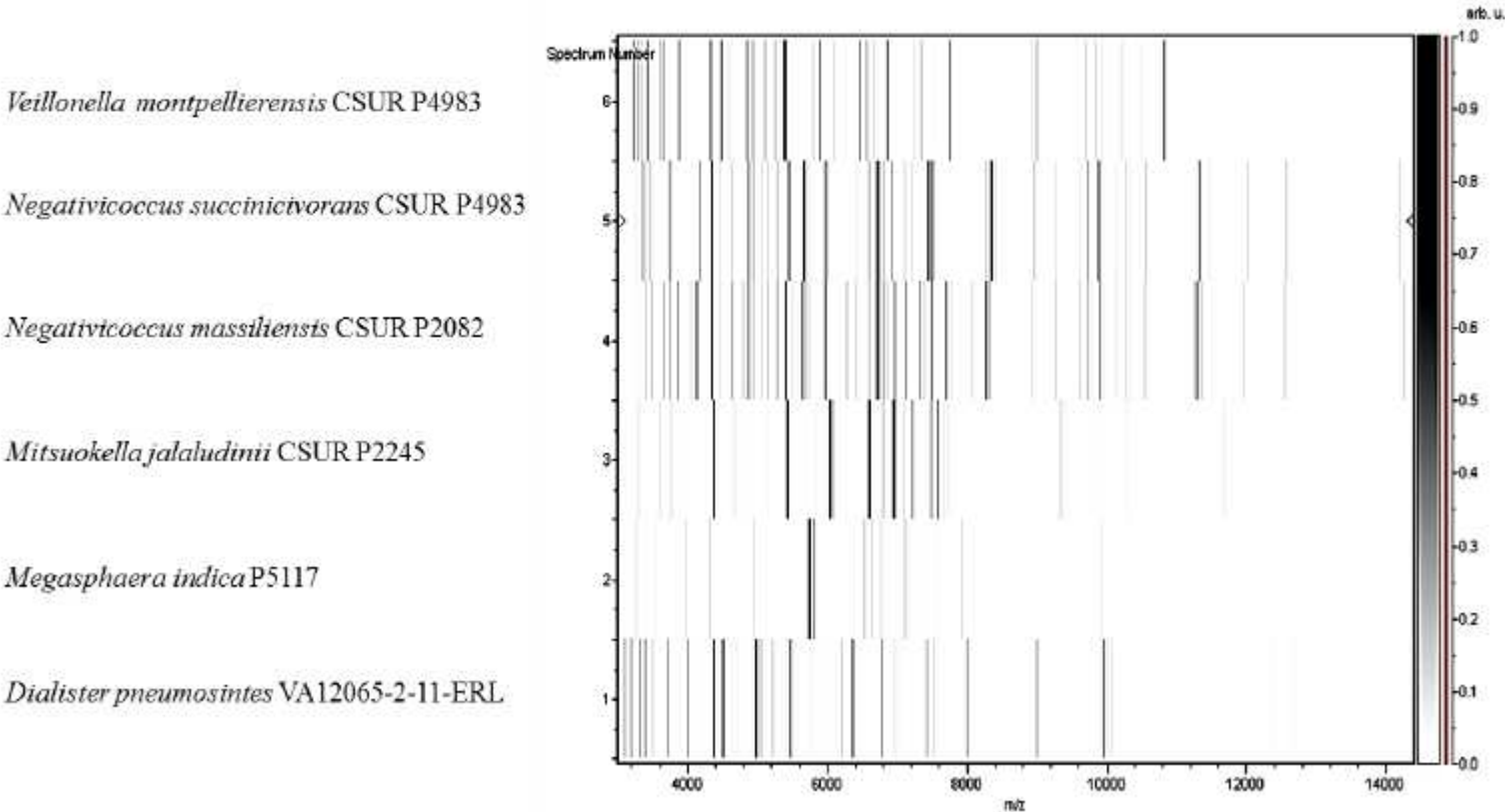
350   **Figure 2: Gel view comparing strain Maseille-P2082^T to other close species**

351   The gel view displays the raw spectra of loaded spectrum files arranged in a pseudo-gel like look. The x-axis records the m/z value. The left y-

352   axis displays the running spectrum number originating from subsequent spectra loading. The peak intensity is expressed by a Grey scale scheme

353   code. The colour bar and the right y-axis indicate the relation between the colour a peak is displayed with and the peak intensity in arbitrary units.

354   Displayed species are indicated on the left.
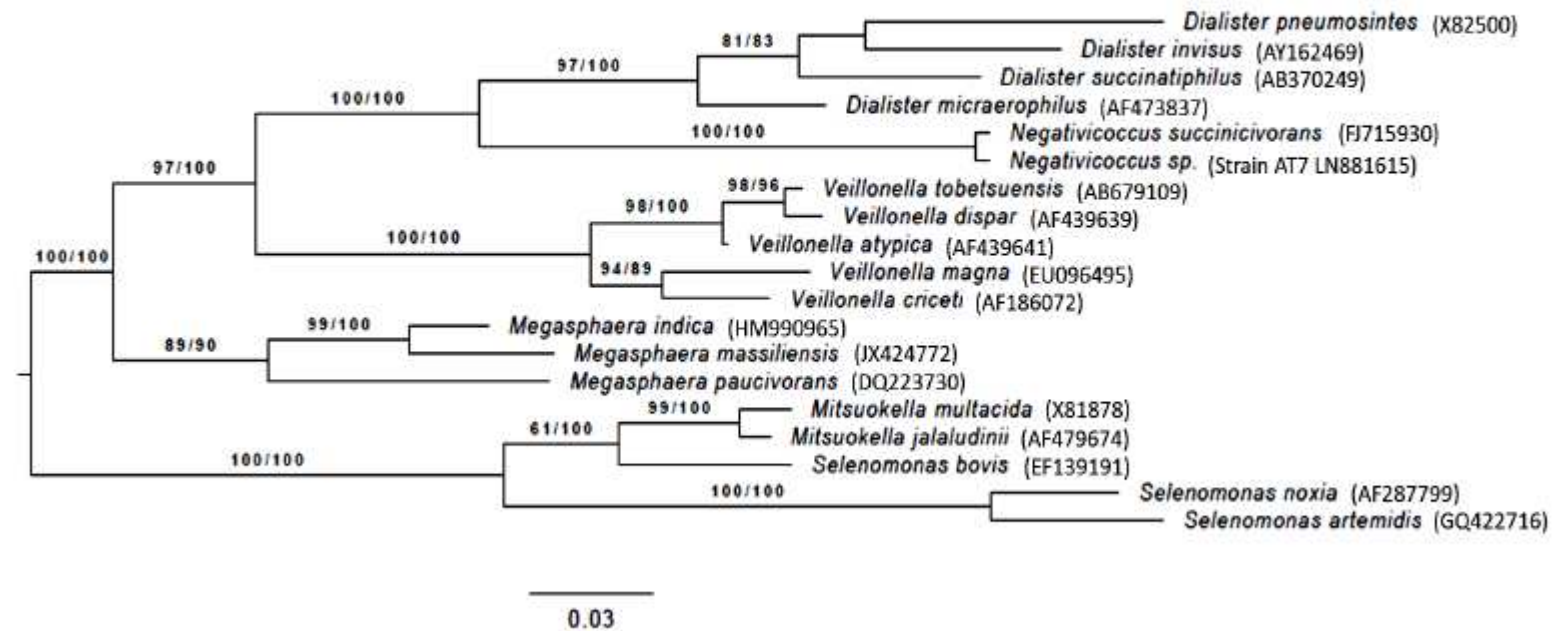
355   **Figure 3: Maximum Likelihood method Phylogenetic tree based on 16S rRNA sequence comparison highlighting the position of strain**

356   **Maseille-P2082^T against other closely related species.**

357   The corresponding GenBank accession numbers for 16S rRNA genes of each strain are indicated in bracket. Sequences were aligned using

358   ClustalW with default parameters and Phylogenies were inferred by the GGDC web server available at http://ggdc.dsmz.de/ using the DSMZ

359   phylogenomics pipeline. The scale bar represents a 3% nucleotide sequence divergence.
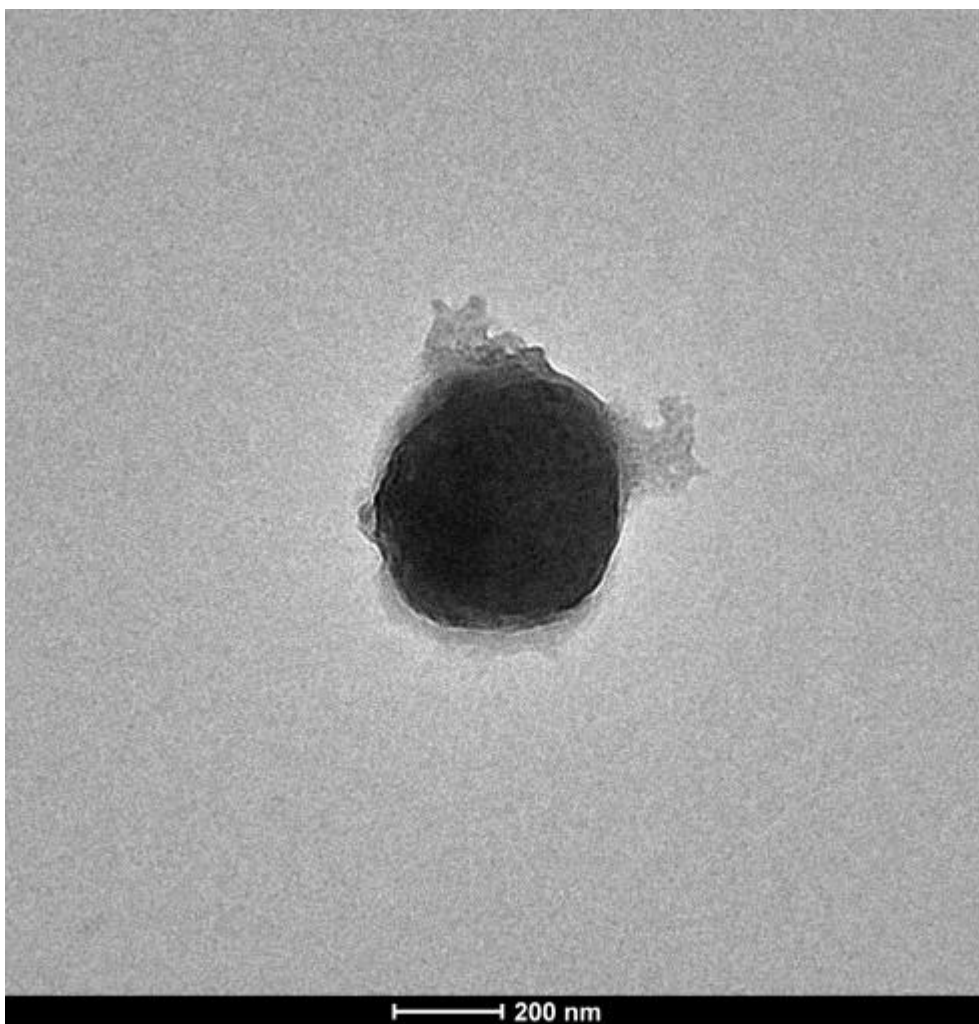
360 **Figure 4: Transmission electron microscopy of strain Maseille-P2082^T** using a Tecnai G20 (FEI company) at an operating voltage of 60kV.

361 The scale bar represents 500 µm
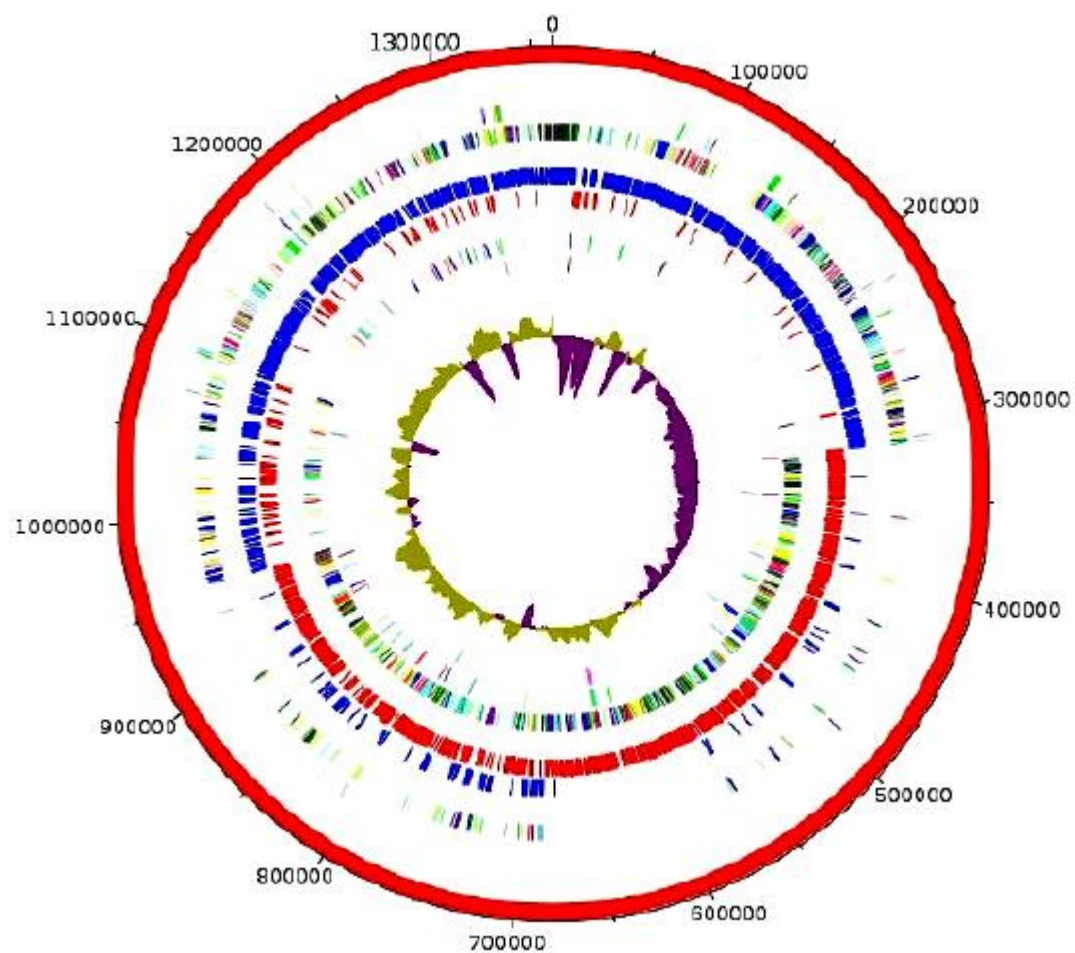
Figure 5: Graphical circular map of the genome of strain Maseille-P2082^T

From outside to the centre: Contigs (red/grey), COG category of genes on the forward strand (three circles), genes on forward strand (blue circle), genes on the reverse strand (red circle), COG category on the reverse strand (three circles), GC content.

