UNIVERSITÉ D'AIX-MARSEILLE
FACULTÉ DE MÉDECINE DE MARSEILLE
ECOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

THÈSE

Présentée et publiquement soutenue devant

LA FACULTÉ DE MÉDECINE DE MARSEILLE

Le 24 Octobre 2016

Par Maryam TIDJANI ALOU

Née le 14 Août 1988 à Bordeaux

Étude du microbiote digestif des enfants atteints de malnutrition sévère aiguë

Pour obtenir le garde de DOCTORAT d'AIX-MARSEILLE UNIVERSITÉ
Spécialité Pathologie Humaine

Membres du Jury de la Thèse :

Mr le Professeur Jean-Louis Mège	Président du jury
Mr le Professeur Ogobara Doumbo	Rapporteur
Mr le Professeur Philippe Gérard	Rapporteur
Mr le Professeur Saïdou Mamadou	Examinateur
Mme le Professeur Bouli Ali Diallo	Co-Directrice de thèse
Mr le Professeur Didier Raoult	Directeur de thèse

Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UMR 7278
Directeur : Pr. Didier Raoult

Avant-propos :

Le format de présentation de cette thèse correspond à une recommandation de la spécialité Maladies Infectieuses et Microbiologie, à l'intérieur 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 envoyée dans un journal 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 article publié, accepté ou soumis associé d'un bref commentaire donnant le sens général du travail. Cette forme de présentation a paru 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

Remerciements

Tout d'abord, je tiens à remercier le Professeur Didier Raoult de m'avoir accueillie dans son laboratoire, de m'avoir encadré et guidé tout au long de ce travail de longue haleine et enfin de m'avoir permis de développer en tant que chercheuse.

Je remercie ma co-directrice de thèse le Professeur Bouli Ali Diallo pour son soutien au cours de ce doctorat.

Je remercie le Professeur Jean-Louis Mège d'avoir accepté de présider mon jury de thèse.

Je remercie le Professeur Ogobara Doumbo et le Professeur Philippe Gérard d'avoir accepter de juger ce travail en tant que rapporteurs.

Je remercie le Professeur Saïdou Mamadou d'avoir accepté de participer à mon jury de soutenance.

Je tiens à remercier toute mon équipe de culturomics sans laquelle ce travail n'aurait pas été possible et plus particulièrement :

Jean-Christophe Lagier, Saber Khelaifia, Matthieu Million et Gregory Dubourg de m'avoir encadré et formé.

Aux résidents permanents ou intermittents de la salle MALDI : Camille Gouiran, El Hadji Seck, Niokhor Dione, Zeina Elsawi

A Khoudia Diop, Sokhna Ndongo, Marcel Sy, Isaac Ngom, Elodie Guilhot

Merci à vous les amis pour tous les rires, toutes les larmes, tout le soutien, toute l'entraide, tous les repas partagés. You guys are the best !

Merci à Sory Ibrahima Traoré et Donia Mouelhi pour avoir participé à ce travail.

Et un grand merci à tous les autres membres de l'équipe culturomics pour l'entraide, les gouters et surtout de m'avoir supporté pendant ces trois années.

Je tiens à remercier le Professeur Philippe Parola pour avoir cru en moi, continer à croire en moi et me faire confiance et sans lequel cette aventure ne serait pas arrivée.

Merci au Professeur Pierre-Edouard Fournier pour son aide et sa gentillesse.

Merci à tous les membres de l'URMITE ayant participé de près ou de loin à ce travail.

Et enfin, merci à ma famille :

A mes parents pour m'avoir soutenue dans tout ce que j'ai entrepris même quand ils n'y croyaient pas. Merci pour tout, les mots ne peuvent pas exprimer toute la gratitude et l'amour que je ressens pour vous.

A ma sœur, à mon homme, à ma deuxième famille (les Rabiou et les Amani) et à mes amis pour m'avoir écouté, soutenu et supporté aux cours de ces trois années. Je vous aime fort mais vous n'avez pas fini de me supporter : D

A toute ma famille de m'aimer, de me soutenir dans les mauvais moments et être fière dans les bons.

A ma sœur Karima, je suis sûre que tu veilles sur moi de l'au-delà.

SOMMAIRE

Résumé

Depuis plusieurs années, il s'avère de plus en plus clair que le microbiote digestif a un impact remarquable sur la santé humaine. Il est affecté par de nombreux facteurs dont l'alimentation. En effet, en fonction du macronutriment majoritaire d'un régime alimentaire, certaines populations et fonctions bactériennes sont stimulées ou inhibées. Plusieurs pathologies de l'intestin ou liées à des troubles nutritionnels ou métaboliques ont un lien causal avec une altération du microbiote digestif parmi lesquelles la malnutrition sévère aiguë. En effet, il a été récemment montré que le microbiote digestif des enfants malnutris était différent et colonisé par des *Proteobacteria*, des *Enterococci*, des bacilles Gram-négatifs et des espèces pathogènes. Au cours de nos travaux, une dysbiose est également observée chez nos patients malnutris par métagénomique et par culturomics avec un enrichissement en bactéries aérobies, en *Proteobacteria* et en espèces potentiellement pathogènes telles que *Streptococcus gallolyticus* et une perte notable en bactéries anaérobies associée à une perte de la capacité antioxydante du tractus gastro-intestinal révélée par une absence totale de *Methanobrevibacter smitii*, archeae méthanogène et un des procaryotes les plus sensibles à l'oxygène du tractus gastro-intestinal ainsi que un potentiel

redox fécal accru. De plus, une perte de la diversité globale, connue et inconnue, est observée. Enfin, par culturomics et métagénomique, nous avons établi un répertoire des bactéries manquantes chez les malnutris dont treize présentent un potentiel probiotique et pourront être testées comme probiotiques dans un modèle expérimental dans un futur proche.

Mots-clés : microbiote digestif ; malnutrition sévère aiguë ; kwashiorkor ; culturomics ; métagénomique ; probiotiques.

Abstract

For the last decade, it has become increasingly clear that the gut microbiota has a tremendous impact on human health. It is affected by several factors among which diet that has a big impact. In fact, according to the major macronutrient in a diet type, specific bacterial populations and functions are stimulated or inhibited. Several pathologies of the gut or linked to nutritional or metabolic disorders among which severe acute malnutrition are causally linked to an alteration of the diversity of the human gut microbiota. In fact, it has recently been shown by several studies that the gut microbiota of malnourished patients was different and colonized by *Proteobacteria*, *Enterococci*, Gram-negative bacilli and pathogenic species.

The analysis of our data regarding the fecal microbiota of children afflicted with severe acute malnutrition from Niger and Senegal showed a dysbiosis observed using metagenomics and culturomics with an increase of aerobic bacteria, *Proteobacteria* and pathogenic species such as *Streptococcus gallolyticus,* and a depletion of anaerobic species associated with a loss of the antioxidant capacity of the gastro-intestinal tract exhibited by a total absence of *Methanobrevibacter smithii*, a methanogenic archaeon and one the most oxygen sensitive prokaryote of the gut microbiota alongside an increased fecal redox potential.

Moreover, a loss of the overall diversity, known and unknown, was observed. Finally, through culturomics and metagenomics, we were able to identify a repertoire of missing microbes in malnourished children among which thirteen presented a probiotic potential and will be tested as such in an experimental model in the near future.

Keywords: gut microbiota; severe acute malnutrition; kwashiorkor; culturomics; metagenomics; probiotics.

Introduction

De la première description de microorganismes dans le tractus gastro-intestinal en 1681 à nos jours, les connaissances sur le microbiote digestif ont exponentiellement augmenté avec l'évolution des méthodes d'exploration des écosystèmes microbiens (1). Ainsi, de nombreuses études ont décrit le microbiote digestif humain qui est composé de bactéries, d'archées, de virus, de champignons et de protozoaires (2). Les bactéries représentent la partie la plus étudiée du microbiote humain avec plus de 10^{14} cellules bactériennes. Ce dernier compte dix fois plus de cellules bactériennes que de cellules humaines ainsi que 150 fois plus de gènes bactériens que de gènes humains (3). Afin de décrire le microbiote digestif, des méthodes basées sur la culture étaient principalement utilisées jusqu'à la décennie précédente qui a vu l'avènement des techniques de séquençage de nouvelle génération et à la vulgarisation des techniques de biologie moléculaire telles que l'hybridation in situ en fluorescence et la PCR en temps réel (2). Cependant, malgré la prédominance d'études métagénomiques, un retour à la culture s'impose et est observée avec "microbial culturomics", technique de culture à haut débit qui consiste en la multiplication des conditions de culture en variant le milieu, la température, l'atmosphère, le pH, la salinité et autres paramètres

physico-chimiques combinée à une identification par MALDI-TOF MS (4,5). Ainsi, le répertoire du microbiote digestif a été amplement élargi avec l'isolement de bactéries jusqu'alors inconnues ou considérées comme incultivables.

La colonisation du tractus digestif commence à la naissance avec une colonisation initiale par des bactéries des microbiotes vaginal et fécal de la mère en cas d'accouchement par voie basse ; en cas d'accouchement par césarienne, les premières bactéries à s'établir appartiennent au microbiote cutané et l'environnement (6,7). Sept phyla bactériens (*Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Proteobacteria, Fusobacteria, Verrucomicrobia* et *Cyanobacteria*-like) ainsi qu'un phylum d'archées (*Euryarchaeota*) représentent la majorité du microbiote digestif mature (8). Plusieurs facteurs tels que l'alimentation, l'environnement, la génétique, la physiologie du tractus gastro-intestinal influencent la composition du microbiote digestif de sorte qu'il existe une grande variabilité inter- et intra-individuelle (8,9). Cependant, le projet MetaHIT a déterminé l'existence de trois entérotypes à travers la population mondiale : l'entérotype1 caractérisé par une forte prévalence en *Bacteroides*, l'entérotype 2 caractérisé par une forte prévalence de *Prevotella* et l'enterotype 3 caractérisé par une forte prévalence en *Ruminococcus* (6,10). Un

autre système de classification, basé sur des analyses métagénomiques, retrouve deux types d'individus dans la population globale, des individus à nombre élevé de gènes et des individus à nombre faible de gènes (11).

Plusieurs projets de large envergure tels que le Human Microbiome Project (12) et le Metagenomics of the Human Intestinal Tract (MetaHIT) (13) ont pour objectif d'étudier de la manière la plus exhaustive possible le microbiote digestif en fonction des nombreux facteurs capables de l'influencer. De ce fait, le rôle majeur du microbiote digestif dans la santé a été mis en évidence de même que des liens entre une altération du microbiote digestif et plusieurs maladies (7). En effet, le microbiote digestif a un rôle avéré dans la polypose, le cancer colorectal, les maladies inflammatoires chroniques de l'intestin, l'entérocolite nécrosante, le diabète ainsi que dans les troubles nutritionnels tels que l'obésité, la malnutrition sévère aiguë, l'anorexie (2,6,14,15).

Plusieurs hypothèses ont été émises au fil des années quant à l'étiologie la malnutrition sévère aiguë (16). Cette pathologie est la cause majeure de mortalité infantile dans le monde avec plus de 20 millions d'enfants affectés et 1 à 6 millions de décès chaque année. Elle est principalement

prévalente en Afrique subsaharienne, en Asie du Sud-est et en Amérique centrale qui sont des zones à insécurité alimentaire récurrente (17). La malnutrition sévère aiguë est diagnostiquée en utilisant des critères anthropométriques tels que l'indice poids-taille inférieur qui doit être inférieur à -3 et la mesure du périmètre brachial qui doit être inférieur à 115 mm d'après les recommandations de l'OMS (18). On en distingue une forme œdémateuse et une forme non-œdémateuse. La malnutrition œdémateuse ou kwashiorkor a été décrite pour la première fois en 1933 par Cicely D. Williams comme affectant principalement des enfants incorrectement sevrés âgés de 1 à 4 ans (19). En plus de l'œdème, d'autres symptômes sont associés au kwashiorkor : des éruptions cutanées, un dépérissement extrême, une stéatose hépatique ainsi qu'une irritabilité des patients. La forme non-œdémateuse, décrite plus tard, en 1967, a été appelée marasme et est caractérisée par un dépérissement extrême des patients entrainé par une perte de la graisse sous-cutanée et une atrophie musculaire (20). Ces deux formes de malnutrition sévère aiguë sont dans certains cas, retrouvées chez certains patients en une forme combinée appelée kwashiorkor marasmique.

Ces pathologies sont communément attribuées à une carence quantitative en protéines dans le cas du marasme et qualitative dans le cas du kwashiorkor. Cependant, ce dernier

n’est pas entièrement expliqué par la carence en protéines. Parmi les nombreuses hypothèses quant à l’étiologie du kwashiorkor, la plus récente suggère une dysbiose du microbiote digestif à l’origine de cette pathologie (16). En effet, une étude de l’équipe de Jeffrey Gordon a montré la possibilité d’une discordance pour le kwashiorkor chez des jumeaux ainsi que la transmission du phénotype du kwashiorkor par la greffe du microbiote fécal chez des souris gnotobiotiques (21). Ainsi, le microbiote fécal serait instrumental dans la pathogénèse de la malnutrition sévère aiguë.

Afin d’explorer cette hypothèse, une revue de la littérature sur l’influence de l’alimentation sur la composition du microbiote digestif ainsi que sur les dysbioses liées à différents types de la malnutrition est nécessaire afin de déterminer les mécanismes selon lesquels les dysbioses peuvent survenir (Partie I). Dans le but de préciser l’altération du microbiote digestif associé à la malnutrition sévère aiguë, nous avons mené une vaste étude cas-contrôles afin de caractériser l’environnement digestif des enfants sévèrement malnutris ainsi que leur microbiote par approche de métagénomique, de PCR en temps réel et mesure de paramètres physico-chimiques (Partie II). Le retour aux approches de culture pour explorer la composition du microbiote digestif s’observe de plus en plus. En

effet, avoir des cultures pures permet de déterminer les capacités fonctionnelles des souches individuellement et en interaction les unes avec les autres. “Microbial culturomics” représente l’approche de culture idéale puisque son efficacité pour étudier le microbiote digestif a été prouvée par de nombreuses études et qu’elle est complémentaire de la métagénomique avec un degré de recouvrement assez faible entre les deux méthodes (4,5,22–24). De plus, culturomics a permis une extension significative du répertoire du microbiote digestif (Partie III).

Nous avons donc associé la métagénomique à “microbial culturomics” afin de caractériser le microbiote digestif des enfants atteints de kwashiorkor par rapport à ceux d’enfants sains et d’identifier les microorganismes manquants chez les malnutris (Partie IV). Au cours de ce travail par culturomics, nous avons élargi le répertoire du microbiote digestif en isolant 45 espèces non précédemment connues dans cet écosystème parmi lesquelles 18 nouvelles espèces et 27 espèces décrites mais non isolées auparavant dans le tractus gastro-intestinal chez les patients atteints de kwashiorkor. De plus, chez les enfants sains, 46 espèces non précédemment connues du microbiote digestif ont été isolées dont 26 nouvelles espèces et 20 espèces décrites mais non précédemment connues du microbiote digestif. Ces nouvelles espèces ont été décrites en

utilisant le concept de taxonogenomics (Partie V). Ce concept associe les technologies de séquençage de nouvelle génération avec le séquençage complet du génome de la bactérie associé à la description phénotypique et protéomique avec le profil protéique de l'espèce obtenue par MALDI-TOF MS.

Partie I : Influence du régime alimentaire sur le microbiote digestif et dysbioses associées aux désordres nutritionnels.

Avant-propos

Cette première partie de mon travail de doctorat montre tout d'abord les méthodes d'études passées et actuelles du microbiote digestif ainsi que la composition d'un microbiote fécal sain. La composition du microbiote digestif varie selon les individus en fonction de différents facteurs dont la zone géographique à différents niveaux comme par exemple le continent, le pays, la zone métropolitaine ou rurale. Ces variations sont principalement dues aux variations de régime alimentaire et donc du nutriment majoritaire en fonction des zones géographique. Nous présentons ici comment le microbiote digestif évolue en fonction de l'apport majeur en un type de nutriment d'un régime alimentaire donné. Des populations bactériennes spécifiques ainsi que les fonctions associées sont stimulées en fonction du macronutriment majoritaire (carbohydrates, lipides ou protéines). Ces modifications du microbiote digestif se retrouvent dans les désordres nutritionnels en fonction de leur pathogénèse et sont ici répertoriés pour l'obésité, la malnutrition sévère aiguë et l'anorexie. La possibilité de modifier le microbiote digestif par des apports nutritionnels ouvre la voie à des études interventionnelles plutôt qu'observationnelles pour manipuler la composition du microbiote digestif de manière non-invasive.

Article 1: Diet Influence on the Gut Microbiota and Dysbiosis related to Nutritional Disorders.

Maryam Tidjani Alou, Jean-Christophe Lagier and Didier Raoult

Publié dans Human Microbiome Journal

Human Microbiome Journal 1 (2016) 3–11

Contents lists available at ScienceDirect

Human Microbiome Journal

journal homepage: www.elsevier.com/locate/humic

Diet influence on the gut microbiota and dysbiosis related to nutritional disorders

Maryam Tidjani Alou [a], Jean-Christophe Lagier [a,b], Didier Raoult [a,c,*]

[a] Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes IRD 198, CNRS 7278, Aix-Marseille Université, Marseille, France
[b] Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes IRD 198, CNRS 7278, Campus Commun UCAD-IRD of Hann, Dakar, Senegal
[c] Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 11 September 2016
Accepted 21 September 2016
Available online 22 September 2016

Keywords:
Gut microbiota
Diet
Dysbiosis
Nutritional disorders

ABSTRACT

Studies concerning the gut microbiota have exponentially increased since the 1970s. A healthy gut microbiota is essential for growth and weight gain in infants as well as for a thorough harvest of energy from diet through a role in digestion. Study techniques include culture-independent and culture-dependent methods aiming at describing the gut microbiota taxonomically and functionally. Healthy gut microbiota plays a role in digestion by metabolizing indigestible macronutrients resulting in short chain fatty acids and other bioactive compounds. Diet was proven to influence the composition of the gut microbiota with specific changes to the major macronutrient contained in the diet. Since diet has an influence on gut microbiota's composition, nutritional disorders such as obesity, severe acute malnutrition and anorexia nervosa are linked to an alteration of the gut microbiota mirroring the physiopathology of the nutritional disorder. These alterations should be the target of future therapeutic interventions in nutritional disorders.

Contents

Abbreviations: BMI, Body Mass Index; FISH, Fluorescent *in situ* Hybridization; FOS, fructo-oligosaccharides; GCPR, G Coupled Protein Receptors; HAZ, height-for-age z-score; HGC, high gene count; HMOs, Human Milk Oligosaccharides; LAB, Lactic Acid Bacteria; LGC, low gene count; SAM, severe acute malnutrition; SCFA, short chain fatty acids; SD, standard deviation; WAZ, weight-for-age z-score; WHZ, weight-for-height z-score; WHO, World Health Organization.

* Corresponding author at: Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes IRD 198, CNRS 7278, Aix-Marseille Université, Marseille, France.
E-mail addresses: tidjani_maryam@hotmail.com (M. Tidjani Alou), jclagier@yahoo.fr (J.-C. Lagier), didier.raoult@gmail.com (D. Raoult).

http://dx.doi.org/10.1016/j.humic.2016.09.001
2452-2317/

Introduction

Over the last decades, interest for the human microbiota and more specifically the gut microbiota has strikingly increased. This diverse ecosystem consists mostly in bacteria alongside archaea, fungi and parasites [1] and is described as the biggest endocrine organ of the human body as it is responsible for the synthesis of several hormones. In fact, it contains ten times more cells than the human body and 150 times more genes than the human genome [2]. The gut microbiota appears to be essential for health and largely implicated in the homeostasis of the human body. A major role of the gut microbiota resides in its support of the digestion all through the gastro-intestinal tract. As such, the gut microbiota is important for energy balance. Many factors such as host genetics, environment, age and diet play a role when it comes to shaping the gut microbiota [3,4]. In this review of the literature, we focus on the diet impact on the healthy gut microbiota and how this microbiota can be affected by nutritional disorders.

Healthy gut microbiota

Exploration techniques

Over the years, the methods of choice to study the gut microbiota have evolved alongside technology. In the early years of gut microbiota studies, culture-based methods were used [5–7]. Because of the anaerobic nature of most of the bacteria in the human gut, culture-based techniques uncovered only 10–25% of the diversity in the gut microbiota [7,8] as proven in the recent years by molecular methods. Nowadays, a majority of studies exploring the gut microbiota diversity are based on 16S rRNA based sequencing and whole genome sequencing of metagenomes [9]. Nevertheless, as a shift of the interest in the gut microbiota is occurring from description to functional capacities, more functional metagenomics and metabolomics studies are being performed on the gut microbiota [9,10]. It should be noted that functional studies of the gut microbiota did not begin with metagenomics studies but with studies of the biochemical activities of the gut microbiota [6]. Other molecular methods to study the gut microbiota include quantitative PCR, reverse transcription quantitative PCR and Fluorescent *in situ* Hybridization (FISH) which in addition to identification, also allow quantification of bacterial cell numbers [2,11,12]. Reproducibility rate between molecular studies is very low, most probably because of sampling, conservation of the samples and DNA extraction methods [2]. Molecular techniques also present a depth bias, detecting species at a minimal concentration of 10^5 bacteria/g. Conversely, the "microbial culturomics" concept allows the detection of species present in low concentrations in addition to the tremendous advantage of having physical strains; the return to culture methods such as culturomics, highly complementary to metagenomics appears to be essential for a thorough exploration of the gut microbiota diversity [7,13,14]. Feces are the most used samples when studying the gut microbiota because of their easy availability. Nevertheless, they do not represent accurately the entire gut microbiota diversity since it varies taxonomically and functionally in each anatomical part of the gastro-intestinal tract [15–17]. This is a factor to be considered when designing studies and interpreting data.

Composition of the healthy gut microbiota

Healthy gut microbiota composition is subject to variation due to age, genetics, environment, diet and gut wall structure [2,12]. Gut colonization starts at birth with a composition depending on the type of delivery and subsequently on the type of diet [18,19].

Infant gut microbiota

Colonization is commonly believed to start at birth with the mother's vaginal microbiota for vaginally delivered newborns. When delivered using C-section, the newborn's gut microbiota is first colonized by the mother's skin microbiota and the environment (Fig. 1) [19,20]. The composition of the gut microbiota of infants evolves rapidly during the first year of life at which point the majority of species belong to the *Clostridium coccoides* group, the *Clostridium botulinum* group, the *Bacteroides* and *Veillonella* genera, the *Verrucomicrobia* phylum represented by *Akkermansia muciniphila* (Fig. 2) [19]. Great inter-individual and even greater intra-individual variations are observed in the diversity of children gut microbiota [21].

Adult gut microbiota

Human gut microbiota presents a low diversity at the phylum level with only one archaeal phylum, *Euryarchaeota* and seven bacterial phyla represented: *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Proteobacteria*, *Fusobacteria*, *Verrucomicrobia* and *Cyanobacteria*-like (Fig. 2) [2,8]. Healthy adult gut microbiota are dominated by three bacterial phyla, *Firmicutes* namely *Lachnospiraceae* and *Ruminococcaceae* families, *Bacteroidetes* namely *Bacteroidaceae*, *Prevotellaceae* and *Rikenellaceae* families and *Actinobacteria* namely *Bifidobacteriaceae* and *Coriobacteriaceae* families (Fig. 2), and one major methanogenic archaeon, *Methanobrevibacter smithii* [2,8,22]. The MetaHIT Consortium suggests a classification of the gut flora into three distinct enterotypes: Enterotype 1 presents a high abundance of *Bacteroides* and a wide saccharolytic potential, Enterotype 2 presents a high abundance of *Prevotella* and a high potential for mucin glycoprotein degradation and Enterotype 3 which presents a high abundance of *Ruminococcus* and potential for mucin degradation and membrane transport of sugars [8]. A more recent classification system based on metagenomics analysis divided the population into carriers of two types of diversity within the gut microbiota: high gene count (HGC) carriers and low gene count (LGC) carriers [12,23]. These classifications systems have been criticized because of the large interpersonal variability in the composition of the gut microbiota at the species level. However, a core group of species seems to be present or absent in each individual alongside a variable microbiome which composition depends on diet and environment [3].

Role of the human gut microbiota in digestion

Composition of the gut microbiota varies through each anatomic site depending on temperature, pH, redox potential, oxygen

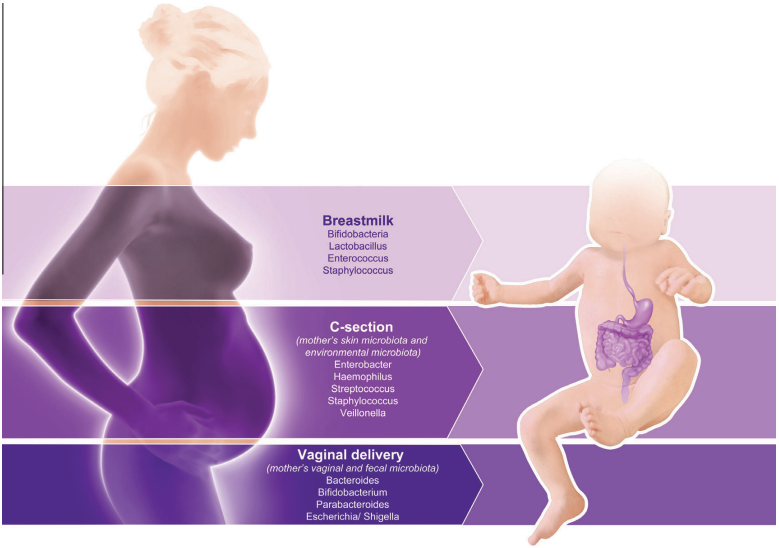


Fig. 1. Early colonization of the gastro-intestinal tract in newborns. Early colonizers of the gastrointestinal tract are specific to the type of delivery. In the case of a vaginal delivery, early colonizers originate from the mother's vaginal and fecal microbiota whereas for C-sections, early colonizers belong to the environment of birth and the mother's skin microbiota. Breast milk microbiota also colonizes the gastrointestinal of newborns.

tension, water activity, salinity and light [2]. The composition of the gut microbiota is also depending on the functional role of each region of the gut microbiota in digestion. The gut microbiota intervenes mainly in the colon where no digestive enzymes are secreted to metabolize macronutrients not digested in the ileum [2,8]. These indigestible macronutrients consist mainly of oligo- and polysaccharides which fermentation by commensal bacteria of the colon results in the synthesis of short chain fatty acids (SCFA), and phenolic compounds which metabolism produces bioactive compounds [8]. Saccharolytic species include species belonging to the *Bacteroides*, *Bifidobacterium*, *Clostridium*, *Eubacterium*, *Lactobacillus* and *Ruminococcus* genera [2]. Fermentation of proteins by the gut microbiota takes also place in the colon via bacterial proteinase and peptidase thanks to species such as *Clostridia*, *Propionibacterium* spp., *Prevotella* spp., *Bifidobacterium* spp. and *Bacteroides* spp. [2,8].

Influence of diet in the gut microbiota composition

Macronutrients related effect

Several studies have explored the link between diet and the gut microbiota because of the potential of dietary interventions to shape the composition of the gut microbiota. Each type of macronutrients (proteins, dietary fibers, fat) influences the gut microbiota specifically (Fig. 3). Changes are observed more at a metabolic level than at a taxonomic level with a quick change in gene expression depending on the macronutrients [12,23,24]. Nevertheless, transient changes can be observed in the diversity of the gut microbiota associated with each macronutrient. These changes affect only specific species which metabolic activity can be affected by the investigated macronutrient [25]. For example, dietary fiber consumption leads to an increase in butyrate-producing species which ferment these fibers in the distal colon (*Roseburia*, *Blautia*, *Eubacterium rectale*, *Faecalibacterium prausnitzii*), in the *Actinobacteria* phylum (*Bifidobacteria*, *Lactobacilli*) and variations in *Bacteroidetes* proportion depending on the type of dietary fiber [2,12,26,27]. A high protein diet is usually a low carbohydrate diet; as such this type of diet stimulates a decrease in butyrate producing species and an increase of species with proteolytic activities such as *Bacteroides* spp [2,12]. Dietary fat has an indirect impact on the gut microbiota diversity: a high fat diet stimulates the production of bile acids which in turn select the growth species with the ability to metabolize biles acids and/or induce the loss of some species due to the antimicrobial activity of bile acids [2,12].

Influence of the type of diet on the gut microbiota composition

Depending on the major macronutrient in a type of diet, it has been highlighted that specific categories of species are stimulated in the gut microbiota (Fig. 3). Vegetarian and vegan diets tend to have a high carbohydrate content associated to a lower protein and fat content. Omnivorous and animal based diets conversely show a high protein and fat content and low carbohydrate content [10]. The latter diets are associated with an increase of bile tolerant bacteria such as *Bacteroides*, *Alistipes* and *Bilophila* [2,23], butyrate producing bacteria, specifically the *Clostridium* cluster XVIa [12]. Since vegetarian and vegan diets have a high carbohydrate content, their gut microbiota are dominated with bacteria with high carbohydrate fermenting bacteria such as the *Prevotella* [23], *Clostridium clostridioforme* and *Faecalibacterium prausnitzii* [12]. Vegetarians

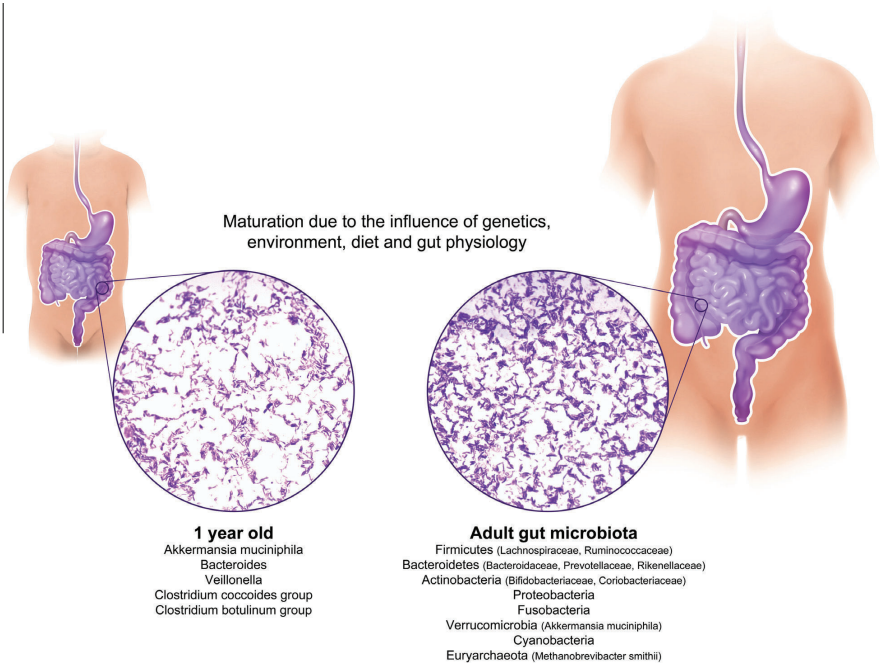


Fig. 2. Maturation of the gut microbiota. The composition of the gut microbiota in children varies greatly during the first year of life before reaching a more stable and mature composition. Maturation into an adult state is due to several factors such as genetics, environment, gut physiology and diet.

show specifically compared to omnivores an increase in the *Clostridium* cluster XVIII (*Lachnospiraceae* and *Clostridium ramosum* group). On the other hand, vegan show a decrease in *Bacteroides*, *Bifidobacteria*, *Enterobacteriaceae* species [8].

Geographic locations such as country, continent or metropolitan areas as opposed to rural areas seem to have an influence on the gut microbiota diversity but this influence to the variation in diet associated with geographic location [21]. Western diets usually present high amounts of fat, processed carbohydrates and low amounts of fibers; as such, they are associated with lower global diversity in the gut microbiota, an increase of *Bacteroides* and a decrease *Prevotella* compared to non-western diets [2]. These diets, especially in rural areas, are characterized by high fiber and complex carbohydrates content and are therefore associated with an abundance of high polysaccharides fermenting bacteria such as *Prevotella, Succinivibrio, and Treponema*. The *Prevotella* genus seems to be a discriminatory taxon between high and low carbohydrate diets and moreover, rural and metropolitan diets, in the gut microbiota of children and adults [8,11,12,21]. The difference between western and non-western diets is mainly due to the presence of processed food in western diet [8].

Bifidobacteria are very abundant in breastfed infants' gut microbiota due to the presence of human milk oligosaccharides (HMOs) which stimulates their growth [2,28] whereas in adults, they seem to be associated with the consumption of agro-pastoral derived products [12].

Influence of prebiotics, probiotics on the gut microbiota

Probiotics are bacterial species with a recognized beneficial effect for health when ingested viable in adequate amounts by individuals. These probiotics are often associated with prebiotics which are non-digestible carbohydrates which metabolization stimulates specific species among which probiotics. The combination of probiotic and prebiotic is called a synbiotic [8,29]. Prebiotics can be used to improve the growth of specific species with probiotics characteristics, modifying the gut microbiota composition in the process and restoring health as a consequence. In fact, the fermentation products of prebiotics are usually SCFA [8] which main beneficial effects for the health include anti-inflammatory and anti-apoptotic activities and prevention of colorectal cancer and colitis [23].