365
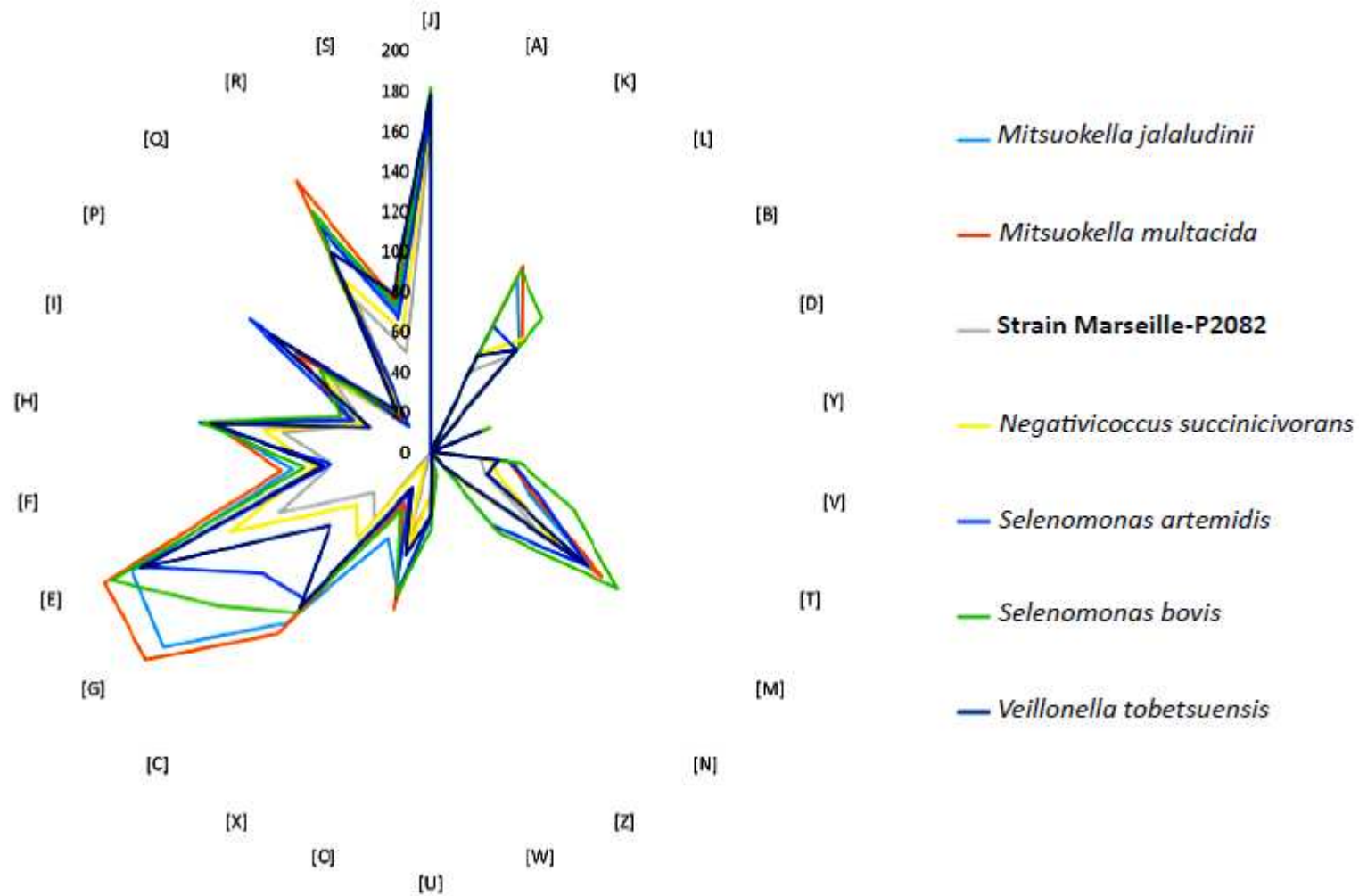
Figure 6: Distribution of functional classes of predicted genes according to the COGs of strain Maseille-P2082^T with its closest species

367 1) Strain Maseille-P2082; 2) *N. succinicivorans* strain ADV; 3) *V. tobetsuensis* strain JCM 17976; 4) *M. jalaludinii* strain M9; 5) *S. artemidis*

368 strain ATCC 43528; 6) *M. multacida* strain DSM 20544; 7) *S. bovis* strain WG

Conclusion et perspectives:

Nous pouvons conclure que le colostrum et le lait maternel abritent un microbiote divers et varié. Ce microbiote est constitué en grande partie de bactéries aérobies, mais également des bactéries anaérobies majeures du tube digestif comme *Faecalibacterium prausnitzii*, *Akkermansia muciniphila* et *Methanobrevibacter smithii*. *Faecalibacterium* et *Akkermansia* qui n'ont pas encore été isolées en culture du microbiote du lait mais détectées par les méthodes moléculaires jouent un rôle important dans la santé humaine, tout comme *Methanobrevibacter smithii*. Le microbiote du colostrum et du lait maternel contient également des bactéries pathogènes comme *Listeria monocytogenes*. *Listeria monocytogenes* est un pathogène responsable des infections graves particulièrement chez les nouveau-nés et les femmes enceintes. Nous lançons un appel aux décideurs sanitaires mondiaux de déployer les ressources nécessaires pour explorer une possible hyperendémicité africaine de *Listeria monocytogenes* et étudier la transmission de cet agent pathogène à travers le lait maternel. La contamination par l'allaitement maternel peut contribuer à la malnutrition aiguë sévère et à la listériose néonatale.

L'approche de «microbial culturomics» nous a permis d'isoler pour la première fois des archées méthanogènes (*Methanobrevibacter smithii* et *Methanobrevibacter oralis*) mais aussi de nouvelles espèces bactériennes et des bactéries pathogènes dans le colostrum et le lait maternel augmentant le répertoire bactériologique associé à l'Homme. Il nous a également permis de clarifier la taxonomie par la reclassification des espèces bactériennes mal classées. La reclassification des espèces mal classées est très importante. Une espèce bactérienne peut avoir les caractéristiques phénotypiques similaires à une autre n'étant ni de la même famille ou ni même du même phylum. C'était le cas de *Faecalibacterium prausnitzii* qui était classé *Fusobacterium prausnitzii*. Actuellement, *Faecalibacterium* constitue l'une des bactéries les plus abondantes du microbiote intestinal humain et joue un rôle important dans la physiologie de l'écosystème du côlon humain alors que les *Fusobacteria* sont des bactéries pathogènes (45,46).

Ainsi, nous proposons en perspective une extension de l'exploration du microbiote du colostrum et du lait maternel à d'autres microorganismes associés à l'homme tels que les archées, les virus, les parasites et champignons. Augmenter le nombre d'échantillons et faire un suivi dynamique sur une longue période. Associer d'autre site anatomique de la mère (bouche, intestin, peau et vagin) et du bébé (bouche, peau et intestin) et voire même à l'alimentation pour répondre à la question de la contamination ou d'interconnexion du

microbiote. Nous proposons aussi de tester l'activité anti-listeria des différentes bactéries probiotiques isolées du lait maternel contre nos huit souches de *Listera monocytogenes*.

Références bibliographiques

1. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The Human Microbiome Project. Nature. 2007 Oct. 18;449(7164):804–10.

2. Vilain A-C. Qu'est-ce que le lait ? Rev Fr Allergol. 2010 Apr; 50 (3):124–7.

3. Gartner LM, Morton J, Lawrence RA, Naylor AJ, O'Hare D, Schanler RJ, et al. Breastfeeding and the use of human milk. Pediatrics. 2005 Feb; 115(2):496–506.

4. Arifeen S, Black RE, Antelman G, Baqui A, Caulfield L, Becker S. Exclusive breastfeeding reduces acute respiratory infection and diarrhea deaths among infants in Dhaka slums. Pediatrics. 2001; 108(4):E67.

5. Duijts L, Jaddoe VWV, Hofman A, Moll HA. Prolonged and exclusive breastfeeding reduces the risk of infectious diseases in infancy. Pediatrics. 2010; 126(1):e18-25.

6. Quigley MA, Kelly YJ, Sacker A. Breastfeeding and hospitalization for diarrheal and respiratory infection in the United Kingdom Millennium Cohort Study. Pediatrics. 2007 Apr; 119(4):e837-842.

7. el-Mohandes AE, Schatz V, Keiser JF, Jackson BJ. Bacterial contaminants of collected and frozen human milk used in an intensive care nursery. Am J Infect Control. 1993; 21(5):226–30.

8. Eidelman AI, Szilagyi G. Patterns of bacterial colonization of human milk. Obstet Gynecol. 1979 May; 53(5):550–2.

9. Urbaniak C, Cummins J, Brackstone M, Macklaim JM, Gloor GB, Baban CK, et al. Microbiota of human breast tissue. Appl Environ Microbiol. 2014; 80(10):3007–14.

10. Martín R, Langa S, Reviriego C, Jimínez E, Marín ML, Xaus J, et al. Human milk is a source of lactic acid bacteria for the infant gut. J Pediatr. 2003; 143(6):754–8.

11. Martín R, Heilig GHJ, Zoetendal EG, Smidt H, Rodríguez JM. Diversity of the *Lactobacillus* group in breast milk and vagina of healthy women and potential role in the colonization of the infant gut. J Appl Microbiol. 2007;103(6):2638–44.

12. Martín R, Jiménez E, Olivares M, Marín ML, Fernández L, Xaus J, et al. *Lactobacillus salivarius* CECT 5713, a potential probiotic strain isolated from infant feces and breast milk of a mother–child pair. Int J Food Microbiol. 2006; 112(1):35–43.

13. Million M, Maraninchi M, Henry M, Armougom F, Richet H, Carrieri P, et al. Obesity-associated gut microbiota is enriched in *Lactobacillus reuteri* and depleted in *Bifidobacterium animalis* and *Methanobrevibacter smithii*. Int J Obes. 2012; 36(6):817–25.

14. Cabrera-Rubio R, Collado MC, Laitinen K, Salminen S, Isolauri E, Mira A. The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. Am J Clin Nutr. 2012; 96(3):544–51.

15. Urbaniak C, Angelini M, Gloor GB, Reid G. Human milk microbiota profiles in relation to birthing method, gestation and infant gender. Microbiome. 2016; 4(1).

16. Hunt KM, Foster JA, Forney LJ, Schütte UME, Beck DL, Abdo Z, et al. Characterization of the diversity and temporal stability of bacterial communities in human milk. PloS One. 2011; 6(6):e21313.

17. Drago L, Toscano M, De Grandi R, Grossi E, Padovani EM, Peroni DG. Microbiota network and mathematic microbe mutualism in colostrum and mature milk collected in two different geographic areas: Italy versus Burundi. ISME J. 2017; 11(4):875–84.

18. Rainey FA, Janssen PH. Phylogenetic analysis by 16S ribosomal DNA sequence comparison reveals two unrelated groups of species within the genus *Ruminococcus*. FEMS Microbiol Lett. 1995; 129(1):69–73.

19. Duncan SH, Hold GL, Harmsen HJM, Stewart CS, Flint HJ. Growth requirements and fermentation products of *Fusobacterium prausnitzii*, and a proposal to reclassify it as *Faecalibacterium prausnitzii* gen. nov., comb. nov. Int J Syst Evol Microbiol. 2002 Nov; 52(Pt 6):2141–6.

20. Barcenilla A, Pryde SE, Martin JC, Duncan SH, Stewart CS, Henderson C, et al. Phylogenetic relationships of butyrate-producing bacteria from the human gut. Appl Environ Microbiol. 2000; 66(4):1654–61.

21. Vital M, Howe AC, Tiedje JM. Revealing the Bacterial Butyrate Synthesis Pathways by Analyzing (Meta) genomic Data. Moran MA, editor. mBio. 2014 Apr 22; 5(2):e00889-14.

22. Dridi B, Henry M, El Khéchine A, Raoult D, Drancourt M. High Prevalence of *Methanobrevibacter smithii* and *Methanosphaera stadtmanae* Detected in the Human Gut Using an Improved DNA Detection Protocol. Dobrindt U, editor. PLoS ONE. 2009; 4(9):e7063.

23. Samuel BS, Gordon JI. A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism. Proc Natl Acad Sci. 2006 Jun 27; 103(26):10011–6.

24. Samuel BS, Hansen EE, Manchester JK, Coutinho PM, Henrissat B, Fulton R, et al. Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. Proc Natl Acad Sci. 2007; 104(25):10643–8.

25. Khelaifia S, Raoult D, Drancourt M. A versatile medium for cultivating methanogenic Archaea. Driessen A, editor. PLoS ONE. 2013; 8(4):e61563.

26. Havelaar AH, Haagsma JA, Mangen M-JJ, Kemmeren JM, Verhoef LPB, Vijgen SMC, et al. Disease burden of foodborne pathogens in the Netherlands, 2009. Int J Food Microbiol. 2012 Jun 1; 156(3):231–8.

27. Oppel KB, Holgate SL, Finlayson H. A retrospective review of *Listeria monocytogenes* infection at Tygerberg Children's Hospital, Cape Town, South Africa, from 2006 to 2016: Is empirical ampicillin still indicated after the first month of life? S Afr Med J. 2018 Oct 26; 108(11):937.

28. Shimbo A, Takasawa K, Nishioka M, Morio T, Shimohira M. Complications of Listeria meningitis in two immunocompetent children. Pediatr Int. 2018 May; 60(5):491–2.

29. Aljičević M, Bešlagić E, Zvizdić Š, Hamzić S, Mahmutović S. *Listeria monocytogenes* as the possible cause of the spontaneous abortion in female of the fertile age. Bosn J Basic Med Sci. 2005 Nov 20; 5(4):89.

30. Riaz A, Noureen S, Liqat I, Arshad M, Arshad N. Antilisterial efficacy of *Lactobacillus brevis* MF179529 from cow: an in vivo evidence. BMC Complement Altern Med. 2019 Dec; 19(1):37.

31. Cavicchioli VQ, Camargo AC, Todorov SD, Nero LA. Novel bacteriocinogenic *Enterococcus hirae* and *Pediococcus pentosaceus* strains with antilisterial activity isolated from Brazilian artisanal cheese. J Dairy Sci. 2017 Apr; 100(4):2526–35.

32. Dubourg G, Lagier JC, Robert C, Armougom F, Hugon P, Metidji S, et al. Culturomics and pyrosequencing evidence of the reduction in gut microbiota diversity in patients with broad-spectrum antibiotics. Int J Antimicrob Agents. 2014 Aug; 44(2):117–24.

33. Lagier J-C, Raoult D. Culturomics : une méthode d'étude du microbiote humain. médecine/sciences. 2016 Nov; 32(11):923–5.

34. Ramasamy D, Mishra AK, Lagier J-C, Padhmanabhan R, Rossi M, Sentausa E, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol. 2014; 64(2):384–91.

35. 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. 2015;36:73–8.

36. Lagier J-C, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect. 2012; 18(12):1185–93.

37. Pfleiderer A, Lagier J-C, Armougom F, Robert C, Vialettes B, Raoult D. Culturomics identified 11 new bacterial species from a single anorexia nervosa stool sample. Eur J Clin Microbiol Infect Dis. 2013 Nov; 32(11):1471–81.

38. Moore ERB, Mihaylova SA, Vandamme P, Krichevsky MI, Dijkshoorn L. Microbial systematics and taxonomy: relevance for a microbial commons. Res Microbiol. 2010 Jul; 161(6):430–8.

39. Trüper H. How to name a prokaryote? Etymological considerations, proposals and practical advice in prokaryote nomenclature. FEMS Microbiol Rev. 1999 Apr; 23(2):231–49.

40. Kim M, Oh H-S, Park S-C, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol. 2014; 64(Pt 2):346–51.

41. Rosselló-Mora R. DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation. In: Molecular identification, systematics, and population structure of prokaryotes. Springer; 2006. p. 23–50.

42. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier P-E, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis. 2009; 49(4):543–51.

43. Lagier J-C, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P, et al. Culture of previously uncultured members of the human gut microbiota by culturomics. Nat Microbiol. 2016; 1:16203.

44. Lee I, Ouk Kim Y, Park S-C, Chun J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol. 2016 Feb 1; 66 (2):1100–3.

45. Amitay EL, Werner S, Vital M, Pieper DH, Höfler D, Gierse I-J, et al. Fusobacterium and colorectal cancer: causal factor or passenger? Results from a large colorectal cancer screening study. Carcinogenesis. 2017 Aug; 38 (8):781–8.

46. Miquel S, Martín R, Rossi O, Bermúdez-Humarán LG, Chatel JM, Sokol H, et al. *Faecalibacterium prausnitzii* and human intestinal health. Curr Opin Microbiol. 2013 Jun; 16 (3):255–61.

Partie III : Annexes

Avant-propos

Dans ces annexes, nous rapportons les autres travaux auxquels nous avons participés au cours de notre thèse. Le premier travail illustre l'importance du concept de culturomics dans le dénombrement du microbiote bactérien du corps humain (**Article XVI**). Notre contribution pour ce travail (article XVI) a été l'isolation de 13 nouvelles bactéries sur les 247 de l'article (5%) dont 5 nouveaux genres (*Bariatricus massiliensis*, *Fournierella massiliensis*, *Hugonella massiliensis*, *Marseillibacillus timonensis* (*Marseillococcus timonensis*) et *Ruminococcus phoceensis* (*Mediterraneibacter massiliensis*) et 8 nouvelles espèces (*Anaerotruncus massiliensis*, *Butyricimonas phoceensis*, *Lachnoclostridium bouchesdurhonense*, *Drancourtella timonensis*, *Eisenbergiella massiliensis*, *Enterococcus massiliensis*, *Negativicoccus massiliensis* et *Paenibacillus ihumii*). Le second fait état de la description de nouvelles bactéries par la méthode plus courte de taxonomie appelée NSA (New Species Announcement) (**Articles XVIII et XIX**) basée essentiellement sur les caractéristiques phénotypiques. *Anaerotruncus rubiinfantis* est une nouvelle bactérie isolée à partir de la selle d'un enfant sénégalais atteint de malnutrition aiguë sévère. Elle est décrite selon la méthode de NSA nouvelle formule. Dans cette nouvelle formule, en plus des caractéristiques phénotypiques une comparaison du génome avec les espèces les plus proches y est ajoutée. Il s'agit de l'identité nucléotidique moyenne des gènes orthologues, calculé par le logiciel en ligne OrthoANI (https://www.ezbiocloud.net/tools/orthoani) (44) (**Article XX**).

Article XVII:

Culture of previously uncultured members of the human gut microbiota by culturomics.

Jean-Christophe Lagier, Saber Khelaifia, Maryam Tidjani Alou, Sokhna Ndongo, Niokhor Dione, Perrine Hugon, Aurelia Caputo, Frédéric Cadoret, Sory Ibrahima Traore, El Hadji Seck, Gregory Dubourg, Guillaume Durand, Gaël Mourembou, Elodie Guilhot, **Amadou Togo**, Sara Bellali, Dipankar Bachar, Nadim Cassir, Fadi Bittar, Jérémy Delerce, Morgane Mailhe, Davide Ricaboni1, Melhem Bilen, Nicole Prisca Makaya Dangui Nieko, Ndeye Mery Dia Badiane, Camille Valles, Donia Mouelhi, Khoudia Diop, Matthieu Million, Didier Musso, Jônatas Abrahão, Esam Ibraheem Azhar, Fehmida Bibi, Muhammad Yasir, Aldiouma Diallo, Cheikh Sokhna, Felix Djossou, Véronique Vitton, Catherine Robert, Jean Marc Rolain, Bernard La Scola, Pierre-Edouard Fournier, Anthony Levasseur and Didier Raoult.

OPEN

Culture of previously uncultured members of the human gut microbiota by culturomics

Jean-Christophe Lagier[1], Saber Khelaifia[1], Maryam Tidjani Alou[1], Sokhna Ndongo[1], Niokhor Dione[1], Perrine Hugon[1], Aurelia Caputo[1], Frédéric Cadoret[1], Sory Ibrahima Traore[1], El Hadji Seck[1], Gregory Dubourg[1], Guillaume Durand[1], Gaël Mourembou[1], Elodie Guilhot[1], Amadou Togo[1], Sara Bellali[1], Dipankar Bachar[1], Nadim Cassir[1], Fadi Bittar[1], Jérémy Delerce[1], Morgane Mailhe[1], Davide Ricaboni[1], Melhem Bilen[1], Nicole Prisca Makaya Dangui Nieko[1], Ndeye Mery Dia Badiane[1], Camille Valles[1], Donia Mouelhi[1], Khoudia Diop[1], Matthieu Million[1], Didier Musso[2], Jônatas Abrahão[3], Esam Ibraheem Azhar[4], Fehmida Bibi[4], Muhammad Yasir[4], Aldiouma Diallo[5], Cheikh Sokhna[5], Felix Djossou[6], Véronique Vitton[7], Catherine Robert[1], Jean Marc Rolain[1], Bernard La Scola[1], Pierre-Edouard Fournier[1], Anthony Levasseur[1] and Didier Raoult[1]*

Metagenomics revolutionized the understanding of the relations among the human microbiome, health and diseases, but generated a countless number of sequences that have not been assigned to a known microorganism[1]. The pure culture of prokaryotes, neglected in recent decades, remains essential to elucidating the role of these organisms[2]. We recently introduced microbial culturomics, a culturing approach that uses multiple culture conditions and matrix-assisted laser desorption/ionization–time of flight and 16S rRNA for identification[2]. Here, we have selected the best culture conditions to increase the number of studied samples and have applied new protocols (fresh-sample inoculation; detection of microcolonies and specific cultures of Proteobacteria and microaerophilic and halophilic prokaryotes) to address the weaknesses of the previous studies[3–5]. We identified 1,057 prokaryotic species, thereby adding 531 species to the human gut repertoire: 146 bacteria known in humans but not in the gut, 187 bacteria and 1 archaea not previously isolated in humans, and 197 potentially new species. Genome sequencing was performed on the new species. By comparing the results of the metagenomic and culturomic analyses, we show that the use of culturomics allows the culture of organisms corresponding to sequences previously not assigned. Altogether, culturomics doubles the number of species isolated at least once from the human gut.

The study of the human gut microbiota has been revived by metagenomic studies[6–8]. However, a growing problem is the gaps that remain in metagenomics, which correspond to unidentified sequences that may be correlated with an identified organism[9]. Moreover, the exploration of relations between the microbiota and human health require—both for an experimental model and therapeutic strategies—the growing of microorganisms in pure culture[10], as recently demonstrated in elucidations of the role of *Clostridium butyricum* in necrotizing enterocolitis and the influence of gut microbiota on cancer immunotherapy effects[11,12]. In recent years, microbial culture techniques have been neglected, which explains why the known microbial community of the human gut is extremely low[13]. Before we initiated microbial culturomics[13] of the approximately 13,410 known bacterial and archaea species, 2,152 had been identified in humans and 688 bacteria and 2 archaea had been identified in the human gut. Culturomics consists of the application of high-throughput culture conditions to the study of the human microbiota and uses matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF) or 16S rRNA amplification and sequencing for the identification of growing colonies, some of which have been previously unidentified[2]. With the prospect of identifying new genes of the human gut microbiota, we extend here the number of recognized bacterial species and evaluate the role of this strategy in resolving the gaps in metagenomics, detailing our strategy step by step (see Methods). To increase the diversity, we also obtained frozen samples from healthy individuals or patients with various diseases from different geographical origins. These frozen samples were collected as fresh samples (stool, small-bowel and colonic samples; Supplementary Table 1). Furthermore, to determine appropriate culture conditions, we first reduced the number of culture conditions used (Supplementary Table 2a–c) and then focused on specific strategies for some taxa that we had previously failed to isolate (Supplementary Table 3).

First, we standardized the microbial culturomics for application to the sample testing (Supplementary Table 1). A refined analysis of our first study, which had tested 212 culture conditions[4], showed that all identified bacteria were cultured at least once using one of the 70 best culture conditions (Supplementary Table 2a). We applied these 70 culture conditions (Supplementary Table 2a) to the study of 12 stool samples (Supplementary Table 1). Thanks to the implementation of the recently published repertoire of human bacteria[13] (see Methods), we determined that the isolated bacteria included 46 bacteria known from the gut but not recovered by culturomics before this work (new for culturomics), 38 that had

[1]Aix Marseille Université URMITE, UM63, CNRS 7278, IRD 198, INSERM 1095, 27 Boulevard Jean Moulin, 13385 Marseille Cedex 5, France. [2]Institut Louis Malardé, Papeete, Tahiti, Polynésie Française. [3]Departamento de Microbiologia Laboratorio de Virus, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil. [4]Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah 21589, Saudi Arabia. [5]Institut de Recherche pour le Développement, UMR 198 (URMITE), Campus International de Hann, IRD, BP 1386, CP, 18524 Dakar, Sénégal. [6]Department of Infectious and Tropical Diseases, Centre Hospitalier de Cayenne, Cayenne, French Guiana. [7]Service de Gastroentérologie, Hôpital Nord, Assistance Publique-Hôpitaux de Marseille, 13915 Marseille, France. *e-mail: didier.raoult@gmail.com
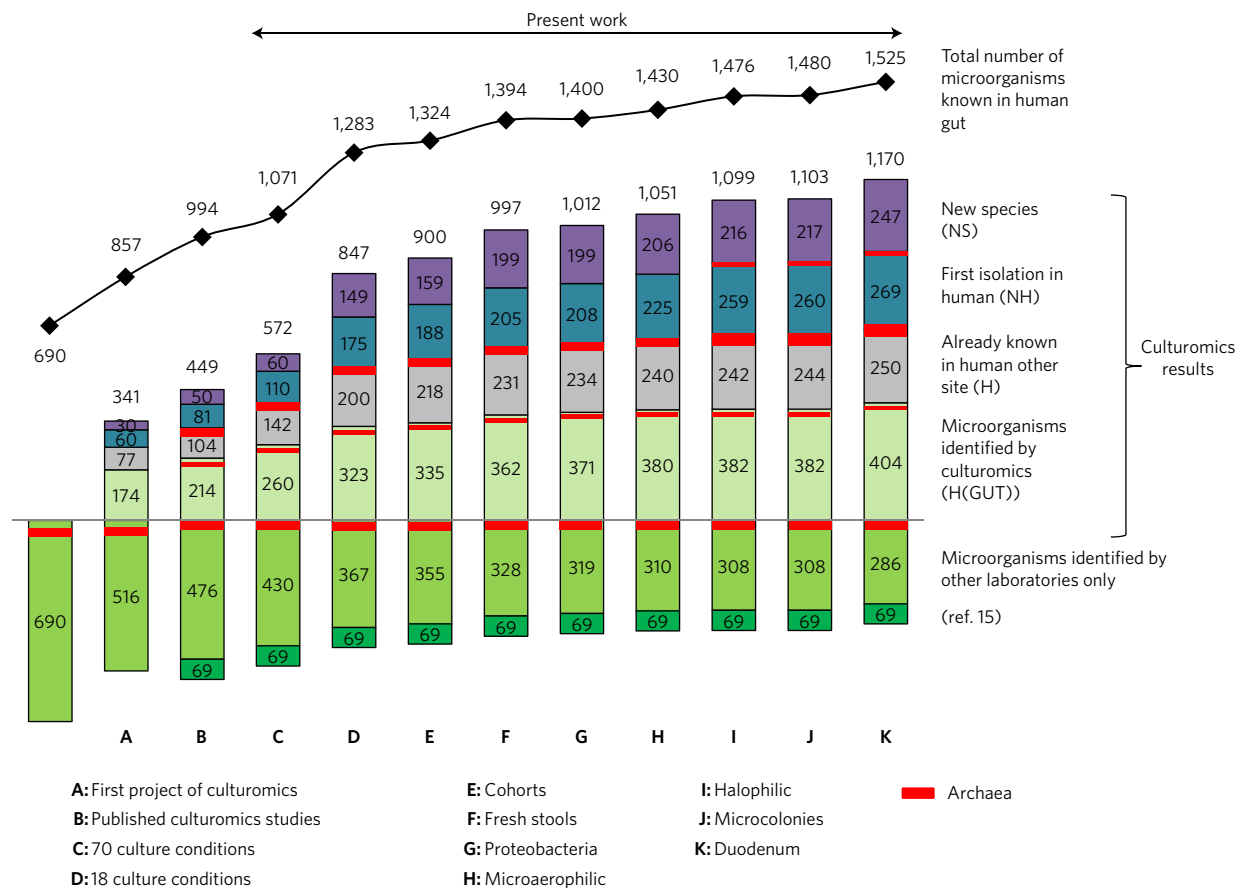
Figure 1 | Number of different bacteria and archaea isolated during the culturomics studies. Columns A and B represent the results from previously published studies, and columns C to K the different projects described herein. The bacterial species are represented in five categories: NS, new species; NH, prokaryotes first isolated in humans; H, prokaryotes already known in humans but never isolated from the human gut; H (GUT), prokaryotes known in the human gut but newly isolated by culturomics; and prokaryotes isolated by other laboratories but not by culturomics.

already been isolated in humans but not from the gut (non-gut bacteria), 29 that had been isolated in humans for the first time (non-human bacteria) and 10 that were completely new species (unknown bacteria) (Fig. 1 and Supplementary Tables 4a and 5).

Beginning in 2014, to reduce the culturomics workload and extend our stool-testing capabilities, we analysed previous studies and selected the 18 best culture conditions[2]. We performed cultures in liquid media in blood culture bottles, followed by subcultures on agar (Supplementary Table 2b). We designed these culture conditions by analysing our first studies. The results of those studies indicated that emphasizing three components was essential: pre-incubation in a blood culture bottle (56% of the new species isolated), the addition of rumen fluid (40% of the new species isolated) and the addition of sheep blood (25% of the new species isolated)[2–5]. We applied this strategy to 37 stool samples from healthy individuals with different geographic provenances and from patients with different diseases (Supplementary Table 1). This new strategy enabled the culture of 63 organisms new to culturomics, 58 non-gut bacteria, 65 non-human bacteria and 89 unknown bacteria (Fig. 1 and Supplementary Tables 4a and 5).

We also applied culturomic conditions (Supplementary Table 2c) to large cohorts of patients sampled for other purposes (premature infants with necrotizing enterocolitis, pilgrims returning from the Hajj and patients before or after bariatric surgery) (Supplementary Table 1). A total of 330 stool samples were analysed. This enabled the detection of 13 bacteria new to culturomics, 18 non-gut bacteria, 13 non-human bacteria and 10 unknown species (Fig. 1 and Supplementary Tables 4a and 5).

Among the gut species mentioned in the literature[13] and not previously recovered by culturomics, several were extremely oxygen-sensitive anaerobes, several were microaerophilic and several were Proteobacteria, and we focused on these bacteria (Supplementary Table 3). Because delay and storage may be critical with anaerobes, we inoculated 28 stools immediately upon collection. This enabled the culture of 27 new gut species for culturomics, 13 non-gut bacteria, 17 non-human bacteria and 40 unknown bacteria (Fig. 1 and Supplementary Tables 3a and 4). When we specifically tested 110 samples for Proteobacteria, we isolated 9 bacteria new to culturomics, 3 non-gut bacteria and 3 non-human bacteria (Fig. 1 and Supplementary Tables 4a and 5). By culturing 242 stool specimens exclusively under a microaerophilic atmosphere, we isolated 9 bacteria new to culturomics, 6 non-gut bacteria, 17 non-human bacteria and 7 unknown bacteria (Fig. 1 and Supplementary Tables 4a and 5). We also introduced the culture of halophilic prokaryotes from the gut and microcolony detection. The culture of halophilic bacteria was performed using culture media supplemented with salt for 215 stool samples, allowing the culture of 48 halophilic prokaryotic species, including one archaea (*Haloferax alexandrinus*), 2 new bacteria for culturomics, 2 non-gut bacteria, 34 non-human bacteria, 10 unknown bacteria and one new halophilic archaea (*Haloferax massiliensis* sp. nov.) (Fig. 1 and Supplementary Tables 4a and 5). Among these 48 halophilic prokaryotic species, 7 were slight halophiles (growing with 10–50 g l^{-1} of NaCl), 39 moderate halophiles (growing with 50–200 g l^{-1} of NaCl) and 2 extreme halophiles (growing with 200–300 g l^{-1} of NaCl).

We also introduced the detection of microcolonies that were barely visible to the naked eye (diameters ranging from 100 to
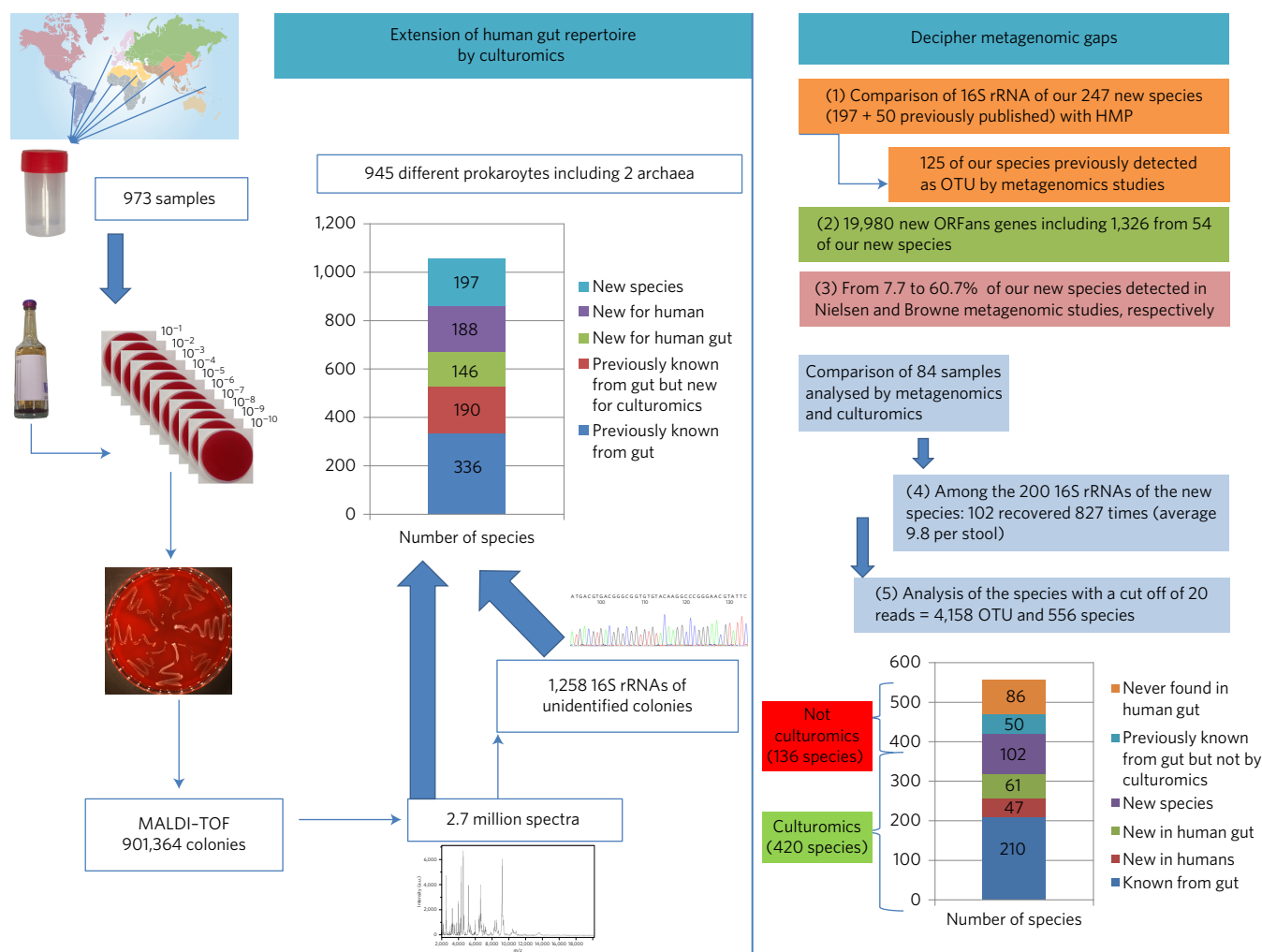
Figure 2 | Summary of the culturomics work that has extended the gut repertoire and filled some of the gaps in metagenomics.

300 µm) and could only be viewed with magnifying glasses. These colonies were transferred into a liquid culture enrichment medium for identification by MALDI–TOF mass spectrometry (MS) or 16S rRNA amplification and sequencing. By testing ten stool samples, we detected two non-gut bacteria, one non-human bacterium and one unknown bacterium that only formed micro-colonies (Fig. 1 and Supplementary Tables 4a and 5). Finally, by culturing 30 duodenal, small bowel intestine and colonic samples, we isolated 22 bacteria new to culturomics, 6 non-gut bacteria, 9 non-human bacteria and 30 unknown bacteria (Fig. 1 and Supplementary Tables 4a and 5). To continue the exploration of gut microbiota, future culturomics studies could also be applied to intestinal biopsies.