In breastfed infants, prebiotics are represented by HMOs from the maternal milk which promote colonization by *Bifidobacterium* species. This effect is also obtained with galacto-oligosaccharides and fructo-oligosaccharides (FOS) supplementation in infant formula. In turn, *Bifidobacterium* species produce SCFA by fermenting prebiotics and participate in the stimulation of the immune system [2]. As for the adults, classical prebiotics include FOS, inulin, galacto-oligosaccharides and lactulose, all naturally present in vegetables such as artichokes, onions, chicory, garlic and leeks [2]. An increase of abundance in *Latobacilli*, *Bacteroides*, *Lachnospiraceae* and *F. prausnitzii* has been observed as a result of supplementation

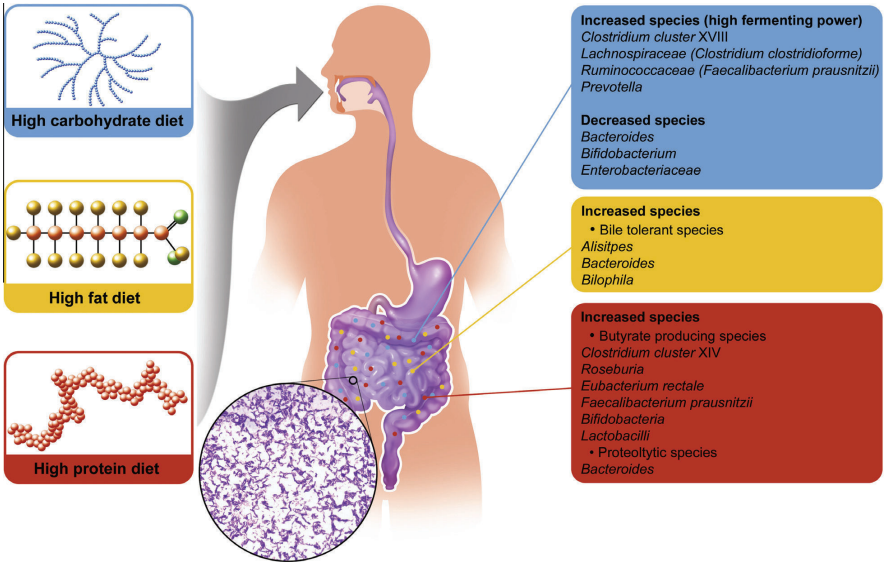


Fig. 3. Diel influences the composition of the gut microbiota. Depending on the major macronutrient in a diet type, the composition of the gut microbiota can be transiently altered to adapt to the digestion pattern for the specific macronutrient. Specific bacterial populations and functions are stimulated accordingly.

Table 1
Nutritional disorders: definition, diagnosis and prevalence.

Nutritional disorder	Type of cellular imbalance	Diagnosis	Prevalence
Obesity and overweight	Positive	• Adults: BMI $\geqslant$25 • Children: BMI-for-age > +1SD	• 1.9 billion adults (2014) • 41 million children (2014)
Undernutrition	Negative	• Stunting: HAZ < −2SD • Wasting: WAZ < −2SD • Underweight: WHZ < −2SD	• 29 million children under five are severely malnourished • 1–6 million deaths every year
Anorexia nervosa	Negative	• BMI <18.5 • Body weight <85% of ideal body weight • Over-evaluation of shape and weight	• 0.3% of adolescents between 15 and 19 years old

HAZ: height-for-age z-score. WAZ: weight-for-height z-score. WHZ: weight-for-height z-score.

of diet with the aforementioned prebiotics [2,23,26]. Several foods are being currently tested for prebiotics capacity [29,30] and testing should continue in order to find probiotics, prebiotics and synbiotics that can modulate the gut microbiota composition and be easily available to the general population.

Nutritional disorders associated with a dysbiosis of the gut microbiota

Classification of nutritional disorders

WHO defines malnutrition as the "cellular imbalance between supply of nutrients and energy and the body's demand for them to ensure growth, maintenance and specific functions" [31]. According to the positive or negative energy balance observed several types of malnutrition are defined. Obesity represents a positive energy balance [32] and is the most prevalent nutritional disorder in western countries; it is also highly prevalent in developing countries, representing a heavier burden on public health than undernutrition worldwide [33]. In fact, according to WHO, the prevalence of obesity has doubled in the last 30 years. In 2014, worldwide 1.9 billion adults were overweight among which 600 million adults were obese while 41 million children under five are obese or overweight [34,35]. WHO defines overweight and obesity in adults with a BMI (Body Mass Index) $\geqslant$ 25 and a BMI $\geqslant$ 30 respectively [34,36]. Obesity can also be defined by the measure of a BMI over 120% of ideal body weight (Table 1) [33]. In children under five, overweight is defined +2 standard deviation (SD) from the median of child growth standards [37] while obesity is defined by +3SD from the aforementioned median. For children aged from 5 to 19 years, WHO recommends to consider as overweight, children with a BMI-for-age >1SD above the WHO growth reference median and as obese, children with a BMI-for-age >2SD above the aforementioned median. Another measure of obesity in children involve a BMI over the 95th percentile which is based on the child's age and sex (Table 1) [36,38]. Common comorbidities include cardiovascular diseases, hypertension, type 2 diabetes,

Table 2
Characteristics of the gut microbiota associated with nutritional disorders.

Nutritional disorder	Diversity alterations	Metabolic alterations
Obesity	Lower global diversity	Increase of genes involved in carbohydrate metabolism
	Increase of the *Firmicutes*: *Bacteroidetes* ratio	Increase of genes involved in phosphotransferase system and membrane transport
	Increase of *Methanobrevibacter smithii*	
	Increase of *Lactobacillus*	Decrease of genes involved in transcription and nucleotide metabolism
	Decrease of *Bifidobacteria*	Decrease of genes involved in cofactors and vitamin metabolism
	Decrease of *Escherichia coli*	
Severe acute malnutrition	Lower global diversity	Increase of virulence factor homologs
	Increase in aerobic species (*Proteobacteria* and pathogenic species)	Increase of genes involved in motility and chemotaxis, respiration, membrane transport and virulence
	Decrease in anaerobic species (*Bacteroidetes*)	Decrease of genes implicated in nutrient uptake and metabolism
	Absence of *Methanobrevibacter smithii*	Inhibition of enzymes implicated in the tricarboxylic acid cycle
Anorexia nervosa	Lower global diversity	Decreased concentrations of the SCFA acetate and propionate
	Increase of *Methanobrevibacter smithii*	
	Decrease of *Lactobacillus*, *Streptococcus*	
	Decrease of the *Clostridium coccoides* and *Bacteroides fragilis* groups	

hypercholesterolemia, non-alcoholic fatty liver disease, cancer and some immune related disorders [33,36,39].

Undernutrition consists of stunting (height-for-age (HAZ) < −2SD), wasting (weight-for-age (WAZ) < −2SD) and underweight (weight-for-height (WHZ) < −2SD). Chronic malnutrition is characterized by stunting while acute malnutrition is characterized wasting, underweight [40,41] and/or nutritional oedema [40,42]. Severe acute malnutrition (SAM) (WHZ < −3SD) affects 29 million under five children and is responsible for the death of 1–6 million children every year (Table 1) [43]. It is mostly prevalent in developing countries of sub-Saharan Africa, Central America and South Asia [44]. Symptoms include, besides nutritional oedema in some cases, diarrhea [45], hepatic steatosis, skin rashes, ulceration, anorexia [46], delayed growth and deficiencies in macronutrients and micronutrients [47]. This type of malnutrition represents a negative energy balance [32] associated with a deficit of macronutrients, more specifically a deficit of protein, quantitative in marasmus and qualitative in kwashiorkor [31,42].

Another example of negative energy balance is anorexia nervosa. Anorexia nervosa is the first described and officially recognized eating disorder with two subtypes, the restricting type and the binge/purging type [48–50]. It has a low prevalence of 0.3% and is observed in developed and developing countries mostly in adolescents aged from 15 to 19 years old [48,49]. Eating disorders have a core symptomatology which consists in over-evaluation of shape and weight. Other symptoms include a significantly low body weight highlighted by a BMI under 18.5 or a body weight under 85% of ideal body weight and in some cases, amenorrhea (Table 1) [48–51]. Given these symptoms, it is clear that anorexia nervosa is a mental disorder that has a physical impact through its manifestation [49,50].

All these nutritional disorders have been proven to be linked to a disruption of the gut microbiota with specific characteristics in the shifting of the bacterial diversity observed in each of the aforementioned nutritional disorders (Table 2).

Alterations of the gut microbiota associated with nutritional disorders

As well described, the gut microbiota is implicated in the regulation of energy metabolism through the digestion of indigestible polysaccharides for the host which fermentation by the microbiota leads to the production of SCFA (propionate, butyrate and acetate); SCFA represent 10% of the daily energy supply in humans [39,52]. They serve as energy supply for the colonocytes as well as ligand for G Coupled Protein Receptors (GCPR) thus influencing insulin sensitivity in adipocytes and peripheral organs, reducing fat accumulation, improving gut motility and nutrient absorption and activating host immunity [39,52,53]. The gut microbiota also regulates energy metabolism by stimulating triglyceride deposition in adipocytes as well as triglycerides and cholesterol synthesis and lipogenesis. Conversely, the gut microbiota inhibits fatty acid oxidation, ketogenesis and glucose consumption [39]. As such, energetic imbalance is linked to gut microbiota alteration.

Alterations of the gut microbiota associated with obesity and the metabolic syndrome

In obese individuals, the gut microbiota alteration stimulates monosaccharides absorption through an increased capillary density in the small intestine epithelium [54] and leads to a greatly improved capacity for harvesting additional energy from diet with an increase of species implicated in indigestible polysaccharide fermentation in the colon [39,55]. In fact, a lower diversity is observed in obese individuals in most studies [35,54–56] with an increase of the *Firmicutes*: *Bacteroidetes* ratio [39,52–54,56,57] alongside an increase in methanogenic archaea such as *M. smithii* which presence improve polysaccharides fermentation by removing H_2 in the gut environment [52,58]. As a result, two major SCFA are found increased in obese subjects [54]: butyrate is the main energy supply for the colonocytes and is negatively correlated with intestinal permeability while acetate is a substrate for hepatic cholesterol synthesis and de novo lipogenesis [57] leading to an increase in adiposity and body weight [59]. Nevertheless, other data suggest a decrease of *M. smithii* in obese subjects compared to lean individuals [60]. A change in the diversity of Lactic Acid Bacteria (LAB) is also observed with a described increase in *Lactobacillus* species [58] and *Lactobacillus reuteri* in particular and low levels of bifidobacteria and *Escherichia coli* [56,60]. Obesogenic species also include *Blautia hydrogenotrophica*, *Coprococcus catuss*, *Eubacterium ventriosim*, *Ruminococcus broomii* and *Ruminococus obeum*. These *Firmicutes* species are known to improve energy harvesting from the diet [56]. The alteration of the gut microbiota associated with obesity composition also leads to a low grade inflammation negatively correlated with the gut microbiota gene count. Obese individuals with a HGC present an increase of anti-inflammatory species such as *F. prausnitzii* [39,52] and *A. muciniphila* and a higher production of organic acids [61] while a decrease in pro-inflammatory species such as *Bacteroides* spp is observed [39]. On the other hand, LGC obese subjects gut microbiota is associated with a higher prevalence of potentially pro-inflammatory species and genes implicated in oxidative stress response [61]. Species which diversity is modified in obesity-associated gut microbiota seem to have

specific functions linked to the physiopathology of obesity. Therefore, the role of the gut microbiota in obesity is more likely due to the genes and metabolites produced by the microbiota. An increase of genes involved in phosphotransferase system, carbohydrate metabolism and membrane transport is observed whereas a decrease in genes mediating transcription and nucleotide metabolism alongside a decrease in cofactors and vitamin metabolism [52].

Alterations of the gut microbiota associated with SAM

Acute malnutrition is associated with wasting in which the gut microbiota seems to have an effect since it has been associated with weight gain and in skeletal growth [62]. SAM-associated gut microbiota is described as immature [63,64], presenting a loss of diversity [65] with an enrichment in *Proteobacteria*, among which aerobic and pathogenic species [45,47,66,67], confirmed by an increase in the number of virulence factors homologs [47] and protein encoding genes functionally related to motility and chemotaxis, respiration, membrane transport and virulence [45]. Aerobic bacterial overgrowth in the small intestine has previously been associated with SAM [67,68]. A decrease in anaerobic species, is also observed [67,69] and moreover, a total absence of *M. smithii*, one of the most oxygen sensitive species in the human gut [69]. The decrease of anaerobic species is consistent with the decrease of the *Bacteroidetes* phylum described by Monira *et al.* which consists mostly of anaerobic species [40,65]. Enriched genera include *Bilophila* [46], *Klebsiella* [65], *Escherichia*, *Streptococcus*, *Shigella*, *Enterobacter*, *Veillonella* and a depletion in genera associated with health such as *Roseburia*, *Faecalibacterium*, *Butyribrio* and the *Synergistetes* phylum [47]. Gupta *et al.* also described an increase of the families *Bacteroidaceae* and *Porphyromonadaceae* [45]. Some depleted species such as *F. prausnizii* have previously been associated with weight gain in gnotobiotic mice colonized with healthy infant microbiota [63]. Depletion in genes implicated in nutrient uptake and metabolism is also observed in the microbiome consistently with the fact that depleted species are members of the *Firmicutes* phylum which are polysaccharides fermenters and SCFA producers [47]. In fact, SCFA availability is reduced in the gut of SAM patients [45]. A selective inhibition by the kwashiorkor-associated gut microbiota of one or more tricarboxylic acid cycle enzymes was observed by Smith *et al.* [46].

Alterations of the gut microbiota associated with anorexia nervosa

Only a few studies have explored the gut microbiota of anorexic patients. Since anorexia nervosa is a nutritional disorder, alterations must be observed in the gut microbiota of anorexic subjects. Studies have reported an increase of the prevalence of *M. smithii* [58,68] and a decrease of the prevalence of *Lactobacillus* species [58] notably a decrease of *Lactobacillus plantarum* alongside a decrease of *Streptococcus* spp., the *Clostridium coccoides* group and the *Bacteroides fragilis* group [70] associated with decreased concentrations of the SCFA, acetate and propionate. Moreover, the gut microbiota of anorexic subjects appears to present a lower total diversity [70]. A study exploring the gut microbiota of an anorexic patient described *Firmicutes*, *Actiobacteria* and *Bacteroidetes* as the most abundant phyla respectively with a high prevalence of anaerobic families [71].

Conclusion and perspectives

The gut microbiota is influenced in its taxonomic composition and its functional abilities by the macronutrients proportion in the diet in order to be able to assist the host in the digestion process. The role of the gut microbiota in digestion makes it an instrumental factor in energy imbalance and consequently in nutritional disorders. Dybiosis of the gut microbiota has in fact been observed in nutritional disorders such as obesity, undernutrition and anorexia nervosa with different characteristics of the gut microbiota associated with each disorder. The gut microbiota of anorexic patients has barely been studied and additional studies are required in order to have a better understanding of how the gut microbiota plays a role in this disease. Conversely, the extensive description of obesity and kwashiorkor associated gut microbiota paves the way to more interventional studies where commensals of the gut microbiota are tested as probiotics in order to reestablish a healthy flora and consequently a healthy individual. The era of microbio-therapeutics is here!

Author agreement

All authors have seen and approved the final version of the manuscript. This work has not been published and is not under consideration for publication elsewhere.

Conflict of interest

No conflict of interest to disclose.

Funding

This work was funded by the Méditerranée Infection foundation. The funding source had no other input in this article.

Acknowledgment

We thank Magdalen Lardière for English reviewing.

References

[1] Hugon P, Lagier J-C, Colson P, Bittar F, Raoult D. Repertoire of human gut microbes. Microb Pathog 2016. http://dx.doi.org/10.1016/j.micpath.2016.06.020.
[2] Maukonen J, Saarela M. Human gut microbiota: does diet matter? Proc Nutr Soc 2015;74:23–36. http://dx.doi.org/10.1017/S0029665114000688.
[3] Derrien M, van Hylckama Vlieg JET. Fate, activity, and impact of ingested bacteria within the human gut microbiota. Trends Microbiol 2015;23:354–66. http://dx.doi.org/10.1016/j.tim.2015.03.002.
[4] Escobar JS, Klotz B, Valdes BE, Agudelo GM. The gut microbiota of Colombians differs from that of Americans, Europeans and Asians. BMC Microbiol 2014;14:311. http://dx.doi.org/10.1186/s12866-014-0311-6.
[5] Savage DC. Microbial ecology of the gastrointestinal tract. Annu Rev Microbiol 1977;31:107–33. http://dx.doi.org/10.1146/annurev.mi.31.100177.000543.
[6] Hill MJ, Drasar BS. The normal colonic bacterial flora. Gut 1975;16:318–23. http://dx.doi.org/10.1136/gut.16.4.318.
[7] 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;28:237–64. http://dx.doi.org/10.1128/CMR.00014-14.
[8] Jandhyala SM, Talukdar R, Subramanyam C, Vuyyuru H, Sasikala M, Nageshwar Reddy D. Role of the normal gut microbiota. World J Gastroenterol 2015;21:8787–803. http://dx.doi.org/10.3748/wjg.v21.i29.8787.
[9] Mandal RS, Saha S, Das S. Metagenomic surveys of gut microbiota. Genomics Proteomics Bioinformatics 2015;13:148–58. http://dx.doi.org/10.1016/j.gpb.2015.02.005.
[10] Wu GD, Compher C, Chen EZ, Smith SA, Shah RD, Bittinger K, et al. Comparative metabolomics in vegans and omnivores reveal constraints on diet-dependent gut microbiota metabolite production. Gut 2016;65:63–72. http://dx.doi.org/10.1136/gutjnl-2014-308209.
[11] Greenhill AR, Tsuji H, Ogata K, Natsuhara K, Morita A, Soli K, et al. Characterization of the gut microbiota of Papua New Guineans using reverse transcription quantitative PCR. PLoS One 2015;10:e0117427. http://dx.doi.org/10.1371/journal.pone.0117427.
[12] Graf D, Di Cagno R, Fåk F, Flint HJ, Nyman M, Saarela M, et al. Contribution of diet to the composition of the human gut microbiota. Microb Ecol Health Dis 2015;26:26164.
[13] 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. http://dx.doi.org/10.1111/1469-0691.12023.

[14] 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;44:117–24. http://dx.doi.org/10.1016/j.ijantimicag.2014.04.020.
[15] Angelakis E, Lagier J-C. Samples and techniques highlighting the links between obesity and microbiota. Microb Pathog 2016. http://dx.doi.org/10.1016/j.micpath.2016.01.024.
[16] Leser TD, Mølbak L. Better living through microbial action: the benefits of the mammalian gastrointestinal microbiota on the host. Environ Microbiol 2009;11:2194–206. http://dx.doi.org/10.1111/j.1462-2920.2009.01941.x.
[17] Raoult D, Henrissat B. Are stool samples suitable for studying the link between gut microbiota and obesity? Eur J Epidemiol 2014;29:307–9. http://dx.doi.org/10.1007/s10654-014-9905-4.
[18] Bäckhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and stabilization of the human gut microbiome during the first year of life. Cell Host Microbe 2015;17:690–703. http://dx.doi.org/10.1016/j.chom.2015.04.004.
[19] Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. PLoS Biol 2007;5:e177. http://dx.doi.org/10.1371/journal.pbio.0050177.
[20] Salazar N, Arboleya S, Valdés L, Stanton C, Ross P, Ruiz L, et al. The human intestinal microbiome at extreme ages of life. Dietary intervention as a way to counteract alterations. Front Genet 2014;5. http://dx.doi.org/10.3389/fgene.2014.00406.
[21] Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature 2012;486:222–7. http://dx.doi.org/10.1038/nature11053.
[22] Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, et al. Metagenomic analysis of the human distal gut microbiome. Science 2006;312:1355–9. http://dx.doi.org/10.1126/science.1124234.
[23] Flint HJ, Duncan SH, Scott KP, Louis P. Links between diet, gut microbiota composition and gut metabolism. Proc Nutr Soc 2015;74:13–22. http://dx.doi.org/10.1017/S0029665114001463.
[24] Aguirre M, Eck A, Koenen ME, Savelkoul PHM, Budding AE, Venema K. Diet drives quick changes in the metabolic activity and composition of human gut microbiota in a validated in vitro gut model. Res Microbiol 2016;167:114–25. http://dx.doi.org/10.1016/j.resmic.2015.09.006.
[25] Costabile A, Deaville ER, Morales AM, Gibson GR. Prebiotic potential of a maize-based soluble fibre and impact of dose on the human gut microbiota. PLoS One 2016;11:e0144457. http://dx.doi.org/10.1371/journal.pone.0144457.
[26] Chung WSF, Walker AW, Louis P, Parkhill J, Vermeiren J, Bosscher D, et al. Modulation of the human gut microbiota by dietary fibres occurs at the species level. BMC Biol 2016;14:3. http://dx.doi.org/10.1186/s12915-015-0224-3.
[27] Dominianni C, Sinha R, Goedert JJ, Pei Z, Yang L, Hayes RB, et al. Sex, body mass index, and dietary fiber intake influence the human gut microbiome. PLoS One 2015;10:e0124599. http://dx.doi.org/10.1371/journal.pone.0124599.
[28] Charbonneau MR, O'Donnell D, Blanton LV, Totten SM, Davis JCC, Barratt MJ, et al. Sialylated milk oligosaccharides promote microbiota-dependent growth in models of infant undernutrition. Cell 2016;164:859–71. http://dx.doi.org/10.1016/j.cell.2016.01.024.
[29] Unno T, Choi J-H, Hur H-G, Sadowsky MJ, Ahn Y-T, Huh C-S, et al. Changes in human gut microbiota influenced by probiotic fermented milk ingestion. J Dairy Sci 2015;98:3568–76. http://dx.doi.org/10.3168/jds.2014-8943.
[30] Liu C, Finegold SM, Song Y, Lawson PA. 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 2008;58:1896–902. http://dx.doi.org/10.1099/ijs.0.65208-0.
[31] Alberda C, Graf A, McCargar L. Malnutrition: etiology, consequences, and assessment of a patient at risk. Best Pract Res Clin Gastroenterol 2006;20:419–39. http://dx.doi.org/10.1016/j.bpg.2006.01.006.
[32] Meijers JMM, van Bokhorst-de van der Schueren MAE, Schols JMGA, Soeters PB, Halfens RJG. Defining malnutrition: mission or mission impossible? Nutr Burbank Los Angel Cty Calif 2010;26:432–40. http://dx.doi.org/10.1016/j.nut.2009.06.012.
[33] Campfield LA, Smith FJ. The pathogenesis of obesity. Baillières Best Pract Res Clin Endocrinol Metab 1999;13:13–30.
[34] Behn A, Ur E. The obesity epidemic and its cardiovascular consequences. Curr Opin Cardiol 2006;21:353–60. http://dx.doi.org/10.1097/01.hco.0000231406.84554.96.
[35] Mishra AK, Dubey V, Ghosh AR. Obesity: an overview of possible role(s) of gut hormones, lipid sensing and gut microbiota. Metabolism 2016;65:48–65. http://dx.doi.org/10.1016/j.metabol.2015.10.008.
[36] Apovian CM. Obesity: definition, comorbidities, causes, and burden. Am J Manage Care 2016;22:s176–85.
[37] WHO | WHO Child Growth Standards: methods and development. WHO n.d. <http://www.who.int/nutrition/publications/childgrowthstandards_technical_report_1/en/>; [accessed 29.06.16].
[38] Kelly AS, Barlow SE, Rao G, Inge TH, Hayman LL, Steinberger J, et al. Severe obesity in children and adolescents: identification, associated health risks, and treatment approaches: a scientific statement from the American Heart Association. Circulation 2013;128:1689–712. http://dx.doi.org/10.1161/CIR.0b013e3182a5cfb3.
[39] Boulangé CL, Neves AL, Chilloux J, Nicholson JK, Dumas M-E. Impact of the gut microbiota on inflammation, obesity, and metabolic disease. Genome Med 2016;8:42. http://dx.doi.org/10.1186/s13073-016-0303-2.
[40] Million M, Diallo A, Raoult D. Gut microbiota and malnutrition. Microb Pathog 2016. http://dx.doi.org/10.1016/j.micpath.2016.02.003.
[41] World Health Organization, UNICEF. WHO child growth standards and the identification of severe acute malnutrition in infants and children: a joint statement by the World Health Organization and the United Nations Children's Fund, 2009.
[42] Cederholm T, Bosaeus I, Barazzoni R, Bauer J, Van Gossum A, Klek S, et al. Diagnostic criteria for malnutrition – an ESPEN consensus statement. Clin Nutr Edinb Scotl 2015;34:335–40. http://dx.doi.org/10.1016/j.clnu.2015.03.001.
[43] Wang H, Liddell CA, Coates MM, Mooney MD, Levitz CE, Schumacher AE, et al. Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014;384:957–79. http://dx.doi.org/10.1016/S0140-6736(14)60497-9.
[44] WHO | Levels and trends in child mortality 2015. WHO n.d. <http://www.who.int/maternal_child_adolescent/documents/levels_trends_child_mortality_2015/en/>; [accessed 21.04.16].
[45] Gupta SS, Mohammed MH, Ghosh TS, Kanungo S, Nair GB, Mande SS. Metagenome of the gut of a malnourished child. Gut Pathog 2011;3:7. http://dx.doi.org/10.1186/1757-4749-3-7.
[46] Smith MI, Yatsunenko T, Manary MJ, Trehan I, Mkakosya R, Cheng J, et al. Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. Science 2013;339:548–54. http://dx.doi.org/10.1126/science.1229000.
[47] Ghosh TS, Gupta SS, Bhattacharya T, Yadav D, Barik A, Chowdhury A, et al. Gut microbiomes of Indian children of varying nutritional status. PLoS One 2014;9:e95547. http://dx.doi.org/10.1371/journal.pone.0095547.
[48] Wilfley DE, Bishop ME, Wilson GT, Agras WS. Classification of eating disorders: toward DSM-V. Int J Eat Disord 2007;40(Suppl: S123–129). http://dx.doi.org/10.1002/eat.20436.
[49] Bulik CM, Reba L, Siega-Riz A-M, Reichborn-Kjennerud T. Anorexia nervosa: definition, epidemiology, and cycle of risk. Int J Eat Disord 2005;37(Suppl: S2-9-21). http://dx.doi.org/10.1002/eat.20107.
[50] Föcker M, Knoll S, Hebebrand J. Anorexia nervosa. Eur Child Adolesc Psychiatry 2013;22(Suppl 1):S29–35. http://dx.doi.org/10.1007/s00787-012-0358-6.
[51] Dellava JE, Thornton LM, Lichtenstein P, Pedersen NL, Bulik CM. Impact of broadening definitions of anorexia nervosa on sample characteristics. J Psychiatr Res 2011;45:691–8. http://dx.doi.org/10.1016/j.jpsychires.2010.10.003.
[52] Gérard P. Gut microbiota and obesity. Cell Mol Life Sci CMLS 2016;73:147–62. http://dx.doi.org/10.1007/s00018-015-2061-5.
[53] Sanchez M, Panahi S, Tremblay A. Childhood obesity: a role for gut microbiota? Int J Environ Res Public Health 2015;12:162–75. http://dx.doi.org/10.3390/ijerph120100162.
[54] Lau E, Carvalho D, Pina-Vaz C, Barbosa J-A, Freitas P. Beyond gut microbiota: understanding obesity and type 2 diabetes. Horm Athens Greece 2015;14:358–69. http://dx.doi.org/10.14310/horm.2002.1571.
[55] Remely M, Tesar I, Hippe B, Gnauer S, Rust P, Haslberger AG. Gut microbiota composition correlates with changes in body fat content due to weight loss. Benef Microbes 2015;6:431–9. http://dx.doi.org/10.3920/BM2014.0104.
[56] Kasai C, Sugimoto K, Moritani I, Tanaka J, Oya Y, Inoue H, et al. Comparison of the gut microbiota composition between obese and non-obese individuals in a Japanese population, as analyzed by terminal restriction fragment length polymorphism and next-generation sequencing. BMC Gastroenterol 2015;15:100. http://dx.doi.org/10.1186/s12876-015-0330-2.
[57] Lecomte V, Kaakoush NO, Maloney CA, Raipuria M, Huinao KD, Mitchell HM, et al. Changes in gut microbiota in rats fed a high fat diet correlate with obesity-associated metabolic parameters. PLoS One 2015;10:e0126931. http://dx.doi.org/10.1371/journal.pone.0126931.
[58] Armougom F, Henry M, Vialettes B, Raccah D, Raoult D. Monitoring bacterial community of human gut microbiota reveals an increase in *Lactobacillus* in obese patients and Methanogens in anorexic patients. PLoS One 2009;4:e7125. http://dx.doi.org/10.1371/journal.pone.0007125.
[59] Janssen AWF, Kersten S. The role of the gut microbiota in metabolic health. FASEB J Off Publ Fed Am Soc Exp Biol 2015;29:3111–23. http://dx.doi.org/10.1096/fj.14-269514.
[60] Million M, Angelakis E, Maraninchi M, Henry M, Giorgi R, Valero R, et al. Correlation between body mass index and gut concentrations of *Lactobacillus reuteri*, *Bifidobacterium animalis*, *Methanobrevibacter smithii* and *Escherichia coli*. Int J Obes 2005;2013(37):1460–6. http://dx.doi.org/10.1038/ijo.2013.20.
[61] Marchesi JR, Adams DH, Fava F, Hermes GDA, Hirschfield GM, Hold G, et al. The gut microbiota and host health: a new clinical frontier. Gut 2016;65:330–9. http://dx.doi.org/10.1136/gutjnl-2015-309990.
[62] Schwarzer M, Makki K, Storelli G, Machuca-Gayet I, Srutkova D, Hermanova P, et al. *Lactobacillus plantarum* strain maintains growth of infant mice during chronic undernutrition. Science 2016;351:854–7. http://dx.doi.org/10.1126/science.aad8588.
[63] Blanton LV, Charbonneau MR, Salih T, Barratt MJ, Venkatesh S, Ilkaveya O, et al. Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children. Science 2016;351:aad3311. http://dx.doi.org/10.1126/science.aad3311.

[64] Subramanian S, Huq S, Yatsunenko T, Haque R, Mahfuz M, Alam MA, et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. Nature 2014. http://dx.doi.org/10.1038/nature13421.
[65] Monira S, Nakamura S, Gotoh K, Izutsu K, Watanabe H, Alam NH, et al. Gut microbiota of healthy and malnourished children in Bangladesh. Front Microbiol 2011;2:228. http://dx.doi.org/10.3389/fmicb.2011.00228.
[66] Kristensen KHS, Wiese M, Rytter MJH, Özçam M, Hansen LH, Namusoke H, et al. Gut microbiota in children hospitalized with oedematous and non-oedematous severe acute malnutrition in Uganda. PLoS Negl Trop Dis 2016;10: e0004369. http://dx.doi.org/10.1371/journal.pntd.0004369.
[67] Mata LJ, Jiménez F, Cordón M, Rosales R, Prera E, Schneider RE, et al. Gastrointestinal flora of children with protein—calorie malnutrition. Am J Clin Nutr 1972;25:1118–26.
[68] Genton L, Cani PD, Schrenzel J. Alterations of gut barrier and gut microbiota in food restriction, food deprivation and protein-energy wasting. Clin Nutr Edinb Scotl 2015;34:341–9. http://dx.doi.org/10.1016/j.clnu.2014.10.003.
[69] Million M, Tidjani Alou M, Khelaifia S, Bachar D, Lagier J-C, Dione N, et al. Increased gut redox and depletion of anaerobic and methanogenic prokaryotes in severe acute malnutrition. Sci Rep 2016;6:26051. http://dx.doi.org/10.1038/srep26051.
[70] Morita C, Tsuji H, Hata T, Gondo M, Takakura S, Kawai K, et al. Gut dysbiosis in patients with anorexia nervosa. PLoS One 2015;10:e0145274. http://dx.doi.org/10.1371/journal.pone.0145274.
[71] 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;32:1471–81. http://dx.doi.org/10.1007/s10096-013-1900-2.

Partie II: Caractérisation du microbiote digestif d'enfants sévèrement malnutris et de leur environnement digestif.

Avant-propos

La malnutrition sévère aiguë est principalement prévalente dans des pays en voie de développement de l'Afrique subsaharienne, de l'Asie du sud-est et de l'Amérique centrale. Dans ces régions du monde, l'alimentation consiste souvent d'aliments à base d'amidon tels que les légumineuses et le manioc ou autres tubercules et de céréales. Un manque cruel de viande, légumes verts et fruits caractérise souvent ce type de régime alimentaire. Ainsi, de faibles apports protéiques et en micronutriments essentiels à la santé humaine, parmi lesquels les antioxydants, sont observés. Plusieurs études ont montré de faibles concentrations en antioxydants conduisant à un stress oxydatif chez les enfants sévèrement malnutris qui joue un rôle majeur dans les mécanismes entrainant le développement des œdèmes chez les patients atteints de kwashiorkor (16,25–27). Une oxydation excessive de l'environnement digestif entrainerait la perte des microorganismes sensibles à l'oxygène et donc une altération de la composition du microbiote digestif.

Dans cette étude, nous avons étudié 80 échantillons fécaux provenant d'enfants sains et d'enfants malnutris provenant du Niger et du Sénégal par métagénomique. Ainsi une perte de diversité globale a été observée chez les malnutris par rapport

aux enfants sains associée à une perte drastique en bactéries anaérobies strictes et à un enrichissement en bactéries aérotolérantes parmi lesquelles certaines bactéries pathogènes. Une absence totale de *Methanobrevibacter smithii*, une archée méthanogène et un des procaryotes les plus sensibles à l'oxygène, est également associée à la malnutrition sévère aiguë de même qu'un potentiel redox fécal accru. Cette altération de la diversité microbienne fécale est cohérente avec le stress oxydatif décrit chez les malnutris probablement dû à un apport insuffisant en antioxydants.

Article 2: Increased Gut Redox and Depletion of Anaerobic and Methanogenic Prokaryotes in Severe Acute Malnutrition

Matthieu Million†, Maryam Tidjani Alou†, Saber Khelaifia, Dipankar Bachar, Jean-Christophe Lagier, Niokhor Dione, Souleymane Brah, Perrine Hugon, Vincent Lombard, Fabrice Armougom, Julien Fromonot, Catherine Robert, Caroline Michelle, Aldiouma Diallo, Alexandre Fabre, Régis Guieu, Cheikh Sokhna, Bernard Henrissat, Philippe Parola & Didier Raoult.

†Co-auteurs à contribution égale.

Publié dans Scientific Reports

SCIENTIFIC REPORTS

OPEN

Increased Gut Redox and Depletion of Anaerobic and Methanogenic Prokaryotes in Severe Acute Malnutrition

Received: 22 October 2015
Accepted: 27 April 2016
Published: 17 May 2016

Matthieu Million[1,*], Maryam Tidjani Alou[1,*], Saber Khelaifia[1], Dipankar Bachar[1], Jean-Christophe Lagier[1], Niokhor Dione[1], Souleymane Brah[2], Perrine Hugon[1], Vincent Lombard[3,4], Fabrice Armougom[1], Julien Fromonot[5], Catherine Robert[1], Caroline Michelle[1], Aldiouma Diallo[6], Alexandre Fabre[7,8], Régis Guieu[5], Cheikh Sokhna[6], Bernard Henrissat[3,4,9], Philippe Parola[1] & Didier Raoult[1,9]

Severe acute malnutrition (SAM) is associated with inadequate diet, low levels of plasma antioxidants and gut microbiota alterations. The link between gut redox and microbial alterations, however, remains unexplored. By sequencing the gut microbiomes of 79 children of varying nutritional status from three centers in Senegal and Niger, we found a dramatic depletion of obligate anaerobes in malnutrition. This was confirmed in an individual patient data meta-analysis including 107 cases and 77 controls from 5 different African and Asian countries. Specifically, several species of the *Bacteroidaceae*, *Eubacteriaceae*, *Lachnospiraceae* and *Ruminococceae* families were consistently depleted while *Enterococcus faecalis*, *Escherichia coli* and *Staphylococcus aureus* were consistently enriched. Further analyses on our samples revealed increased fecal redox potential, decreased total bacterial number and dramatic *Methanobrevibacter smithii* depletion. Indeed, *M. smithii* was detected in more than half of the controls but in none of the cases. No causality was demonstrated but, based on our results, we propose a unifying theory linking microbiota specificity, lacking anaerobes and archaea, to low antioxidant nutrients, and lower food conversion.