In addition, we performed five studies to evaluate the role of culturomics for deciphering the gaps in metagenomics[9]. First, we compared the 16S rRNA sequences of the 247 new species (the 197 new prokaryotic species isolated here in addition to the 50 new bacterial species isolated in previous culturomic studies[3–5]) to the 5,577,630 reads from the 16S rRNA metagenomic studies listed by the Human Microbiome Project (HMP) (http://www.hmpdacc.org/catalog). We found sequences, previously termed operational taxonomic units (OTUs), for 125 of our bacterial species (50.6%). These identified bacterial species included *Bacteroides bouchedurhonense*, which was recovered in 44,428 reads, showing that it is a common bacterium (Supplementary Table 6). Second, because the genome sequencing of 168 of these new species allowed the generation of 19,980 new genes that were previously unknown (ORFans genes)

(Supplementary Table 7), we blasted these with 13,984,809 contigs/scaffolds from the assembly of whole metagenomic studies by HMP, enabling the detection of 1,326 ORFans (6.6%) from 54 of our new bacterial species (including 45 detected also from 16S) (Supplementary Table 8). Therefore, at least 102 new bacterial species were found but not identified in previous metagenomic studies from the HMP. Third, we searched for our 247 new species in the 239 human gut microbiome samples from healthy individuals described by Browne *et al.*, in which 137 bacterial species were isolated[15]. We captured 150 of our new species in these metagenomics data, representing 60.7% (Supplementary Table 9). Moreover, we also identified 19 of our species (7.7%) from 396 human stool individuals described by Nielsen *et al.*, from which 741 metagenomic species and 238 unique metagenomic genomes were identified[16] (Supplementary Table 9). Fourth, we analysed the 16S rRNA metagenomic sequences of 84 stools also tested by culturomics (Supplementary Table 10). We compared the OTUs identified by blast with a database including the 16S rRNA of all species isolated by culturomics. Among the 247 16S rRNA of the new species, 102 were recovered 827 times, with an average of 9.8 species per stool. Finally, analysis of these species using a cutoff threshold of 20 reads identified 4,158 OTUs and 556 (13.4%) species (Supplementary Table 11), among which 420 species (75.5%) were recovered by culturomics. Of these, 210 (50%) were previously found to be associated with the human gut, 47 were not previously found in humans (11.2%), 61 were found in humans but not in the gut (14.5%) and 102 (24.3%) were new species.
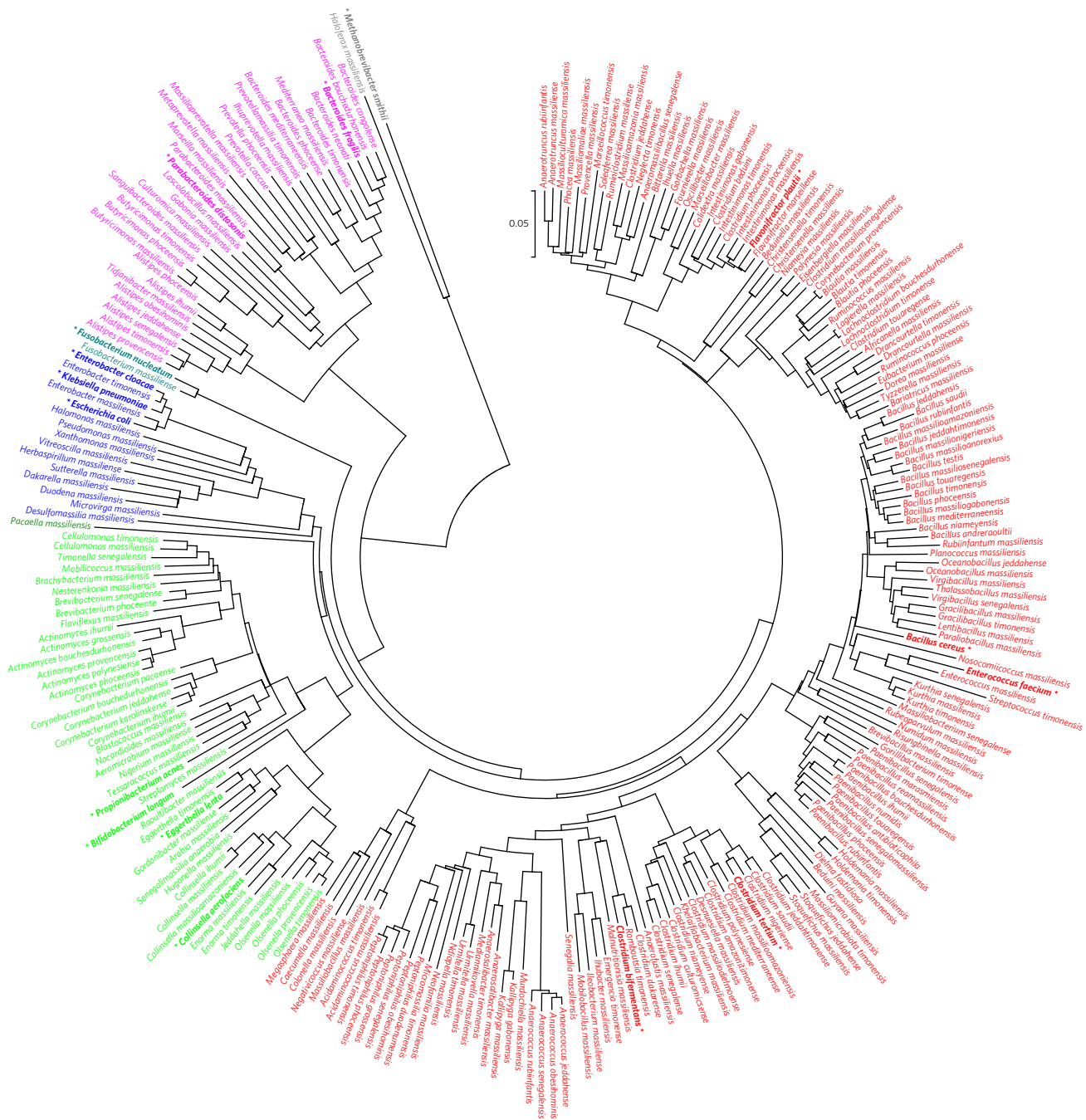
Figure 3 | Phylogenetic tree of the 247 new prokaryote species isolated by culturomics. Bacterial species from Firmicutes are highlighted in red, Actinobacteria (light green), Proteobacteria (blue), Bacteroidetes (purple), Synergistetes (green), Fusobacteria (dark green) and Archaea (grey), respectively. The sequences of 16 prokaryotic species belonging to six phyla previously known from the human gut and more frequently isolated by culture in human gut are highlighted in bold and by an asterisk.

Interestingly, among the 136 species not previously found by culturomics, 50 have been found in the gut and 86 have never previously been found in the human gut (Fig. 2 and Supplementary Table 11).

Overall, in this study, by testing 901,364 colonies using MALDI–TOF MS (Supplementary Table 1), we isolated 1,057 bacterial species, including 531 newly found in the human gut. Among them, 146 were non-gut bacteria, 187 were non-human bacteria, one was a non-human halophilic archaeon and 197 were unknown bacteria, including two new families (represented by *Neofamilia massiliensis* gen. nov., sp. nov. and *Beduinella massiliensis* gen. nov., sp. nov.) and one unknown halophilic archaeon (Fig. 1 and Supplementary Table 4a). Among these, 600 bacterial species belonged to Firmicutes, 181 to Actinobacteria, 173

to Proteobacteria (a phylum that we have under-cultured to date; Supplementary Table 5), 88 to Bacteroidetes, 9 to Fusobacteria, 3 to Synergistetes, 2 to Euryarchaeota, 1 to Lentisphaerae and 1 to Verrucomicrobia (Supplementary Table 4a). Among these 197 new prokaryotes species, 106 (54%) were detected in at least two stool samples, including a species that was cultured in 13 different stools (*Anaerosalibacter massiliensis*) (Supplementary Table 4a). In comparison with our contribution, a recent work using a single culture medium was able to culture 120 bacterial species, including 51 species known from the gut, 1 non-gut bacterium, 1 non-human bacterium and 67 unknown bacteria, including two new families (Supplementary Table 12).

To obtain these significant results we tested more than 900,000 colonies, generating 2.7 million spectra, and performed 1,258 molecular identifications of bacteria not identified through MALDI–TOF, using 16S rRNA amplification and sequencing. The new prokaryote species are available in the Collection de Souches de l'Unité des Rickettsies (CSUR) and Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) (Supplementary Tables 4a and 5). All 16S sequences of the new species and the species unidentified by MALDI–TOF, as well as the genome sequences of the new species, have been deposited in GenBank (Supplementary Tables 5 and 13). In addition, thanks in part to an innovative system using a simple culture for the archaea without an external source of hydrogen[17], among these prokaryotes we isolated eight archaeal species from the human gut, including two new ones for culturomics, one non-gut archaea, four non-human archaea and one new halophilic species.

We believe that this work is a key step in the rebirth of the use of culturing in human microbiology[2–5,16] and only the efforts of several teams around the world in identifying the gut microbiota repertoire will allow an understanding and analysis of the relations between the microbiota and human health, which could then participate in adapting Koch's postulates to include the microbiota[21]. The rebirth of culture, termed culturomics here, has enabled the culturing of 77% of the 1,525 prokaryotes now identified in the human gut (Fig. 1 and Supplementary Table 4b). In addition, 247 new species (197 cultured here plus 50 from previous studies) and their genomes are now available (Fig. 3). The relevance of the new species found by culturomics is emphasized because 12 of them were isolated in our routine microbiology laboratory from 57 diverse clinical samples (Supplementary Table 14). In 2016, 6 of the 374 (1.6%) different identifications performed in the routine laboratory were new species isolated from culturomics. As 519 of the species found by culturomics in the gut for the first time (Fig. 1) were not included in the HMP (Supplementary Table 15) and because hundreds of their genomes are not yet available, the results of this study should prompt further genome sequencing to obtain a better identification in gut metagenomic studies.

Methods

Samples. To obtain a larger diversity of gut microbiota, we analysed 943 different stool samples and 30 small intestine and colonic samples from healthy individuals living or travelling in different geographical regions (Europe, rural and urban Africa, Polynesia, India and so on) and from patients with diverse diseases (for example, anorexia nervosa, obesity, malnutrition and HIV). The main characteristics are summarized in Supplementary Table 1. Consent was obtained from each patient, and the study was approved by the local Ethics Committee of the IFR48 (Marseille, France; agreement no. 09–022). Except for the small intestine and stool samples that we directly inoculated without storage (see sections 'Fresh stool samples' and 'Duodenum and other gut samples'), the faecal samples collected in France were immediately aliquoted and frozen at −80 °C. Those collected in other countries were sent to Marseille on dry ice, then aliquoted and frozen at −80 °C for between 7 days and 12 months before analysis.

Culturomics. Culturomics is a high-throughput method that multiplies culture conditions in order to detect higher bacterial diversity. The first culturomics study concerned three stool samples, 212 culture conditions (including direct inoculation in various culture media), and pre-incubation in blood culture bottles incubated aerobically and anaerobically[4]. Overall, 352 other stool samples, including stool samples from patients with anorexia nervosa[3], patients treated with antibiotics[5], or Senegalese children, both healthy and those with diarrhoea[22], were previously studied by culturomics, and these results have been comprehensively detailed in previous publications[3–5]. In this work, we only included the genome sequences of the 50 new bacterial species isolated in these previous works to contribute to our analysis of culturomics and to fill some of the gaps left by metagenomics. In addition, these previously published data are clearly highlighted in Fig. 1, illustrating the overall contribution of culturomics in exploring the gut microbiota.

Bacterial species isolated from our new projects and described here were obtained using the strategy outlined in the following sections.

Standardization of culturomics for the extension of sample testing. A refined analysis allowed the selection of 70 culture conditions (Supplementary Table 2a) for

the growth of all the bacteria[4]. We applied these culture conditions to 12 more stool samples and tested 160,265 colonies by MALDI–TOF (Supplementary Table 1). The 18 best culture conditions were selected using liquid media enrichment in a medium containing blood and rumen fluid and subculturing aerobically and anaerobically in a solid medium (Supplementary Table 2b)[2]. Subcultures were inoculated every three days on solid medium, and each medium was kept for 40 days. We applied these culture conditions to 40 stool samples, ultimately testing 565,242 colonies by MALDI–TOF (Supplementary Table 1).

Cohorts. In parallel to these main culturomics studies, we used fewer culture conditions to analyse a larger number of stool samples. We refer to these projects as cohorts. Four cohorts were analysed (pilgrims returning from the Hajj, premature infants with necrotizing enterocolitis, patients before and after bariatric surgery, and patients for acidophilic bacterial species detection). A total of 330 stool samples generated the 52,618 colonies tested by MALDI–TOF for this project (Supplementary Table 1).

Pilgrims from the Hajj. A cohort of 127 pilgrims was included and 254 rectal swabs were collected from the pilgrims: 127 samples were collected before the Hajj and 127 samples were collected after the Hajj. We inoculated 100 µl of liquid sample in an 8 ml bottle containing Trypticase Soy Broth (BD Diagnostics) and incubated the sample at 37 °C for 1 day. We inoculated 100 µl of the enriched sample into four culture media: Hektoen agar (BD Diagnostics), MacConkey agar+Cefotaxime (bioMérieux), Cepacia agar (AES Chemunex) and Columbia ANC agar (bioMérieux). The sample was diluted 10^{-3} before being plated on the MacConkey and Hektoen agars and 10^{-4} before being plated on the ANC agar. The sample was not diluted before being inoculated on the Cepacia agar. Subcultures were performed on Trypticase Soy Agar (BD Diagnostics) and 3,000 colonies were tested using MALDI–TOF.

Preterm neonates. Preterm neonates were recruited from four neonatal intensive care units (NICUs) in southern France from February 2009 to December 2012 (ref. 12). Only patients with definite or advanced necrotizing enterocolitis corresponding to Bell stages II and III were included. Fifteen controls were matched to 15 patients with necrotizing enterocolitis by sex, gestational age, birth weight, days of life, type of feeding, mode of delivery and duration of previous antibiotic therapy. The stool samples were inoculated on 54 preselected culture conditions (Supplementary Table 2c). The anaerobic cultures were performed in an anaerobic chamber (AES Chemunex). A total of 3,000 colonies were tested by MALDI–TOF for this project.

Stool analyses before and after bariatric surgery. We included 15 patients who had bariatric surgery (sleeve gastrectomy or Roux-en-Y gastric bypass) from 2009 to 2014. All stool samples were frozen before and after surgery. We used two different culture conditions for this project. Each stool sample was diluted in 2 ml of Dulbecco's phosphate-buffered saline, then pre-incubated in both anaerobic (BD Bactec Plus Lytic/10 Anaerobic) and aerobic (BD Bactec Plus Lytic/10 Aerobic) blood culture bottles, with 4 ml of sheep blood and 4 ml of sterile rumen fluid being added as previously described[4]. These cultures were subcultured on days 1, 3, 7, 10, 15, 21 and 30 in 5% sheep blood Columbia agar (bioMérieux), and 33,650 colonies were tested by MALDI–TOF.

Acidophilic bacteria. The pH of each stool sample was measured using a pH meter: 1 g of each stool specimen was diluted in 10 ml of neutral distilled water (pH 7) and centrifuged for 10 min at 13,000g; the pH values of the supernatants were then measured. Acidophilic bacteria were cultured after stool enrichment in a liquid medium consisting of Columbia Broth (Sigma-Aldrich) modified by the addition of (per litre) 5 g $MgSO_4$, 5 g $MgCl_2$, 2 g KCl, 2 g glucose and 1 g $CaCl_2$. The pH was adjusted to five different values: 4, 4.5, 5, 5.5 and 6, using HCl. The bacteria were then subcultured on solid medium containing the same nutritional components and pH as the culture enrichment. They were inoculated after 3, 7, 10 or 15 incubation days in liquid medium for each tested pH condition. Serial dilutions from 10^{-1} to 10^{-10} were then performed, and each dilution was plated on agar medium. Negative controls (no inoculation of the culture medium) were included for each condition.

Overall, 16 stool samples were inoculated, generating 12,968 colonies, which were tested by MALDI–TOF.

Optimization of the culturomics strategy. In parallel with this standardization period, we performed an interim analysis in order to detect gaps in our strategy. Analysing our previously published studies, we observed that 477 bacterial species previously known from the human gut were not detected. Most of these species grew in strict anaerobic (209 species, 44%) or microaerophilic (25 species, 5%) conditions, and 161 of them (33%) belonged to the phylum Proteobacteria, whereas only 46 of them (9%) belonged to the phylum Bacteroidetes (Supplementary Table 3). The classification was performed using our own database: (http://www.mediterranee-infection.com/article.php?laref=374&titre=list-of-prokaryotes-according-to-their-aerotolerant-or-obligate-anaerobic-metabolism). Focusing on these bacterial species, we designed specific strategies with the aim of cultivating these missing bacteria.