Malnutrition is estimated to cause 3 million child deaths annually or 45% of all child deaths[1] and reflects the first Millennium Development Goal[2]. Severe acute malnutrition (SAM) is defined as a very low weight-for-height z-score (WHZ < -3 SD, corresponding to a weight below the average weight of healthy controls of the same height minus 3 standard deviations calculated according to World Health Organization (WHO) standards[3]), a mid-upper-arm circumference (MUAC) of less than 115 mm in children aged 6–59 months, or bilateral nutritional edema[2]. Its prevalence is estimated at 29 million children under 5 years of age[2,4]. SAM has been associated with high mortality (10–30%) and increased risk of diarrhea, pneumonia and systemic infections including bacteremia with *Staphylococcus aureus*, *Streptococcus*, *Salmonella*, *Klebsiella*, and *Escherichia coli*.

SAM is not only associated with inadequate quantitative intake of energy and protein but also with a qualitative lack of micronutrients. The diet is restricted to cooked but contaminated starchy foods (rice, millet, sorghum) and lacks milk, meat, leafy vegetables and fruits (mango, orange, guava), which are natural sources of micronutrients and dietary antioxidants (vitamins A, C, E)[5–8]. This inadequate diet leads to low levels of antioxidants[9],

[1]Aix-Marseille Université, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM63, CNRS 7278, IRD 198, INSERM 1095, Marseille, France. [2]Hôpital National de Niamey, Niamey, Niger. [3]Centre National de la Recherche Scientifique, UMR 7257, 13288 Marseille, France. [4]Aix-Marseille Université, Architecture et Fonctions des Macromolécules Biologiques, 163, avenue de Luminy, 13288 Marseille cedex 9, France. [5]Laboratoire de Biochimie, hôpital de la Timone, UMR MD2, IRBA, Marseille, France. [6]Institut de Recherche pour le Développement (IRD), URMITE, Dakar, Sénégal. [7]Service de pédiatrie multidisciplinaire, hôpital de la Timone Enfant, APHM, Marseille France. [8]UMR_S 910, Aix-Marseille Université, Marseille, France. [9]Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to D.R. (email: didier.raoult@gmail.com)

including uric acid, ascorbic acid[10] and glutathione[11], in the plasma and liver of these children. The resulting oxidative stress[12] is a key feature in the pathophysiology of complicated (edematous) SAM[11], confirmed through the beneficial effects of antioxidants[13]. However, only glutathione and alpha-lipoic acid supplementation have been shown to have positive effects on survival, while N-acetylcysteine, riboflavin, vitamin E and selenium were not shown to have a statistically significant effect[13,14]. Ready-to-use therapeutic food (RUTF), including several antioxidants, has recently revolutionized the management and prognosis of SAM in outpatient settings for children >6 months of age, without complications[15], suggesting that the redox status could be altered in malnutrition.

Age, diet, and geography were first identified as major determinants of the gut microbiota[16]. While the link between gut microbiota and malnutrition has long been suspected, studies using metagenomic methods have revealed that specific gut microbiota alterations predict therapeutic failure[17]. In addition, gut microbiota immaturity persists despite therapeutic foods[18]. A recent review evidenced that the early depletion in gut *Bifidobacterium longum* represents the first step in gut microbiota alteration associated with severe acute malnutrition (SAM) before a disruption of the Healthy Mature Anaerobic Gut Microbiota (HMAGM), leading to diarrhea, malabsorption, deficient energy harvest and vitamin biosynthesis, and systemic invasion by microbial pathogens[19].

Archaea are typical representatives of the HMAGM worldwide[16]. Unlike the broad diversity of Bacteria in the gut, the Archaea comprised mainly members of the *Methanobacteriales* order, which are H2-oxidizing methanogens with *Methanobrevibacter smithii* being the leading representative species in healthy volunteers[20]. Significantly higher numbers of H2-utilizing methanogenic Archaea were observed in obese individuals than in normal-weight or post-gastric-bypass individuals leading to the hypothesis that interspecies H2 transfer between bacterial and archaeal species increases energy uptake by the human large intestine in obese persons[21]. As Archaea are vulnerable to trace amounts of free radicals, surviving only in a strongly reduced environment (redox potential < -200mV[22]), and, as such, are defined as extremely oxygen-sensitive prokaryotes, we hypothesized that a gut redox alteration, expected in SAM, should result in a major change in *Methanobrevibacter smithii* concentrations in the gut of severely undernourished children.

In this study, we tested whether gut obligate anaerobes are prone to depletion and determined whether aerotolerant species are prone to enrichment in the guts of severely undernourished children. We confirmed the putative depletion of extremely oxygen-sensitive prokaryotes, specifically targeting *Methanobrevibacter smithii,* representing the host-bacterial-archaeal mutualism[23]. We also analyzed the gut redox potential as a first step linking the lack of antioxidant and altered redox status, previously reported in malnutrition, to gut microbiota alterations. The confounding role of age[16,18] was specifically ruled out through a sensitivity analysis including a matched-pairs study and meta-regression.

Results

Recruiting center bias. We screened 86 West African children of varying nutritional status (Dataset 1) and sequenced the fecal metagenome of 79 children from Dakar, Dielmo and N'Diop in Senegal and from Niamey in Niger. The samples from 10 children associated with a poor sequencing depth (<100,000 reads) were excluded and all other 69 samples were normalized to a 100,000 reads depth. To explore factors impacting gut microbial diversity among these 69 children with comparable gut metagenomes, we first performed a Pearson exploratory principal component analysis including relative abundances of the 2,245 identified bacterial species as active variables and center of recruitment, age, sex, and nutritional status as supplementary (explanatory) variables. Recruiting center alone differentiated children into distinct groups (Fig. 1), suggesting that geography is the main explanatory variable of overall gut microbiota variability. In a sensitivity analysis to test the robustness of our results, a second PCA was performed excluding 10 outliers (6 healthy children and 2 children with SAM from Dakar, Senegal, and 2 children with SAM from Niger) identified by the ROUT method[24] (Q = 1%) based on the Euclidian distance to the framework origin (Supplementary Figure 1). This second PCA confirmed geography as the major determinant of the global variability of the gut microbiota among our samples. Strikingly, the children from Dielmo and N'Diop (south of Senegal, green) clustered with those of Niamey (Niger, red) at 2,124 km apart, but not with those from Dakar (Senegal, blue), which were much closer at 160 km. This dramatic center bias suggests that cases and controls should be recruited from the same center (and not only the same country) to identify gut microbiota alterations associated with any disease (malnutrition) and that multicenter data should not be pooled but require meta-analysis. Age, sex or nutritional status were not satisfactory explanatory variables of overall gut microbiota variability. As no children with SAM from Dielmo and N'Diop were included, these centers were excluded from subsequent analyses.

Gut microbiota maturation is associated with an enrichment in obligate anaerobes. After identifying the recruiting center bias and the need for meta-analysis, we sought to identify whether age has a significant impact on the enrichment or depletion of anaerobic bacteria in the digestive tract. Meta-regression appeared as the most suitable approach to identify whether age (gut microbiota maturation) was associated with enrichment or depletion of obligate anaerobes independently of nutritional status and recruiting center. We also decided to only include children with severe acute malnutrition (SAM) and healthy controls (CTL), strictly based on the WHO anthropometric criteria[2,3], in order to improve the consistency among studies and to increase the power of the meta-analysis. Consequently, all children with other forms of malnutrition (moderate wasting, moderate to severe underweight and moderate to severe stunting) were excluded from this analysis.

After strict ascertainment of cases and controls, 10 cases and 11 controls from Dakar, Senegal and 10 cases and 9 controls from Niamey, Niger were included in 2 distinct comparisons in our meta-analysis. We also used this meta-analysis to test whether our results were similar to those of the published case-control studies with publicly available metagenomic data. Preliminary searches identified 26 records focusing on gut microbiota alterations associated with malnutrition; 21 records were immediately excluded however (Fig. 2). Thus, 5 full-text articles were assessed for eligibility, and 2 articles were further excluded because the assignment of metagenomics

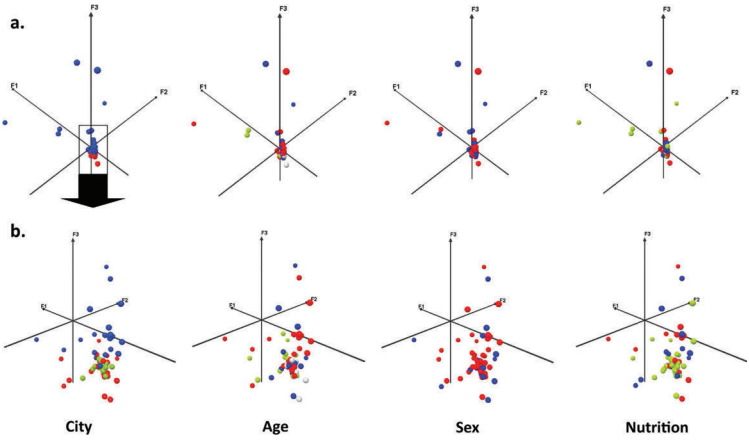


Figure 1. Explanatory variables of global variability of the 16S rRNA gut microbiota among 69 West African children of varying nutritional status. (**a**) A Pearson Principal component analysis (PCA) of the 16S rRNA gut metagenomes of 69 West African children revealed that geography was the main factor in gut microbiota variability, as only recruiting centers differentiated children into distinct groups (up). (**b**) In a sensitivity analysis to test the robustness of the results, a second PCA was performed excluding 10 outliers (6 healthy children and 2 children with SAM from Dakar, Senegal, and 2 children with SAM from Niamey, Niger) identified by the ROUT method (Q = 1%) based on the Euclidian distance to the framework origin[24]. This second PCA confirmed geography as the major determinant of the global variability of the gut microbiota among our samples (down). *City*: blue: Dakar (Senegal), green: Dielmo and N'Diop (contiguous villages in Senegal), red: Niamey (Niger). *Age*: red: up to 12 months, blue: >12 months and up to 24 months, green: >24 months, grey: missing data, *Sex*: red: female, blue: male, *Nutritional status*: green: healthy controls, red: severe acute malnutrition, blue: children with other forms of malnutrition.

reads could not be obtained at the prokaryotic species level[25,26]. Consequently, our two comparisons, along with three comparisons from the literature[17,18,27] were included from 5 different countries (Bangladesh, India, Malawi, Niger and Senegal) corresponding to 264 children (Table 1). After case and control ascertainment, only 107 children with SAM and 77 controls were included. Although individual assessment could not be performed for one study[18], the risk of bias was considered to be low; therefore, these participants were included (Table 1).

Our 2 comparisons and 2 of the 3 included comparisons from the literature were consistent with a positive summary aerotolerant odds ratio (AOR = (aerotolerant prokaryotes enriched in SAM x obligate anaerobic prokaryotes depleted in SAM)/(aerotolerant prokaryotes depleted in SAM x obligate anaerobic prokaryotes enriched in SAM), see methods) suggesting a relative depletion of obligate anaerobes and enrichment of aerotolerant organisms in SAM (AOR 5.00, 95% confidence interval (95%CI) 0.82–30.49, p = 0.08, Fig. 3a). Heterogeneity was however substantial ($I^2 = 81\%$) and significant (p = 0.0003), with one contradictory study[17], prompting the detection of explanatory covariates. As suspected based on the literature[16], meta-regression revealed that the age difference between cases and controls significantly altered the aerotolerant odds ratio (slope −0.06, 95%CI −0.12 to −0.0003, p = 0.048, Fig. 3b). Indeed, this confirmed that gut microbiota maturation is associated with an enrichment in obligate anaerobes[16].

Relative gut anaerobic depletion and aerotolerant enrichment in SAM. To check for the confounding role of age, we identified ten matched pairs, including six matched pairs in the Dakar study, along with two twin pairs from the Smith *et al.* study[17]. Strikingly, in this study, 9/11 'healthy' co-twins presented severe stunting and did not fulfill our control selection criteria, so these pairs were excluded. Two matched pairs from the Ghosh *et al.* study[27] were also included (Supplementary Table S1). Pairing was perfect for sex and highly effective for age (rs (Spearman) 0.98, p < 0.0001). The mean age ± SD was 20.3 ± 17.3 and 19.2 ± 15.3 for cases and controls, respectively (Wilcoxon matched-pairs signed rank test, p = 0.43), so that the confounding role of age was excluded, as the controls were slightly younger than cases, and obligate anaerobes increased with age[16].

The meta-analysis of these ten matched pairs, and participants of the Subramanian study (ruling out the confounding role of age using a linear mixed model, see Supplementary Table 15a of[18]), revealed a significantly increased AOR (10.59, 95%CI 2.17–51.69, p = 0.004). A subgroup analysis by continent reduced heterogeneity between studies (I^2 reduced from 75% to 0%) with a significant subgroup difference (I2 = 91.3% − p = 0.0007, Fig. 3c). The direction of the effect was consistent in all four comparisons after ruling out the confounding roles of age, sex and bias associated with geography. These results suggest that SAM is associated with a dramatic relative

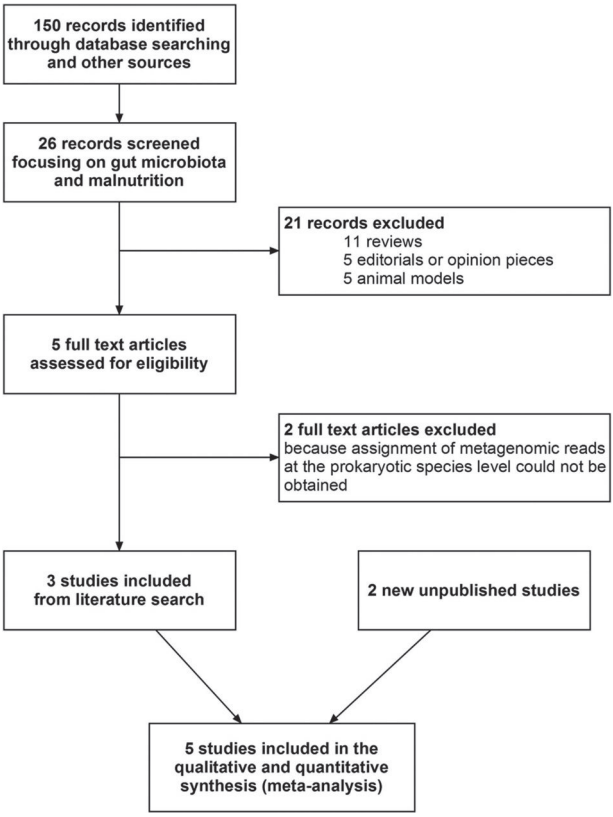


Figure 2. Meta-analysis study flow diagram. A literature review of case-control studies with available metagenomic data at the species level identified 5 comparisons eligible for meta-analysis. Study selection was performed according to the MOOSE checklist and guidelines. The 2 excluded full-text articles were studies from Monira[25] and Gupta[26].

anaerobic depletion regardless of age, sex and recruitment center (geography). Moreover, this relative anaerobic depletion may be more profound in Asia than in Africa.

The number of depleted anaerobes was several times higher than the number of enriched aerotolerant species in SAM in all comparisons [2.8 in Dakar (present study), 2.4 in India[27], 2.9 in Malawi[17] and 9.3 in Bangladesh[18]], suggesting that malnutrition was more closely associated with decreased anaerobic than increased aerotolerant diversity. A qualitative meta-analysis confirmed the depletion of several obligate anaerobes, including several *Firmicutes* (3 *Eubacteriaceae*, 3 *Lachnospiraceae*, 2 *Ruminococcaceae, 1 Erysipelotrichaceae), Bacteroidetes* (2 *Bacteroidaceae) and Actinobacteria (2 Eggerthella, 1 Coriobacteriaceae). Conversely,* some aerotolerant bacteria were enriched in SAM, including *Enterococcus faecalis, Escherichia coli,* and *Staphylococcus aureus,* which all represent potential pathogens (Fig. 4).

Gut microbial and M. smithii depletion and increased redox potential in SAM. To clarify whether only relative, or also absolute abundance of obligate anaerobes, was depleted in SAM, we quantified the total bacterial number by flow cytometry. *M. smithii*, as one of the most oxygen-sensitive prokaryotes in the gut, was also assessed by quantitative real-time PCR to confirm anaerobic depletion. Redox potential was investigated in the fecal samples as a putative explanatory variable representing oxidation of the gut environment. The total bacterial concentration was significantly decreased in SAM (mean $\pm$ SD, 10.9 ± 0.30 log10 bacteria/g of stool in SAM versus 12.12 ± 0.42 in controls, two-tailed paired t-test $p = 0.008$, Fig. 5). With the aforementioned results, this confirmed

Reference	Dates	Country, city, center	Study design	Method of fecal metagenomics analysis, sequencing depth	Number of screened children (number of 16S metagenomes)	Individual data available	Number of ascertained SAM cases included in the meta-analysis (nb of 16S metagenomes)	Number of ascertained controls included in the meta-analysis (nb of 16S metagenomes)
Present study, Niger	2014	Niger, Niamey	Observational study	V3v4 Illumina MiSeq, 200,000 reads	34 (28)	Yes	10 (10)	9 (9)
Present study, Senegal	2014	Senegal, Dakar	Observational study	V3v4 Illumina MiSeq, 200,000 reads	52 (51)	Yes	10 (10)	11 (11)
Smith[17]	2013	Malawi, Southern region, 5 centers (Mayaka, Mbiza, Chamba, Mitondo, Makhwira)	longitudinal comparative study of monozygotic and dizygotic twin pairs	V4 16S Illumina HiSeq instrument, 116,414 reads	44 (183)	Yes	14 - Mayaka 3, Mbiza 3, Chamba 4, Mitondo 1, Makhwira 3 –(15)	5 - Mbiza 4, Chamba 1 (6)
Subramanian[18]	2014	Bangladesh, Dhaka	Cohort study	V4 16S Illumina HiSeq instrument, unknown metagenomics depth	114 (1074)[a]	No but species enriched or depleted in SAM calculated using a mixed linear regression model controlling for age available in online supplementary material	64 (78)[a]	50 (996)[a]
Ghosh[27]	2014	India, Birbhum district	Observational studies of children with varying nutritional status	Untargeted sequencing GS-FLX, 10,000 reads[b]	20 (20)	Yes	9 (9)	2 (2)
Total					264 screened children (1356)		107 (122)	77 (1024)

Table 1. Studies comparing the gut metagenomes of children with severe acute malnutrition to healthy controls. [a]Information kindly communicated by the authors[18]. [b]Raw data were reanalyzed by one of us (BD) to obtain species assignement (see supplementary methods). SAM: severe acute malnutrition.

the absolute decrease in gut obligate anaerobes in SAM. This was further confirmed by the fact that *M. smithii* was detected in five out of six controls but not in the cases (McNemar test, $p = 0.04$). Gut redox was abnormally positive and significantly increased in SAM (mean ± SD (mV), 27.0 ± 47.8 vs. -26.5 ± 61.8, $p = 0.001$).

Including all participants from Niger and Senegal, *M. smithii* was not detected in the 20 children with SAM, although this microbe was detected in 40% and 75% of controls ($p < 10^{-4}$) based on v3v4 16S rRNA large-scale sequencing and qPCR, respectively (Supplementary Table S2). We compared all cases and ascertained controls recruited in Dakar in a linear regression model, with age, sex, and SAM as the independent variables. Finally, SAM was the only independent predictor of the total bacterial count (SAM, −0.78, −1.31 to −0.25, $p = 0.007$ – $R^2 = 0.39$), *M. smithii* concentration (−4.28, −6.41 to −2.15, $p = 0.0004$, $R^2 = 0.47$) and fecal redox potential (coefficient 51.38, 0.71 – 102.05, $p = 0.047$, $R^2 = 0.20$).

Discussion

In the present study, we observed global bacterial depletion, a depletion in obligate anaerobes and a lack of methanogens in the gut microbiota of children with SAM. Quantifying the fecal redox potential as an explanatory parameter, we found for the first time a gut environment oxidation associated with childhood SAM. We identified the depletion of many oxygen-intolerant prokaryotes, while a small number of aerotolerant and potentially pathogenic prokaryotes were enriched in the gut of children with SAM. In particular, the well-known potential pathogens *Enterococcus faecalis*, *Escherichia coli* and *Staphylococcus aureus* were consistently enriched in SAM while *Methanobrevibacter smithii*, the prokaryote most sensitive to oxygen and the main representative of Archaea in the gut[20], was not detected in children with SAM. These results found in two West African centers (Niger and Senegal) were reproduced by reanalysis of previously published studies[17,18,27] from different countries (Bangladesh, India and Malawi) and different continents (Africa and Asia). Using principal component analysis and meta-analysis, the geographical bias[16] was confirmed and controlled for. Using matched-pairs, meta-regression, and multivariate linear regression analyses, the confounding role for age[16,18] was confirmed and controlled for.

The results of our study led us to propose a unifying theory linking gut environment oxidation to depletion of anaerobic and methanogenic prokaryotes in severe acute malnutrition[28]. The lack of adequate breastfeeding and nutrients[5,6,29], maintaining the natural fecal anti-oxidant capacity[30], is the most probable cause. Indeed, as a diet selected to raise the intake of dietary antioxidants increases the fecal antioxidant capacity[30–33], a diet depleted in antioxidants, as previously reported in SAM[6], is expected to decrease the fecal antioxidant capacity.

Abnormal inversion of the ratio of anaerobes to aerobes has been reported for a long time in SAM, with incomplete correction after treatment[19,34]. *In vitro*, antioxidants are able to keep highly oxygen-sensitive bacteria alive in ambient air[35] and facilitate the *in vitro* aerobic culture of *Bacteroides*[36]. In humans or in a batch-culture fermentation system reflective of the distal region of the human large intestine, flavonoids, with a 3–5 times

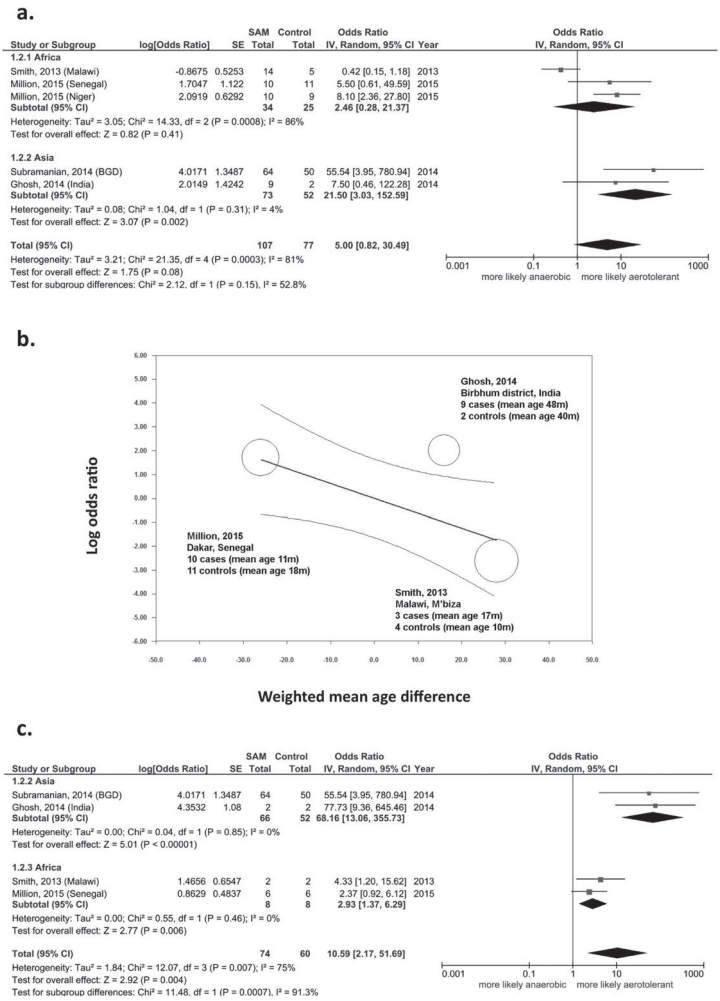


Figure 3. Meta-analysis of all ascertained cases and controls. A positive aerotolerant odds ratio (AOR) correspond to a relative aerotolerant enrichment/anaerobic depletion in SAM. (**a**) Before controlling for age, there was a trend for an increased AOR in severe acute malnutrition (SAM), but heterogeneity was substantial, (**b**) Meta-regression confirmed that age difference between cases and controls significantly altered the AOR. (**c**) Matched pairs analysis (mixed linear regression for Subramanian controlling for age) found a dramatic and significant anaerobic gut prokaryote depletion in SAM independently of geography and age. SAM: severe acute malnutrition, BGD: Bangladesh.

greater antioxidant capacity than vitamins C or E *in vitro* and being poorly absorbed in the gut[37,38], dramatically increase microbial members of the HMAGM[16,19] including *Eubacterium rectale*[16] and *Blautia coccoides*[39,40]. Strikingly, *Eubacterium rectale* was found consistently depleted in SAM (Fig. 4). HMAGM maintenance by antioxidants could also explain why glutathione dramatically reduces *Helicobacter*-induced gastric pathologies[41].

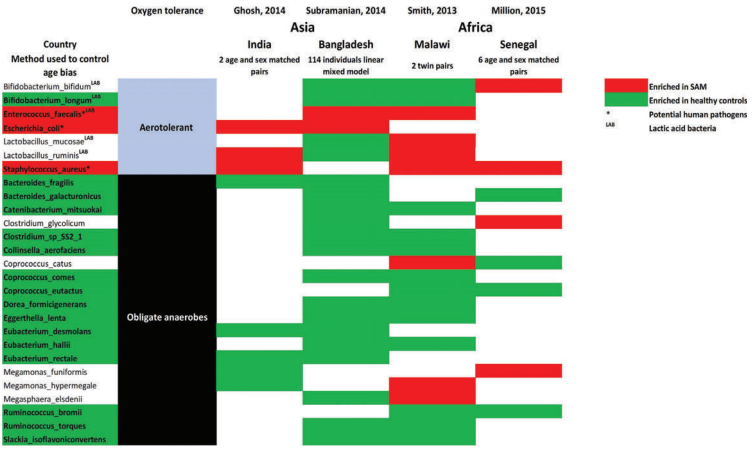


Figure 4. Species consistently enriched and/or depleted in SAM. Qualitative meta-analysis identified the depletion of several obligate anaerobes and an enrichment of the aerotolerant *Enterococcus faecalis*, *Escherichia coli* and *Staphylococcus aureus* in SAM. Red: enriched in SAM. Green: depleted in SAM. *Potential human pathogens. SAM: severe acute malnutrition, LAB: Lactic acid bacteria.

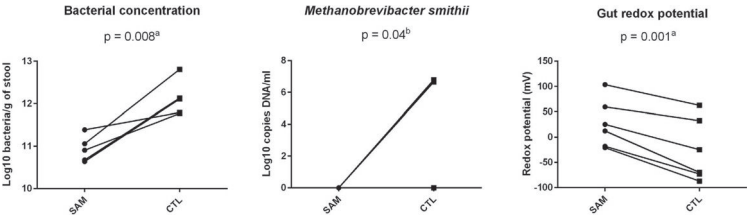


Figure 5. Alteration of gut total bacterial number, *M. smithii* and redox potential in SAM. (**a**) Two-sided paired t-test. (**b**) McNemar test. Total gut bacterial concentration and *M. smithii* were significantly depleted in SAM while the gut redox potential was increased in SAM. These results, obtained after analysis of 6 age- and sex-matched pairs, were confirmed by a linear regression model including all 21 individuals from Dakar, Senegal with age, sex and SAM as the independent variables. SAM: severe acute malnutrition, CTL: healthy controls.

Host-bacterial-archeal mutualism was recently demonstrated as a key in improving food conversion (the mass of food eaten divided by the weight gain): In gnotobiotic animals, gut colonization by the anaerobe *Bacteroides thetaiotaomicron* increased food conversion limited by H2 accumulation[23]. Cocolonization by *B. thetaiotaomicron* and *M. smithii* resulted in optimal food conversion via the removal of the H2 excess by methanogenesis[23]. While this process was considered to be impossible in ambient air, we recently demonstrated that bacteria considered unable to grow in aerobic atmospheres (anaerobic), are indeed able to grow in such an environment provided that the medium contains adequate antioxidants[36]. Moreover, we were also able to produce methane (methanogenesis) by *B. thetaiotaomicron* and *M. smithii* in aerobic atmosphere using ascorbate, uric acid and gluthatione[42]. This experimental evidence led us to suspect a causal link between the lack of dietary antioxidants, anaerobic depletion (including *B. thetaiotaomicron* and *M. smithii*), and decreased energy harvest and food conversion.

Despite the reported associations and *in vitro* experimental model, no definite evidence for a causal role of diet or gut redox on anaerobic depletion is presented and other factors should be discussed. Age, determining gut colonization by obligate anaerobes[16,18] and methanogens including *M. smithii*[16,18,20], was controlled for using matched-pairs and multivariate linear regression analyses. Diet is not the only factor as human submucosal gland secretions contain uric acid that serves as an antioxidant[43]. A contaminated environment may play a role through injection of pathogens in the gut microbiota. Subsequent proliferation of pathogens in the gut in association with gut microbiota alteration could result in a host disease defined as a 'gut microbiota infection'[44]. Among

such pathogens, *Salmonella enterica* and *Clostridium difficile* pathogenic strains have been shown to deplete the global and more specifically the anaerobic gut microbiota, regardless of diet[45,46]. Other factors able to deplete gut anaerobes include the lack of nutrients devoid of antioxidants but shaping the gut microbiota (prebiotics[47]), host genetic factors[48], altered innate lymphoid cells[49] and deficient fucosylation[50]. Here, we propose that gut redox is a key parameter shaping the gut microbiota. The relative role of gut redox alteration among all factors influencing the gut microbiota (including intestinal inflammation, which itself could alter the gut redox) will be elucidated by future studies.

Our findings provide new and important insights into the vicious cycle of malnutrition, correlating altered redox metabolism with a possibly irreversible disruption of host-archaeal-bacterial mutualism, via a depletion of anaerobes and methanogens. Parallel invasion by aerotolerant bacterial pathogens highlights the need for antibiotics as previously demonstrated[51]. Additional studies are needed to confirm the oxidation of the gut environment in SAM and to test the beneficial effect of RUTF and antioxidants in the restoration of gut obligate anaerobes and methanogens. Different antioxidants may have different effects[13,14] with no effect of antioxidants completely absorbed in the proximal gut[52] or not associated with adequate breastfeeding, nutrition and gut microbiota infection control. Based on our results, new therapeutic options could be tested in severe or refractory SAM cases in association with antibiotics and up-to-date management[15], including higher levels of antioxidants, unabsorbed antioxidants (carotenoids and flavonoids)[37], supplementation of lacking microorganisms and transplantation of fecal microbiota.

Materials and Methods

Experimental design. We conducted two prospective case-control studies in Niger (Niamey) and Senegal (Dakar, Dielmo (13°43′N 16°24′W) and NDiop (13°41′N 16°22′W)). Dielmo and N'Diop are two contiguous villages included in a longitudinal malaria survey[53]. The gut microbiota of children with varying nutritional status were analyzed using a principal component analysis before comparing children with SAM (according to the 2009 WHO criteria[2,3]) to ascertained healthy controls (CTL) according to the STROBE statements[54]. Recruitment of children <60 months attending the clinic 'Notre Dame de L'Esperance' for malnutrition in Thiaroye, Dakar, Senegal, occurred in April, 2014. Children from Dielmo and Ndiop were recruited between September and December 2014 and recruitment of children from the National Hospital, Niamey, Niger ranged from February to November 2014. The exclusion criteria included the following: absence of parent or patient consent, antibiotic administration <2 months before stool collection and insufficient fecal amount. Exclusion criteria for controls included any form of malnutrition, presence of fever, acute respiratory infection (cough), acute or chronic diarrhea in the previous 4 weeks. Symptoms and complications of malnutrition (cough, fever, diarrhea, vomiting), or other forms of malnutrition (moderate to severe stunting and underweight) were not a reason for exclusion for cases since they are considered as a direct consequence and part of severe acute malnutrition. Clinical data (sex, date of birth, date of sampling, clinical history, weight, height, mid-upper arm circumference and antibiotic use) were recorded using a standardized questionnaire. Growth monitoring parameters (weight-for-length (WHZ), weight-for-age (WAZ) and height-for-age (HAZ) z scores) were assessed using the WHO Anthro v3.2.2, January 2011. Length was measured recumbently for children less than 12 months of age and height was measured standing in children greater than 12 months of age. WHZ and WAZ, however, are not appropriate in the presence of edematous SAM. After collection using sterile plastic containers, the stool samples were aliquoted and frozen at −80 °C until manipulations.

All aspects of the study were approved by the local ethics committee 'Comité d'éthique de l'IFR 48, Service de Médecine Légale (Faculté de Médecine, Marseille, France) under accession number 09–022, 2010. Informed consent was obtained and the nature and possible consequences of the studies were explained. Only verbal consent from patients or parents was required for this study according to French bioethics decree Number 2007–1220, published in the official journal of the French Republic and to article L1211-2 of the French Code of Public Health. Professor DIALLO and Professor ADEHOSSI certified that this study was not in opposition to the declaration of Helsinki and in accordance with Senegalese and Nigerien laws respectively (certificates available on request).

Case and control definition and ascertainment. Cases were children with SAM defined as a WHZ <−3 standard deviations (SD), calculated according to WHO standards, a MUAC of less than 115 mm in children aged 6–59 months, or bilateral nutritional edema[2,3]. Control children were asymptomatic (no cough, fever, diarrhea or vomiting) without any form of malnutrition (WHZ > −2 SD, Height-for-age Z-score >−2 SD, weight-for-age z-score >−2 SD, MUAC >125 mm in children aged 6–59 months) attending the centers for non-nutritive disorders. Local nutritional experts (AD, SB) performed the case ascertainment and control selection, and the clinical and anthropometric data were verified (MM).

Treatment of undernourished children. None of the undernourished children were in feeding programs prior to hospital admission. Fecal sampling preceded anti-infectious treatment in all cases. In Senegal, an energetic milk (milk, oil, sugar) was given immediately along with amoxicillin, most often in connection with a skin infection. An Emmel test (HbS detection), a thick smear, and a direct stool parasitological examination were systematically performed. Specific treatment was given based on the results (Artemisinin-based combination therapy, mebendazole and/or metronidazole), but none of the included children were found to be positive for parasites. Vitamin A was given only after stabilization, always after fecal sampling. If complicated malnutrition was diagnosed, the child was referred to the University Hospital. In Niger, children were recruited immediately on diagnosis, before any treatment or lipid-based nutritional supplements was administered. Self-medication was also excluded, so that bias in reporting of antibiotic administration <2 months before stool collection was

minimized. Overall, the anti-anaerobic effect of anti-infectious drugs (metronidazole), frequently used in the malnutrition setting, played no confounding role in this study.

Outcomes, exposure and predictors. The principal summary measure was the 'aerotolerant' odds ratio (AOR), calculated as follows: AOR = (number of aerotolerant species enriched in SAM × number of obligate anaerobes depleted in SAM)/(number of aerotolerant species depleted in SAM × number of obligate anaerobes enriched in SAM).

Sequencing and biochemical methods. Sequencing was performed using 16S v3v4 amplification prior to large-scale sequencing using an Illumina MiSeq Engine. Assignment was performed down to the species level. To control for depth bias, samples yielding less than 100,000 reads at the species level were subsequently excluded prior to the normalization of the remaining samples (all relative abundance $<10^{-5}$ were converted to 0). Specific PCR analyses targeting *M. smithii* were performed as previously described[20]. Fecal redox was measured, and total bacterial concentration was assessed using flow cytometry (methods are fully detailed in Supplementary Data).

Database of obligate anaerobes. A bacterial oxygen tolerance database was generated based on the literature (Supplementary Methods, available online at http://www.mediterranee-infection.com/article.php?laref=374). Each phylotype was assigned as 'obligate anaerobe', 'aerotolerant' or 'unknown' according to oxygen tolerance.

Statistical analysis. After normalization, an exploratory Pearson principal component analysis was first performed using the relative abundance of bacterial species as active variable and age, sex, and nutritional status as supplementary elements using XLSTAT v2014.3.07 (Addinsoft, Paris, France). The ROUT method (Q = 1%)[24] was used to robustly identify outliers based on the Euclidian distance to the framework origin (PCA centroid) calculated using the first three components of each sample. Thus, in each center, species were considered to be enriched or depleted in cases or controls based on univariate linear regression calculated using MINE[55] with the R package version 3.1.0. Pairing was conducted to control for age and sex, selecting cases and controls with the same sex, center recruitment and closest age. A paired t-test and McNemar test were used to compare the total number of bacteria, fecal redox potential and presence of *M. smithii* between age- and sex-matched pairs, respectively. Different multivariable stepwise linear regression models were created, including either total bacterial concentration, *M. smithii* concentration or fecal redox potential as the dependent variable and age, sex and SAM as the independent variables using SPSS v24.0 (IBM, Paris, France). All tests were two-sided, and a p-value < 0.05 was considered significant.

Meta-analysis of individual patient data. We performed a meta-analysis of individual patient data with meta-regression controlling for age according to the MOOSE guidelines[56] including our two comparisons (Dakar (Senegal) and Niamey (Niger) – no children with SAM were recruited in Dielmo & N'Diop). Eligibility criteria for other studies included analyzing the fecal prokaryotes of children with SAM and healthy controls using metagenomic methods at the species level. The search terms included severe acute malnutrition, kwashiorkor, marasmus AND metagenome or microbiota. There was no restriction according to date or language. The selection and ascertainment criteria for the cases and controls were applied as described above when individual data were available. The protocols (PICOS), exclusion criteria, data preparation and bias assessment are detailed in the Supplementary Methods and Supplementary Table S3. The meta-analysis was performed using the Mantel-Haenszel method and a random-effects model using Review Manager version 5.2 (Cochrane Collaboration, Copenhagen, Denmark). Heterogeneity was assessed using I^2 analysis. Meta-regression controlling for age was performed using comprehensive meta-analysis version 2.2.064 (Biostat, Englewood, NJ, USA) as reported by Borenstein *et al.*[57].

References

1. R. E. Black *et al.* Maternal and Child Nutrition Study Group. Maternal and child undernutrition and overweight in low-income and middle-income countries. *Lancet*. **382,** 427–451 (2013).
2. World Health Organization, United Nations Children's Fund. WHO child growth standards and the identification of severe acute malnutrition in infants and children, a joint statement. http://www.who.int/nutrition/publications/severemalnutrition/9789241598163/en/, (2009) Date of access: 14/04/2016.
3. WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards based on length/height, weight and age. *Acta Paediatr.* **95,** 76–85 (2006).
4. United Nations Children's Fund (UNICEF). Improving child nutrition, the achievable imperative for global progress. http://www.unicef.org/publications/index_68661.html, (2013) Date of access: 05/04/2016.
5. C. D. Williams. A nutritional disease of childhood associated with a maize diet. *Arch Dis Child.* **8,** 423–33 (1933).
6. H. Kismul, J. Van den Broeck & T. M. Lunde. Diet and kwashiorkor: a prospective study from rural DR Congo. *PeerJ.* **2,** e350 (2014).
7. N. F. Carvalho, R. D. Kenney, P. H. Carrington & D. E. Hall. Severe nutritional deficiencies in toddlers resulting from health food milk alternatives. *Pediatrics*. **107,** E46 (2001).
8. F. Mori *et al.* A kwashiorkor case due to the use of an exclusive rice milk diet to treat atopic dermatitis. *Nutr J*. **14,** 83 (2015).
9. A. Fechner *et al.* Antioxidant status and nitric oxide in the malnutrition syndrome kwashiorkor. *Pediatr Res*. **49,** 237–243 (2001).
10. A. Raoult. Aspects of malnutrition in big child in French East Africa: biochemical data; proteinemia & blood electrophoresis. II. *Bull Soc Pathol Exot Filiales*. **1,** 114–125 (1959).
11. M. H. Golden & D. Ramdath. Free radicals in the pathogenesis of kwashiorkor. *Proc Nutr Soc*. **46,** 53–68 (1987).
12. G. A. Preidis *et al.* The undernourished neonatal mouse metabolome reveals evidence of liver and biliary dysfunction, inflammation, and oxidative stress. *J Nutr.* **144,** 273–81 (2014).
13. K. Becker *et al.* Effects of antioxidants on glutathione levels and clinical recovery from the malnutrition syndrome kwashiorkor – a pilot study. *Redox Rep*. **10,** 215–226 (2005).
14. H. Ciliberto *et al.* Antioxidant supplementation for the prevention of Kwashiorkor in Malawian children : randomised, double blind, placebo controlled trial. *BMJ*. **14,** 1109 (2005).

15. The World Health Organization, the World Food Programme, the United Nations System Standing Committee on Nutrition and the United Nations Children's Fund. Community-based management of severe acute malnutrition. http://www.who.int/nutrition/topics/statement_commbased_malnutrition/en/, (2007) Date of access: 14/04/2016.
16. T. Yatsunenko *et al.* Human gut microbiome viewed across age and geography. *Nature*. **486,** 222–227 (2012).
17. M. I. Smith *et al.* Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. *Science*. **339,** 548–554 (2013).
18. S. Subramanian *et al.* Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature*. **510,** 417–421 (2014).
19. M. Million, A. Diallo & D. Raoult. Gut microbiota and malnutrition. *Microb Pathog.* Epub ahead of print. doi: 10.1016/j.micpath.2016.02.003 (2016).
20. B. Dridi, M. Henry, A. El Khechine, D. Raoult & M. Drancourt. High prevalence of *Methanobrevibacter smithii* and *Methanosphaera stadtmanae* detected in the human gut using an improved DNA detection protocol. *PLoS One*. **4,** e7063 (2009).
21. H. Zhang *et al.* Human gut microbiota in obesity and after gastric bypass. *Proc Natl Acad Sci USA* **106,** 2365–2370 (2009).
22. S. Hirano, N. Matsumoto, M. Morita, K. Sasaki & N. Ohmura. Electrochemical control of redox potential affects methanogenesis of the hydrogenotrophic methanogen *Methanothermobacter thermautrotrophicus*. *Lett Appl Microbiol*. **56,** 315–21 (2013).
23. B. S. Samuel & J. I. Gordon. A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism. *Proc Natl Acad Sci USA* **103,** 10011–10006 (2006).
24. H. J. Motulsky & R. E. Brown. Detecting outliers when fitting data with nonlinear regression - a new method based on robust nonlinear regression and the false discovery rate. *BMC Bioinformatics*. **7,** 123 (2006).
25. S. Monira *et al.* Gut microbiota of healthy and malnourished children in Bangladesh. *Front Microbiol*. **2,** 228 (2011).
26. S. S. Gupta *et al.* Metagenome of the gut of a malnourished child. *Gut Pathog*. **3,** 7 (2011).
27. T. S. Ghosh *et al.* Gut microbiomes of Indian children of varying nutritional status. *PLoS One*. **9,** e95547 (2014).
28. M. G. Espey. Role of oxygen gradients in shaping redox relationships between the human intestine and its microbiota. *Free Radic Biol Med*. **55,** 130–140 (2013).
29. C. D. Williams. Kwashiorkor: a nutritional disease of children associated with a maize diet. *Lancet*. **2,** 1151–1152 (1935).
30. M. Garsetti, N. Pellegrini, C. Baggio & F. Brighenti. Antioxidant activity in human faeces. *Br J Nutr*. **84,** 705–710 (2000).
31. M. A. Bianchi *et al.* Ability of a high-total antioxidant capacity diet to increase stool weight and bowel antioxidant status in human subjects. *Br J Nutr*. **104,** 1500–1507 (2010).
32. M. N. Orozco, N. W. Solomons, K. Schumann, J. K. Friel & A. L. de Montenegro. Antioxidant-rich oral supplements attenuate the effects of oral iron on *in situ* oxidation susceptibility of human feces. *J Nutr.* **140,** 1105–1110 (2010).
33. I. R. Record, J. K. McInerney, M. Noakes & A. R. Bird. Chocolate consumption, fecal water antioxidant activity, and hydroxyl radical production. *Nutr Cancer.* **47,** 131–135 (2003).
34. L. J. Mata *et al.* Gastrointestinal flora of children with protein--calorie malnutrition. *Am J Clin Nutr*. **25,** 118–126 (1972).
35. M. T. Khan, J. M. van Dijl & H. J. Harmsen. Antioxidants keep the potentially probiotic but highly oxygen-sensitive human gut bacterium *Faecalibacterium prausnitzii* alive at ambient air. *PLoS One*. **9,** e96097 (2014).
36. B. La Scola, S. Khelaifia, J. C. Lagier & D. Raoult. Aerobic culture of anaerobic bacteria using antioxidants: a preliminary report. *Eur J Clin Microbiol Infect Dis*. **33,** 1781–1783 (2014).
37. B. Halliwell, K. Zhao & M. Whiteman. The gastrointestinal tract: a major site of antioxidant action? *Free Radic Res*. **33,** 819–830 (2000).
38. S. B. Lotito & B. Frei. Consumption of flavonoid-rich foods and increased plasma antioxidant capacity in humans: cause, consequence, or epiphenomenon? *Free Radic Biol Med*. **15,** 1727–1746 (2006).
39. T. Clavel *et al.* Isoflavones and functional foods alter the dominant intestinal microbiota in postmenopausal women. *J Nutr.* **135,** 2786–92 (2005).
40. X. Tzounis *et al.* Flavanol monomer-induced changes to the human faecal microflora. *Br J Nutr*. **99,** 782–792 (2008).
41. E. De Bruyne *et al.* Oral glutathione supplementation drastically reduces *Helicobacter*-induced gastric pathologies. *Sci Rep.* **6,** 20169, doi: 10.1038/srep20169 (2016).
42. S. Khelaifia *et al.* Aerobic culture of methanogenic archaea without an external source of hydrogen. *Eur J Clin Microbiol Infect Dis.* Epub ahead of print. doi: 10.1007/s10096-016-2627-7 (2016).
43. D. B. Peden *et al.* Uric acid is a major antioxidant in human nasal airway secretions. *Proc Natl Acad Sci USA* **87,** 7638–7642 (1990).
44. A. L. Byrd & J. A. Segre. Infectious disease. Adapting Koch's postulates. *Science*. **351,** 224–226 (2016).
45. V. Shankar *et al.* Species and genus level resolution analysis of gut microbiota in *Clostridium difficile* patients following fecal microbiota transplantation. *Microbiome*. **2,** 13 (2014).
46. M. Barman *et al.* Enteric salmonellosis disrupts the microbial ecology of the murine gastrointestinal tract. *Infect Immun*. **76,** 907–915 (2008).
47. G. R. Gibson, H. M. Probert, J. V. Loo, R. A. Rastall & M. B. Roberfroid. Dietary modulation of the human colonic microbiota: updating the concept of prebiotics. *Nutr Res Rev.* **17,** 259–275 (2004).
48. Z. A. Khachatryan *et al.* Predominant role of host genetics in controlling the composition of gut microbiota. *PLoS One.* **3,** e3064 (2008).
49. Y. Goto *et al.* Innate lymphoid cells regulate intestinal epithelial cell glycosylation. *Science*, **12;** 345(6202):1254009 (2014).
50. P. C. Kashyap *et al.* Genetically dictated change in host mucus carbohydrate landscape exerts a diet-dependent effect on the gut microbiota. *Proc Natl Acad Sci USA* **15,** 17059–17064 (2013).
51. I. Trehan *et al.* Antibiotics as part of the management of severe acute malnutrition. *N Engl J Med*. **368,** 425–435 (2013).
52. B. S. Samuel *et al.* Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. *Proc Natl Acad Sci USA* **105,** 16767–16772 (2008).
53. F. Fenollar *et al. Tropheryma whipplei* bacteremia during fever in rural West Africa. *Clin Infect Dis*. **51,** 515–521 (2010).
54. E. E. von Elm *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. **370,** 1453–1457 (2007).
55. D. N. Reshef *et al.* Detecting novel associations in large data sets. *Science*. **334,** 1518–1524 (2011).
56. D. F. Stroup *et al.* Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. **283,** 2008–2012 (2000).
57. M. Borenstein, L. V. Hedges, J. P. T. Higgins & H. R. Rothstein. Meta-Regression. In *Introduction to meta-analysis*. pp. 187–203 (Wiley, Chichester, United Kingdom, 2009).

Acknowledgements

We would like to thank E. Adehossi, A. Moulaye, S. Moussa, F. Amadou, H. Sibiri, D. Alhousseini and B. Ali Diallo for their help in the collection of clinical data and samples in Niger, along with F. Fenollar her for helpful contribution collecting samples from Senegal. We would sincerely like to thank Gilbert LOCATELLI, airline pilot, for having helped us with the transport of fecal samples from Niger to France. We would like to thank the anonymous reviewers who helped us to dramatically improve the value of the manuscript. This work has been funded by the Institut Hospitalo-Universitaire Méditerranée Infection. The funding source had no role in the writing of the manuscript, or the decision to submit it for publication, nor in data collection, analysis,

interpretation, trial design, patient recruitment or any aspect pertinent to the study. None of the authors were paid by a pharmaceutical company or other agency. MM and DR (corresponding author) had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Author Contributions

M.M. and D.R. analyzed the data and wrote the paper, S.B., A.D. and C.S. coordinated the clinical cohort and the collection of the samples in Niger (S.B.) and Senegal (A.D. and C.S.), M.T., S.K., J.C.L., N.D., P.H. and J.F. performed fecal sample analyses, C.R. and C.M. performed the metagenomics analysis, D.B., V.L. and F.A. performed the bioinformatics analysis. A.F. (pediatrician) and A.D. (Nutrition expert) helped with data analysis and interpretation. P.P. managed the relationships with partners in Niger. R.G. and B.H. contributed to data interpretation. D.R. designed the study. All authors reviewed the manuscript.

Additional Information

Accession codes: Metagenomic datasets have been deposited at the EBI (http://www.ebi.ac.uk/) under accession codes PRJEB7053.

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: SK and DR are co-inventors of a patent on the culture of anaerobic bacteria using anti-oxidants (1H53 316 CAS 9 FR).

How to cite this article: Million, M. *et al.* Increased Gut Redox and Depletion of Anaerobic and Methanogenic Prokaryotes in Severe Acute Malnutrition. *Sci. Rep.* **6**, 26051; doi: 10.1038/srep26051 (2016).

Partie III: Culturomics, le nouveau tournant dans l'exploration du microbiote digestif.

Avant-propos

La vulgarisation des techniques de séquençage de nouvelle génération il y'a plus d'une décennie a fait des techniques de biologie moléculaire, les méthodes de référence pour explorer la diversité et les fonctions du microbiote digestif (1). Grâce à ces techniques, des liens de causalité ont été établis entre certaines pathologies comme la malnutrition sévère aiguë et des altérations du microbiote digestif (14). Dans le cas de la malnutrition sévère aiguë, il s'avère évident qu'une réhabilitation du microbiote digestif de ces patients est nécessaire. Autant pour établir des stratégies thérapeutiques que pour étudier de manière plus précise les interactions entre espèces par la mise au point de modèles expérimentaux, il est important d'avoir des cultures des microorganismes identifiés en culture pure. Un retour aux approches d'exploration par culture est donc requis. Pour cela, "microbial culturomics", technique de culture à haut débit, est la méthode idéale à l'efficacité testée et prouvée par plusieurs études (4,5,22–24). Par la multiplication de conditions de culture en variant les milieux de culture avec parfois une supplémentation par du sang et/ ou du rumen et les conditions physico-chimiques, cette méthode tente de reproduire les conditions les plus proches de l'environnement digestif afin d'isoler des bactéries jusqu'à alors incultivées.

Dans cette partie, nous décrivons comment culturomics a permis non seulement, d'étendre le répertoire de bactéries cultivées du microbiote digestif mais aussi d'identifier des nouvelles espèces correspondant à des séquences non assignées appartenant à des ensembles de données métagénomiques d'échantillons fécaux. Ainsi, culturomics permet d'améliorer le savoir collectif sur le microbiote digestif.

Article 3: 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 Ricaboni, Melhem Bilen, Nicole Prisca Makaya Dangui, Souleymane Ndeye Mery Dia Badiane, Camille Valles, Donia Mouelhi, Khoudia Diop, Matthieu Million, Didier Musso, Jõnatas Abrahao, 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*

Accepté dans Nature Microbiology- In press.

Culture of previously uncultured members of 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[1], Souleymane Ndeye Mery Dia Badiane[1], Camille Valles[1], Donia Mouelhi[1], Khoudia Diop[1], Matthieu Million[1], Didier Musso[2], Jõnatas Abrahao[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*]

1. Aix Marseille Université URMITE, UM63, CNRS 7278, IRD 198, INSERM 1095, 27 Bd 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, 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

*To whom correspondence should be addressed
Prof. Didier Raoult (corresponding author)
Email: didier.raoult@gmail.com
Phone: 33 4 91 38 55 17
Fax: 33 4 91 38 77 72
Text word count: 2,011
Abstract: 200
Keywords: gut microbiota, bacterial culture, culturomics, metagenomics

Abstract

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 which uses multiple culture conditions and MALDI-TOF and 16S rRNA for identification[2]. Here, we selected the best culture conditions to increase the number of the studied samples and 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 was 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, growing 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]. Currently, in the approximately 13,410 known bacterial and archaea species, 2,152 have been identified in humans, and 688 bacteria and 2 archaea had been identified in the human gut before we initiated microbial culturomics [13]. Culturomics consists of the application of high-throughput culture conditions to the study of the human microbiota and uses MALDI-TOF or 16SrRNA amplification and sequencing for identifying 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 (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 had been 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, 2B, and 2C) 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] (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 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, 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 for culturomics, 58 non-gut bacteria, 65 non-human bacteria and 89 unknown bacteria (Fig. 1, Supplementary Tables 4A and 5).

In addition, we 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 for culturomics, 18 non-gut bacteria, 13 non-human bacteria and 10 unknown species (Fig. 1, 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, Supplementary Tables 3A and 4). When we specifically tested 110 samples for *Proteobacteria*, we isolated 9 bacteria new for culturomics, 3 non-gut bacteria and 3 non-human bacteria (Fig. 1, Supplementary Tables 4A and 5). By culturing 242 stool specimens exclusively under a microaerophilic atmosphere, we isolated 9 bacteria new for culturomics, 6 non-gut bacteria, 17 non-human bacteria and 7 unknown bacteria (Fig. 1, 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, Supplementary Tables 4A and 5). Among these 48 halophilic prokaryotic species, 7 were slight halophiles (growing with 10-50 g/L of NaCl), 39 moderate halophiles (growing with 50-200 g/L of NaCl), and 2 extreme halophiles (growing with 200-300 g/L of NaCl).

We also introduced the detection of microcolonies that were barely visible to the naked eye (diameters ranging from 100 µm to 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 MS or 16S rRNA amplification and sequencing. By testing 10 stool samples, we detected 2 non-gut bacteria, 1 non-human bacterium and 1 unknown bacterium that only formed microcolonies (Fig. 1, Supplementary Tables 4A and 5). Finally, by culturing 30 duodenal, small bowel intestine and colonic

samples, we isolated 22 bacteria new for culturomics, 6 non-gut bacteria, 9 non-human bacteria and 30 unknown bacteria (Fig. 1, 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 HMP (http://www.hmpdacc.org/catalog). We found sequences, previously termed Operational Taxonomic Units (OTU), 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 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 396 and 239 human gut microbiome samples described by Nielsen et al.[15] and Browne et al.[16], respectively. We captured 19 and 150 of our 247 new species in these metagenomics data, representing 7.7% and 60.7%, respectively (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 cut-off 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. 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, 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 non-human halophilic archaea, 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 archaea (Fig. 1, 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 CSUR collection and DSMZ collection (Supplementary Table 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 Table 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 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 herein, has enabled culturing of 77% of the 1,525 prokaryotes now identified in the human gut (Fig. 1, 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 lab from 57 diverse clinical samples (Supplementary Table 14). In 2016, 6 of the 374 (1.6%) different identifications performed in the routine lab 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 Human Microbiome Project (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, etc.) and from patients with diverse diseases (anorexia nervosa, obesity, malnutrition, HIV, etc.). The main characteristics are summarized in Supplementary Table 1. Each patient's consent was obtained, and the study was approved by the local Ethics Committee of the IFR48 (Marseille, France) (Agreement number 09-022, Marseille, France). Except for the small intestine and stool samples that we directly inoculated without storage (see below), 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 7 days to 12 months before analysis.

Culturomics

Culturomics is a high-throughput method which multiplies culture conditions in order to detect higher bacterial diversity. The first culturomics study concerned three stool samples; 212 culture conditions, including both 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 (see below). 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 following the strategy subsequently outlined.

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). Then, we selected the 18 best culture conditions, using liquid media enrichment in a medium containing blood and rumen fluid and subculturing aerobically and anaerobically in a solid media (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, Heidelberg,

Germany) and incubated the sample at 37°C for 1 day. We inoculated 100 µl of the enriched sample into 4 culture media: Hektoen agar (BD Diagnostics, Heidelberg, Germany), MacConkey agar + Cefotaxime (bioMérieux, Marcy l'Etoile, France), Cepacia agar (AES Chemunex, Bruz, France) and Columbia ANC agar (bioMérieux, Marcy l'Etoile, France). 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, Heidelberg, Germany), and 3,000 colonies were tested using MALDI-TOF.

Preterm neonates

Preterm neonates were recruited from 4 neonatal intensive care units (NICU) in southern France from February 2009 through December 2012[12]. Only patients with definite or advanced necrotizing enterocolitis corresponding to Bell stage 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 into 54 preselected culture conditions (Supplementary Table 2C). The anaerobic cultures were

performed in an anaerobic chamber (AES Chemunex, Combourg, France). We tested 3,000 colonies 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 2 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, Pont de Claix, France) and aerobic (BD Bactec Plus Lytic/10 Aerobic, Pont de Claix, France) 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, Marcy l'Etoile, France), 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 minutes at 13,000 ×g; the pH of the supernatants were then measured. Acidophilic

bacteria were cultured after the stool enrichment in a liquid medium consisting of Columbia Broth (Sigma-Aldrich, Saint-Quentin Fallavier, France) modified by the addition of (per litre) 5 g of $MgSO_4$, 5 g of $MgCl_2$, 2 g of KCl, 2 g of glucose and 1 g of $CaCl_2$. The pH was adjusted to five different values: 4, 4.5, 5, 5.5 and 6, using HCl. Then, the bacteria were 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. Then, serial dilutions from 10^{-1} to 10^{-10} were performed, and each dilution was then 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 as 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, Pont de Claix, France) and followed on days 2, 5, 10 and 15. Conditions tested were anaerobic Columbia with 5% sheep blood (bioMérieux, Marcy l'Etoile, France) at 37°C with or without thermic shock (20 min/80°C), 28°C, anaerobic Columbia with 5% sheep blood agar (bioMérieux, Marcy l'Etoile, France) and 5% rumen fluid, and R-medium (ascorbic acid 1 g/L, uric acid 0.4 g/L, and glutathione 1 g/L, 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, Pont de Claix, France) supplemented with vancomycin (100 µg/L) (Sigma-Aldrich, Saint-Quentin Fallavier, France). The subcultures were performed on 8 different selective solid media for the growth of *Proteobacteria*. We inoculated onto MacConkey agar (Biokar-Diagnostics, Beauvais, France), Buffered Charcoal Yeast Extract (BD Diagnostic, Pont de Claix, France), Eosine-Methylene Blue agar (Biokar-Diagnostics, Beauvais, France), Salmonella-Shigella agar (Biokar-Diagnostics, Beauvais, France), Drigalski agar (Biokar-Diagnostics, Beauvais, France), Hektoen agar (Biokar-Diagnostics, Beauvais, France), Thiosulfate-Citrate-Bile-Sucrose (BioRad, Paris, France) and Yersinia agar (BD Diagnostic, Pont de Claix, France) 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, Pont de Claix, France). We tested 15 different culture conditions using Pylori agar (bioMérieux, Marcy l'Etoile, France), Campylobacter agar (BD, Pont de Claix, France), Gardnerella agar (bioMérieux, Marcy l'Etoile, France), 5% sheep blood agar (bioMérieux, Marcy l'Etoile, France) and our own R-medium as previously described [23]. We incubated Petri dishes only in microaerophilic conditions using GENbag microaer systems (bioMérieux, Marcy l'Etoile, France) or CampyGen agar (bioMérieux, Marcy l'Etoile, France), except the R-medium, which was incubated aerobically at 37°C. These culture conditions generated 41,392 colonies, and these 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, Saint-Quentin Fallavier, France), modified by adding (per litre): MgCl2·6H2O, 5 g; MgSO4·7H2O, 5 g; KCl, 2 g; CaCl2·2H2O, 1 g; NaBr, 0.5 g; NaHCO3, 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 (Saint-Quentin

Fallavier, France). Four concentrations (100 g/L, 150 g/L, 200 g/L and 250 g/L) of NaCl were used.

We tested 215 different stool samples. One gram of each stool specimen was inoculated aerobically into 100 ml of liquid medium in flasks at 37°C while stirring at 150 rpm. Subcultures were inoculated after 3, 10, 15 and 30 incubation days for each culture condition. Then, serial dilutions from 10^{-1} to 10^{-10} were 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 begin to appear after the fifteenth day. All colonies were picked and re-streaked several times to obtain pure cultures, which were sub-cultured on a solid medium consisting of Colombia Agar medium (Sigma-Aldrich, Saint-Quentin Fallavier, France) 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, Paris, France). 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, Saint-Quentin Fallavier, France) 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 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 5 duodenum samples previously frozen at -80°C. We tested 25,000 colonies 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, Pont de Claix, France), 5% sheep blood agar (bioMérieux, Marcy l'Etoile, France) incubated in both microaerophilic and anaerobic conditions, R-

medium [23] and Pylori agar (bioMérieux, Marcy l'Etoile, France). 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 7 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 microorganisms 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, in order to culture halophilic archaea, we designed specific culture conditions described above in the 'halophilic bacteria' section.

Identification methods

The colonies were identified using mass spectrometry 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 [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/arkotheque/client/ihu_bacteries/recherche/index.php) to classify all isolated prokaryotes into 4 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, in order 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, the 16S rRNA accession numbers and the genome sequencing accession number

Most of the strains isolated in this study were deposited in the Collection de Souches de l'Unité des Rickettsies (CSUR, WDCM 875) and are easily available at (http://www.mediterranee-infection.com/article.php?laref=14&titre=collection-de-souches&PHPSESSID=cncregk417fl97gheb8k7u7t07) (Supplementary Table 4A and 4B). All the new prokaryote species were deposited into 2 international collections: the CSUR and the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) (Supplementary Table 5). Importantly, among the 247 new prokaryotes species (197 in the present study + 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

accession numbers are available in Supplementary Table 5; among these new species, 174 already have a DSMZ number, whereas for the other 64 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, among the genomes of the new prokaryotes already sequenced, 163 draft genomes are already deposited with an available GenBank accession number, and all other genome sequencing is still in progress, as the culturomics are still running in our lab.

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 Inc., San Diego, CA, USA) 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 [Qiagen, Courtaboeuf, France]). For the first 24 metagenomes, we used GS FLX Titanium (Roche Applied Science, Meylan, France). Primers were designed to produce an amplicon length (576 bp) that was approximately equivalent to the average length of reads produced by the GS FLX Titanium (Roche Applied Science, Meylan, France), 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 V6 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 sequenced 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 TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTA CGGGNGGCWGCAG; ReOvAd_785RGTCTCGTGGGCT CGGAGATGTGTATAAGAGAC AGGACTACHVGGGTATC TAATCC). Samples were amplified individually for the 16S 'V3–V4' regions by Phusion High Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA, USA) 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 with greatest accuracy, robust reactions, and high tolerance for inhibitors, and finally by an error rate which was approximately 50-fold lower than that of DNA polymerase and 6-fold 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 x 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 Inc., San Diego, CA, USA) 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 of DNA was mechanically fragmented on the Covaris device (KBioScience-LGC Genomics, Queens Road, Teddington, Middlesex, TW11 0LY, UK) 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 the 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 per conditions with 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 the gsRunBrowser and gsAssembler_Roche. Sequences obtained with Roche were assembled on the 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 the high-sensitivity kit (Life Technologies, Carlsbad, CA, USA). In the first approach, 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 profile 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, 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 the Covaris device S2 in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies Inc, Santa Clara, CA, USA). 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 Inc, Fullerton, CA, USA), 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 x 250 bp being 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 less than 80 amino acids. These threshold parameters have been used in previous studies to define ORFans [12-14]. The 168 genomes considered in this study are listed in

Supplementary Table 7. These genomes represent a total of 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 Human Microbiome Project 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 Human Microbiome Project 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 of metagenomics

The raw fastq files of paired end reads from Illumina Miseq of 84 metagenomes analysed concomitantly by culturomics were filtered and analysed in the following steps (accession number 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 which contained both primers (forward and reverse). In the following filtering steps, the sequences containing N were removed. Sequences with the length shorter than 200 nts were removed, and the sequences longer than 500 nts were trimmed. Both the forward and reverse primers were also removed from each of the sequences. An additional filtering step was applied to remove the chimeric sequences by using UCHIME [32] of USEARCH [33]. The filtering

steps were performed using the QIIME pipeline [34]. Strict dereplication (clustering of duplicate sequences) were performed on the filtered sequences, and then, they were sorted by decreasing number of abundance [35-37]. For each metagenome, the clustering of OTUs was done 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 released 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 sub-databases 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 contains 374 sequences, the non-human gut species database contains 256 sequences, and the human species not reported in gut database contains 237 sequences.

Taxonomic assignments

For the taxonomic assignments, we applied at least 20 reads per OTU. The OTUs were then searched against each database by using the BLASTN [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.

Reference list

1. Lagier, J.C. et al. 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**: 1185-93 (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. et al. 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 U S A* 2005; **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**:543-6 (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**:822-8 (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. et al. Analyses of Intestinal Microbiota: Culture versus Sequencing. *ILAR J* **56(2)**: 228-40 (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).

Supplementary information reference list:

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. et al. 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* 2014; **2(1). pii**: e01181-13 (2014).

27. Dridi, B. et al. 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. et al. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* **13**: 31-13 (2012).

32. Edgar, R.C. et al. 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. et al. Basic local alignment search tool. *J Mol Biol* **215**: 403-10 (1990).

Figure and Table legends

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 5 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 labs but not by culturomics.

Figure 2: Summary of the culturomics work that has extended the gut repertoire and filled some of the gaps in metagenomics.

Figure 3: Phylogenetic tree of the 247 new prokaryote species isolated by culturomics. Bacterial species from Firmicutes are highlighted in red, Actinobacteria (green), Proteobacteria (blue), Bacteroidetes (purple), Synergistetes, Fusobacteria and Archaea (black), respectively. The sequences of 16 prokaryotic species belonging to 6 phyla, previously known from human gut, and more frequently isolated by culture in human gut are highlighted by a star.

Correspondence and requests for materials should be addressed to D.R. (didier.raoult@gmail.com)

Acknowledgements:

We thank Prof. René Valero, Dr. Asif Ahmad Jiman-Fatani, Bouli Ali Diallo, Jean-Bernard Lekana-Douki, Bruno Senghor, Aurore Derand, Laurène Gandois, Floriane Tanguy, Sophie Strouk, Coraline Tamet, Florent Lunet, Myriam Kaddouri, Louisa Ayoub, Lucien Frégère, Noémie Garrigou, Anne Pfleiderer, Anne Farina and Vanessa Ligonnet for technical support. This work was funded by IHU Méditerranée Infection as a part of a Foundation Louis D grant and by the Deanship of Scientific Research (DSR), King Abdulaziz University, under grant No. (1-141/1433 HiCi).

Author contributions

D.R. conceived and designed the experiments; J.C.L., S.K., M.T.A., N.S., N.D., P.H., A.C., F.C., S.I.T., E.S., G.D., G.D., G.M., E.G., A.T., S.B., D.B., N.C., F.B., J.D., M.M., D.R., M.B., N.P.M.D., S.D.B., C.V., D.M., K.D., M.M., C.R., J.M.R., B.L., 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; and J.C.L., A.L. and D.R. wrote the manuscript. All authors read and approved the final manuscript.