Fresh stool samples. As the human gut includes extremely oxygen-sensitive bacterial species, and because frozen storage kills some bacteria[10], we tested 28 stool samples from healthy individuals and directly cultivated these samples on collection and without storage. Each sample was directly cultivated on agar plates, enriched in blood culture bottles (BD Bactec Plus Lytic/10 Anaerobic) and followed on days 2, 5, 10 and 15. Conditions tested were anaerobic Columbia with 5% sheep blood (bioMérieux) at 37 °C with or without thermic shock (20 min/80 °C), 28 °C, anaerobic Columbia with 5% sheep blood agar (bioMérieux) and 5% rumen fluid and R-medium (ascorbic acid 1 g l^{-1}, uric acid 0.4 g l^{-1}, and glutathione 1 g l^{-1}, pH adjusted to 7.2), as previously described[23]. For this project, 59,688 colonies were tested by MALDI–TOF.

Proteobacteria. We inoculated 110 stool samples using pre-incubation in blood culture bottles (BD Bactec Plus Lytic/10 Anaerobic) supplemented with vancomycin (100 μg l^{-1}; Sigma-Aldrich). The subcultures were performed on eight different selective solid media for the growth of Proteobacteria. We inoculated onto MacConkey agar (Biokar-Diagnostics), buffered charcoal yeast extract (BD Diagnostic), eosine-methylene blue agar (Biokar-Diagnostics), Salmonella–Shigella agar (Biokar-Diagnostics), Drigalski agar (Biokar-Diagnostics), Hektoen agar (Biokar-Diagnostics), thiosulfate-citrate-bile-sucrose (BioRad) and Yersinia agar (BD Diagnostic) and incubated at 37 °C, aerobically and anaerobically. For this project, 18,036 colonies were tested by MALDI–TOF.

Microaerophilic conditions. We inoculated 198 different stool samples directly onto agar or after pre-incubation in blood culture bottles (BD Bactec Plus Lytic/10 Anaerobic bottles, BD). Fifteen different culture conditions were tested using Pylori agar (bioMérieux), Campylobacter agar (BD), Gardnerella agar (bioMérieux), 5% sheep blood agar (bioMérieux) and our own R-medium as previously described[23]. We incubated Petri dishes only in microaerophilic conditions using GENbag microaer systems (bioMérieux) or CampyGen agar (bioMérieux), except the R-medium, which was incubated aerobically at 37 °C. These culture conditions generated 41,392 colonies, which were tested by MALDI–TOF.

Halophilic bacteria. In addition, we used new culture conditions to culture halophilic prokaryotes. The culture enrichment and isolation procedures for the culture of halophilic prokaryotes were performed in a Columbia broth medium (Sigma-Aldrich), modified by adding (per litre): MgCl$_2$·6H$_2$O, 5 g; MgSO$_4$·7H$_2$O, 5 g; KCl, 2 g; CaCl$_2$·2H$_2$O, 1 g; NaBr, 0.5 g; NaHCO$_3$, 0.5 g and 2 g of glucose. The pH was adjusted to 7.5 with 10 M NaOH before autoclaving. All additives were purchased from Sigma-Aldrich. Four concentrations of NaCl were used (100 g l^{-1}, 150 g l^{-1}, 200 g l^{-1} and 250 g l^{-1}).

A total of 215 different stool samples were tested. One gram of each stool specimen was inoculated aerobically into 100 ml of liquid medium in flasks at 37 °C while stirring at 150 r.p.m. Subcultures were inoculated after 3, 10, 15 and 30 incubation days for each culture condition. Serial dilutions from 10^{-1} to 10^{-10} were then performed in the culture medium and then plated on agar medium. Negative controls (no inoculation of the culture medium) were included for each culture condition. After three days of incubation at 37 °C, different types of colonies appeared: yellow, cream, white and clear. Red and pink colonies began to appear after the 15th day. All colonies were picked and re-streaked several times to obtain pure cultures, which were subcultured on a solid medium consisting of Colombia agar medium (Sigma-Aldrich) NaCl. The negative controls remained sterile in all culture conditions, supporting the authenticity of our data.

Detection of microcolonies. Finally, we began to focus on microcolonies detected using a magnifying glass (Leica). These microcolonies, which were not visualized with the naked eye and ranged from 100 to 300 μm, did not allow direct identification by MALDI–TOF. We subcultured these bacteria in a liquid medium (Columbia broth, Sigma-Aldrich) to allow identification by MALDI–TOF after centrifugation. Ten stool samples were inoculated and then observed using this magnifying glass for this project, generating the 9,620 colonies tested.

Duodenum and other gut samples. Most of the study was designed to explore the gut microbiota using stool samples. Nevertheless, as the small intestine microbiota are located where the nutrients are digested[24], which means there are greater difficulties in accessing samples than when using stool specimens, we analysed different levels of sampling, including duodenum samples (Supplementary Table 1). First, we tested five duodenum samples previously frozen at −80 °C. A total of 25,000 colonies were tested by MALDI–TOF. In addition, we tested samples from the different gut levels (gastric, duodenum, ileum and left and right colon) of other patients. We tested 25,048 colonies by MALDI–TOF for this project. We tested 15 culture conditions, including pre-incubation in blood culture bottles with sterile rumen fluid and sheep blood (BD Bactec Plus Lytic/10 Anaerobic), 5% sheep blood agar (bioMérieux), and incubation in both microaerophilic and anaerobic conditions, R-medium[23] and Pylori agar (bioMérieux). Overall, we tested 50,048 colonies by MALDI–TOF for this project.

Archaea. The culture of methanogenic archaea is a fastidious process, and the necessary equipment for this purpose is expensive and reserved for specialized laboratories. With this technique, we isolated seven methanogenic archaea through culturomic studies as previously described[25–27]. In addition, we propose here an affordable alternative that does not require specific equipment[17]. Indeed, a simple double culture aerobic chamber separated by a microfilter (0.2 μm) was used to grow two types of microorganism that develop in perfect symbiosis. A pure culture of *Bacteroides thetaiotaomicron* was placed in the bottom chamber to produce the hydrogen necessary for the growth of the methanogenic archaea, which was trapped in the upper chamber. A culture of *Methanobrevibacter smithii* or other hydrogenotrophic methanogenic archaea had previously been placed in the chamber. In the case presented here, the methanogenic archaea were grown aerobically on an agar medium supplemented with three antioxidants (ascorbic acid, glutathione and uric acid) and without the addition of any external gas. We subsequently cultured four other methanogenic archaeal species for the first time aerobically, and successfully isolated 13 strains of *M. smithii* and 9 strains of *Methanobrevibacter oralis* from 100 stools and 45 oral samples. This medium allows aerobic isolation and antibiotic susceptibility testing. This change allows the routine study of methanogens, which have been neglected in clinical microbiology laboratories and may be useful for biogas production. Finally, to culture halophilic archaea, we designed specific culture conditions (described in the 'Halophilic bacteria' section).

Identification methods. The colonies were identified using MALDI–TOF MS. Each deposit was covered with 2 ml of a matrix solution (saturated α-cyano acid-4-hydroxycinnamic in 50% acetonitrile and 2.5% trifluoroacetic acid). This analysis was performed using a Microflex LT system (Bruker Daltonics). For each spectrum, a maximum of 100 peaks was used and these peaks were compared with those of previous samples in the computer database of the Bruker Base and our homemade database, including the spectra of the bacterial species identified in previous works[28,29]. An isolate was labelled as correctly identified at the species level when at least one of the colonies' spectra had a score ≥1.9 and another of the colonies' spectra had a score ≥1.7 (refs 28,29).

Protein profiles are regularly updated based on the results of clinical diagnoses and on new species providing new spectra. If, after three attempts, the species could not be accurately identified by MALDI–TOF, the isolate was identified by 16S rRNA sequencing as previously described. A threshold similarity value of >98.7% was chosen for identification at the species level. Below this value, a new species was suspected, and the isolate was described using taxonogenomics[30].

Classification of the prokaryotes species cultured. We used our own online prokaryotic repertoire[13] (http://hpr.mediterranee-infection.com/arktheque/client/ihu_bacteries/recherche/index.php) to classify all isolated prokaryotes into four categories: new prokaryote species, previously known prokaryote species in the human gut, known species from the environment but first isolated in humans, and known species from humans but first isolated in the human gut. Briefly, to complete the recent work identifying all the prokaryotes isolated in humans[13], we examined methods by conducting a literature search, which included PubMed and books on infectious diseases. We examined the Medical Subject Headings (MeSH) indexing provided by Medline for bacteria isolated from the human gut and we then established two different queries to automatically obtain all articles indexed by Medline dealing with human gut isolation sites. These queries were applied to all bacterial species previously isolated from humans as previously described, and we obtained one or more articles for each species, confirming that the bacterium had been isolated from the human gut[13].

International deposition of the strains, 16S rRNA accession numbers and genome sequencing accession number. Most of the strains isolated in this study were deposited in CSUR (WDCM 875) and are easily available at http://www.mediterranee-infection.com/article.php?laref=14&titre=collection-de-souches&PHPSESSID=cncregk417fl97gheb8k7u7t07 (Supplementary Tables 4a and b). All the new prokaryote species were deposited into two international collections: CSUR and DSMZ (Supplementary Table 5). Importantly, among the 247 new prokaryotes species (197 in the present study and 50 in previous studies), we failed to subculture 9 species that were not deposited, of which 5 were nevertheless genome sequenced. Apart from these species, all CSUR accession numbers are available in Supplementary Table 5. Among these viable new species, 189 already have a DSMZ number. For the other 49 species, the accession number is not yet assigned but the strain is deposited. The 16S rRNA accession numbers of the 247 new prokaryotes species are available in Supplementary Table 5, along with the accession number of the known species needing 16S rRNA amplification and sequencing for identification (Supplementary Table 14). Finally, the 168 draft genomes used for our analysis have already been deposited with an available GenBank accession number (Supplementary Table 5) and all other genome sequencing is still in progress, as the culturomics are still running in our laboratory.

New prokaryotes. All new prokaryote species have been or will be comprehensively described by taxonogenomics, including their metabolic properties, MALDI–TOF spectra and genome sequencing[30]. Among these 247 new prokaryote species, 95 have already been published (PMID available in Supplementary Table 5), including 70 full descriptions and 25 'new species announcements'. In addition, 20 are under

review and the 132 others are ongoing (Supplementary Table 5). This includes 37 bacterial species already officially recognized (as detailed in Supplementary Table 5). All were sequenced successively with a paired-end strategy for high-throughput pyrosequencing on the 454-Titanium instrument from 2011 to 2013 and using MiSeq Technology (Illumina) with the mate pair strategy since 2013.

Metagenome sequencing. Total DNA was extracted from the samples using a method modified from the Qiagen stool procedure (QIAamp DNA Stool Mini Kit). For the first 24 metagenomes, we used GS FLX Titanium (Roche Applied Science). Primers were designed to produce an amplicon length (576 bp) that was approximately equivalent to the average length of reads produced by GS FLX Titanium (Roche Applied Science), as previously described. The primer pairs commonly used for gut microbiota were assessed *in silico* for sensitivity to sequences from all phyla of bacteria in the complete Ribosomal Database Project (RDP) database. Based on this assessment, the bacterial primers 917F and 1391R were selected. The V6 region of 16S rRNA was pyrosequenced with unidirectional sequencing from the forward primer with one-half of a GS FLX Titanium PicoTiterPlate Kit 70×75 per patient with the GS Titanium Sequencing Kit XLR70 after clonal amplification with the GS FLX Titanium LV emPCR Kit (Lib-L).

Sixty other metagenomes were amplified for 16S rRNA sequencing using MiSeq technology. PCR-amplified templates of genomic DNA were produced using the surrounding conserved regions' V3–V4 primers with overhang adapters (FwOvAd_341F TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGG NGGCWGCAG; ReOvAd_785RGTCTCGTGGGCTCGGAGATG TGTATAAGA GACAGGACTACHVGGGTATCTAATCC). Samples were amplified individually for the 16S V3–V4 regions by Phusion High Fidelity DNA Polymerase (Thermo Fisher Scientific) and visualized on the Caliper Labchip II device (Illumina) by a DNA 1K LabChip at 561 bp. Phusion High Fidelity DNA Polymerase was chosen for PCR amplifications in this biodiversity approach and deep sequencing: a thermostable DNA polymerase characterized by the greatest accuracy, robust reactions and high tolerance for inhibitors, and finally by an error rate that is approximately 50-fold lower than that of DNA polymerase and sixfold lower than that of Pfu DNA polymerase. After purification on Ampure beads (Thermo Fisher Scientific), the concentrations were measured using high-sensitivity Qbit technology (Thermo Fisher Scientific). Using a subsequent limited-cycle PCR on 1 ng of each PCR product, Illumina sequencing adapters and dual-index barcodes were added to each amplicon. After purification on Ampure beads, the libraries were then normalized according to the Nextera XT (Illumina) protocol. The 96 multiplexed samples were pooled into a single library for sequencing on the MiSeq. The pooled library containing indexed amplicons was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and paired-end sequencing with dual index reads of 2 × 250 bp were performed in a single 39-hour run. On the instrument, the global cluster density and the global passed filter per flow cell were generated. The MiSeq Reporter software (Illumina) determined the percentage indexed and the clusters passing the filter for each amplicon or library. The raw data were configured in fasta files for R1 and R2 reads.

Genome sequencing. The genomes were sequenced using, successively, two high-throughput NGS technologies: Roche 454 and MiSeq Technology (Illumina) with paired-end application. Each project on the 454 sequencing technology was loaded on a quarter region of the GS Titanium PicoTiterPlate and sequenced with the GS FLX Titanium Sequencer (Roche). For the construction of the 454 library, 5 μg DNA was mechanically fragmented on the Covaris device (KBioScience-LGC Genomics) through miniTUBE-Red 5Kb. The DNA fragmentation was visualized through the Agilent 2100 BioAnalyser on a DNA LabChip7500. Circularization and fragmentation were performed on 100 ng. The library was then quantified on Quant-it Ribogreen kit (Invitrogen) using a Genios Tecan fluorometer. The library was clonally amplified at 0.5 and 1 cpb in 2 emPCR reactions according to the conditions for the GS Titanium SV emPCR Kit (Lib-L) v2 (Roche). These two enriched clonal amplifications were loaded onto the GS Titanium PicoTiterPlates and sequenced with the GS Titanium Sequencing Kit XLR70. The run was performed overnight and then analysed on the cluster through gsRunBrowser and gsAssembler_Roche. Sequences obtained with Roche were assembled on gsAssembler with 90% identity and 40 bp of overlap. The library for Illumina was prepared using the Mate Pair technology. To improve the assembly, the second application in was sometimes performed with paired ends. The paired-end and the mate-pair strategies were barcoded in order to be mixed, respectively, with 11 other genomic projects prepared with the Nextera XT DNA sample prep kit (Illumina) and 11 others projects with the Nextera Mate Pair sample prep kit (Illumina). The DNA was quantified by a Qbit assay with high-sensitivity kit (Life Technologies). In the first approach, the mate pair library was prepared with 1.5 μg genomic DNA using the Nextera mate pair Illumina guide. The genomic DNA 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) with a DNA 7500 LabChip. The DNA fragments, which ranged in size, had an optimal size of 5 kb. No size selection was performed, and 600 ng of 'tagmented' fragments measured on the Qbit assay with the high-sensitivity kit were circularized. The circularized DNA was mechanically sheared to small fragments, with optimal fragments being 700 bp, on a Covaris S2 device in microtubes. The library profile was visualized on a High

Sensitivity Bioanalyzer LabChip (Agilent Technologies). The libraries were normalized at 2 nM and pooled. After a denaturation step and dilution at 15 pM, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. To prepare the paired-end library, 1 ng of genome as input was required. DNA was fragmented and tagged during the tagmentation step, with an optimal size distribution at 1 kb. Limited-cycle PCR amplification (12 cycles) completed the tag adapters and introduced dual-index barcodes. After purification on Ampure XP beads (Beckman Coulter), the library was normalized and loaded onto the reagent cartridge and then onto the instrument along with the flow cell. For the 2 Illumina applications, automated cluster generation and paired-end sequencing with index reads of 2 × 250 bp were performed in single 39-hour runs.

ORFans identification. Open reading frames (ORFs) were predicted using Prodigal with default parameters for each of the bacterial genomes. However, the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial sequences were searched against the non-redundant protein sequence (NR) database (59,642,736 sequences, available from NCBI in 2015) using BLASTP. ORFans were identified if their BLASTP E-value was lower than 1e-03 for an alignment length greater than 80 amino acids. We used an E-value of 1e-05 if the alignment length was <80 amino acids. These threshold parameters have been used in previous studies to define ORFans (refs 12–14). The 168 genomes considered in this study are listed in Supplementary Table 7. These genomes represent 615.99 Mb and contain a total of 19,980 ORFans. Some of the ORFans from 30 genomes were calculated in a previous study[4] with the non-redundant protein sequence database containing 14,124,377 sequences available from NCBI in June 2011.

Metagenomic 16S sequences. We collected 325 runs of metagenomic 16S rRNA sequences available in the HMP data sets that correspond to stool samples from healthy human subjects. All samples were submitted to Illumina deep sequencing, resulting in 761,123 Mo per sample on average, and a total of 5,970,465 high-quality sequencing reads after trimming. These trimmed data sets were filtered using CLC Genomics Workbench 7.5, and reads shorter than 100 bp were discarded. We performed an alignment of 247 16S rRNA sequences against the 5,577,630 reads remaining using BLASTN. We used a 1e-03 e-value, 100% coverage and 98.7% cutoff, corresponding to the threshold for defining a species, as previously described. Finally, we reported the total number of aligned reads for each 16S rRNA sequence (Supplementary Table 8).