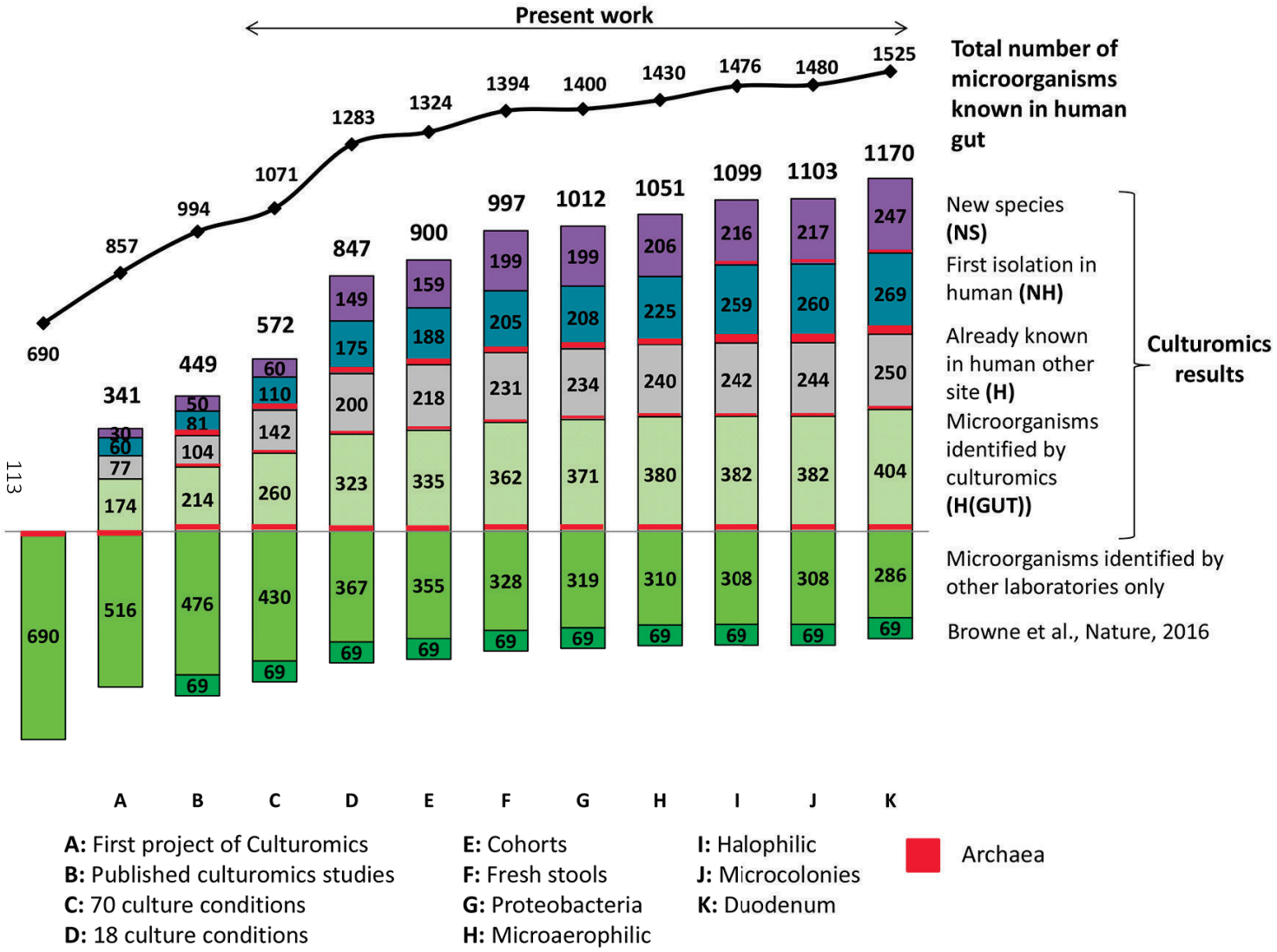
Additional Information

Supplementary information is available online. Reprints and permissions information is

available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.R.

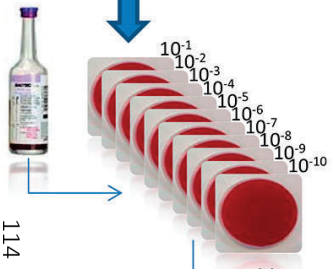
Competing interests

The authors declare no competing financial interest


Present work
Total number of microorganisms known in human gut
New species (NS)
First isolation in human (NH)
Already known in human other site (H)
Microorganisms identified by culturomics (H(GUT))
Culturomics results
Microorganisms identified by other laboratories only
Browne et al., Nature, 2016
690
857
994
1071
1283
1324
1394
1400
1430
1476
1480
1525
690
341
449
572
847
900
997
1012
1051
1099
1103
1170
113
A
B
C
D
E
F
G
H
I
J
K
A: First project of Culturomics
B: Published culturomics studies
C: 70 culture conditions
D: 18 culture conditions
E: Cohorts
F: Fresh stools
G: Proteobacteria
H: Microaerophilic
I: Halophilic
J: Microcolonies
K: Duodenum
Archaea

Extension of human gut repertoire by culturomics

973 samples

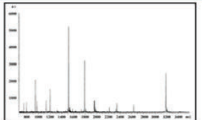


114

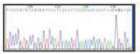
J1

MALDI-TOF
901,364 colonies

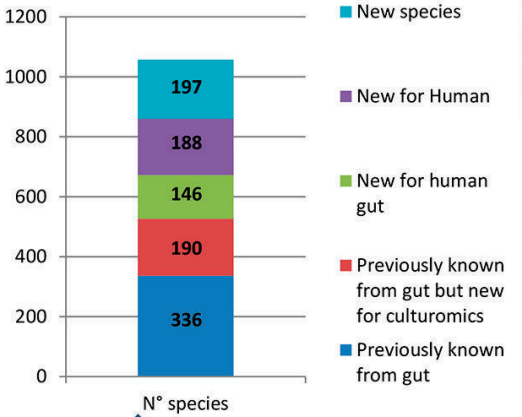
2.7 million spectra



1,258 16S rRNAs of unidentified colonies



945 different prokaryotes including 2 archaea



Decipher metagenomic gaps

1) Comparison of 16S rRNA of our 247 new species (197 + 50 previously published) with HMP

125 of our species previously detected as OTU by metagenomics studies

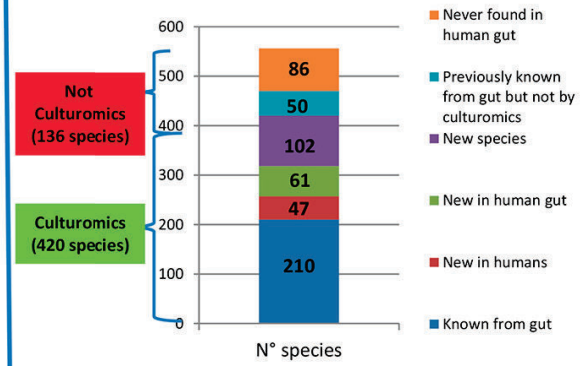
2) 19,980 new ORFans genes including 1,326 from 54 of our new species

3) From 7.7 to 60.7% of our new species detected in Nielsen and Browne metagenomic studies, respectively

Comparison of 84 samples analysed by metagenomics and culturomics

4) Among the 200 16S rRNAs of the new species : 102 recovered 827 times (average 9.8 per stool)

5) Analysis of the species with a cut-off of 20 reads = 4,158 OTU and 556 species



Partie IV: Identification de probiotiques potentiels dans la malnutrition sévère aiguë par culturomics et métagénomique.

Avant-propos

Plusieurs études ont exploré le microbiote digestif des enfants malnutris principalement par approche de métagénomique et pour les études pionnières, par approche de culture. Ainsi, il a été mis en évidence un enrichissement en bactéries aérotolérantes qui s'accompagne d'un enrichissement en *Proteobacteria* et en espèces pathogènes (14,28,29). Cet enrichissement en bactéries pathogènes est cohérent avec la susceptibilité accrue aux infections observées chez les patients malnutris. Une perte drastique de la diversité anaérobie stricte avec une absence totale de l'archée méthanogène, *Methanobrevibacter smithii* est également observée dans le microbiote digestif des malnutris (30,31).

Le traitement de la malnutrition sévère aiguë se fait par renutrition grâce à un aliment thérapeutique prêt à être utilisé en ambulatoire d'après les recommandations de l'OMS (32) associée à une antibiothérapie à large spectre dans les cas les plus extrêmes (33). La dysbiose observée dans la malnutrition sévère aiguë et le taux de décès inévitable malgré le traitement suggère la nécessité d'optimiser le traitement actuel en réhabilitant un microbiote sain. Dans le but d'identifier les espèces manquantes dans le microbiote digestif, nous avons étudié par culturomics et par métagénomique, deux approches complémentaires, le microbiote

fécal d'enfants sains et d'enfants malnutris. La comparaison des deux phénotypes a permis de mettre en évidence, par culturomics et par métagénomique, une perte globale de la diversité connue comme inconnue (nombre de nouvelles espèces et nombre d'OTU non assignées) qui s'accompagne un enrichissement significatif en *Proteobacteria* et de l'espèce potentiellement pathogène, *Streptococcus gallolyticus* ainsi qu'une diminution significative de la diversité anaérobie stricte. De plus, parmi les bactéries identifiées par les deux techniques, 45 espèces sont manquantes chez les malnutris parmi lesquelles 13 espèces ayant un potentiel probiotique.

La réhabilitation du microbiote digestif par la greffe fécale a révolutionné dans les pays occidentalisés la prise en charge des infections à *Clostridium difficile* et est une voie prometteuse de la décolonisation digestive des patients porteurs de bactéries multirésistantes. Cependant, la greffe fécale est difficilement applicable à large échelle (34). La possibilité d'administrer un complexe multi-microbien composé de probiotiques adapté à la malnutrition sévère aiguë en remplacement de la greffe fécale pour rétablir un microbiote sain. Pour cela, des études interventionnelles de microbio-thérapeutique représentent le futur de la recherche dans ce domaine.

Article 4: Identification of Probiotics Candidates for the Treatment of Severe Acute Malnutrition through Culturomics and Metagenomics.

Maryam TIDJANI ALOU†, Matthieu MILLION†, Sory Ibrahima TRAORE, Donia MOUELHI, Saber KHELAIFIA, Dipankar BACHAR, Aurelia CAPUTO, Jeremy DELERCE, Souleymane BRAH, Daouda ALHOUSSEINI, Cheikh SOKHNA, Catherine ROBERT, Bouli Ali DIALLO, Aldiouma DIALLO, Philippe PAROLA, Jean-Christophe LAGIER, Didier RAOULT.

†Co-auteurs à contribution égale.

Under review chez Microbiome

Microbiome

Identification of Probiotics Candidates for the Treatment of Severe Acute Malnutrition through Culturomics and Metagenomics

--Manuscript Draft--

Manuscript Number:	MBIO-D-16-00247
Full Title:	Identification of Probiotics Candidates for the Treatment of Severe Acute Malnutrition through Culturomics and Metagenomics
Article Type:	Research
Funding Information:	Fondation Méditerranée Infection — Not applicable
Abstract:	Background Severe acute malnutrition (SAM) is the world-leading cause of children under-five's death. Recent metagenomics studies have established a link between gut microbiota and SAM, describing an immaturity with a striking depletion in oxygen sensitive prokaryotes and an increased fecal redox potential. In this study, samples of SAM patients and healthy children were collected from Niger and Senegal. Using culturomics with metagenomics as a complementary approach to unravel the SAM-associated gut microbiota specificity, we explored the gut microbial diversity of kwashiorkor patients and healthy infants in order to identify the missing microbes in kwashiokor patients. Results We found a global decreased diversity, a decrease in the hitherto unknown diversity (new species isolation), a depletion in oxygen sensitive prokaryotes including Methanobrevibacter smithii and an enrichment in potentially pathogenic Proteobacteria, Fusobacteria and Streptococcus gallolyticus; 13 species identified only in healthy children using culturomics and metagenomics were identified as probiotics candidates, providing a possible safe and convenient alternative to fecal transplant to restore healthy gut microbiota in malnourished children. Conclusions A dysbiosis of the gut microbiota is highlighted in this study. The importance of culture-dependent methods in addition of metagenomics, to study the gut microbiota is also evidenced since 13 possible probiotics in SAM were isolated. Microbiotherapy based on selected strains is expected to improve the current treatment of SAM with an improved recovery rate and a reduced mortality rate by reestablishing a healthy gut microbiota.
Corresponding Author:	Didier Raoult FRANCE
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	
Corresponding Author's Secondary Institution:	
First Author:	Maryam Tidjani Alou
First Author Secondary Information:	
Order of Authors:	Maryam Tidjani Alou Matthieu Million Sory Ibrahima Traore Donia Mouelhi Saber Khelaifia Dipankar Bachar

	Aurelia Caputo
	Souleymane Brah
	Daouda Alhousseini
	Cheikh Sokhna
	Jeremy Delerce
	Bouli Ali Diallo
	Aldiouma Diallo
	Catherine Robert
	Philippe Parola
	Jean-Christophe Lagier
	Didier Raoult
Order of Authors Secondary Information:	

Identification of Probiotics Candidates for the Treatment of Severe Acute Malnutrition through Culturomics and Metagenomics

Maryam TIDJANI ALOU[1,5†], Matthieu MILLION[1†], Sory Ibrahima TRAORE[1,2], Donia MOUELHI[1], Saber KHELAIFIA[1], Dipankar BACHAR[1], Aurelia CAPUTO[1], Jeremy DELERCE[1], Souleymane BRAH[3], Daouda ALHOUSSEINI[3], Cheikh SOKHNA[4], Catherine ROBERT[1], Bouli Ali DIALLO[5], Aldiouma DIALLO[4], Philippe PAROLA[1], Jean-Christophe LAGIER[1], Didier RAOULT[1*].

[1]Aix-Marseille Université, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM63, CNRS 7278, IRD 198, INSERM 1095, 27 Boulevard Jean Moulin, 13385 Marseille, Provence-Alpes-Côte d'Azur, France.

[2]Département d'Epidémiologie des Affections Parasitaires, Faculté de Médecine, Université des Sciences, des Techniques et Technologies de Bamako, BP 1805 Bamako, Mali

[3]Hôpital National de Niamey, BP 238 Niamey, Niger

[4]Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes IRD 198, CNRS 7278, Aix-Marseille Université, Marseille, France; Campus Commun UCAD-IRD of Hann, Dakar, Senegal

[5]Laboratoire de microbiologie, Département de biologie, Université Abdou Moumouni de Niamey, BP 237, Niamey, Niger.

[†]Equally contributing authors

[*]Corresponding author: Didier Raoult (didier.raoult@gmail.com)

Email addresses:

MTA: tidjani_maryam@hotmail.com

MM: matthieumillion@gmail.com

SIT: tisory@yahoo.fr

DM: mouelhidonia@gmail.com

SK: khelaifia_saber@yahoo.fr

DB : dipankar3@gmail.com

AC : aurelia.caputo@univ-amu.fr

JD: jeremy.delerce@gmail.com

SB: brahsouleymane@yahoo.fr

DA: amd145@yahoo.fr

CS: cheikh.sokhna@ird.fr

CR: catherine.robert@univ-amu.fr

BAD : bouli_diallo48@yahoo.fr

AD: aldiouma.diallo@ird.fr

PP: philippe.parola@univ-amu.fr

JCL: jclagier@yahoo.fr

Abstract

Background

Severe acute malnutrition (SAM) is the world-leading cause of children under-five's death. Recent metagenomics studies have established a link between gut microbiota and SAM, describing an immaturity with a striking depletion in oxygen sensitive prokaryotes and an increased fecal redox potential. In this study, samples of SAM patients and healthy children were collected from Niger and Senegal. Using culturomics with metagenomics as a complementary approach to unravel the SAM-associated gut microbiota specificity, we explored the gut microbial diversity of kwashiorkor patients and healthy infants in order to identify the missing microbes in kwashiokor patients.

Results

We found a global decreased diversity, a decrease in the hitherto unknown diversity (new species isolation), a depletion in oxygen sensitive prokaryotes including *Methanobrevibacter smithii* and an enrichment in potentially pathogenic *Proteobacteria*, *Fusobacteria* and *Streptococcus gallolyticus*; 13 species identified only in healthy children using culturomics and metagenomics were identified as probiotics candidates, providing a possible safe and convenient alternative to fecal transplant to restore healthy gut microbiota in malnourished children.

Conclusions

A dysbiosis of the gut microbiota is highlighted in this study. The importance of culture-dependent methods in addition of metagenomics, to study the gut microbiota is also evidenced since 13 possible probiotics in SAM were isolated. Microbiotherapy based on selected strains is expected to improve the current treatment of SAM with an improved recovery rate and a reduced mortality rate by reestablishing a healthy gut microbiota.

Keywords: Severe acute malnutrition; kwashiorkor; gut microbiota; culturomics; metagenomics; probiotics; *Methanobrevibacter smithii*; *Streptococcus gallolyticus.*

Abbreviations

CTL: Control

HAZ: Height-for-age z-score

HMO: Human Milk Oligosaccharides

KWA: Kwashiorkor

MALDI TOF MS: Matrix Assisted Laser Desorption/ Ionization Mass Spectrometry

MUAC: Mid-Upper Arm Circumference

OTU: Operational Taxonomic Unit

RUTF: Ready To Use Therapeutic Food

SAM: Severe Acute Malnutrition

SD: Standard Deviation

WAZ: Weigth-for-Age z-score

WHZ: Weigth-for-Height z-score

1. Background

Undernutrition is the worldwide leading cause of mortality for children under five accounting for 1 to 6 million deaths every year [1]. Moreover, severe acute malnutrition (SAM) affects 20 million children, mostly from developing countries of sub-Saharan African, Central America and South Asia [2]. WHO defines SAM using anthropometric indicators such as mid-upper arm circumference (MUAC), height-for-age z-score (HAZ), weight-for-age z-score (WAZ) and weight-for-height z-score (WHZ) assessing stunting, wasting, and underweight[3,4]. SAM is diagnosed by a -3SD below WHO median WHZ score and/or a mid-upper arm circumference under 115 mm [4]. In addition, clinical signs such as bilateral oedema allows a distinction between oedematous and non-oedematous SAM [4].

Oedematous SAM was first described by Cicely D. Williams in 1933 who named it kwashiorkor, a disease usually seen in improperly weaned children between 1 and 4 years old. Clinical signs included nutritional oedema, wasting, diarrhea, irritability and skin rashes [5]. Clinical signs of non-oedematous SAM (marasmus) were described later as severe wasting with loss of subcutaneous fat, muscle atrophy and a shrunken and wizened face [6]. Both type of malnutrition used to be regrouped under the term protein energy malnutrition because of the alleged association with an inadequate intake of protein, qualitative in the case of kwashiorkor and quantitative in the case of marasmus [7,8]. Nevertheless, the etiology of kwashiorkor was incompletely explained by the deficiency of protein in the diet thus causing a recent differentiation from marasmus and the emergence of several other hypothesis concerning its physiopathology [9].

The presence of oedema in SAM has been associated with absence of breastfeeding, lower household dietary diversity score and lower fish, nuts, dairy products, green leafy vegetables and fresh fruits consumption [10]. Nowadays, gut microbiota is considered an instrumental factor in kwashiorkor pathogenesis [11]. Exploration of the gut microbiota of malnourished

children using metagenomics and molecular approaches revealed an altered microbiota with a loss of diversity [11], more specifically a loss of strict anaerobic species among which *Methanobrevibacter smithii*, the most oxygen-sensitive prokaryote of the human gut which was totally absent. This loss of anaerobic diversity was associated with a high redox potential and an enrichment in aerobic species [12].

Gold standard treatment of uncomplicated SAM has been revolutionized by the use of Ready-to-Use-Therapeutic-Food (RUTF) in an outpatient setting [2]; RUTF is made of peanut paste, milk, vegetable oil and sugar providing adequate levels of energy, proteins, vitamins and minerals. However, about 5% children still die even with optimal updated treatment including antibiotics [13,14].

In order to improve the current treatment protocol and therefore improve recovery and mortality rates of kwashiorkor patients, a restoration of the gut microbiota of malnourished patients to the normal composition in infants would be essential. Fecal transplant is the answer to restoration of a healthy microbiota but its mainstream application is not yet possible [15]. A study conducted on twins revealed the possibility of a twin discordance for kwashiorkor associated with a characteristic fecal bacterial community in the malnourished child [11]. This community associated with local diet caused weight loss in gnotobiotic mice associated with profound metabolic changes in the gut microbiota. This study opens perspectives for replacement of fecal transplant with specific species used as probiotics. There seem to be missing microbes in the gut of kwashiorkor patients which identification is essential [13]. In fact, utilization of the missing repertoire as probiotics requires viable isolated strains which can only be obtained using a culture approach such as “microbial culturomics” which efficacy for high throughput culture and gut microbiota exploration has been proven [16,17].

In this study, we described the gut microbiota of kwashiorkor patients and healthy children using metagenomics and culturomics approaches in order to identify prokaryotes that could be possible probiotics that could restore a healthy gut microbiota in malnourished children. The complementarity between metagenomics and culturomics [16] makes these two techniques particularly adapted to this study, enabling a better exploration of the fecal samples and the possibility to have probiotics available at the end of this study. We here assign a more active role to the gut microbiota by making it part of the solution (treatment) rather than being part of the problem (cause). With such an objective, this study differs from previous studies of our team concerning the gut microbiota which had a more descriptive approach.

2. Results

2.1. Population and descriptive culturomics and metagenomics results

Ten kwashiorkor patients (KWA) and 5 controls (CTL) were selected. Six patients and 3 CTL were recruited in Senegal and 4 patients and 2 CTL were recruited in Niger, so that the proportion of KWA patients originated from Niger and Senegal was identical (60%). No significant difference was observed for the age and sex of the KWA patients and CTL (Table 1). All 15 samples were analyzed using 18 culture conditions. For KWA patients, 144,330 colonies were isolated and tested using MALDI TOF MS with a mean of 14,433 colonies per sample (standard deviation (SD), 2975). For CTL samples, 59,578 colonies were tested using MALDI-TOF MS with an average of 11,916 ± 404 colonies per sample. A total of 335 species were isolated in KWA samples whereas 281 species were isolated in CTL samples. All 15 samples were analyzed by 16S rRNA-targeted metagenomics. A total of 2,933,416 and 1,842,831 reads were generated from the 10 KWA and 5 CTL samples, respectively.

2.2.Decreased diversity in kwashiorkor patients

With culturomics, taking into account the hitherto unknown diversity (see section 4.4) and the species previously reported in the human gut, the alpha and beta diversities were estimated in the KWA and CTL groups. The β-diversity [18] estimated by the U/T ratio was used to compare KWA and CTL groups. The U/T ratio (see Section 4.5) was significantly lower in the KWA group indicating a lower β-diversity (151/335 (45%) in KWA samples vs 185/281 (66%) in CTL group, uncorrected bilateral chi squared test, $p < 3 \times 10^{-7}$, Table 2). 45 species not known from the human gut were isolated from the KWA group (n = 10) including 9 new species and 9 new genera (Table 3), 15 are known but not previously found in humans and 12 are already known in humans but not previously found in the gut. In our group of 5 CTL, we isolated 46 species unknown from the human gut including 26 new species (Table 3), among which 8 new genera and 1 new family (*Neofamiliaceae* fam. nov.), 14 are known but not previously found in humans and 6 are already known in humans but not previously found in the gut. The hitherto unknown diversity assessed by culturomics was dramatically decreased in the KWA group (Table 4). However, the difference was significant only for new species (mean number of new species found by sample ± standard deviation, 1.8 ± 1.5 in KWA vs 5.0 ± 2.6 in CTL, unpaired bilateral student test, p = 0.009, Fig. 1) and for previously known species but not previously found in humans (1.5 ± 1.2 for KWA vs 2.8 ± 0.4 for CTL, bilateral Mann Whitney test, p = 0.02, Table 4) suggesting a decreased α-diversity.

With metagenomics, calculating the mean Shannon Index ± SD by group showed that the global diversity was decreased in KWA even if not significantly (3.2 ± 0.8 vs 3.8 ± 0.8 in CTL, bilateral student test, p = 0.19, Table 5), consistently with previous studies [19]. The hitherto unknown diversity assessed was consistently and significantly decreased in KWA as unidentified OTUs were lower in the KWA group (Fig. 1). Accordingly, at the prokaryotic level, only 5% of all reads in the KWA group were not assigned versus 26% in CTL patients

(percentage of reads unassigned at the prokaryotic level ± SD, 0.05 ± 0.02 for KWA vs 0.26 ± 0.22 for CTL, p = 0.009).

When cumulating the results obtained with culturomics and metagenomics, 771 species were identified in the KWA samples among which only 148 (19%) species were isolated with the two techniques. As for CTL samples, 651 species were identified among which 110 (17%) by the two techniques. The proportion of species identified with the two techniques was not significantly different (148/771 (19%) in KWA vs 110/651 (17%) in CTL, uncorrected bilateral Chi square test, p= 0.26), suggesting a complementarity between metagenomics and culturomics approaches and a congruence of this complementarity among KWA and CTL.

2.3. Loss of anaerobic species in kwashiorkor patients associated with reduced fecal redox potential.

Anaerobic species are lost in Kwashiorkor patients

According to the 'culturomics' results, both aerotolerant and anaerobic β-diversities were significantly lower in KWA (Table 2) but the decrease of anaerobic β-diversity (-29%, U/T ratio, 43/111 (39%) in KWA vs 76/112 (68%) in CTL, p = 0.00001) was more important than the decrease in aerotolerant β-diversity (-16%, 108/224 (48%) in KWA vs 109/169 (64%) in CTL, p = 0.001). The difference was significant (Δ aerotolerant β-diversity/Δ anaerobic β-diversity, One-sample test for binomial proportion, normal-theory method, $p < 10^{-7}$).

The metagenomics analysis showed an increased aerotolerant α-diversity but not significantly (Shannon index ± standard deviation, 2.07 ± 0.8 in KWA vs 1.35 ± 0.6 in CTL, unpaired bilateral student test, p = 0.1, Table 5) while the anaerobic diversity was significantly different and decreased in KWA (1.05 ± 0.98 vs 3.1 ± 1.5, bilateral Mann

Whitney test, p = 0.02, Table 5). These results confirmed the drastic decrease in anaerobic diversity found by culturomics.

Absence of Methanobrevibacter smithii associated with a reduced gut redox

Real-time PCR quantification showed also an overwhelming absence of *M. smithii* in KWA patients although it was detected in all CTL patients (Fig. 2). The redox potential per sample was also higher than in CTL patients though the difference between the two groups was not significant (-0.6 ± 57.6 in KWA patients vs -17.1 ±54.5 in CTL patients, unpaired bilateral student test, p = 0.6) suggesting a more oxidized gut environment in KWA patients (Fig. 2). These results are consistent with the previously stated results and the fact that *Euryarchaeota* are among the most oxygen-sensitive prokaryotes in the gut.

2.4. *Proteobacteria* and *Streptococcus gallolyticus* increase in kwashiorkor

With culturomics, five bacterial phyla were isolated in KWA with a majority of *Firmicutes* (208 species), followed by 56 *Proteobacteria*, 47 *Actinobacteria*, 21 *Bacteroidetes* and 3 *Fusobacteria*. A total of 108 genera were isolated including *Clostridium* (46 species), *Bacillus* (28), *Streptococcus* (18), *Staphylococcus* (17), *Enterococcus* (13), *Paenibacillus* (11), *Lactobacillus* (13) and *Corynebacterium* (9). In CTL samples, only 4 phyla were isolated: *Firmicutes* (192 species), *Actinobacteria* (42), *Proteobacteria* (25) and *Bacteroidetes* (21). Strikingly, there was no *Fusobacteria* species. A total of 91 genera were identified among these 4 phyla. The most represented were *Clostridium* (45 species), *Bacillus* (30), *Paenibacillus* (18), *Streptococcus* (12), *Staphylococcus* (9) and *Lactobacillus* (8) (Table 4). A significant increase in the frequency of *Proteobacteria* was found in KWA (56/335 (17%) vs 25/281 (9%) in CTL, uncorrected bilateral chi-square test, p = 0.004, Table 2). There was no significant difference for other phyla (Table 2). At the species level, significantly enriched species in KWA (Fig. 3) included *Bacteroides thetaiotaomicron, Bifidobacterium breve, Bifidobacterium catenulatum, Gemella haemolysans, Hafnia alvei, Rothia aeria,*

Staphylococcus hominis, *S. gallolyticus* and *Streptococcus lutetiensis*. All of these species are susceptible to β-lactam antibiotics [20–24] which are used in kwashiorkor treatment [14,25]. Species such as *G. haemolysans*, *H. alvei, S. gallolyticus* and *S. lutetiensis* are potentially pathogenic and have been previously associated with diarrhea, gastroenteritis and endocarditis [26–28].

The metagenomics analysis showed that in KWA samples, 2,783,881 reads assigned at a prokaryotic species level were distributed into 9 phyla (*Actinobacteria, Bacteroidetes, Chloroflexi, Firmicutes, Fusobacteria, Planctomycetes, Proteobacteria, Synergistetes and Verrucomicrobia*). CTL samples generated 1,333,589 reads assigned at a prokaryotic species level and divided into the following 8 phyla: *Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria, Lentisphaerae, Proteobacteria, Tenericutes* and *Verrucomicrobia*. *Chloroflexi, Planctomycetes* and *Synergistetes* were detected only in KWA patients whereas *Lentisphaerae* and *Tenericutes* were detected only in CTL patients. These reads matched 589 species in the KWA group and 486 in the CTL group. *Proteobacteria* were also detected more frequently in KWA (131/589 (22%) vs 75/486 (15%) in CTL, uncorrected bilateral chi square test, p = 0.004, Table 6) thus confirming the culturomics results. Interestingly, among *Proteobacteria*, the difference was significant only for alpha-*Proteobacteria* (22/589 (3.7%) vs 2/486 (0.4%), p = 0.0002, Table 6). *Firmicutes* were detected less frequently (319/589 (54%) vs 296/486 (61%), p = 0.026, Table 6) and *Euryarchaeota* were not detected in KWA (0/589 (0%) vs 4/486 (0.8%), p = 0.027, Table 6).

At the genus level, *Streptococcus* were more frequent in KWA (40/589 (6.8%) vs 23/486 (4.7%) in CTL, uncorrected bilateral chi square, p = 0.029) whereas *Prevotella* and *Bacillus* were less frequent in KWA (2/589 (0.3%) vs 14/486 (2.9%) in CTL, p = 0.0006 and 11/589 (1.9%) vs 28/486 (5.7%), p = 0.0006, respectively, Table 6). At the species level, the two species with the largest increased frequency in KWA belong to the *Streptococcus* genus

including *Streptococcus peroris* (9/10 (90%) in KWA vs 1/5 (20%) in CTL, bilateral Barnard test p = 0.009) and *S. gallolyticus* (7/10 (70%) in KWA vs 0/5 (0%) in CTL, bilateral Barnard test p = 0.014, Fig. 4).

Interestingly, *S. gallolyticus*, a well-known pathogen associated both with endocarditis and colon cancer [13,29–31], was previously associated with the malnourished children's gut microbiota in the largest metagenomics study to date [19] and was the only species increased in KWA by both culturomics and metagenomics.

2.5. Comparison of metagenomics results with the literature.

When comparing metagenomics data from the literature on malnourished and healthy children (including a study performed by our team), we only found 9 core species among malnourished children and 10 core species among healthy children (Fig. 5). Four species were common to both cores: *Bacteroides fragilis, Lactobacillus ruminis, Megasphaera elsdenii* and *Prevotella copri*. This data set highlights a very low reproducibility between metagenomics studies.

2.6. Missing repertoire in kwashiorkor patients

Species identified by metagenomics and culturomics in KWA and CTL patients were compared and 45 species were identified only in CTL samples (Table 7). These species belonged overwhelmingly to the *Firmicutes* phylum (32) followed by a few species from the *Actinobacteria* (5), the *Bacteroidetes* (4) and the *Proteobacteria* (4) phyla. Among the missing repertoire, strikingly, 23 species (51%) were strictly anaerobic.

For each of these species, we searched through the literature to find a possible probiotic use. Thirteen species , 10 *Firmicutes*, 2 *Bacteroidetes* and 1 *Actinobacteria* were found to have possible probiotic features or were representative of a healthy flora: *Alistipes indistinctus, Anaerostipes caccae, Bacillus licheniformis, Bacillus subtilis, Bacteroides salyersiae, Bifidobacterium adolescentis, Clostridium hylemonae, Intestinimonas*

butyriciproducens, Lactobacillus perolens, Lactobacillus parabuchneri, Lactobacillus vaccinostercus, Terrisporobacter glycolicus and *Weissella confusa* (Table 8). Probiotic features included short chain fatty acid production, antioxidant metabolism, bile acids detoxification and antibacterial potential. Each of these species was isolated in our CTL group of samples and is readily available in the CSUR collection of our laboratory.

3. Discussion

In this culturomics and metagenomics study, we were able to identify a missing gut repertoire in kwashiorkor patients through a characterization of the gut microbiota of kwashiorkor patients as opposed to healthy children microbiota. Thirteen species emerged from this analysis as a possible microbial complex which could be utilized as probiotics in SAM. Characteristics of the kwashiorkor-associated gut microbiota included a depletion of the hitherto unknown, global and anaerobic diversity and an enrichment in potentially pathogenic and oxidative stress resistant *Fusobacteria*, *Proteobacteria* and *S. gallolyticus* in kwashiorkor. The depletion of anaerobic diversity was associated with an absence of *M. smithii* and a higher redox potential. The only consistent and significant result at the species level obtained by both culturomics and metagenomics was an enrichment in *S. gallolyticus*, a potentially highly virulent pathogen responsible of colon cancer and endocarditis [29–31].