We collected the sequences of the 3,871,657 gene non-redundant gene catalogue from the 396 human gut microbiome samples (https://www.cbs.dtu.dk/projects/CAG/)[15]. We performed an alignment of 247 16S rRNA sequences against the 3,871,657 gene non-redundant gene catalogue using BLASTN with a threshold of 1e-03 e-value, 100% coverage and 98.7% cutoff. The new species identified in these data are reported in Supplementary Table 9. We collected the raw data sets of 239 runs deposited at EBI (ERP012217)[16]. We used the PEAR software (PMID 24142950) for merging raw Illumina paired-end reads using default parameters. We performed an alignment of 247 16S rRNA sequences against the 265,864,518 merged reads using BLASTN. We used a 1e-03 e-value, 100% coverage and 98.7% cutoff. The list of the new species identified in these data is included in Supplementary Table 9.

Whole metagenomic shotgun sequences. We collected the contigs/scaffolds from the assembly of 148 runs available in the HMP data sets. The initial reads of these samples were assembled using SOAPdenovo v.1.04 (PMID 23587118). These assemblies correspond to stool samples from healthy human subjects and generated 13,984,809 contigs/scaffolds with a minimum length of 200 bp and a maximum length of 371,412 bp. We aligned the 19,980 ORFans found previously against these data sets using BLASTN. We used a 1e-05 e-value, 80% coverage and 80% identity cutoff. Finally, we reported the total number of unique aligned ORFans for each species (Supplementary Table 8).

Study of the gaps in metagenomics. The raw fastq files of paired-end reads from an Illumina Miseq of 84 metagenomes analysed concomitantly by culturomics were filtered and analysed in the following steps (accession no. PRJEB13171).

Data processing: filtering the reads, dereplication and clustering. The paired-end reads of the corresponding raw fastq files were assembled into contigs using Pandaseq[31]. The high-quality sequences were then selected for the next steps of analysis by considering only those sequences that contained both primers (forward and reverse). In the following filtering steps, the sequences containing N were removed. Sequences with length shorter than 200 nt were removed, and sequences longer than 500 nt were trimmed. Both forward and reverse primers were also removed from each of the sequences. An additional filtering step was applied to remove the chimaeric sequences using UCHIME (ref. 32) of USEARCH (ref. 33). The filtering steps were performed using the QIIME pipeline[34]. Strict dereplication (clustering of duplicate sequences) was performed on the filtered sequences, and they were then sorted by decreasing number of abundance[35–37]. For each metagenome, the clustering of OTUs was performed with 97% identity. Total OTUs from the 84 metagenomes (Supplementary Table 10) clustered with 93% identity.

Building reference databases. We downloaded the Silva SSU and LSU database1 and release 123 from the Silva website and, from this, a local database of predicted amplicon sequences was built by extracting the sequences containing both primers. Finally, we had our local reference database containing a total of 536,714 well-annotated sequences separated into two subdatabases according to their gut or non-gut origin. We created four other databases containing 16S rRNA of new species sequences and species isolated by culturomics separated into three groups (human gut, non-human gut, and human not reported in gut). The new species database contains 247 sequences, the human gut species database 374 sequences, the non-human gut species database 256 sequences and the human species not reported in gut database 237 sequences.

Taxonomic assignments. For taxonomic assignments, we applied at least 20 reads per OTU. The OTUs were then searched against each database using BLASTN (ref. 38). The best match of ≥97% identity and 100% coverage for each of the OTUs was extracted from the reference database, and taxonomy was assigned up to the species level. Finally, we counted the number of OTUs assigned to unique species.

References

1. Lagier, J. C., Million, M., Hugon, P., Armougom, F. & Raoult, D. Human gut microbiota: repertoire and variations. *Front. Cell Infect. Microbiol.* **2,** 136 (2012).
2. 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,** 237–264 (2015).
3. Pfleiderer, A. et al. Culturomics identified 11 new bacterial species from a single anorexia nervosa stool sample. *Eur. J. Clin. Microbiol. Infect. Dis.* **32,** 1471–1481 (2013).
4. Lagier, J. C. et al. Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin. Microbiol. Infect.* **18,** 1185–1193 (2012).
5. Dubourg, G. et al. Culturomics and pyrosequencing evidence of the reduction in gut microbiota diversity in patients with broad-spectrum antibiotics. *Int. J. Antimicrob. Agents* **44,** 117–124 (2014).
6. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444,** 1022–1023 (2006).
7. Ley, R. E. et al. Obesity alters gut microbial ecology. *Proc. Natl Acad. Sci. USA* **102,** 11070–11075 (2005).
8. Gill, S. R. et al. Metagenomic analysis of the human distal gut microbiome. *Science* **312,** 1355–1359 (2006).
9. Rinke, C. et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499,** 431–437 (2013).
10. Lagier, J. C. et al. Current and past strategies for bacterial culture in clinical microbiology. *Clin. Microbiol. Rev.* **28,** 208–236 (2015).
11. Vetizou, M. et al. Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. *Science* **350,** 1079–1084 (2015).
12. Cassir, N. et al. *Clostridium butyricum* strains and dysbiosis linked to necrotizing enterocolitis in preterm neonates. *Clin. Infect. Dis.* **61,** 1107–1115.
13. Hugon, P. et al. A comprehensive repertoire of prokaryotic species identified in human beings. *Lancet Infect. Dis.* **15,** 1211–1219 (2015).
14. The Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* **486,** 215–221 (2012).
15. Browne, H. P. et al. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* **533,** 543–546 (2016).
16. Nielsen, H. B. et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32,** 822–828 (2014).
17. Khelaifia, S. et al. Aerobic culture of methanogenic archaea without an external source of hydrogen. *Eur. J. Clin. Microbiol. Infect. Dis.* **35,** 985–991 (2016).
18. Rettedal, E. A., Gumpert, H. & Sommer, M. O. Cultivation-based multiplex phenotyping of human gut microbiota allows targeted recovery of previously uncultured bacteria. *Nat. Commun.* **5,** 4714 (2014).
19. Hiergeist, A., Gläsner, J., Reischl, U. & Gessner, A. Analyses of intestinal microbiota: culture versus sequencing. *ILAR J.* **56,** 228–240 (2015).
20. Rajilic-Stojanovic, M. & de Vos, W. M. The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiol. Rev.* **38,** 996–1047 (2014).
21. Byrd, A. L. & Segre, J. A. Infectious disease. Adapting Koch's postulates. *Science* **351,** 224–226 (2016).
22. Samb-Ba, B. et al. MALDI–TOF identification of the human gut microbiome in people with and without diarrhea in Senegal. *PLoS ONE* **9,** e87419 (2014).
23. Dione, N., Khelaifia, S., La Scola, B., Lagier, J.C. & Raoult D. A quasi-universal medium to break the aerobic/anaerobic bacterial culture dichotomy in clinical microbiology. *Clin. Microbiol. Infect.* **22,** 53–58 (2016).
24. Raoult, D. & Henrissat, B. Are stool samples suitable for studying the link between gut microbiota and obesity? *Eur. J. Epidemiol.* **29,** 307–309 (2014).
25. Khelaifia, S., Raoult, D. & Drancourt, M. A versatile medium for cultivating methanogenic archaea. *PLoS ONE* **8,** e61563 (2013).
26. Khelaifia, S. et al. Draft genome sequence of a human-associated isolate of methanobrevibacter arboriphilicus, the lowest-G+C-content archaeon. *Genome Announc.* **2,** e01181 (2014).
27. Dridi, B., Fardeau, M.-L., Ollivier, B., Raoult, D. & Drancourt, M. *Methanomassiliicoccus luminyensis* gen. nov., sp. nov., a methanogenic archaeon isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* **62,** 1902–1907 (2012).
28. 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. *J. Clin. Microbiol.* **51,** 2182–2194 (2013).
29. 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,** 543–551 (2009).
30. Ramasamy, D. et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. *Int. J. Syst. Evol. Microbiol.* **64,** 384–391 (2014).
31. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: paired-end assembler for Illumina sequences. *BMC Bioinformatics* **13,** 31 (2012).
32. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27,** 2194–2200 (2011).
33. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–2461 (2010).
34. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7,** 335–336 (2010).
35. Stoeck, T. et al. Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. *BMC Biol.* **7,** 72–77 (2009).
36. Mondani, L. et al. *Microbacterium lemovicicum* sp. nov., a bacterium isolated from a natural uranium-rich soil. *Int. J. Syst. Evol. Microbiol.* **63,** 2600–2606 (2013).
37. Boissiere, A. et al. Midgut microbiota of the malaria mosquito vector *Anopheles gambiae* and interactions with *Plasmodium falciparum* infection. *PLoS Pathog.* **8,** e1002742 (2012).
38. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).

Author contributions

D.R. conceived and designed the experiments. J.-C.L., S.K., M.T.A., S.N., N.D., P.H., A.C., F.C., S.I.T., E.H.S., G.Dub., G.Dur., G.M., E.G. A.T., S.B., D.B., N.C., F.B., J.D., M.Ma., D.R., M.B., N.P.M.D.N., N.M.D.B., C.V., D.M., K.D., M.Mi., C.R., J.M.R., B.L.S., P.-E.F. and A.L. performed the experiments. D.M., J.A., E.I.A., F.B., M.Y., A.D., C.S., F.D. and V.V. contributed materials/analysis tools. J.-C.L., A.C., A.L. and D.R. analysed the data. J.-C.L., A.L. and D.R. wrote the manuscript. All authors read and approved the final manuscript.

Additional information

Supplementary information is available for this paper. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.R.

Competing interests

The authors declare no competing financial interests.

256

Article XVIII:

"*Bariatricus massiliensis*" as a new bacterial species from human gut microbiota.

Simon Bessis, **Amadou Togo**, Grégory Dubourg, Didier Raoult and Pierre-Edouard Fournier.

257

257

"Bariatricus massiliensis" as a new bacterial species from human gut microbiota

S. Bessis, T. Amadou, G. Dubourg, D. Raoult and P.-E. Fournier

Aix-Marseille Université, URMITE, UM63, CNRS7278, IRD198, Inserm 1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine, Marseille, France

Abstract

We report here the main phenotypic characteristics of *"Bariatricus massiliensis"* strain AT12 (CSUR P2179), isolated from the stool of a 58-year-old woman who underwent bariatric surgery.
© 2016 The Authors. Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases.

Corresponding author: P.-E. Fournier, Aix-Marseille Université, URMITE, UM63, CNRS7278, IRD198, Inserm 1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine, 27 Boulevard Jean Moulin, 13385, Marseille cedex 05, France
E-mail: pierre-edouard.fournier@univ-amu.fr

In 2015, as part of our culturomics project aiming at identifying all bacterial species within the human microbiota [1,2], we analysed a stool specimen from a 58-year-old-woman who benefitted from bariatric surgery for obesity, with a body mass index of 36.65 kg/m^2 before surgery. The patient provided signed informed consent, and the study was validated by the ethics committee of the Institut Federatif de Recherche IFR48 under number 09-022.

We isolated the strictly anaerobic strain AT12, which could not be identified by matrix-assisted desorption ionization−time of flight mass spectrometry (MALDI-TOF MS; Microflex; Bruker Daltonics, Leipzig, Germany) [3]. The MALDI-TOF MS spectra are available in our database (http://www.mediterraneeinfection.com/article.php?laref=256&titre=urms-database). The initial growth was obtained in anaerobic conditions at 37°C after 10 days of culture on 5% sheep blood−enriched Columbia agar (bioMérieux, Marcy l'Etoile, France). Agar-grown colonies were greyish and ranged in diameter from 0.5 to 1.5 mm. Strain AT12 is a rod-shaped and polymorphic Gram-positive bacillus, ranging in length from 1.25 to 2.5 μm. Strain AT12 is catalase and oxidase negative. The complete 16S rRNA gene was sequenced using a 3130-XL sequencer (Applied Biotechnologies, Villebon sur Yvette, France). Strain AT12 exhibited a sequence identity of 94.37% with *Clostridium nexile* strain ATCC 27757 (GenBank accession no. X73443), its closest phylogenetic neighbour with a validly published name (Fig. 1). This putatively classified strain AT12 within a new genus within the order Clostridiales.

Clostridium nexile strain ATCC 27757 is a strictly anaerobic Gram-positive bacterium initially detected in human faeces in 1974. It was later isolated again from the faecal flora of 20 healthy Japanese Hawaiians [4]. In addition, it has been shown that *Clostridium nexile* possesses a gene encoding a trypsin-dependent lantibiotic ruminococcin A (*Rum*A) [5]. This protein may be an effective bacteriocin against *Clostridium perfringens* [5].

Another bacterium closely related to *"B. massiliensis"* strain AT12 is *Lactonifactor longoviformis,* another member of the order *Clostridiales*. It was isolated for the first time in 2007 from a fresh stool sample from a healthy man. The first isolate was a strictly anaerobic Gram-positive, helically coiled rod [6].

Because of a 16S rRNA difference greater than 5% with its closest phylogenetic neighbour [7], we propose that strain AT12 is the representative strain of a new genus with the order *Clostridiales*, for which we propose the name *"Bariatricus"* gen. nov. after bariatrics, the medical specialty that deals with the causes, prevention and treatment of obesity. Strain AT12^T is the type species of *"Bariatricus massiliensis"* sp. nov.

http://dx.doi.org/10.1016/j.nmni.2016.04.003
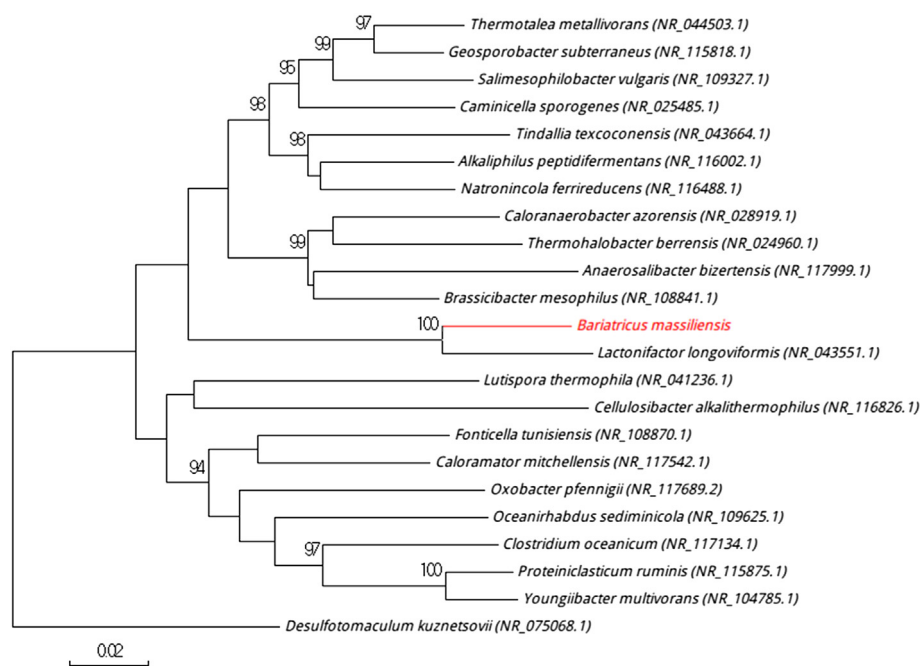
FIG. 1. Phylogenetic tree showing position of *"Bariatricus massiliensis"* strain AT12^T relative to other phylogenetically close members of order Clostridiales. GenBank accession numbers are indicated in parentheses. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using maximum-likelihood method within MEGA software. Numbers at nodes are percentages of bootstrap values obtained by repeating analysis 500 times to generate majority consensus tree. Only bootstraps scores at least 90% were retained. Scale bar = 2% nucleotide sequence divergence.

Nucleotide sequence accession number

The 16S RNA gene sequence was deposited in GenBank under accession number LN898273.

Deposit in a culture collection

Strain AT12^T was deposited in the Collection de Souches de l'Unité des Rickettsies (CSUR; WDCM 875) under number P2179.

Conflict of Interest

None declared.

References

[1] Lagier JC, Hugon P, Khelaifia S, Fournier PE, La Scola B, Raoult D. The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. Clin Microbiol Rev 2015;28:237–64.

[2] Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 2012;18:1185–93.

[3] Seng P, Abat C, Rolain JM, Colson P, Lagier JC, Gouriet F, et al. Identification of rare pathogenic bacteria in a clinical microbiology laboratory: impact of matrix-assisted laser desorption ionization–time of flight mass spectrometry. J Clin Microbiol 2013;51:2182–94.

[4] Moore WEC, Holdeman LV. Human fecal flora: the normal flora of 20 Japanese-Hawaiians. Appl Microbiol 1974;27:961–79.

[5] Marcille F, Gomez A, Joubert P, Ladire M, Veau G, Clara A, et al. Distribution of genes encoding the trypsin-dependent lantibiotic ruminococcin A among bacteria isolated from human fecal microbiota. Appl Environ Microbiol 2002;68:3424–31.

[6] Clavel T, Lippman R, Gavini F, Doré J, Blaut M. *Clostridium saccharogumia* sp. nov. and *Lactonifactor longoviformis* gen. nov., sp. nov., two novel human faecal bacteria involved in the conversion of the dietary phytoestrogen secoisolariciresinol diglucoside. Syst Appl Microbiol 2007;30:16–26.

[7] Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol 2014;64(Pt 2):346–51.