As determined by the strict application of the WHO criteria to define child growth standards [4], each of our kwashiorkor patients was a textbook case and each of our controls was a healthy child under five. Cases and controls included in this study originated from 2 different geographic locations thus allowing a generalization of the results. Moreover, every sample was analyzed using the same protocol for metagenomics with a specific DNA extraction protocol [32] and used with or without prior deglycosylation of the sample [33]; the same 18 culture conditions were also used for culturomics. The “microbial culturomics” method which efficacy for exploring the gut microbiota is no longer to be proven

[16,17,34,35] presents like other culture approaches [36] the tremendous advantage over metagenomics of providing a physical collection of strains on which further analysis can be carried out. Metagenomics studies are the preferential technique for the exploration of the gut microbiota diversity but have a very low reproducibility among studies probably due to the differences in sampling, sample conservation, DNA extraction protocol, sequencing method and data analysis strategy [37]. Moreover, culture approaches allow the extension of the gut microbiota known diversity and functions [36]. A complementarity between microbial culturomics and metagenomics has also been highlighted in several studies [16,17,34]. Furthermore, congruence between the two techniques is observed with similar characteristics in the gut microbiota of malnourished children such as a loss of global diversity associated with a loss of anaerobic diversity and unknown diversity alongside an enrichment of species belonging to the *Proteobacteria* phylum despite a low overlap between the species identified using the two techniques. In fact, only 17 to 19% of species were identified with both techniques in KWA and CTL samples respectively (see Section 2.1).

Recent studies have suggested an influence of the gut microbiota on the growth of infants. In fact, the mere presence of the gut microbiota seems to be necessary for weight gain and skeletal growth in post-natal infants [38,39]. Indeed, based on previous studies, fecal transplantation was suggested to treat kwashiorkor [40]. Composition of the gut microbiota is greatly variable during the first two to three years of an infant life before reaching its mature state [41,42] and it has recently been shown that the influence of the gut microbiota on growth and weight gain is age-specific [38] and more specifically strain-specific [39]. Characterizing the gut microbiota of malnourished and healthy children through a cultivation approach thus becomes a necessity in order to, first, have a better knowledge of the composition of their gut microbiota and, second, to identify species associated with children health and missing in KWA patients. These species could serve as probiotics which are defined as live

microorganisms which when administrated in adequate amounts can have a beneficial effect on the health of an individual [43]. Identifying and isolating probiotic species can revolutionize the actual treatment for SAM by adding an antibiotic course to eradicate pathogenic species and transplanting probiotic species to the current renutrition protocol. Thirteen bacterial species identified by culturomics and metagenomics and specific to healthy children were revealed. These species (Table 4) are dominant in the normal gut microbiota of infants [44,45] and adults [37,46]. The other species exhibit other useful functions such as production of antimicrobial compounds [47,48], metabolization of antioxidants [49], bile acids to improve the small intestine as a suitable growth environment for bacteria [50] and mutualistic association with commensals of the gut microbiota [51]. Probiotic species should be able to survive even transiently in the gastrointestinal tract and therefore resist low pH values and bile acids. Additionally, these strains should be non-pathogenic, produce antibacterial species and be able to adhere to gut epithelial tissue. *W. confusa* (Table 8) was described as a probiotic fulfilling all the aforementioned conditions [52–55]; so were *B. licheniformis* and *B. subtilis* (Table 8) in animals [56]. Administration of these potential probiotics associated with *M. smithii* can replace fecal transplant as a way to restore a healthy gut microbiota in malnourished children and provide an easy addition to current SAM treatment as opposed to fecal transplantation [15].

A rehabilitation of the gut microbiota of malnourished children is necessary since a dysbiosis is observed with a significant increase in *Proteobacteria* in KWA patients observed in this study through culturomics and metagenomics approaches. The abundance of *Proteobacteria* was observed in other studies using both a cultivation approach (Mata et al., 1972; Million et al., 2016b) and a metagenomics approach (Ghosh et al., 2014; Monira et al., 2011; Subramanian et al., 2014). These studies have also shown an increase of pathogenic bacteria like by *S. gallolyticus* in the metagenomics approach (Subramanian et al., 2014) and

of members of the *Shigella, Edwardsiella* and *Salmonella* genera by a cultivation approach (Mata et al., 1972; Million et al., 2016b) in the gut of malnourished children. *S. gallolyticus*, which was also increased in the gut of our patients, are sensible to large spectrum β-lactam antibiotics, such as ampicillin and cephalosporins as well as many other pathogenic species belonging to the *Proteobacteria* phylum and *Fusobacteria* species (Nomoto et al., 2013; Poulsen et al., 2012; Roberts et al., 2006; Stock et al., 2005), showing how these antibiotics drastically improve recovery and mortality rates in malnourished patients (Million et al., 2016b, 2016c; Trehan et al., 2013).

The depletion of the hitherto unknown and anaerobic diversity most likely represents a depletion of micronutrients (minerals, prebiotics, anti-oxidants) and a loss of the reduced environment which both are culture requirements for most strict anaerobes. The depletion of anaerobic species in the gut microbiota of KWA patients is here observed for the first time using a culture-based approach, except for a study by Mata *et al* in 1972, realized at a low scale, which reported a decrease in anaerobic species in the proximal gut and the feces of malnourished children. This depletion in anaerobic species has previously been linked to a default in the normal colonization of the gut by *M. smithii* which growth depends heavily on hydrogen producing species and a minimal to absent oxygen tension in the environment [12,57,58]. It was also linked to an increased redox potential in malnourished patients and therefore to a reduced pH [12]. A shift in pH values has previously been associated with an abnormal flora [59] as observed in KWA patients.

The low inter-individual variability observed in KWA patients is probably due to a similarly low variability in the diet of kwashiorkor patients. In fact, the diet of children in regions with a high SAM prevalence is often made of starchy food like cassava, yam, maize and banana among others and characterized by a low intake in animal products [60] resulting in low uric acid production and therefore low antioxidant content. Abnormal breastfeeding is

also a risk factor for SAM [5,13] and specifically oedematous malnutrition [10]. As breast milk has a high antioxidant capacity [61] and include Human-specific highly diverse Milk Oligosaccharides (HMOs) which are prebiotics shaping the infant gut microbiota [62], it can be suspected that a deficient breastfeeding prevents the implantation of the Healthy Mature Anaerobic Gut Microbiota [13] resulting in a microbial aerotolerant and pathogenic chaos. Indeed, sialylated and fucosylated HMOs which seem to promote infant growth are less represented in the breast milk of mothers whose children develop SAM [63].

4. Conclusions

In this study, in addition to a characterization of the gut microbiota of KWA patients, missing species were identified as potential probiotics. In order to test the efficacy of these possible probiotics, we aim to develop an experimental model with axenic mice on which this cocktail of probiotics would be tested. We would expect this cocktail to result in weight gain in these mice. All studies concerning the gut microbiota of SAM patients are passive, exploratory studies. It is time to apply the knowledge we gained on SAM gut microbiota on interventional studies investigating the possibility of associating microbiotherapy with renutrition to treat SAM.

5. Methods

5.1. Population and samples

10 severely undernourished children with nutritional edema (kwashiorkor) and 5 healthy children without any criteria of malnutrition (no wasting, no stunting and not underweight), according to the 2009 WHO criteria [4], were recruited in Senegal and Niger (Table 9). Two African centers were included to test geographical generalization (Fig. 6). The aforementioned number of samples was chosen because of the time requirement for culturomics analysis of each sample (12,000 colonies isolated by 18 culture conditions after inoculation at day 1, 3, 7, 10, 15, 21 and 30 of the pre-incubated fecal sample corresponding

to a 6 weeks' protocol for each sample). More cases than CTL were included since we favored the characterization of the gut microbiota of malnourished children. 6 and 4 cases were included in the Campus International UCAD/IRD of Hann, Dakar, Senegal and in the Pediatrics emergency room of the National Hospital of Niamey, Niger, respectively. 3 and 2 CTL of same sex and similar age were recruited by a snowball approach in Dakar and Niamey respectively. Oral consent was given by the parents and the study was approved by the local ethics committee IFR48 under agreement number 09-022.

5.2. Microbial culturomics: high-throughput bacterial culture

High-throughput bacterial culture

Samples were analyzed using 18 culture conditions (Fig. 7) with identification of 12,000 colonies by sample as described previously [16,17]. For each condition, 1 gram of each sample was diluted in 9 mL of PBS (ThermoFisher Scientific, Illkirch, France); the suspension was then inoculated in liquid medium and incubated at the chosen temperature atmosphere. At day 1, day 3, day 7, day 10, day 15, day 21 and day 30, 10-fold serial dilutions of the culture were inoculated on 5% sheep blood Agar (Becton Dickinson, Le Pont de Claix, France). A subculture of the resulting colonies was performed for purification. The colonies were then identified using MALDI-TOF/MS [64,65].

High-throughput bacterial identification: MALDI-TOF/MS

The MALDI-TOF identification was performed with a microflex from Bruker Daltonics (Bremen, Germany) according to the manufacturer's instructions. Each colony was deposited in duplicate on a 96 MSP microplate and covered with 2 µL matrix solution. The solution was made of saturated α-cyano-4-hydroxycinnamic acid, 50% acetonitrile and 2.5% trifluoroacetic acid. Each obtained spectrum was matched against the 7562 spectra of Bruker's and Laboratory La Timone's home database (as of January, 2016). A score > 1.9 allowed identification at the species level [64].

Identification of strain unidentified by MALDI-TOF/MS

DNA from the unidentified colonies (MALDI-TOF score < 1.9) was extracted using EZ1 DNA Tissue Kit from Qiagen in an automated EZ1 advanced instrument according to the manufacturer's instructions. The DNA was amplified with primers FD1 and RP2 targeting all bacteria at an annealing temperature of 52°C. The amplified product was purified 96-well purification plate and then re-amplified using the BigDye Terminator v1.1 cycle sequencing kit (Qiagen) at an annealing temperature of 52°C with primers FD1, RP2, 536F, 536R, 800F, 800R, 1050 F and 1050R. Sequences of all the primers are recorded in Table 10. The product was purified and analyzed using an ABI PRISM 3130x Genetic Analyzer (Applied Biosystems). The resulting sequences were analyzed using the software ChromasPro and compared to the NCBI database with the Blast software. Sequences with a similarity percentage under 98.65% or 95% [66] were identified respectively as new species or new genera, respectively, and described according to the taxonogenomics concept [67].

5.3. High throughput sequencing

Total DNA was extracted of the stool samples using 2 protocols [32]. The first one is based on a physical lysis with glass powder followed by an overnight chemical lysis proteinase K. The resulting solution is washed and eluted according to the Macherey-Nagel DNA Tissue extraction kit [32]. For the second protocol, glycoprotein lysis and deglycosylation steps were added to the first protocol [33]. Sequencing of the resulting DNA from these 2 protocols was performed targeting the v3v4 region of the 16s rRNA as previously described [12]. All reads from the two protocols were grouped and clusterized with a threshold of 97 % identity to obtain Operational Taxonomic Units (OTU). All OTUs made of less than 20 reads were removed. Remaining OTU were blasted against Silva123 and assigned to a species if they matched one with at least 97 % identity. OTU not assigned to any species were considered 'unidentified'. As several OTUs matched identical species,

evidencing the absence of biological relevancy of OTUs, the total number of identified species + the number of unidentified OTU was expected to be smaller than the total number of OTUs.

5.4. Definition of the hitherto unknown diversity

As mentioned above, microbial culturomics give an unprecedented opportunity to access and quantify the hitherto unknown microbial diversity. Indeed, 16S targeted metagenomics gives us the possibility to assess and quantify richness and abundance of previously undescribed bacteria, evidencing that most of the richness and diversity of the gut microbiota was unknown to date [68]. We defined the hitherto unknown diversity as the number of new species added to the number of species not previously known from the human gut by sample for culturomics analysis and as the number of unidentified OTU for metagenomics analysis.

5.5. Strategizing

β-diversity is a measure of biodiversity comparing the species diversity between ecosystems along environmental gradients. This involves comparing the number of taxa that are unique to each ecosystem. To assess the microbial diversity by culturomics, we defined the 'unique microbiota' as the group of species found in only one child of each group. Accordingly, we calculated the unique/total microbiota richness (the U/T ratio) of each children group (kwashiorkor or controls) to access this β-diversity. This ratio was not expected to be biased by the different sample size between the 2 groups. Accordingly, higher unique/total microbiota richness (increased U/T ratio) represents a higher diversity. For the metagenomics results, the diversity was estimated by calculating Shannon indexes for all the species, aerotolerant and anaerobic species using the following formula: $H' = \sum pi * \log_2 pi$ where pi is the proportion of each species in the sample which diversity is being estimated [69].

Metagenomics and culturomics results were compared in order to highlight the complementarity between these two methods.

The hitherto unknown global and strict anaerobic diversity of cases and controls was compared using a culturomics approach. The ultimate goal of this study was to determine which microbes were missing in KWA patients. The species identified by both culturomics and metagenomics for each group of samples were compared and the species specific to healthy children were identified and their biological interest investigated to assess their potential usefulness as probiotics in SAM.

5.6. Statistical analysis

For each quantitative analysis, normality was tested using either D'Agostino-Pearson or Kolmogorov-Smirnov tests for a small number of samples. Bilateral unpaired student test or Mann-Whitney test were used according to normality (Gaussian distribution). Bilateral exact Fisher test or uncorrected Chi-squared test were used to compare proportions. Barnard bilateral test was used when sample size was very small [70]. All statistical analysis were performed with SPSS v21.0 (IBM, Paris, France). Correction for false discovery rate was not necessary since this is an exploratory study [71].

6. Declarations

Author contributions

Conceptualization, J.C.L. and D.R. Methodology, S.K. and D.R. Validation, M.M. and D.R. Formal analysis, M.T.A., M.M and D.R. Investigation, M.T.A., S.I.T., D.M. and C.R. Resources, S.B., D.A, C.S and A.D. Data curation, B.D., A.C. and J.D. Writing- Original draft: M.T.A., M.M. and D.R. Writing- Review & Editing: M.T.A., M.M., J.C.L and D.R. Visualization: M.T.A and M.M. Supervision: M.M., S.K., J.C.L, B.A.D and D.R. Project Administration: P.P. and D.R. Funding Acquisition: D.R.

Acknowledgements

We would like to thank E. Adehossi for his help with the sample and clinical data collection in Niger. We would also like to thank all the members of the culturomics team that contributed to this high throughput culture effort, F. Bittar and P-E Fournier for new species validation. Finally, we would like to thank M. Lardière for English reviewing. None of the authors declared a conflict of interest.

Funding

This work was funded by the Mediterranée Infection Foundation which gave no other contribution to the process that led to the submission of this paper.

Availability of data and materials

The datasets during and/or analysed during the current study available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

The children parents gave their oral informed consent to the participation in this study. The study was approved by the Institut Fédératif de Recherche 48 (Faculty of Medecine, Marseille, France) under agreement number 09-022.

Reference list

1. WHO | Levels and trends in child mortality 2015 [Internet]. [cited 2016 Apr 21]. Available from: http://www.who.int/maternal_child_adolescent/documents/levels_trends_child_mortality_2015/en/

2. UNICEF, Nutrition Section, United Nations System Standing Committee on Nutrition, World Food Programme, World Health Organization. Community-based management of severe acute malnutrition [Internet]. 2007 [cited 2016 Jun 29]. Available from: http://www.unicef.org/publications/files/Community_Based_Management_of_Sever_Acute__Malnutirtion.pdf

3. Blössner M, Onís M de. Malnutrition: quantifying the health impact at national and local levels. Geneva: World Health Organization; 2005.

4. World Health Organization, UNICEF. WHO child growth standards and the identification of severe acute malnutrition in infants and children: a joint statement by the World Health Organization and the United Nations Children's Fund. [Internet]. 2009 [cited 2016 Jun 29]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK200775/

5. Williams CD. A nutritional disease of childhood associated with a maize diet. Arch. Dis. Child. 1933; 8:423–33.

6. Mukherjee KL. Classification of protein-calorie undernutrition in children. Arch. Dis. Child. 1967; 42:647–51.

7. Waterlow JC. Classification and definition of protein-calorie malnutrition. Br. Med. J. 1972; 3:566–9.

8. Whitehead RG, Alleyne GA. Pathophysiological factors of importance in protein-calorie malnutrition. Br. Med. Bull. 1972; 28:72–9.

9. Manary MJ, Heikens GT, Golden M. Kwashiorkor: more hypothesis testing is needed to understand the aetiology of oedema. Malawi Med. J. Volume 21 numéro 3. 2009 Sep;106–7.

10. Rytter MJH, Namusoke H, Babirekere-Iriso E, Kæstel P, Girma T, Christensen VB, et al. Social, dietary and clinical correlates of oedema in children with severe acute malnutrition: a cross-sectional study. BMC Pediatr. 2015;15: 25.

11. Smith MI, Yatsunenko T, Manary MJ, Trehan I, Mkakosya R, Cheng J, et al. Gut Microbiomes of Malawian Twin Pairs Discordant for Kwashiorkor. Science. 2013; 339:548–54.

12. Million M, Tidjani Alou M, Khelaifia S, Bachar D, Lagier J-C, Dione N, et al. Increased Gut Redox and Depletion of Anaerobic and Methanogenic Prokaryotes in Severe Acute Malnutrition. Sci. Rep. 2016; 6:26051.

13. Million M, Diallo A, Raoult D. Gut microbiota and malnutrition. Microb. Pathog. 2016;

14. Trehan I, Goldbach HS, LaGrone LN, Meuli GJ, Wang RJ, Maleta KM, et al. Antibiotics as part of the management of severe acute malnutrition. N. Engl. J. Med. 2013; 368:425–35.

15. Khoruts A, Sadowsky MJ, Hamilton MJ. Development of fecal microbiota transplantation suitable for mainstream medicine. Clin. Gastroenterol. Hepatol. Off. Clin. Pract. J. Am. Gastroenterol. Assoc. 2015; 13:246–50.

16. 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.

17. 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; 28:237–64.

18. Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Freestone AL, et al. Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. Ecol. Lett. 2011; 14:19–28.

19. Subramanian S, Huq S, Yatsunenko T, Haque R, Mahfuz M, Alam MA, et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. Nature [Internet]. 2014 [cited 2015 Sep 7]; Available from: http://www.nature.com/doifinder/10.1038/nature13421

20. Carlier M, Stove V, Wallis SC, De Waele JJ, Verstraete AG, Lipman J, et al. Assays for therapeutic drug monitoring of β-lactam antibiotics: A structured review. Int. J. Antimicrob. Agents. 2015; 46:367–75.

21. Marrie TJ, Kwan C. Antimicrobial susceptibility of Staphylococcus saprophyticus and urethral staphylococci. Antimicrob. Agents Chemother. 1982; 22:395–7.

22. Michon J, Jeulin D, Lang J-M, Cattoir V. Rothia aeria acute bronchitis: the first reported case. Infection. 2010; 38:335–7.

23. Moubareck C, Gavini F, Vaugien L, Butel MJ, Doucet-Populaire F. Antimicrobial susceptibility of bifidobacteria. J. Antimicrob. Chemother. 2005; 55:38–44.

24. Stock I, Rahman M, Sherwood KJ, Wiedemann B. Natural antimicrobial susceptibility patterns and biochemical identification of Escherichia albertii and Hafnia alvei strains. Diagn. Microbiol. Infect. Dis. 2005; 51:151–63.

25. Lazzerini M, Tickell D. Antibiotics in severely malnourished children: systematic review of efficacy, safety and pharmacokinetics. Bull. World Health Organ. 2011; 89:594–607.

26. Abbott SL, Moler S, Green N, Tran RK, Wainwright K, Janda JM. Clinical and laboratory diagnostic characteristics and cytotoxigenic potential of Hafnia alvei and Hafnia paralvei strains. J. Clin. Microbiol. 2011; 49:3122–6.

27. Helft G, Tabone X, Metzger JP, Vacheron A. Gemella haemolysans endocarditis with colonic carcinoma. Eur. J. Med. 1993; 2:369–70.

28. Jin D, Chen C, Li L, Lu S, Li Z, Zhou Z, et al. Dynamics of fecal microbial communities in children with diarrhea of unknown etiology and genomic analysis of associated Streptococcus lutetiensis. BMC Microbiol. 2013; 13:141.

29. Amado C, Hidalgo MJ, Sedano C, Hebel A, Porte L, Braun S, et al. [Streptococcus gallolyticus (ex S. bovis) bacteremia and its relationship with colonic or hepatobiliary disease and endocarditis]. Rev. Chil. Infectol. Órgano Of. Soc. Chil. Infectol. 2015; 32:430–4.

30. Butt J, Romero-Hernández B, Pérez-Gómez B, Willhauck-Fleckenstein M, Holzinger D, Martin V, et al. Association of Streptococcus gallolyticus subspecies gallolyticus with colorectal cancer: Serological evidence. Int. J. Cancer. 2016; 138:1670–9.

31. Rusniok C, Couvé E, Da Cunha V, El Gana R, Zidane N, Bouchier C, et al. Genome sequence of Streptococcus gallolyticus: insights into its adaptation to the bovine rumen and its ability to cause endocarditis. J. Bacteriol. 2010; 192:2266–76.

32. 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. PloS One. 2009; 4: e7063.

33. Angelakis E, Bachar D, Henrissat B, Armougom F, Audoly G, Lagier J-C, et al. Glycans affect DNA extraction and induce substantial differences in gut metagenomic studies. Sci. Rep. 2016; 6:26276.

34. 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; 44:117–24.

35. 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; 32:1471–81.

36. 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; 1:16131.

37. Maukonen J, Saarela M. Human gut microbiota: does diet matter? Proc. Nutr. Soc. 2015; 74:23–36.

38. Blanton LV, Charbonneau MR, Salih T, Barratt MJ, Venkatesh S, Ilkaveya O, et al. Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children. Science. 2016; 351: aad3311.

39. Schwarzer M, Makki K, Storelli G, Machuca-Gayet I, Srutkova D, Hermanova P, et al. Lactobacillus plantarum strain maintains growth of infant mice during chronic undernutrition. Science. 2016; 351:854–7.

40. Subramanian S, Blanton LV, Frese SA, Charbonneau M, Mills DA, Gordon JI. Cultivating healthy growth and nutrition through the gut microbiota. Cell. 2015; 161:36–48.

41. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. PLoS Biol. 2007; 5: e177.

42. Salazar N, Arboleya S, ValdÃ©s L, Stanton C, Ross P, Ruiz L, et al. The human intestinal microbiome at extreme ages of life. Dietary intervention as a way to counteract alterations. Front. Genet. 2014; 5: 406.

43. Unno T, Choi J-H, Hur H-G, Sadowsky MJ, Ahn Y-T, Huh C-S, et al. Changes in human gut microbiota influenced by probiotic fermented milk ingestion. J. Dairy Sci. 2015; 98:3568–76.

44. Bäckhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe. 2015; 17:690–703.

45. Jost T, Lacroix C, Braegger C, Chassard C. Impact of human milk bacteria and oligosaccharides on neonatal gut microbiota establishment and gut health. Nutr. Rev. 2015; 73:426–37.

46. Engels C, Ruscheweyh H-J, Beerenwinkel N, Lacroix C, Schwab C. The Common Gut Microbe Eubacterium hallii also Contributes to Intestinal Propionate Formation. Front. Microbiol. 2016; 7:713.

47. Hong HA, Khaneja R, Tam NMK, Cazzato A, Tan S, Urdaci M, et al. Bacillus subtilis isolated from the human gastrointestinal tract. Res. Microbiol. 2009; 160:134–43.

48. Shobharani P, Padmaja RJ, Halami PM. Diversity in the antibacterial potential of probiotic cultures Bacillus licheniformis MCC2514 and Bacillus licheniformis MCC2512. Res. Microbiol. 2015; 166:546–54.

49. Zhang Z, Peng X, Li S, Zhang N, Wang Y, Wei H. Isolation and identification of quercetin degrading bacteria from human fecal microbes. PloS One. 2014; 9:e90531.

50. Narushima S, Itoha K, Miyamoto Y, Park S-H, Nagata K, Kuruma K, et al. Deoxycholic acid formation in gnotobiotic mice associated with human intestinal bacteria. Lipids. 2006; 41:835–43.

51. Hunger S, Gössner AS, Drake HL. Trophic links between the acetogen Clostridium glycolicum KHa and the fermentative anaerobe Bacteroides xylanolyticus KHb, isolated from Hawaiian forest soil. Appl. Environ. Microbiol. 2011; 77:6281–5.

52. Goh HF, Philip K. Purification and Characterization of Bacteriocin Produced by Weissella confusa A3 of Dairy Origin. PloS One. 2015; 10:e0140434.

53. Lee KW, Park JY, Jeong HR, Heo HJ, Han NS, Kim JH. Probiotic properties of Weissella strains isolated from human faeces. Anaerobe. 2012; 18:96–102.

54. Malik A, Sumayyah S, Yeh C-W, Heng NCK. Identification and sequence analysis of pWcMBF8-1, a bacteriocin-encoding plasmid from the lactic acid bacterium Weissella confusa. FEMS Microbiol. Lett. 2016;363.

55. Nam H, Ha M, Bae O, Lee Y. Effect of Weissella confusa strain PL9001 on the adherence and growth of Helicobacter pylori. Appl. Environ. Microbiol. 2002; 68:4642–5.

56. Gaggìa F, Mattarelli P, Biavati B. Probiotics and prebiotics in animal feeding for safe food production. Int. J. Food Microbiol. 2010;141 Suppl 1: S15-28.

57. Dione N, Khelaifia S, La Scola B, Lagier JC, Raoult D. A quasi-universal medium to break the aerobic/anaerobic bacterial culture dichotomy in clinical microbiology. Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis. 2016; 22:53–8.

669 58. Samuel BS, Gordon JI. A humanized gnotobiotic mouse model of host–archaeal–bacterial
670 mutualism. Proc. Natl. Acad. Sci. U. S. A. 2006; 103:10011–6.

671 59. Donders GGG, Gonzaga A, Marconi C, Donders F, Michiels T, Eggermont N, et al.
672 Increased vaginal pH in Ugandan women: what does it indicate? Eur. J. Clin. Microbiol.
673 Infect. Dis. Off. Publ. Eur. Soc. Clin. Microbiol. 2016; 35:1297–303.

674 60. Kismul H, Van den Broeck J, Lunde TM. Diet and kwashiorkor: a prospective study from
675 rural DR Congo. PeerJ. 2014; 2:e350.

676 61. Tijerina-Sáenz A, Innis SM, Kitts DD. Antioxidant capacity of human milk and its
677 association with vitamins A and E and fatty acid composition. Acta Paediatr. Oslo Nor. 1992.
678 2009; 98:1793–8.

679 62. Bode L. The functional biology of human milk oligosaccharides. Early Hum. Dev.
680 2015;91: 619–22.

681 63. Charbonneau MR, O'Donnell D, Blanton LV, Totten SM, Davis JCC, Barratt MJ, et al.
682 Sialylated Milk Oligosaccharides Promote Microbiota-Dependent Growth in Models of Infant
683 Undernutrition. Cell. 2016; 164:859–71.

684 64. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier P-E, Rolain JM, et al. Ongoing
685 revolution in bacteriology: routine identification of bacteria by matrix-assisted laser
686 desorption ionization time-of-flight mass spectrometry. Clin. Infect. Dis. 2009; 49:543–51.

687 65. Seng P, Abat C, Rolain JM, Colson P, Lagier J-C, Gouriet F, et al. Identification of rare
688 pathogenic bacteria in a clinical microbiology laboratory: impact of matrix-assisted laser
689 desorption ionization-time of flight mass spectrometry. J. Clin. Microbiol. 2013; 51:2182–94.

66. 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; 64:346–51.

67. 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.

68. Hayashi H, Sakamoto M, Benno Y. Phylogenetic analysis of the human gut microbiota using 16S rDNA clone libraries and strictly anaerobic culture-based methods. Microbiol. Immunol. 2002; 46:535–48.

69. Spellerberg IF, Fedor PJ. A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the "Shannon–Wiener" Index. Glob. Ecol. Biogeogr. 2003; 12:177–9.

70. Barnard GA. A New Test for 2 × 2 Tables. Nature. 1945; 156:177.

71. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiol. Camb. Mass. 1990; 1:43–6.

Figure legends

Figure 1: Comparison of the Hitherto Unknown Diversity between kwashiorkor patients and controls.

With culturomics, the hitherto unknown diversity is represented by the number of new species and the number of species not previously known in the human gut while with metagenomics, the hitherto unknown diversity is represented by the number of unassigned OTU. The hitherto unknown diversity was compared in kwashiorkor and CTL groups. A significant loss of the hitherto unknown diversity is observed in kwashiorkor patients.

Figure 2: Oxidation of the gut environment in kwashiorkor patients.

Presence of M. smithii and fecal redox potential were assessed in each sample and compared between the kwashiorkor and the control groups. An absence of M. smithii is observed in kwashiorkor patients associated to an increased redox potential therefore suggesting an oxidized gut environment in kwashiorkor patients.

Figure 3: Increased frequency of species in Kwashiorkor samples and control samples for the culturomics approach.

Each bar represents the relative frequency difference for each species with red bars representing an increased frequency in kwashiorkor patients and green bars representing an increased frequency in CTL. A majority of oxygen tolerant species are increased in the gut of kwashiorkor patients. *: *P* value ranging from 0.01 and 0.05

Figure 4: Increased frequency of species in Kwashiorkor samples and control samples for the metagenomics approach.

Each bar represents the relative frequency difference for each species with red bars representing an increased frequency in kwashiorkor patients and green bars representing an increased frequency in CTL. Only four species were significantly increased in kwashiorkor

patients among which three were aerotolerant. *: *P* value ranging from 0.01 and 0.05. **: *P* value ranging from 0.001 and 0.01.

Figure 5: Low reproducibility among metagenomics studies

Results from metagenomics studies from the literature were compared exhibiting very little reproducibility between studies with on 9 core species for kwashiorkor gut microbiota and 10 core species for control gut microbiota.

Figure 6: Sample repartition according to geographic origin.

Four samples were collected from kwashiorkor patients in Niger while six samples were collected from kwashiorkor patients in Senegal. As for controls, stool samples were collected from 3 healthy children from Senegal and 2 healthy children from Niger.

Figure 7: The eighteen culture conditions of standardized culturomics.

The eighteen culture conditions are here represented according to the preincubation liquid medium, mode of treatment of the stool sample, temperature and atmosphere. The red bars represent in the third column an active filtration (5 µm) applied to the stool sample, in the fourth column a thermic shock (80°C during 20 minutes) applied to the stool sample, in the fifth column a 28°C incubation temperature and in the sixth column an anaerobic atmosphere of incubation. No coloration represents in the third column no active filtration applied to the stool sample, in the fourth column no thermic shock applied to the stool sample, in the fifth column a 37°C incubation temperature and in the sixth column an aerobic atmosphere of incubation.

Table 1. Baseline characteristics

	Kwashiorkor (n = 10)	Controls (n = 5)	*P*-value
Age (months, mean ± SD)	13.4 ± 17.8[a]	25.1 ± 7.6	0.20[b]
Sex	3/6 (50%)[a]	3 (60%)	0.99[c]
Oedema	10 (100%)	0 (0%)	
Poids (kg)	5.2 ± 0.8	12.2 ±1.9	0.004[d]
Taille (cm)	61.2 ± 3.8	89.0 ±8.7	0.01[d]
WHZ	NR	-0.4 ± 0.25	
WAZ	NR	-0,13±1,02	
HAZ	-4.0 ± 1.3[a]	0.5 ± 2.0	0.07[d]

SD: Standard deviation, WHZ: Weight-for height z-score, WAZ: Weight-for-age z-score, HAZ: Height-for-age z-score, NR: Not relevant in the presence of edema, [a]Age and sex missing for four samples from Niger, [b]Bilateral Mann-Whitney test, [c]Bilateral Barnard test, [d]Bilateral unpaired t-test.

Table 2. Culturomics highlights an altered diversity in kwashiorkor.

All samples	Kwashiorkor (n=10)	Controls (n=5)	Δ Diversity	*P* value
All species[a]	151/335 (45%)	185/281 (66%)	+21%	$2.5*10^{-7}$ [c]
Anaerobic species[a]	43/111 (39%)	76/112 (68%)	+29%	0.00001 [c]
Aerotolerant species[a]	108/224 (48%)	109/169 (64%)	+16%	0.001 [c]
Actinobacteria[b]	47/335 (14%)	42/281 (15%)	+1%	0,74 [c]
Bacteroidetes[b]	21/335 (6%)	21/281 (7%)	+1%	0,55 [c]
Firmicutes[b]	208/335 (62%)	192/281 (68%)	+6%	0,1 [c]
Fusobacteria[b]	3/335 (0,8%)	0/281 (0%)	-0.8%	0,32 [d]
Proteobacteria[b]	56/335 (17%)	25/281 (9%)	-8%	0,004 [c]

[a]Beta-diversity was assessed using the U/T ratio (U/T: Unique/ Total). [b]The diversity at the phylum level was assessed by the proportion in each phylum among the total number of species isolated. [c]Uncorrected bilateral Chi square test. [d]exact bilateral Fisher test.

Δ Diversity = Diversity in controls – Diversity in kwashiorkor.

Global diversity, assessed by the U/T ratio (see methods), is significantly decreased alongside the anaerobic and aerotolerant diversity in kwashiorkor patients. *Proteobacteria* species were significantly enriched in kwashiorkor. No *Fusobacteria* species were isolated in controls.

Table 3. Repartition of new species according to their phylum, their tolerance to oxygen and their origin.