Article XIX:

"*Lachnoclostridium bouchesdurhonense*," a new bacterial species isolated from human gut microbiota.

Togo Amadou, Michel Hosny, Bernard La Scola and Nadim Cassir.

"*Lachnoclostridium bouchesdurhonense,*" a new bacterial species isolated from human gut microbiota

T. Amadou, M. Hosny, B. La Scola and N. Cassir

Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), UM63 CNRS 7278 IRD 198 INSERM U1095, IHU Méditerranée Infection, Pôle des Maladies Infectieuses, Assistance Publique-Hôpitaux de Marseille, Faculté de Médecine, Marseille, France

Abstract

We report the main characteristics of "*Lachnoclostridium bouchesdurhonense*" strain AT5[T] (=CSUR P2181), a new bacterial species isolated from the gut microbiota of an obese patient from Marseille.
© 2016 The Author(s). Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases.

Corresponding author: N. Cassir, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), UM63 CNRS 7278 IRD 198 INSERM U1095, IHU Méditerranée Infection, Pôle des Maladies Infectieuses, Assistance Publique-Hôpitaux de Marseille, Faculté de Médecine, 27 Bd Jean 9 Moulin, 13005 Marseille, France
E-mail: cassirnadim@gmail.com

In order to describe the bacterial flora of the gastrointestinal tract, a stool sample was collected from a 38-year-old obese patient from France. The patient, who provided informed oral consent, had a body mass index of 33 kg/m^2. The study was approved by the ethics committee of the Institut Fédératif de Recherche 48, Marseille. The stool was cultivated using the culturomics approach [1,2]. The culture of strain AT5[T] was achieved on 5% sheep's blood–enriched Columbia agar (bioMérieux, Marcy l'Etoile, France). The colonies obtained were small, circular and smooth, with a mean diameter of 0.5 mm. Bacterial cells were Gram negative and rod shaped. Strain AT5[T] did not exhibit catalase and oxidase activities.

Colonies were not identified by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) using a Microflex spectrometer (Bruker Daltonics, Bremen, Germany) [3,4]. Therefore, we sequenced the 16S rRNA gene of strain AT5[T] using the fD1-rP2 primers as previously described [5] and a 3130-XL sequencer (Applied Biosciences, Saint Aubin, France). The obtained sequence was 94.79% similar to the 16S rRNA gene sequence of *Lachnoclostridium bolteae* strain 16351 (GenBank accession number NR_025567) (Figure 1) [6]. According to the 16S rRNA gene sequence similarity for species demarcation of prokaryotes [7,8], we propose that strain AT5[T] is representative of a new species within the recently described *Lachnoclostridium* genus [9] for which we propose the name "*Lachnoclostridium bouchesdurhonense*" sp. nov. (bouch.du.rhon.ense, L. fem. adj. *bouchedurhonense,* for Bouches-du-Rhône, the department where the city of Marseille is located, where strain AT5[T] was isolated).

MALDI-TOF MS spectrum

The MALDI-TOF MS spectrum of "*Lachnoclostridium bouchesdurhonense*" is available online (http://www.mediterranee-infection.com/article.php?laref=256&titre=urms-database).

Nucleotide sequence accession number

The 16S rRNA gene sequence was deposited in GenBank under accession number LN881614.
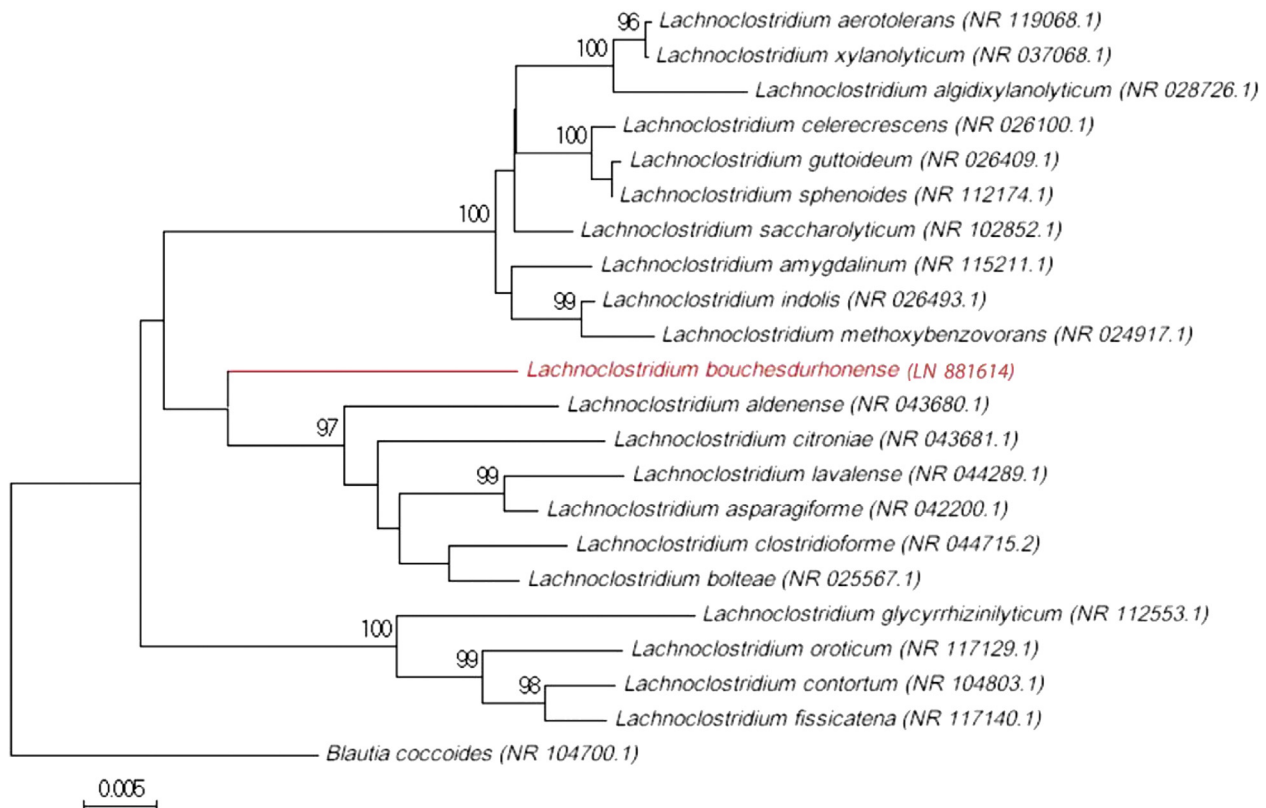
FIG. 1. Phylogenetic tree showing position of "*Lachnoclostridium bouchesdurhonense*" strain AT5[T] relative to other phylogenetically close species with validly published name. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using maximum-likelihood method within MEGA software. Numbers at nodes are percentages of bootstrap values (>95%) obtained by repeating analysis 500 times to generate majority consensus tree. *Blautia coccoides* was used as outgroup. Scale bar indicates 5% nucleotide sequence divergence.

Deposit in a culture collection

Strain AT5[T] was deposited in the Collection de Souches de l'Unité des Rickettsies (CSUR) under number P2181.

Acknowledgement

Conflict of Interest

None declared.

References

[1] Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 2012;18:1185–93.

[2] Lagier JC, Hugon P, Khelaifia S, Fournier PE, La Scola B, Raoult D. The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. Clin Microbiol Rev 2015;28:237–64.

[3] Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 2009;49:543–51.

[4] Seng P, Abat C, Rolain JM, Colson P, Lagier JC, Gouriet F, et al. Identification of rare pathogenic bacteria in a clinical microbiology laboratory: impact of matrix-assisted laser desorption ionization–time of flight mass spectrometry. J Clin Microbiol 2013;51:2182–94.

[5] Drancourt M, Bollet C, Carlioz A, Martelin R, Gayral JP, Raoult D. 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. J Clin Microbiol 2000;38:3623–30.

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

[7] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152.

[8] Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol 2014;64:346–51.

[9] Yutin N, Galperin MY. A genomic update on clostridial phylogeny: gram-negative spore formers and other misplaced clostridia. Environ Microbiol 2013;15:2631–41.

Article XX:

Anaerotruncus rubiinfantis **sp. nov., a bacterium isolated from the stool of a child**
suffering from severe acute malnutrition: Kwashiorkor.

Amadou Hamidou Togo, Maryam Tidjani Alou, Pierre-Edouard Fournier, Didier Raoult,

Grégory Dubourg.

Anaerotruncus rubiinfantis sp. nov, a bacterium isolated from the stool of a child suffering from severe acute malnutrition: Kwashiorkor.

Amadou Hamidou TOGO[1, 2], Maryam TIDJANI ALOU[1], Cheikh Ibrahima LO[2, 3], Matthieu MILLION[1, 2], Jean-Christophe LAGIER[1, 2], Cheikh SOKHNA[2, 4], Pierre-Edouard FOURNIER[2, 3], Didier RAOULT[1, 2] and Grégory DUBOURG[1, 2*]

[1] Aix Marseille Univ, IRD, AP-HM, MEΦI, Marseille, France

[2] IHU-Méditerranée Infection, Marseille, France

[3] Aix Marseille Univ, IRD, AP-HM, SSA, VITROME, Marseille, France

[4] Campus Commun UCAD-IRD of Hann, Dakar, Senegal

*Corresponding author: Grégory Dubourg, Institut Hospitalo-Universitaire Méditerranée Infection, 19-21 Boulevard Jean Moulin 13385 Marseille Cedex 05, France. Phone: + 33 (0) 4 13 73 24 01. Fax: + 33 (0) 4 13 73 24 02.

E-mail address: greg.dubourg@gmail.com

Running title: *Anaerotruncus rubiinfantis* sp. nov.

Abstract

We report here the main characteristics of a new anaerobic bacterium; strain Marseille-P2276, isolated from the stool of child suffering from kwashiorkor. The isolate display 94.07% 16S rRNA gene sequence similarity with *Anaerotruncus colihominis* strain WAL 14565[T]. The draft genome of this new bacterium is 3.41 Mb long, with 55.1 mol% G+C content. Based on phenotypic, phylogenetic and genomic characteristics, we propose the creation of the species *Anaerotruncus rubiinfantis* sp. nov., strain Marseille-P2276 (=CSURP2276[T] =DSM 101766[T]).

Introduction

Microbial culturomics is an approach aiming increase the bacterial repertoire associated to human beings [1] through isolation new bacteria and relies on the multiplication of culture and atmosphere conditions in order to mimic the natural conditions of bacterial growth. When a new bacterium is isolated, it is described according to a polyphasic strategy including phenotypic characteristics, MALDI-TOF MS, phylogenetic and genomic analysis [2]. Microbial culturomics has been used to study the gut microbiota of children with severe acute malnutrition in Senegal and Niger. Here we present the main phenotypic, phylogenetic characteristics and the draft genome sequencing of a new member within the genus *Anaerotruncus*.

Isolation and growth conditions

Growth and isolation were done as previously described [1]. The initial growth was obtained after culture on 5% sheep blood Columbia Agar (bioMérieux, Marcy l'Etoile, France) medium in anaerobic conditions at 37°C following two days of incubation. This bacterium was isolated from the stool samples collected from a child suffering from severe acute malnutrition (Kwashiorkor). The study and the assent procedure were approved by the local ethics committee and the ethics committee of the IFR48 (Faculty of Medicine, Marseille, France) under number (09-022). The initial growth was obtained after culture on 5% sheep blood Columbia Agar (bioMérieux) in anaerobic conditions at 37°C. A screening have been made by MALDI-TOF MS on a Microflex LT spectrometer (Bruker Daltonics, Bremen, Germany) as previously described [3]. The obtained spectra (Figure 1) were imported into MALDI Biotyper 3.0 software (Bruker Daltonics) and analysed against the main spectra of the bacteria included in the Bruker database constantly updated with MEPHI database (http://www.mediterranee-infection.com/article.php?larub=280&titre=urms-database).

Strain identification

In order to classify this bacterium, its 16S rRNA gene was sequenced, as previously described [4]. The 16S rRNA nucleotide sequences were assembled and corrected using CodonCode Aligner software (http://www.codoncode.com). The strain Marseille-P2276 exhibit 94.07% 16S RNA sequence similarity with *Anaerotruncus colihominis* strain WAL 14565[T], (Type strain GenBank accession number AJ315980) [5], the phylogenetically closest species with standing in nomenclature (Figure 2). We consequently classify this strain as a member of a new species within *Ruminococcaceae* family and *Firmicutes* phylum.

Phenotypic characteristics

Phenotypic characteristics of strain Marseille-P2276 were studied as previously described [6,7]. Colonies grown on 5% sheep blood agar (bioMérieux) after 48 hours of incubation were translucent with 1-2 mm of diameter. Cells were Gram-negative staining, non-motile, non-spore-forming and rod-shaped, ranging in width from 0.1 to 0.2µm and in length from 0.5 to 1.2 µm (Figure 3). The strain has neither catalase nor oxidase activities. Growth temperature of strain Marseille-P2276 ranged from 28°C to 45°C, but optimal growth temperature is 37°C in anaerobic condition. Using API ZYM strips (bioMérieux), positive reactions were observed with acid phosphatase, alkaline phosphatase, esterase (C4), lipase (C14), cystine arylamidase, leucine arylamidase, valine arylamidase, naphtol-AS-BI-phosphohydrolase, trypsin, α-chymotrypsin, α-galactosidase, β-galactosidase, β-glucuronidase and α-glucosidase. Using API 20A strip (bioMérieux), positive reactions were observed with; D-glucose, D-saccharose, D-maltose and glycerol. Biochemical characteristics of the strain Marseille-P2276 are summarized in Table 1.

Genome sequencing

The genome of strain Marseille-P2276 was sequenced, assembled and analysed as previously described [8]. The genome of strain Marseille-P2276 is about 3.41 Mb long with a 55.1 mol%

of G+C content and contains 3,388 predicted genes. The degree of genomic similarity of strain Marseille-P2276 with closely related species was estimated using the OrthoANI software [9]. Values among closely related species (Figure 4) ranged from 60.80% between *Faecalibacterium prausnitzii* strain ATC27768[T] and *Fournierella massiliensis* strain AM2[T] to 74.32% between *Anaerotruncus massiliensis* strain AT3[T] and *Fournierella massiliensis* strain AT2[T]. When strain Marseille-P2276 was compared to these closely species, values ranged from 65.05% with *Hydrogenoanaerobacterium saccharovorans* strain CGMCC[T] to 73.25% with *Anaerotruncus massiliensis* strain AT3[T]. We note that the strain Marseille-P2276 shared an average nucleotide identity of 72.46% with *A. colihominis* strain 14565[T], the type species of the genus *Anaerotruncus*.

Conclusion

The new isolate, Marseille-P2276, has a 16S rRNA sequence identity of 94.07% with 5.93% sequence divergence and OrthoANI value of less than 95% (72.46%) with *Anaerotruncus colihominis*, the closest species with standing in nomenclature. Based on these phenotypic, phylogenetic and genomic characteristics, the name *Anaerotruncus rubiinfantis* sp. nov. was proposed. The type strain is Marseille-P2276[T] which isolated from the stool of a child suffering with kwashiorkor.

Deposit in culture collections: Strain Marseille-P2276 was deposited in two different strain collections under number (=CSURP2276; = DSM 101766).

Nucleotide sequence accession number: The 16S rRNA gene and genome sequences of strain Marseille-P2276 were deposited in GenBank under accession number, LN881593 and FKLA00000000 respectively.

The Digital Protologue database Taxon Number for strain Marseille-P2276 is TA00944.

269

References:

1. Lagier J-C, Armougom F, Million M, Hugon P, Pagnier I, Robert C, *et al*. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect. 2012 Dec;18(12):1185–93.

2. Ramasamy D, Mishra AK, Lagier J-C, Padhmanabhan R, Rossi M, Sentausa E, *et al.* A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. International Journal of Systematic and Evolutionary Microbiology. 2014; 64(2):384–91.

3. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier P-E, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis. 2009; 49(4):543–51.

4. Morel A-S, Dubourg G, Prudent E, Edouard S, Gouriet F, Casalta J-P, *et al*. Complementarity between targeted real-time specific PCR and conventional broad-range 16S rDNA PCR in the syndrome-driven diagnosis of infectious diseases. 2015 Mar; 34(3):561–70.

5. Lawson PA, Song Y, Liu C, Molitoris DR, Vaisanen M-L, Collins MD, *et al*. *Anaerotruncus colihominis* gen. nov., sp. nov., from human faeces. Int J Syst Evol Microbiol. 2004; 54(Pt 2):413–7.

6. Tidjani Alou M, Khelaifia S, Michelle C, Andrieu C, Armstrong N, Bittar F, et al. *Anaerococcus rubiinfantis* sp. nov., isolated from the gut microbiota of a Senegalese infant with severe acute malnutrition. Anaerobe. 2016 Aug; 40:85–94.

7. Alou MT, Rathored J, Michelle C, Dubourg G, Andrieu C, Armstrong N, *et al*. *Inediibacterium massiliense* gen. nov., sp. nov., a new bacterial species isolated from the

gut microbiota of a severely malnourished infant. Antonie van Leeuwenhoek. 2017 Jun; 110(6):737–50.

8. Diop A, Khelaifia S, Armstrong N, Labas N, Fournier P-E, Raoult D, et al. Microbial culturomics unravels the halophilic microbiota repertoire of table salt: description of *Gracilibacillus massiliensis* sp. nov. Microb Ecol Health Dis. 2016; 27:32049.