Species	Family	Oxygen tolerant	Species origin	16S rRNA accession number[a]
Africanella massiliensis[b]	*Ruminococcaceae*	No	Control	LT223700
Bacillus massilionigeriensis	*Bacillaceae*	Yes	Control	LT161887
Bacillus mediterraneensis	*Bacillaceae*	Yes	Control	LT161888
Bacillus phoceensis	*Bacillaceae*	Yes	Control	LN881595
Bacillus rubiinfantis	*Bacillaceae*	Yes	Control	LK021113
Bacillus testis	*Bacillaceae*	Yes	Control	LN827531
Bacillus touaregensis	*Bacillaceae*	Yes	Control	LT223701
Brachybacterium massiliense	*Dermabacteraceae*	Yes	Control	LN906631
Brevibacterium phoceense	*Brevibacteriaceae*	Yes	Control	LN998064
Clostridium massiliodielmoense	*Clostridiaceae*	No	Control	LN998063
Clostridium massiliosenegalense	*Lachnospiraceae*	No	Control	LT161890
Clostridium nigeriense	*Clostridiaceae*	No	Control	LT161894
Clostridium touaregense	*Lachnospiraceae*	No	Control	LT161895
Enterobacter timonensis	*Enterobacteriaceae*	Yes	Control	LN906632
Khelaifiabacterium massiliensis[b]	*Clostridiaceae*	No	Control	LN850733
Lagierella massiliensis[b]	*Peptoniphilaceae*	No	Control	LN870299
Lascolabacillus massiliensis[b]	*Porphyromonadaceae*	No	Control	LN827535
Massiliobacillus massiliensis[b]	*Sporomusaceae*	No	Control	LT161896
Murdochiella massiliensis	*Peptoniphilaceae*	No	Control	LN866998
Ndiopella massiliensis[b]	*Peptoniphilaceae*	No	Control	LN866993
Neofamilia massiliensis[c]	*Neofamiliaceae*	No	Control	LN866999
Niameyia massiliensis[b]	*Lachnospiraceae*	No	Control	LN850735
Paenibacillus phoceensis	*Paenibacillaceae*	Yes	Control	LN998053
Paenibacillus senegalomassiliensis	*Paenibacillaceae*	Yes	Control	LN890284
Paenibacillus touaregensis	*Paenibacillaceae*	Yes	Control	LT223571
Peptoniphilus phoceensis	*Peptoniphilaceae*	No	Control	LN881605
Senegalia massiliensis[b]	*Clostridiaceae*	No	Control	LN881608
Anaerococcus rubiinfantis	*Peptoniphilaceae*	No	Kwashiorkor	LN881592
Anaeromassilibacillus senegalensis[b]	*Ruminococcaceae*	No	Kwashiorkor	LN866991
Anaerotruncus rubiinfantis	*Ruminococcaceae*	No	Kwashiorkor	LN881593

Bacillus andreraoultii	*Bacillaceae*	Yes	Kwashiorkor	LK021120
Bacillus niameyensis	*Bacillaceae*	Yes	Kwashiorkor	LK985389
Desnuesiella massiliensis[b]	*Clostridiaceae*	Yes	Kwashiorkor	LN846906
Flaviflexus massiliensis	*Actinomycetaceae*	Yes	Kwashiorkor	LK985390
Malnutritionisia massiliensis[b]	*Clostridiaceae*	No	Kwashiorkor	LN850734
Massilibacterium senegalense[b]	*Bacillaceae*	Yes	Kwashiorkor	LN828943
Mediannikovella massiliensis[b]	*Clostridiaceae*	No	Kwashiorkor	LN849776
Mobilicoccus massiliensis	*Dermatophilaceae*	Yes	Kwashiorkor	LK985391
Nigerium massiliensis[b]	*Propionibacteriaceae*	No	Kwashiorkor	LK985392
Olsenella massiliensis	*Atopobiaceae*	No	Kwashiorkor	LN827536
Paenibacillus rubiinfantis	*Paenibacillaceae*	Yes	Kwashiorkor	LN881603
Rubeoparvulum massiliense[b]	*Bacillaceae*	No	Kwashiorkor	LN828926
Rubiinfantum massiliense[b]	*Bacillaceae*	Yes	Kwashiorkor	LK985393
Tessaracoccus massiliensis	*Propionibacteriaceae*	Yes	Kwashiorkor	LK985394

[a]GenBank accession number Genbank (http://www.ncbi.nlm.nih.gov/nuccore). [b]New genus.

[c]New family.

Table 4. Comparison of the cultured gut bacterial diversity between kwashiorkor and control children

Per sample	Kwashiorkor (n = 10)	Controls (n = 5)	*P* value
Global diversity (mean ± SD)			
Nb of phyla	4.2 ± 0.6	3.8 ± 0.4	0.2 [a]
Nb of genera	36 ± 7	34 ± 12	0.67[b]
Total Nb of species	90 ± 22	92 ± 20	0.82[b]
Nb of HG species	86 ± 21	85 ± 18	0.92[b]
Nb of species H but not G	1.2 ± 0.8	1.4 ± 0.9	0.75[a]
Nb of NH species	1.5 ± 1.2	2.8 ± 0.4	0.02[a]
Nb of new species	1.8 ± 1.5	5 ± 2.6	0.009[b]
Diversity by phylum (Nb species in each phylum (mean ± SD))			
Firmicutes	60 ± 18	68 ± 13	0.39[b]
Proteobacteria	11 ± 5	8 ± 7	0.3[b]
Actinobacteria	12 ± 4	11 ± 4	0.7[b]
Bacteroidetes	6.6 ± 4.2	5.4 ± 7.7	0.32[a]
Fusobacterium	0.5 ± 0.8	0 ± 0	0.21[b]
Diversity by genus (Nb species in each genus (mean ± SD))			
Clostridium	14 ± 6	18 ± 9	0.27[b]
Bacillus	9 ± 5	10 ± 3	0.6[b]
Paenibacillus	7 ± 1	8 ± 2	0.23[b]

Streptococcus	6 ± 1	5 ± 2	0.39[b]
Staphylococcus	6 ± 3	4 ± 2	0.10[b]
Lactobacillus	2.1 ± 2.1	1.6 ± 2	0.67[b]
Bifidobacterium	2.4 ± 1.7	0.8 ± 1.3	0.09[b]

SD: Standard deviation, HG species: species previously isolated in the human gut, species H but not G: species previously isolated in humans but not in the gut, NH species: species not previously isolated in humans. [a]Bilateral Mann-Whitney test. [b]Bilateral unpaired t-test.

The unknown diversity is lower in the gut microbiota of kwashiorkor patients (significantly less new species and NH species isolated per sample of kwashiorkor).

Table 5. Metagenomics evidenced a decreased fecal anaerobic diversity in kwashiorkor

Per sample	Kwashiorkor (n=10)	Controls (n=5)	*P* value[a]
Global diversity[b] (mean ± SD)	3.2 ± 0.8	3.8 ± 0.8	0.19
Aerotolerant diversity[b]	2.0 ±0.8	1.3 ± 0.6	0.1
Anaerobic diversity[b]	1.0 ± 1.0	3.1 ± 1.5	0.02

SD: Standard deviation. [a]Bilateral unpaired t-test. [b]Shannon indexes were calculated (see methods) for each sample. Using metagenomics, only anaerobic diversity was significantly decreased in kwashiorkor.

Table 6. Comparison of the metagenomics gut bacterial diversity between kwashiorkor and control children

	Kwashiorkor (n = 10)	Controls (n = 5)	*P* value
Global bacterial abundance Proportion of reads assigned at the species level[a] (mean ± SD)	0.95 ± 0.02	0.74 ± 0.22	0.009[b]
Proportion per phylum[c]			
Actinobacteria	86/589 (14%)	61/486 (12%)	0.37[d]
Bacteroidetes	44/589 (7%)	46/486 (9%)	0.24[e]
Chloroflexi	1/589 (0.2%)	0/486 (0%)	0.36[f]
Euryarchaeota	0/589 (0%)	4/486 (0.8%)	0.027[e]
Firmicutes	319/589 (54%)	296/486 (61%)	0.026[d]
Fusobacteria	7/589 (1.2%)	1/486 (0.2%)	0.12[e]
Lentisphaerae	0/589 (0%)	1/486 (0.2%)	0.90[e]
Proteobacteria	131/589 (22%)	75/486 (15%)	0.004[d]
Tenericutes	0/589 (0%)	1/486 (0.2%)	0.90[e]
Verrucomicrobia	1/589 (0.2%)	1/486 (0.2%)	>0.99[e]
Proportion per class			
Actinobacteria	69/589 (12%)	38/486 (8%)	0.033[d]
Coriobacteriia	16/589 (3%)	22/486 (4%)	0.109[d]
Bacteroidia	38/589 (6%)	46/486 (9%)	0.066[d]
Flavobacteriia	0/589 (0%)	4/486 (0.8%)	0.082[e]
Sphingobacteriia	2/589 (0.3%)	0/486 (0%)	0.59[e]
Bacilli	158/589 (27%)	124/486 (25%)	0.63[d]
Clostridia	115/589 (19%)	132/486 (27%)	0.003[d]
Erysipelotrichia	11/589 (2%)	14/486 (3%)	0.27[d]
Negativicutes	31/589 (5%)	25/486 (5%)	>0.99[d]
Tissierellia	4/589 (0.7%)	1/486 (0.2%)	0.50[e]
Alphaproteobacteria	22/589 (4%)	2/486 (0.4%)	0.0002[d]
Betaproteobacteria	12/589 (2%)	6/486 (1%)	0.31[d]
Deltaproteobacteria	5/589 (0.8%)	3/486 (0.6%)	0.94[d]
Epsilonproteobacteria	2/589 (0.3%)	2/486 (0.4%)	>0.99[e]
Gammaproteobacteria	91/589 (15%)	63/486 (13%)	0.25[d]
Fusobacteriia	7/589 (1%)	1/486 (0.2%)	0.12[d]
Lentisphaeria	0/589 (0%)	1/486 (0.2%)	0.90[e]
Mollicutes	0/589 (0%)	1/486 (0.2%)	0.90[e]
Verrucomicrobiae	1/589 (0.2%)	1/486 (0.2%)	>0.99[e]
Chloroflexia	1/589 (0.2%)	0/486 (0%)	>0.99[e]

SD: Standard deviation. [a]Result per sample. [b]Bilateral unpaired t-test. [c]Number of species belonging to this taxonomical group divided by the total number of species in this group. [d]Uncorrected bilateral chi square test. [e]Exact bilateral Fisher test.

The proportion of reads was estimated per sample. Diversity was estimated for all samples by calculating proportions of species in each phylum or class. *Firmicutes* species were significantly decreased in kwashiorkor, specifically species belonging to the *Clostridia* class. *Proteobacteria* species were significantly increased in kwashiorkor, specifically species belonging to the *Alphaproteobacteria* class. No *Euryarchaeota* was found in kwashiorkor.

Table 7. Missing microbes in Kwashiorkor identified both by culturomics and metagenomics.

Species	Obligate anaerobes	Phylum	Class	Order	Family	Genera	Possible probiotic
Acinetobacter lwoffii	0	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Moraxellaceae*	*Acinetobacter*	No
Alistipes indistinctus	1	*Bacteroidetes*	*Bacteroidia*	*Bacteroidales*	*Rikenellaceae*	*Alistipes*	Yes
Alistipes putredinis	1	*Bacteroidetes*	*Bacteroidia*	*Bacteroidales*	*Rikenellaceae*	*Alistipes*	No
Alistipes senegalensis	1	*Bacteroidetes*	*Bacteroidia*	*Bacteroidales*	*Rikenellaceae*	*Alistipes*	No
Alloscardovia omnicolens	0	*Actinobacteria*	*Actinobacteria*	*Bifidobacteriales*	*Bifidobacteriaceae*	*Alloscardovia*	No
Anaerostipes caccae	1	*Firmicutes*	*Clostridia*	*Clostridiales*	*Lachnospiraceae*	*Anaerostipes*	Yes
Arthrobacter agilis	0	*Actinobacteria*	*Actinobacteria*	*Micrococcales*	*Micrococcaceae*	*Arthrobacter*	No
Asaccharospora irregularis	1	*Firmicutes*	*Clostridia*	*Clostridiales*	*Peptostreptococcaceae*	*Asaccharospora*	No
Bacillus cereus	0	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*	*Bacillus*	No
Bacillus firmus	0	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*	*Bacillus*	No
Bacillus idriensis	0	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*	*Bacillus*	No
Bacillus licheniformis	0	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*	*Bacillus*	Yes
Bacillus niabensis	0	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*	*Bacillus*	No
Bacillus subtilis	0	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*	*Bacillus*	Yes
Bacillus thermoamylovorans	0	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*	*Bacillus*	No
Bacteroides salyersiae	1	*Bacteroidetes*	*Bacteroidia*	*Bacteroidales*	*Bacteroidaceae*	*Bacteroides*	Yes
Bifidobacterium adolescentis	1	*Actinobacteria*	*Actinobacteria*	*Bifidobacteriales*	*Bifidobacteriaceae*	*Bifidobacterium*	Yes
Clostridium amygdalinum	1	*Firmicutes*	*Clostridia*	*Clostridiales*	*Lachnospiraceae*	*Lachnoclostridium*	No
Clostridium cadaveris	1	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Clostridium*	No
Clostridium glycolicum	1	*Firmicutes*	*Clostridia*	*Clostridiales*	*Peptostreptococcaceae*	*Terrisporobacter*	Yes
Clostridium hylemonae	1	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Clostridium*	Yes
Clostridium neonatale	1	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Clostridium*	No
Clostridium oroticum	1	*Firmicutes*	*Clostridia*	*Clostridiales*	*Lachnospiraceae*	*Lachnoclostridium*	No
Clostridium paraputrificum	1	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Clostridium*	No
Clostridium saccharolyticum	1	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Clostridium*	No

Clostridium sordellii	1	*Firmicutes*	*Clostridia*	*Clostridiales*	*Peptostreptococcaceae*	*Peptoclostridium*	No
Dialister pneumosintes	1	*Firmicutes*	*Negativicutes*	*Selenomonadales*	*Veillonellaceae*	*Dialister*	No
Enterococcus dispar	0	*Firmicutes*	*Bacilli*	*Lactobacillales*	*Enterococcaceae*	*Enterococcus*	No
Faecalitalea cylindroides	1	*Firmicutes*	*Erysipelotrichia*	*Erysipelotrichales*	*Erysipelotrichaceae*	*Faecalitalea*	No
Gemella sanguinis	0	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillales Family XI Incertae Sedis*	*Gemella*	No
Intestinimonas butyriciproducens	1	*Firmicutes*	*Clostridia*	*Clostridiales*	*Unclassified clostridiales*	*Intestinimonas*	Yes
Lactobacillus parabuchneri	0	*Firmicutes*	*Bacilli*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus*	Yes
Lactobacillus perolens	0	*Firmicutes*	*Bacilli*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus*	Yes
Lactobacillus vaccinostercus	0	*Firmicutes*	*Bacilli*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus*	Yes
Micrococcus lylae	0	*Actinobacteria*	*Actinobacteria*	*Micrococcales*	*Micrococcaceae*	*Micrococcus*	No
Neisseria flavescens	0	*Proteobacteria*	*Betaproteobacteria*	*Neisseriales*	*Neisseriaceae*	*Neisseria*	No
Pantoea septica	0	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacteriales*	*Enterobacteriaceae*	*Pantoea*	No
Paraclostridium bifermentans	1	*Firmicutes*	*Clostridia*	*Clostridiales*	*Peptostreptococcaceae*	*Paraclostridium*	No
Slackia exigua	1	*Actinobacteria*	*Coriobacteriia*	*Eggerthellales*	*Eggerthellaceae*	*Slackia*	No
Staphylococcus haemolyticus	0	*Firmicutes*	*Bacilli*	*Bacillales*	*Staphylococcaceae*	*Staphylococcus*	No
Staphylococcus hominis	0	*Firmicutes*	*Bacilli*	*Bacillales*	*Staphylococcaceae*	*Staphylococcus*	No
Streptococcus vestibularis	0	*Firmicutes*	*Bacilli*	*Lactobacillales*	*Streptococcaceae*	*Streptococcus*	No
Sutterella wadsworthensis	1	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Sutterellaceae*	*Sutterella*	No
Veillonella dispar	1	*Firmicutes*	*Negativicutes*	*Selenomonadales*	*Veillonellaceae*	*Veillonella*	No
Weissella confusa	0	*Firmicutes*	*Bacilli*	*Lactobacillales*	*Leuconostocaceae*	*Wesseila*	Yes

792

Table 8. Potential probiotics identified by culturomics and metagenomics and their possible function.

Species	Obligate anaerobe*	Phylum	Function
Alistipes indistinctus	1	*Bacteroidetes*	Common member of the healthy gut microbiota (Nagai et al., 2010)
Anaerostipes caccae	1	*Firmicutes*	Common member of the healthy gut microbiota (Maukonen and Saarela, 2015)
Bacillus licheniformis	0	*Firmicutes*	Antibacterial potential (Shobharani et al., 2015)
Bacillus subtilis	0	*Firmicutes*	Antibacterial potential (Hong et al., 2009)
Bacteroides salyersiae	1	*Bacteroidetes*	Mutualistic association with *T. glycolicus* for polysaccharides fermentation (Hunger et al., 2011)
Bifidobacterium adolescentis	1	*Actinobacteria*	Common member of the healthy gut microbiota in breastfed infants (Jost et al., 2015)
Clostridium hylemonae	1	*Firmicutes*	Bile acids detoxification for a more suitable growth environment for bacteria (Hunger et al., 2011)
Intestinimonas butyriciproducens	1	*Firmicutes*	SCFA producer and common member of the healthy gut microbiota (Engels et al., 2016)
Lactobacillus parabuchneri	0	*Firmicutes*	Common member of the healthy gut microbiota in breastfed infants (Jost et al., 2015)
Lactobacillus perolens	0	*Firmicutes*	Common member of the healthy gut microbiota in breastfed infants (Jost et al., 2015)
Lactobacillus vaccinostercus	0	*Firmicutes*	Common member of the healthy gut microbiota in breastfed infants (Jost et al., 2015)
Terrisporobacter glycolicus	1	*Firmicutes*	Mutualistic association with acetogenic Bacteroides for polysaccharides fermentation (Jost et al., 2015)
Weissella confusa	0	*Firmicutes*	Antioxidant metabolism (Zhang et al., 2014)

*1: obligate anaerobic species. 0: facultative anaerobe or aerobic species.

Table 9. List of samples according to geographic origin, sex and age.

Sample name	Origin	Sex	Age
Kwashiorkor 1	Niger	NA	NA
Kwashiorkor 8	Niger	NA	NA
Kwashiorkor 12	Niger	NA	NA
Kwashiorkor 14	Niger	NA	NA
Kwashiorkor 01	Senegal	Female	7 months old
Kwashiorkor 02	Senegal	Female	49 months old
Kwashiorkor 04	Senegal	Female	4 months old
Kwashiorkor 05	Senegal	Male	12 months old
Kwashiorkor 06	Senegal	Male	6 months old
Kwashiorkor 010	Senegal	Male	2.2 months old
Control N6	Niger	Female	7 months old
Control N12	Niger	Female	44 months old
Control S05	Senegal	Male	28.6 months old
Control S07	Senegal	Female	38 months old
Control S50/04	Senegal	Female	8 months old

NA: non-available data

Table 10. List of primers and probes used in our study.

Organism	Assay	Primer/probe name and sequence (5'-3')	Dye
Methanobrevibacter smithii [1]	16S rDNA	Smit.16S-740F,CCGGGTATCTAATCCGGTTC Smit.16S-862R,CTCCCAGGGTAGAGGTGAAA Smit.16SFAM,CCGTCAGAATCGTTCCAGTCAG	FAM (MGB)
16S rRNA amplification[2]	16S rRNA	fD1, AGAGTTTGATCATGGCTCAG rP2, ACGGCTACCTTGTTACGACTT	
16S rRNA sequencing[2]	16S rRNA	357F, TACGGGAGGCAGCAG 357R, CTGCTGCCTCCCGTA 536F, CAGCAGCCGCGGTAATAC 536R, GTATTACCGCGGCTGCTG 800F, ATTAGATACCCTGGTAG 800R, CTACCAGGGTATCTAAT 1050F,TGTCGTCAGCTCGTG 1050R,CACGAGCTGACGACA	

Figure 1

Click here to download Figure Figure 1.tif

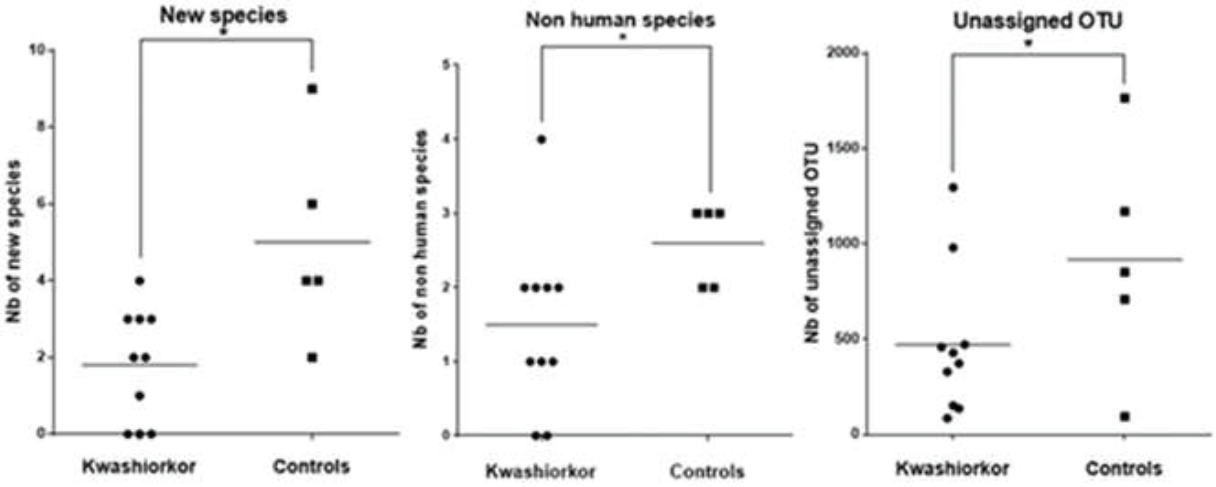
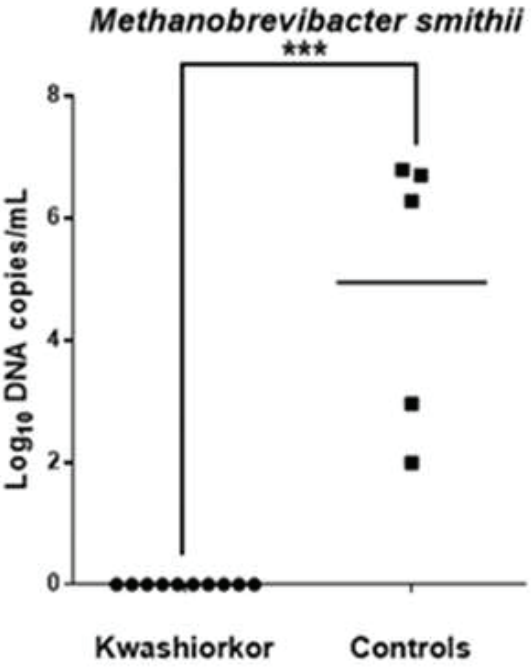
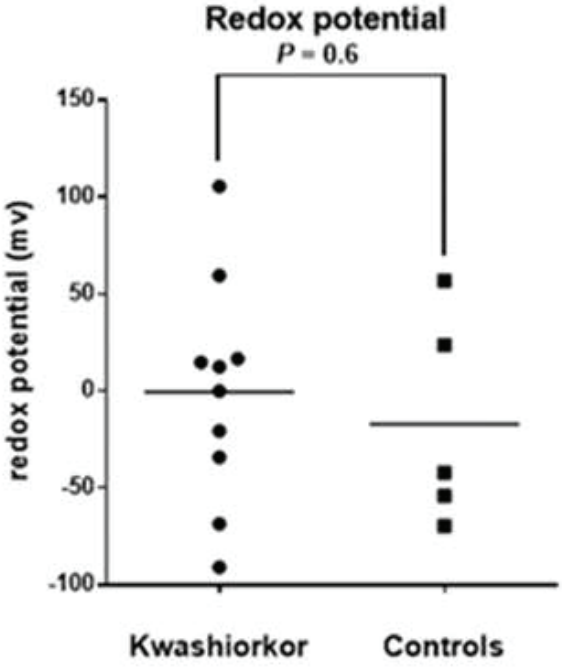


Figure 2

Click here to download Figure Figure 2.tif





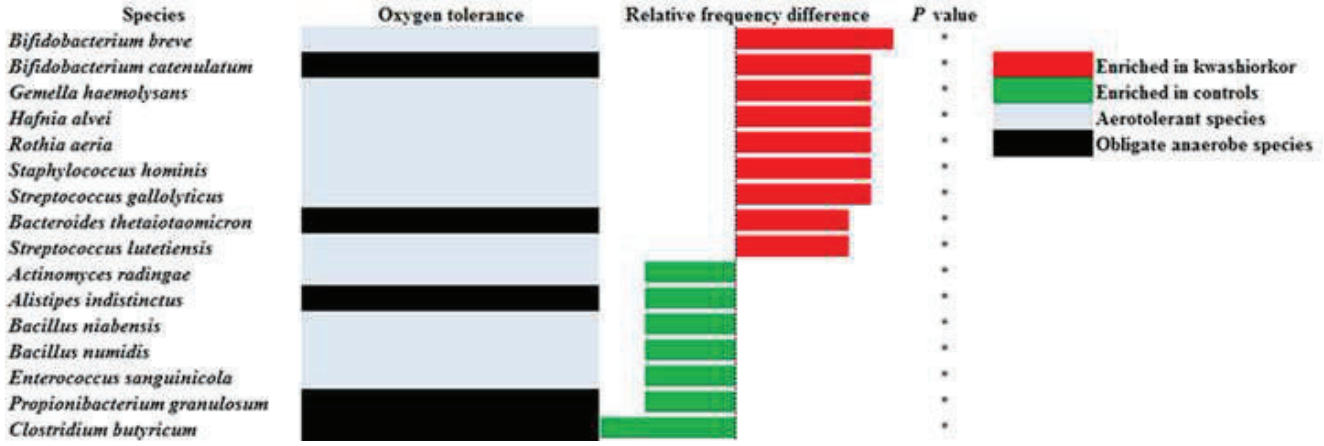
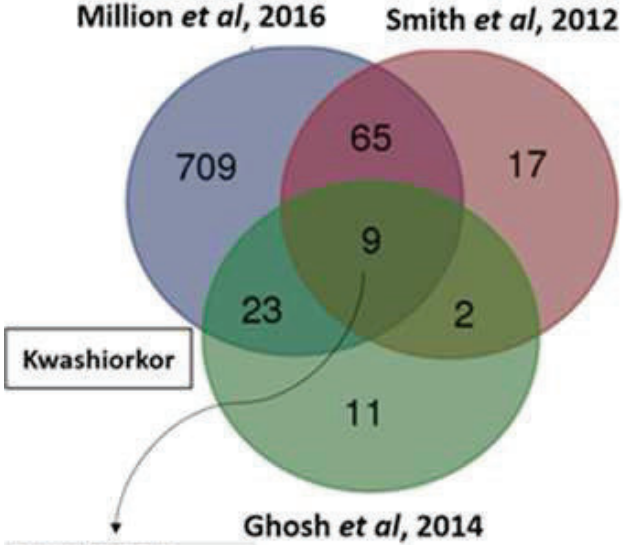


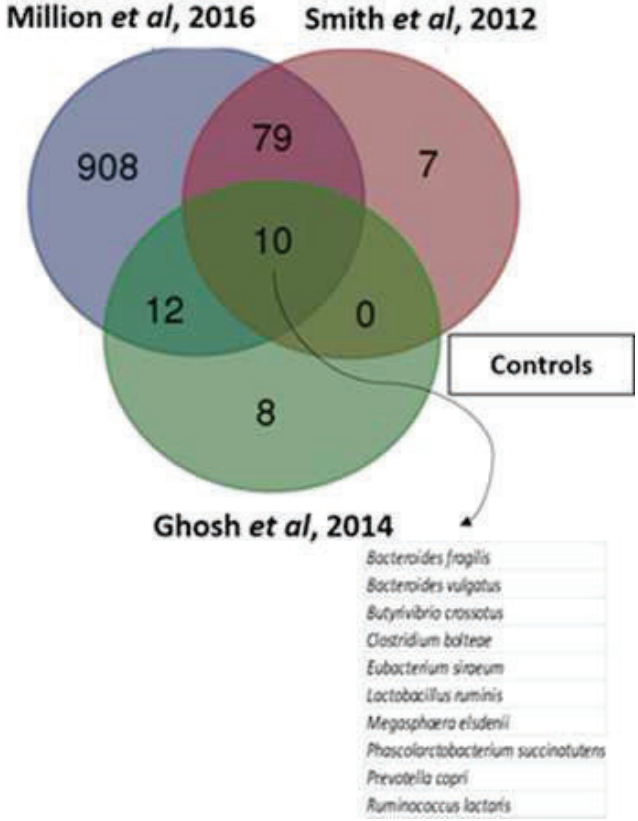
Figure 4

Click here to download Figure Figure 4.tif

Figure 5

Click here to download Figure Figure 5.tif





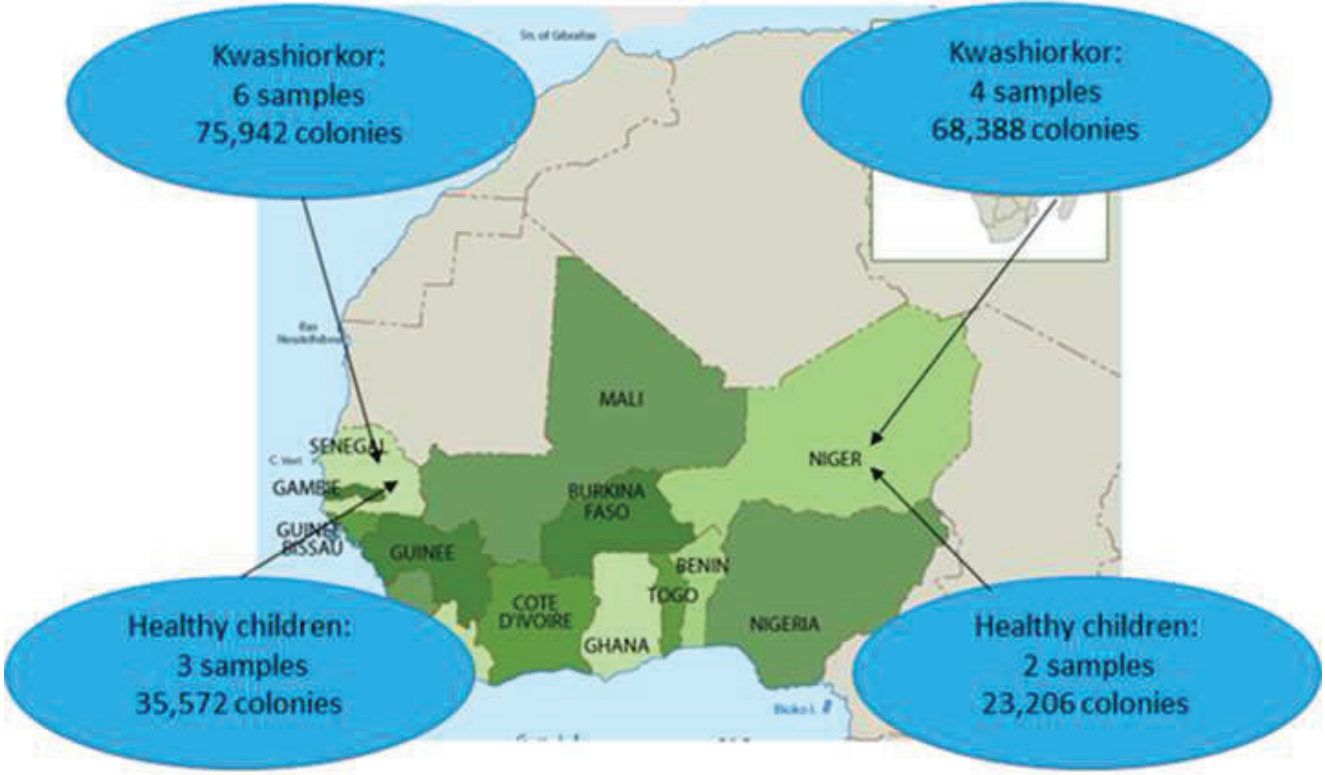
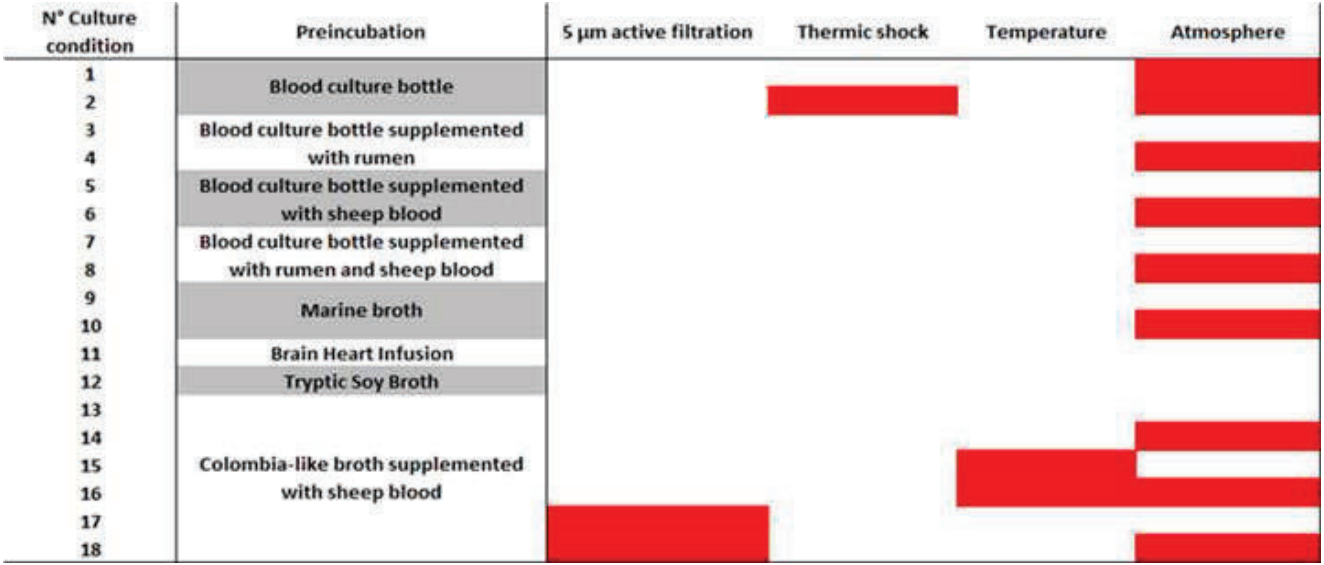


Figure 7

Click here to download Figure Figure 7.tif



Partie V: Description des nouvelles espèces isolées au cours de ce travail par taxonogenomics.

Avant-propos

L'exploration du microbiote digestif par culturomics a permis l'isolation de 280 nouvelles espèces depuis la naissance de cette méthode parmi lesquelles 30 ont été primo-isolées par mes soins au cours de l'étude du microbiote digestif des enfants sains et malnutris. La description de nouvelles espèces se fait actuellement selon un modèle classique basé sur les caractéristiques phénotypiques (forme, sporulation, atmosphère, description de la membrane cellulaire, capacités métaboliques) et génotypiques (hybridation ADN-ADN) ainsi que sur la construction de relations phylogénétiques basées sur l'ARN ribosomal 16S (35–37). Cependant, avec l'essor des technologies de séquençage de nouvelle-génération et leur coût décroissant et le temps révolutionnaire d'identification et de comparaison des bactéries grâce au MALDI-TOF MS (38,39), une évolution du modèle de description des nouvelles espèces bactériennes s'impose. A cet effet, le concept de description taxonogenomics a été créé au sein de notre laboratoire ; ce concept intègre la description du génome à la description phénotypique ainsi qu'à une description protéique par le profil MALDI-TOF (5,40). Le concept "taxonogenomics" représente le futur de la taxonomie avec l'actualisant par l'ajout des dernières technologies. Ainsi, un nouveau puit de connaissance

est créé avec le séquençage de génomes bactériens enrichissant les banques de données et réduisant la part de séquences non assignées des métagénomes.