9. 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. 2016 Feb; 66(2):1100–3.

10. Togo AH, Diop A, Dubourg G, Khelaifia S, Richez M, Armstrong N, et al. *Anaerotruncus massiliensis* sp. nov., a succinate-producing bacterium isolated from human stool from an obese patient after bariatric surgery. New Microbes New Infect. 2019 May; 29:100508.

11. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980 Dec; 16(2):111–20.

12. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016; 33(7):1870–4.

Table 1: Differential characteristics of strain Marseille-P2276 (**1**) compared with
Anaerotruncus massiliensis strain AT3[T][10] (**2**) and *Anaerotruncus colihominis* strain WAL
14565[T][5] (**3**) the closest species.

Properties	1	2	3
Cells size (µm)	0.2/1.2	0.3/2.9	0.5/2-5
Genome size (Mb)	3.41	3.14	3.75
G+C content (%)	55.1	63.7	54
Gram-staining	–	–	+
Motility	–	–	–
Spore forming	–	–	–
Esculin hydrolysis	–	+	–
Gelatin hydrolysis	–	+	–
Urease	–	–	–
Indole	–	+	+
Catalase	–	+	–
Acid from:			
Arabinose	–	–	–
Cellobiose	–	–	+
Glucose	+	+	+
Lactose	–	–	–
Maltose	+	V	+
Mannose	–	V	+
Rafinose	–	–	–
Sucrose	+	–	–

Salicin	–	–	–
Trehalose	–	+	–
Alkaline phosphatase	+	+	–
Isolated from	Human feces	Human feces	Human feces

+, positive ; -, negative ; v, variable.
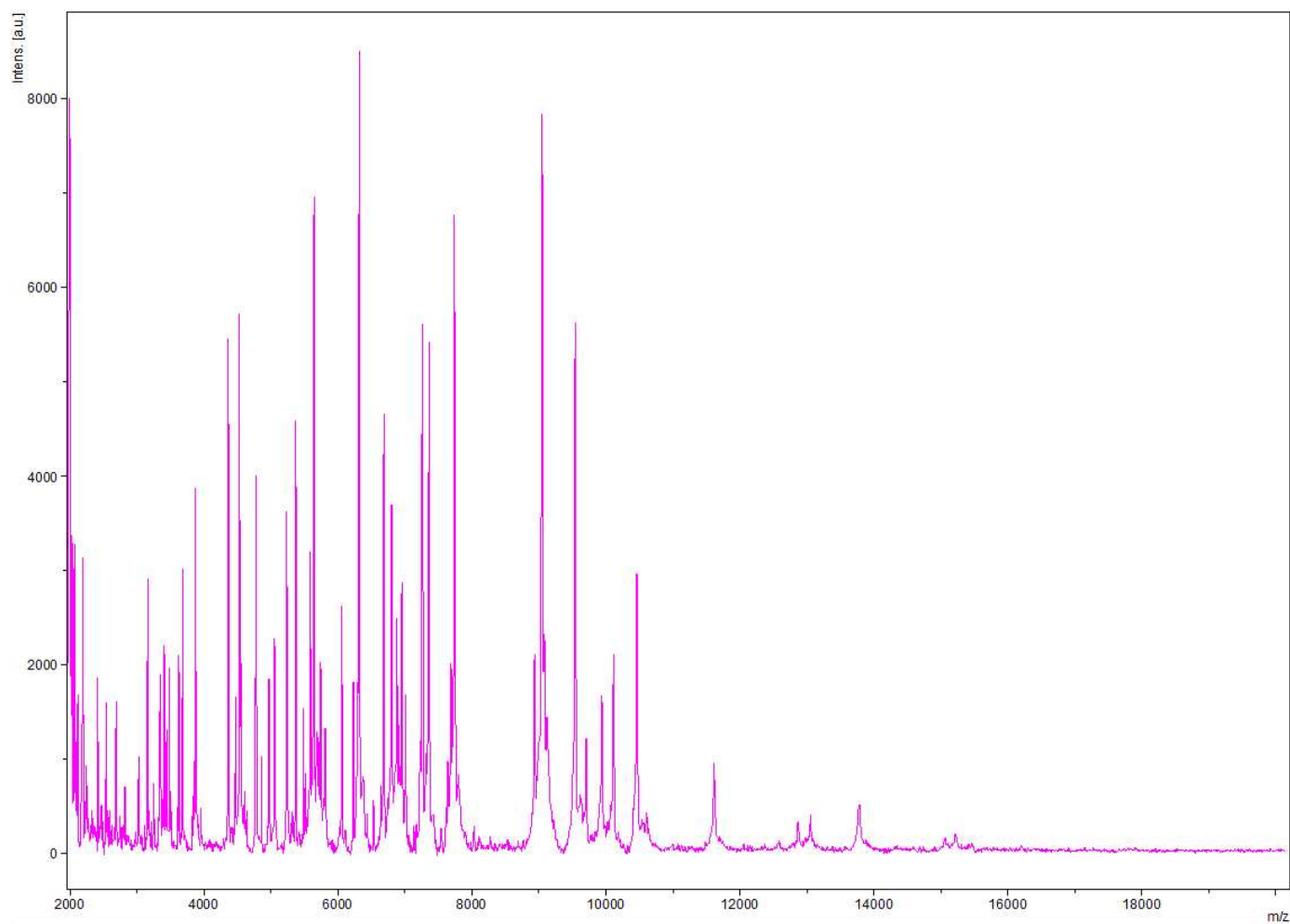
Figure 1: MALDI-TOF MS Reference mass spectrum. Spectra from 12 individual colonies were compared and a reference spectrum was generated.
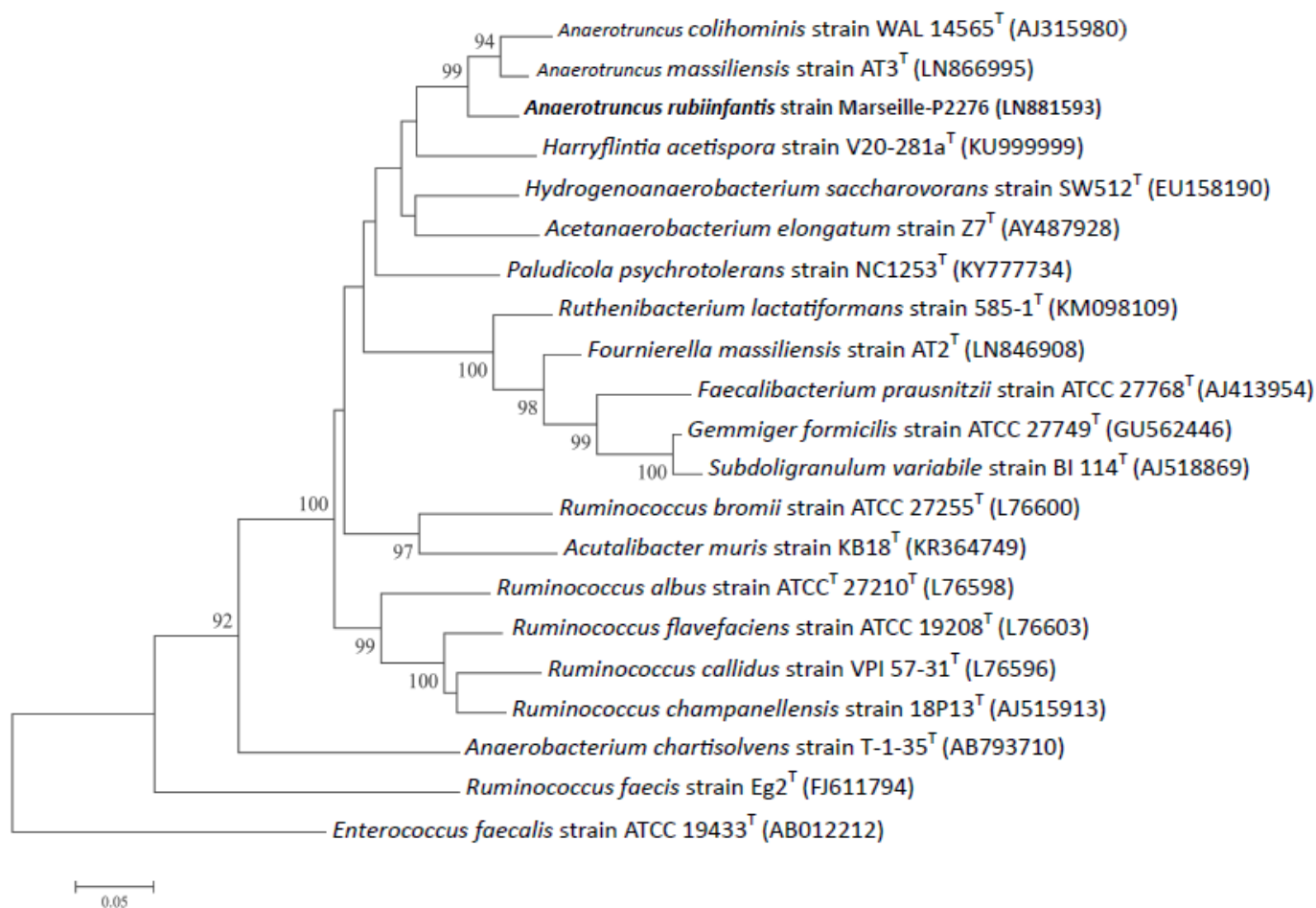
Figure 2: Maximum Likelihood phylogenetic tree highlighting the position of strain Marseille-P2276 with other phylogenetically-close neighbours. The respective GenBank accession numbers for 16S rRNA genes are indicated in parenthesis. The evolutionary history was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model [11]. Evolutionary analyses were conducted in MEGA7 [12]. *Enterococcus faecalis* was used as outgroup. The scale bar indicates a 5% nucleotide sequence divergence between species.
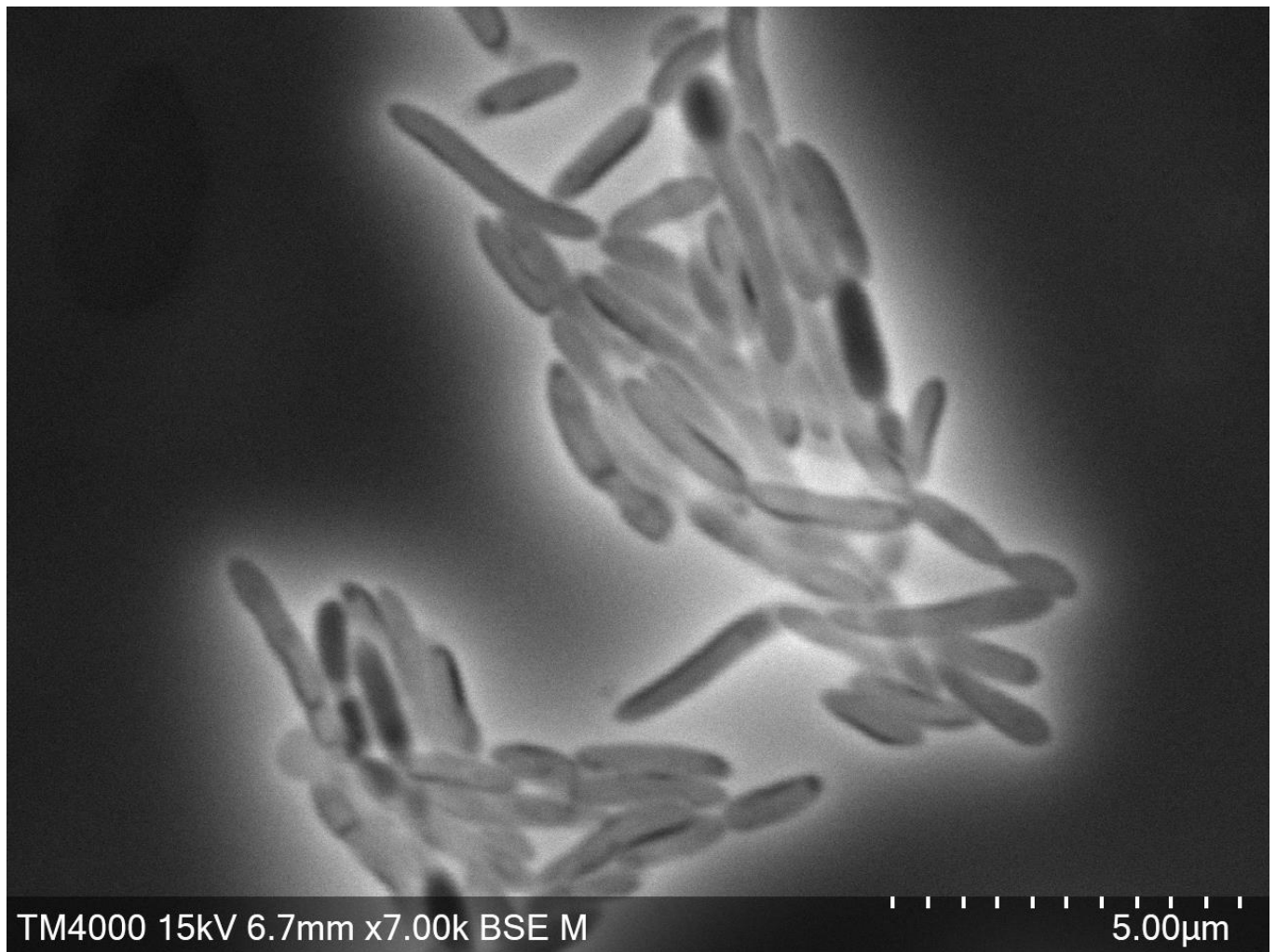
Figure 3: Electron micrograph of strain Marseille-P2276 was acquired with a TM4000.

Colonies were collected from agar and immersed into 500µl of 2.5 % glutaraldehyde fixative

solution.
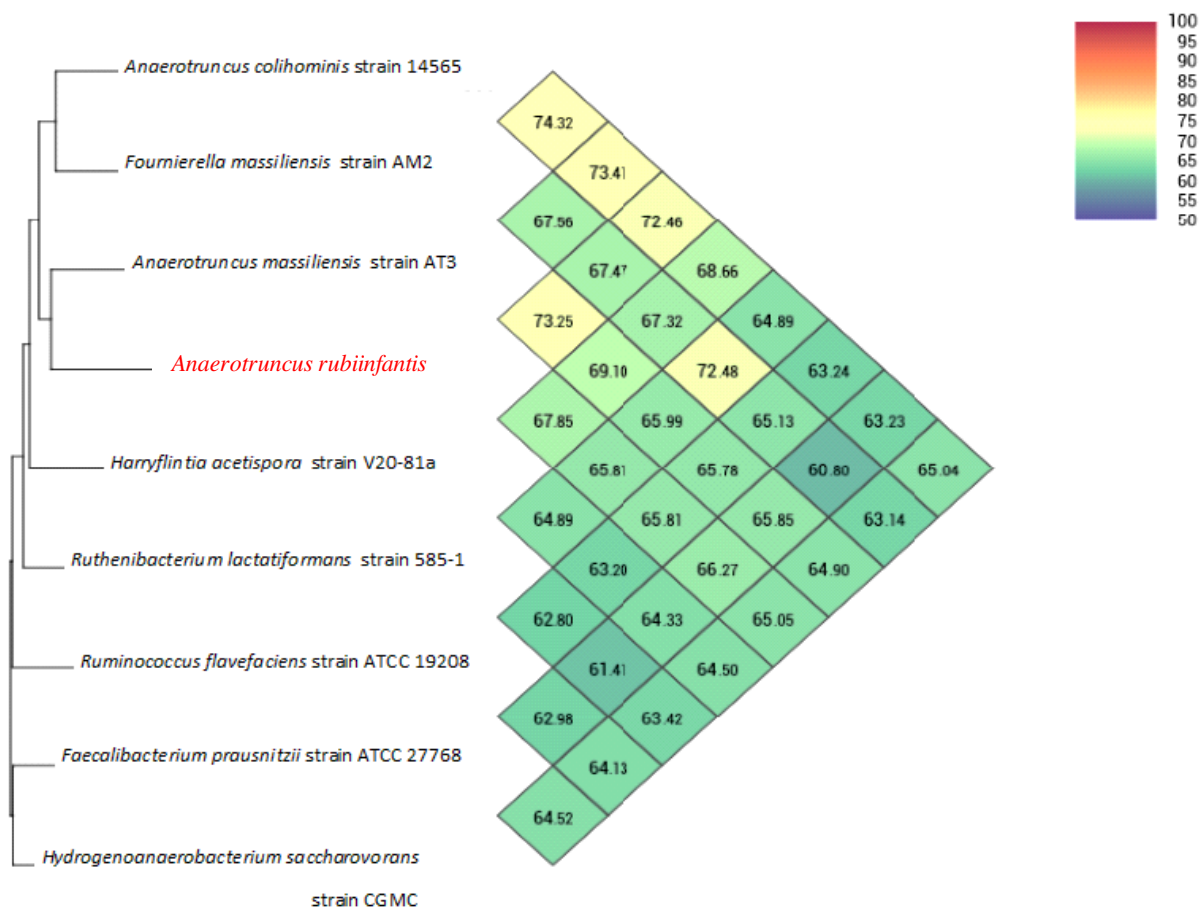
Figure 4: Heatmap generated with OrthoANI values calculated using the OAT software between strain Marseille-P2276 and other closely related species with standing in nomenclature.