Cependant, le séquençage et l'analyse d'un génome complet prennent un temps non négligeable qui s'ajoute au délai considérable entre la publication et l'officialisation des noms proposés pour les nouvelles espèces. Tous ces délais entrainent une disponibilité à la communauté scientifique retardée et de ce fait, un retard des futures avancées scientifiques liées à ces nouvelles espèces. Le format de description intermédiaire, new species announcement, permet de pallier à ces problèmes en mettant à disposition de la communauté scientifique les caractéristiques de base de la potentielle nouvelle espèce. Ce format permet également d'éviter le chaos taxonomique actuellement observé dû à la publication et officialisation tardives de ces potentielles nouvelles espèces (41).

Dans cette partie sont présentées les descriptions et "new species announcements" réalisés au cours de ma thèse. La description des nouvelles espèces non incluses dans ce manuscript est en cours.

Article 5: *Anaerococcus rubiinfantis* sp. nov., isolated from the gut microbiota of a Senegalese infant with severe acute malnutrition.

Maryam Tidjani Alou, Saber Khelaifia, Caroline Michelle, Claudia Andrieu, Nicholas Armstrong, Fadi Bittar, Cheikh Sokhna, Aldiouma Diallo, Pierre-Edouard Fournier, Didier Raoult, Matthieu Million.

Publié dans Anaerobe

Anaerobe 40 (2016) 85–94

Contents lists available at ScienceDirect

Anaerobe

journal homepage: www.elsevier.com/locate/anaerobe

Anaerobes in the microbiome

Anaerococcus rubiinfantis sp. nov., isolated from the gut microbiota of a Senegalese infant with severe acute malnutrition

Maryam Tidjani Alou [a], Saber Khelaifia [a], Caroline Michelle [a], Claudia Andrieu [a], Nicholas Armstrong [a], Fadi Bittar [a], Cheikh Sokhna [b], Aldiouma Diallo [b], Pierre-Edouard Fournier [a], Didier Raoult [a, b, c], Matthieu Million [a, *]

[a] Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes IRD 198, CNRS 7278, Aix-Marseille Université, Marseille, France
[b] Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes IRD 198, CNRS 7278, Campus Commun UCAD-IRD of Hann, Dakar, Senegal
[c] Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 5 March 2016
Received in revised form
16 June 2016
Accepted 17 June 2016
Available online 18 June 2016

Handling Editor: Elisabeth Nagy

Keywords:
Anaerococcus rubiinfantis
Gut microbiota
Culturomics
Taxono-genomics
Malnutrition

ABSTRACT

Anaerococcus rubiinfantis sp. nov. strain mt16^T is a new species within the genus *Anaerococcus*, which was isolated by the culturomics approach from the gut microbiota of an infant suffering from kwashiorkor. A phenotypic, biochemical and proteomic description of this strain is hereby presented alongside a complete annotation of its genome. This strictly anaerobic species forms Gram-positive non-sporeforming cocci. The major fatty acid was hexadecanoic acid. The phylogenetic analysis of strain mt16^T showed a 97.9% similarity level with *Anaerococcus vaginalis*, the closest validly published species. Its genome is 1,929,161 bp long with 29.5% G + C content and contains 1808 protein-coding genes and 56 RNA genes, among which are six rRNA genes. Genomic analysis identified 41/1864 coding genes as ORFans (2.2%) and at least 620/1808 (34.9%) orthologous proteins which are not shared with the closest phylogenetic species. We believe that the extension of the human anaerobic gut compendium by culturomics is one of the first steps that will improve the understanding of the links between the microbiome and health or disease.

1. Introduction

Severe acute malnutrition, as the leading cause of child mortality, is a major public health issue [1]. Recently, several studies have highlighted a dysbiosis of the gut microbiota of children affected with severe acute malnutrition [2,3]. In order to characterize the gut microbiota of severely undernourished children, we used the "culturomics" approach, which consists in the multiplication of culture conditions with a variation of media, temperature, pH and atmosphere conditions to describe a bacterial ecosystem as exhaustively as possible [4]. A new member of the *Anaerococcus* genus was thus isolated. This genus was created by Ezaki in 2001 and contains eight validated species, the type species being *Anaerococcus prevotii* [5]. These Gram-positive, obligate anaerobic, and non-spore forming cocci were mostly isolated from human vaginal discharges and various purulent secretions [6].

The recent breakthrough in molecular techniques has allowed additional data such as the complete description of the genome, to be added to classical bacterial classification, which is based on phenotypic and genotypic characteristics and phylogeny based on the 16s rRNA gene [7–9]. Thanks to the decreasing cost of genome sequencing, a new concept of bacterial description [10], taxono-genomics, was created in our laboratory allying classical bacterial classification features to a proteomic description with the MALDI-TOF profile [11] and a description of the complete annotated genome of the described species [12].

Here follows the description of a new member of the

Abbreviations: AGIOS, average of genomic identity of orthologous gene sequences; bp, base pairs; COG, Clusters of Orthologous Groups; CSUR, Collection de souches de l'Unité des Rickettsies; dDDH: digital, DNA-DNA hybridization; DSM, Deutsche Sammlung von Mikroorganismen; EI, electron impact; FAME, fatty acid methyl ester; FTP, file transfer protocol; GC/MS, gaz chromatography/mass spectrometry; INSDC, International Nucleotide Sequence Database Collaboration; kb, kilobases; MALDI-TOF MS, matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry; MIC, minimal inhibitory concentration; MS, mass spectrometry; NRPS, non-ribosomal peptide synthase; ORF, open reading frame; PKS, polyketide synthase; TE buffer, tris-EDTA buffer; SDS, sodium dodecyl sulfate; URMITE, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes.

* Corresponding author.
E-mail address: matthieumillion@gmail.com (M. Million).

http://dx.doi.org/10.1016/j.anaerobe.2016.06.007
1075-9964/

Anaerococcus genus isolated from the fecal sample of a Senegalese girl suffering from severe acute malnutrition (kwashiorkor), strain mt16^T (=CSUR P2032 = DSM 101186).

2. Materials and methods

2.1. Sample information

The fecal sample of a 7-month-old severely malnourished girl was collected with the consent of the child's parent. At the time of the collection, no antibiotic was administered to the patient and the sample was frozen at −80 °C upon reception at the hospital La Timone in Marseille, France. Approval for this study was given by the Institut Fédératif de Recherche 48 (Faculty of Medicine, Marseille, France) under agreement number 09–022. The collected sample was cultivated using the eighteen conditions of standardized culturomics, as described previously [13].

2.2. Growth conditions

The ideal growth environment of strain mt16^T was determined by testing different temperatures (25, 28, 37, 45 and 56 °C) under aerobic, anaerobic and microaerophilic growth. Aerobic growth was tested with or without 5% CO_2. Anaerobic and microaerophilic atmospheres were created using respectively GENbag Anaer and GENbag microair systems (BioMerieux, Marcy l'Etoile, France).

2.3. Morphologic, biochemical and antibiotic susceptibility tests

Determination of phenotypic characteristics (oxidase, catalase, gram staining, motility, sporulation) was realized as carried out previously [14]. For electron microscopy with negative staining, a 40 μL bacterial suspension drop was deposited on formvar coated grids, and incubated at 37 °C for 30 min. It was then followed by a 10 s incubation on ammonium molybdate 1%. The grids were dried on blotting paper and finally observed using a Tecnai G20 transmission electron microscope (FEI Company, Limeil-Brevannes, France).

Biochemical features were described using API 20A, ZYM and 50CH strips according to the manufacturer's instructions (BioMérieux). Since strain mt16 was isolated in anaerobic condition, the strips were all incubated in anaerobic atmosphere using GENbag Anaer (BioMerieux) for 4, 24 and 48 h for API ZYM, 20A and 50CH strips respectively. In order to test nitrate reduction, we also used an API 20NE strip in the same conditions as the API 20A strip.

Strain mt16 was grown on 5% sheep blood enriched Colombia agar (BioMerieux) for the fatty acids analysis which was carried out by GC/MS. For this purpose, two samples were prepared with approximately 15 mg of bacterial biomass each harvested from several culture plates. Cellular fatty acid methyl esters (FAME) were prepared as described by Sasser [15]. Briefly, GC/MS analyses were realized by using a Clarus 500 gas chromatograph equipped with a SQ8S MS detector (Perkin Elmer, Courtaboeuf, France). 2 μL of FAME extracts were volatilized at 250 °C (split 20 mL/min) in a Focus liner with wool and separated on an Elite-5MS column (30 m, 0.25 mm i.d., 0.25 mm film thickness) using a linear temperature gradient (70–290 °C at 6 °C/min), allowing the detection of C4 to C24 fatty acid methyl esters. Helium flowing at 1.2 mL/min was used as carrier gas. The MS inlet line was set at 250 °C and EI source at 200 °C. Full scan monitoring was performed from 45 to 500 *m/z*. All data were collected and processed using Turbomass 6.1 (Perkin Elmer). FAMEs were identified by a spectral database search using MS Search 2.0 operated with the Standard Reference Database 1A (National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA) and the FAMEs mass spectral database (Wiley, Chichester, UK). Retention time correlations with estimated nonpolar retention indexes from the NIST database were obtained using a 37-component FAME mix (Supelco; Sigma-Aldrich, Saint-Quentin Fallavier, France); FAME identifications were confirmed using this index. Our results were compared to that of *Anaerococcus* strains readily available in our in house collection, CSUR, namely *Anaerococcus vaginalis* CSUR P727, *Anaerococcus senegalensis* CSUR P156, *Anaerococcus murdochii* CSUR P1063, *Anaerococcus tetradius* CSUR P1064 and *Anaerococcus octavius* CSUR P179.

Antibiotic susceptibility was determined using the antibiotic gradient strip test method with Etest strips (BioMerieux) thus allowing to obtain the minimal inhibitory concentration (MIC) for each tested antibiotic. Inoculum suspension was prepared with a culture of strain mt16^T grown on sheep blood enriched Colombia agar (BioMerieux) suspended in a sterile saline solution (0.85% NaCl) to obtain a turbidity of 0.5 McFarland. The inoculum was then plated with cotton swabs on 5% horse blood and β-NAD enriched Mueller Hinton Agar (BioMerieux) as per EUCAST recommendations [16,17]. Etest strips (amoxicillin, amoxicillin/clavulanic acid, clindamycin, imipenem, metronidazole, penicillin G, vancomycin) were deposited. The plates were incubated for 48 h in an anaerobic atmosphere. Elliptical zones of inhibition are formed around the strip and the intersection with the strip indicated the MIC [17]. *Escherichia coli* strain DSM 1103 was used as a quality control strain. MICs were interpreted according to the EUCAST recommendations [18].

2.4. Strain identification by MALDI-TOF MS

Strain mt16 was identified using MALDI-TOF MS technology [11,19]. This proteomic analysis was conducted using a Microflex spectrometer (Bruker Daltonics, Leipzig, Germany) with a MTP 96 MALDI-TOF target plate (Bruker). The obtained spectra for a colony are imported into the MALDI BioTyper software (version 2.0, Bruker) and analyzed by standard pattern matching (with default parameter settings) against the 7567 references contained in our database (Bruker database incremented with our data). An identification score was obtained: a score ≥1.9 allowed an identification at the species level; a score between 1.9 and 1.7 allowed identification only at the genus level and finally a score <1.7 gave no identification.

2.5. Strain identification by 16S rRNA sequencing

Due to an identification of <1.7 for the MALDI-TOF MS analysis, the 16S rRNA gene was sequenced using fD1 and rP2 primers as previously described [20]. Stackebrandt and Ebers [21] determined a 98.65% similarity level threshold to define a new species without performing DNA-DNA hybridization.

2.6. Genomic DNA preparation

Strain mt16^T was cultured on 5% sheep blood-enriched Columbia agar (BioMerieux) at 37 °C anaerobically. Bacteria grown on three Petri dishes were resuspended in 400 μL of TE buffer. Then, 200 μL of this suspension was diluted in 1 mL TE buffer for lysis treatment that included a 30- minute incubation with 2.5 μg/μL lysozyme at 37 °C, followed by an overnight incubation with 20 μg/μL proteinase K at 37 °C. Extracted DNA was then purified using 3 successive phenol-chloroform extractions and ethanol precipitations at −20 °C overnight. After centrifugation, the DNA was resuspended in 160 μL TE buffer.

2.7. Genome sequencing and assembly

Genomic DNA (gDNA) of strain mt16 was sequenced on the 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). gDNA was quantified by a Qubit assay with the high sensitivity kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA) to 66.2 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 (tagmentation) 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 kb up to 11 kb with an optimal size at 3.927 kb. No size selection was performed and 505 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal length at 597 bp on the Covaris device S2 in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies Inc, Santa Clara, CA, USA) and the final concentration library was measured at 59.2 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. An automated cluster generation and sequencing run was performed in a single 39-h run in a 2 × 251-bp.

2.8. Genome annotation and analysis

Open Reading Frames (ORFs) were predicted using Prodigal [22] 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 against the Clusters of Orthologous Groups (COG) 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 $1e^{-05}$. The tRNAScanSE tool [23] was used to find tRNA genes, whereas ribosomal RNAs were found by using RNAmmer [24]. Lipoprotein signal peptides and the number of transmembrane helices were predicted using Phobius [25]. 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).

Genomes were automatically retrieved from the 16S RNA tree using XEGEN software (Phylopattern) [26]. For each selected genome, the complete genome sequence, proteome genome sequence (all gene sequences encoding proteins in a genome) and Orfeome genome sequence (all gene sequences encoding orphan proteins in a genome) were retrieved from the FTP of NCBI. All proteomes were analyzed with proteinOrtho [27]. Then for each couple of genomes, a similarity score was computed. This score is the mean value of nucleotide similarity between all couples of orthologues between the two genomes studied (AGIOS) [28]. 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). To evaluate the genomic similarity among studied *Anaerococcus* strains, we determined two parameters; dDDH, which exhibits a high correlation with DDH [29,30], and AGIOS [28], which was designed to be independent from DDH. dDDH was determined using the Genome-to-Genome Distance Calculator (GGDC) server (http://ggdc.dsmz.de/distcalc2.php). Genomes were aligned by the GGDC 2.0 Blast + method and results were interpreted with the formula 2 as recommended [30]. Annotation and comparison processes were performed in the Multi-Agent software system DAGOBAH [31] that include Figenix [32] libraries to provide pipeline analysis.

3. Results

3.1. Phenotypic description

Growth was observed at 28, 37 and 45 °C in anaerobic atmosphere. Ideal growth was observed after 48 h at 37 °C in anaerobic atmosphere. No growth occurred in aerobic and microaerophilic conditions. Non-sporeforming and non-motile cells formed small and white colonies with a mean diameter of 2 mm. Gram-staining showed Gram-positive cocci (Fig. 1). Negative staining of strain mt16 observed in electron microscopy showed cocci arranged in pairs with a mean diameter 0.9 μm (Fig. 2).

No oxidase activity was exhibited. Nevertheless, a positive catalase activity was observed for strain mt16. An API ZYM strip also showed positive activity for other enzymes: alkaline phosphatase, leucine arylamidase, valine arylamidase, acid phosphatase and Naphthol-AS-BI-phosphohydrolase. No activity was found for the following enzymes: esterase (C4), lipase esterase (C8), lipase (C14), cysteine arylamidase, trypsin, α-chymotrypsin, α-galactosidase, β-galactosidase, β-glucuronidase, α-glucosidase, β-glucosidase, *N*-acetyl-β-glucosaminidase, α-mannosidase and α-fucosidase. Nitrate reductase and nitrite reductase activities were absent according to the API 20A strip. Using an API 20A strip, indole formation was positive and acid was produced from D-glucose, D-sucrose, D-maltose and D-mannose. No acid production was observed from D-lactose, D-salicine, D-xylose, L-arabinose, glycerol, D-cellobiose, D-melezitose, D-raffinose, D-sorbitol, L-rhamnose and D-trehalose. Urease, β-glucosidase and protease activities were absent. An API 50 CH also showed the metabolism of the following carbohydrates: tagatose and potassium-5-ketogluconate. The other carbohydrates sampled in the 50CH strip were not metabolized by strain mt16: glycerol, erythritol, D-arabinose, L-arabinose, D-ribose, D-xylose, L-xylose, D-adonitol, methyl-βD-xylopyranoside, D-galactose, D-glucose, D-fructose, D-mannose, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl-αD-mannopyranoside, methyl-αD-glucopyranoside, amygdalin, arbutin, esculin ferric citrate, salicin, D-cellobiose, D-maltose, D-lactose, D-melibiose, D-sucrose, D-trehalose, inulin, D-melezitose, D-raffinose, starch, glycogen, xylitol, gentiobiose, D-turanose, D-lyxose, D-fucose, L-fucose, D-arabitol, L-arabitol, potassium gluconate and potassium 2-ketogluconate.

The most abundant fatty acid was hexadecanoic acid (53%).

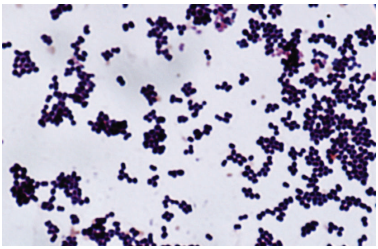


Fig. 1. Gram staining. Gram staining of *Anaerococcus rubiinfantis* strain mt16^T.

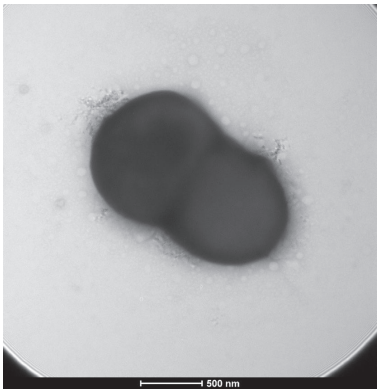


Fig. 2. Transmission electron microscopy. Transmission electron microscopy of *Anaerococcus rubiifantis* strain mt16^T using a Tecnai G20 transmission electron microscope (FEI Company). The scale bar represents 500 nm.

Unsaturated fatty acids 13-octadecenoic and 9,12-octadecadienoic were also measured with significant quantities. Minor amounts of other, mainly saturated, fatty acids, were also detected (Table 1). The comparison to other *Anaerococcus* species showed a relative similarity in the fatty acid profiles with 16:0 being the major fatty acid except in *A. murdochii* (major fatty acid: 18:1n9). All compared *Anaeroccoccus* species present core abundant fatty acids that are 16:0, 18:1n9, 18:2n6, 18:0 and 14:0 (Table 1).

Strain mt16 was susceptible to benzylpenicillin (MIC < 0.002 μg/mL), amoxicillin (MIC < 0.016 μg/mL), amoxicillin/clavulanic acid (MIC < 0.016 μg/mL), clindamycin (MIC < 0.023 μg/mL), imipenem (MIC < 0.002 μg/mL), metronidazole (MIC < 0.016 μg/mL) and vancomycin (MIC < 0.016 μg/mL). For *Escherichia coli* strain DSM 1103 which was used as a quality control strain, the following MICs (4, 4, 0.125 μg/mL) were obtained for amoxicillin, amoxicillin/clavulanic acid and imipenem respectively. For this quality control strain, no growth inhibition was observed with the following antibiotics: benzylpenicillin, clindamycin, metronidazole and vancomycin.

3.2. Strain identification by MALDI-TOF MS

Strain mt16 was first isolated in April 2015 after a 10 days' anaerobic pre-incubation in liquid Colombia broth supplemented with 5% sheep blood at 37 °C. The liquid culture was then inoculated on 5% sheep blood Colombia agar (BioMerieux) and incubated in anaerobic condition at 37 °C. According to the MALDI-TOF analysis, strain mt16's highest match was *A. vaginalis* with a score of 1.575. This score under 1.7 did not allow the identification of strain mt16 at the species nor at the genus level and suggested that the closest species might be *A. vaginalis*. Strain mt16 was deposited on a MTP 96 target plate twelve times and the reference spectra obtained (Fig. 3) were incremented into our database. The spectrum of strain mt16 was also compared to other *Anaerococcus* species as represented in Fig. 4. The intensity of the peaks corresponding to strain mt16 was generally lower than that of compared *Anaerococcus* species, including its closest species, *A. vaginalis*.

3.3. Strain identification by 16S rRNA sequencing

In order to identify strain mt16, the 16S rRNA gene was sequenced and the obtained sequence (Accession number LN881592) showed a 97.9% similarity level with *A. vaginalis* strain CCUG 31349^T (Accession number AF542229.1), the phylogenetically closest validated species (Fig. 5). The general characteristics of strain mt16^T are presented in Table 2. According to the similarity level threshold defined by Stackebrandt and Ebers [20], we propose strain mt16^T as the type strain of a new species within the *Anaerococcus* genus for which we suggest the name *Anaerococcus rubiinfantis* sp. nov.

3.4. Genome properties

Strain mt16^T's genome (Accession number FAVH00000000) is 1,929,161 bp with a 29.5% G + C content (Table 3, Fig. 6). At the final stage of assembly, it is composed of 3 scaffolds (composed of 5 contigs). There were 1864 predicted genes including 1808 protein-coding genes and 56 RNA genes. Among the RNA genes, there were 3 5S rRNA genes, 2 23S rRNA genes, a unique 16s rRNA gene and 50 tRNA genes. A putative function was assigned by COGS or NR blast to 1427 genes. 41 genes were identified as ORFans and the

Table 1
Cellular fatty acid profiles (%) of *Anaerococcus rubiinfantis* strain mt16 compared with other *Anaerococcus* species.

Fatty acids	Name	*A. rubiinfantis*	*A. vaginalis*[a]	*A. senegalensis*[a]	*A. tetradius*[a]	*A. octavius*[a]	*A. murdochii*[a]
5:0 anteiso	2-methyl-butanoic acid	TR	ND	ND	ND	ND	ND
5:0 iso	3-methyl-butanoic acid	ND	ND	ND	ND	9.0 ± 1.6	ND
10:0	Decanoic acid	ND	1.2 ± 0.2	TR	TR	ND	ND
12:0	Dodecanoic acid	TR	TR	TR	TR	TR	TR
14:0	Tetradecanoic acid	2.1 ± 0.03	6.9 ± 0.2	4.4 ± 0.2	2.7 ± 0.1	4.1 ± 0.2	1.9 ± 0.0
15:0	Pentadecanoic acid	TR	1.0 ± 0.1	TR	TR	TR	TR
16:0	Hexadecanoic acid	**52.7 ± 0.4**	**59.2 ± 1.1**	**54.6 ± 0.3**	**35.7 ± 1.2**	**49.6 ± 0.6**	28.6 ± 0.2
16:1n7	9-Hexadecenoic acid	1.1 ± 0.1	TR	ND	TR	ND	ND
17:0	Heptadecanoic acid	TR	ND	TR	TR	ND	TR
17:0 anteiso	14-methyl-Hexadecanoic acid	ND	ND	ND	TR	ND	ND
18:0	Octadecanoic acid	3.6 ± 0.3	5.7 ± 0.4	11.5 ± 1.4	16.7 ± 0.3	9.2 ± 0.3	13.4 ± 0.4
18:1n6	12-Octadecenoic acid	ND	ND	TR	1.9 ± 0.2	ND	1.9 ± 0.1
18:1n7	11-Octadecenoic acid	ND	ND	ND	1.0 ± 0.3	ND	TR
18:1n9	9-Octadecenoic acid	28.2 ± 0.04	18.8 ± 0.8	17.6 ± 0.7	29.9 ± 0.5	23.5 ± 1.4	**35.9 ± 0.3**
18:2n6	9,12-Octadecadienoic acid	10.1 ± 0.5	6.5 ± 0.3	9.2 ± 0.6	11.0 ± 0.2	3.8 ± 0.3	17.0 ± 0.7
20:4n6	5,8,11,14-Eicosatetraenoic acid	ND	ND	ND	ND	ND	TR

These values represent the mean peak area percentage calculated from the analysis of FAMEs in 3 sample preparations ± standard deviation (n = 3); TR = trace amounts < 1%. Bold represents the majority fatty acid for this species.

[a] These strains are provided by the Collection des Souches de l'Unité des Rickettsies (CSUR).

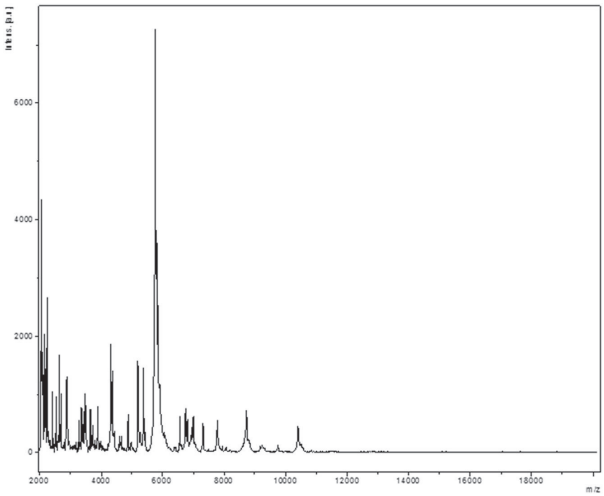


Fig. 3. Reference mass spectrum from *Anaerococcus rubiinfantis* strain mt16^T. Spectra from 12 individual colonies were compared and a reference spectrum was generated.

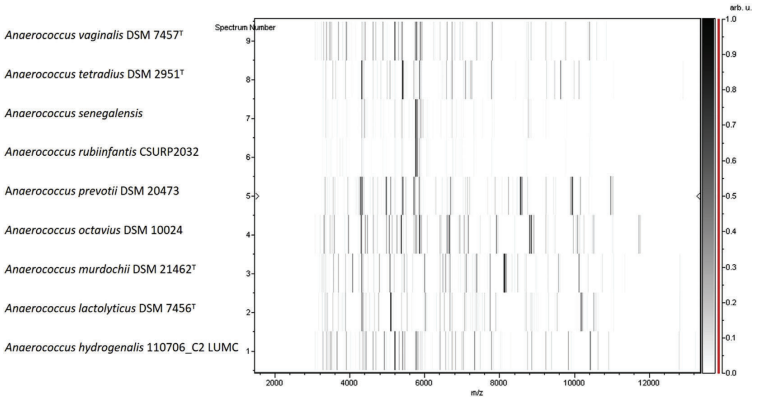


Fig. 4. MALDI-TOF MS comparative analysis of *Anaerococcus* species. Gel view comparing *Anaerococcus rubiinfantis* strain mt16^T to other species within the *Anaerococcus* genus. The gel view displays the raw spectra of loaded spectrum files arranged in a pseudo-gel like look. The x-axis records the m/z value. The left y-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.

remaining 274 genes were identified as encoding hypothetical proteins. The gene distribution into COG functional categories is reported in Table 4.

3.5. Genome comparison

Strain mt16 genomic characteristics were compared to the characteristics of closely related species (Table 5).

The size of the draft genome sequence of strain mt16 (1.93 MB) is smaller than those of *A. prevotii* (2 MB), *A. tetradius* (2.10 MB) and *Anaerococcus lactolyticus* (2.18 MB) but larger than the genomes of *Peptoniphilus harei* (1.84 MB), *Anaerococcus hydrogenalis* (1.89 MB) and *A. vaginalis* (1.87 MB). The G + C content of strain mt16 (29.45%) is smaller than most of the compared genomes: *P. harei* (34.44%), *A. tetradius* (34.15%), *A. hydrogenalis* (29.64%), *A. prevotii* (35.64%) and *A. lactolyticus* (34.94%). Only *A. vaginalis* G + C content (28.87%) is smaller than that of strain mt16.

The gene content of strain mt16 is larger than *P. harei*, *A. vaginalis* and *A. prevotii*'s gene contents (1,724, 1764 and 1806 respectively) but smaller than *A. tetradius*, *A. hydrogenalis* and

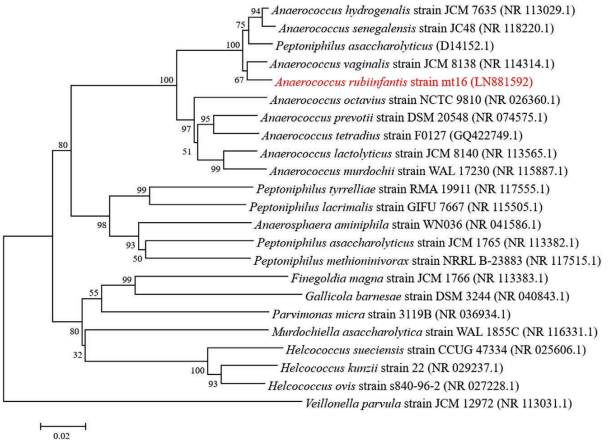


Fig. 5. Phylogenetic tree highlighting the position of *Anaerococcus rubiinfantis* strain mt16^T relative to other close strains. The respective GenBank accession numbers for 16S rRNA genes are indicated in parenthesis. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using the maximum-likelihood method within the MEGA6 software. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 1000 times to generate a majority consensus tree. *Parvimonas micra* strain 3119B was used as an outgroup. The scale bar represents a 1% nucleotide sequence divergence.

Table 2
Classification and general features of *Anaerococcus rubiinfantis* strain mt16^T.

Property	Term
Current classification	Domain: *Bacteria* Phylum: *Firmicutes* Class: *Tissierellia* Order: *Tissierellales* Family: *Peptoniphilaceae* Genus: *Anaerococcus* Species: *Anaerococcus rubiinfantis* Type strain: mt16
Gram stain	Positive
Cell shape	Cocci
Motility	Non-motile
Sporulation	Non-sporulating
Temperature range	Mesophilic
Optimum temperature	37 °C

Table 3
Nucleotide content and gene count levels of the genome.

Attribute	Genome (total)	
	Value	% of total[a]
Size (bp)	1,929,161	100
G + C content (%)	568,166	29.45
Coding region (bp)	1,737,861	90.08
Total genes (total ORFs)	1864	100
RNA genes	56	3
Protein-coding genes	1808	97
Genes with function prediction	1427	78.9
Genes assigned to COGs	1243	68.75
Genes with signal peptides	163	9.01
Genes with transmembrane helices	412	22.78
ORFan genes	41	2.26
Genes associated with PKS or NRPS[b]	2	0.11
N° of antibiotic resistance 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.
[b] PKS: Polyketide Synthase; NRPS: Non-Ribosomal Peptide Synthase.

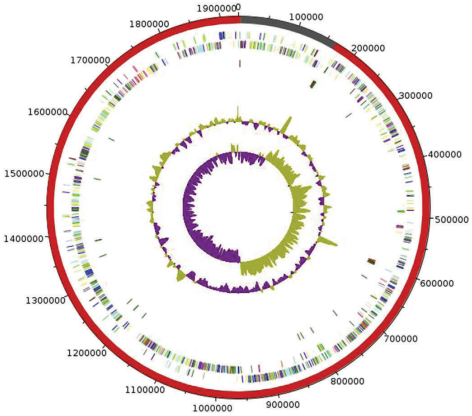


Fig. 6. Graphical circular map of the chromosome. From outside to the center: Contigs (red/grey), genes on the forward strand colored by COG categories (only gene assigned to COG), RNA genes (tRNAs green, rRNAs red), G + C content and G + C skew (purple: value below the mean, yellow: value above the mean). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

A. lactolyticus's gene contents (2,079, 2069 and 2253 respectively).

The distribution of genes into COG categories was similar in all compared genomes except for the carbohydrates metabolism and transport for *P. harei*, for which this function is reduced (Fig. 7). Strain mt16 shared 1,003, 738, 1,061, 1,073, 1188 and 1091 orthologous genes with *A. lactolyticus, P. harei, A. tetradius, A. hydrogenalis, A. vaginalis and A. prevotii* respectively (Table 6). Among species with standing in nomenclature, AGIOS values ranged from 58 to

Table 4
Number of genes associated with the 25 general COG functional categories.

Code	Value	% of total[a]	Description
[J]	167	9.23	Translation
[A]	0	0	Rna processing and modification
[K]	90	4.97	Transcription
[L]	63	3.48	Replication, recombination and repair
[B]	0	0	Chromatin structure and dynamics
[D]	19	1.05	Cell cycle control, mitosis and meiosis
[Y]	0	0	Nuclear structure
[V]	43	2.37	Defense mechanisms
[T]	47	2.59	Signal transduction mechanisms
[M]	58	3.20	Cell wall/membrane biogenesis
[N]	5	0.27	Cell motility
[Z]	0	0	Cytoskeleton
[W]	3	0.16	Extracellular structures
[U]	16	0.88	Intracellular trafficking and secretion
[O]	57	3.15	Posttranslational modification, protein turnover,chaperones
[X]	20	1.10	Mobilome: prophages, transposons
[C]	88	4.86	Energy production and conversion
[G]	106	5.86	Carbohydrate transport and metabolism
[E]	121	6.69	Amino acid transport and metabolism
[F]	62	3.43	Nucleotide transport and metabolism
[H]	77	4.26	Coenzyme transport and metabolism
[I]	33	1.82	Lipid transport and metabolism
[P]	79	4.37	Inorganic ion transport and metabolism
[Q]	14	0.77	Secondary metabolites biosynthesis, transport and catabolism
[R]	121	6.69	General function prediction only
[S]	66	3.65	Function unknown
–	565	31.25	Not in COGs

[a] The total is based on the total number of protein coding genes in the annotated genome.

Table 5
Genome comparison of closely related species to *Anaerococcus rubiinfantis* strain mt16^T.

Name of organisms	INSDC identifier[a]	Size (Mb)	G + C (%)	Protein coding genes	Total genes (total ORFs)
Anaerococcus rubiifantis Strain mt16^T	FAVH00000000	1.92	29.45	1808	1864
Anaerococcus hydrogenalis DSM 7454	ABXA00000000	1.89	29.64		2069
Anaerococcus vaginalis ATCC 51170	ACXU00000000.1	1.87	28.87		1764
Peptoniphilus harei ACS-146-V-Sch2b	AENP00000000.1	1.84	34.44		1724
Anaerococcus lactolyticus ATCC 51172	JRMW00000000.1	2.18	34.94		2253
Anaerococcus prevotii DSM 20548	CP001708.1	2.00	35.64		1806
Anaerococcus tetradius ATCC 35098	ACGC00000000.1	2.10	34.15		2079

[a] INSDC: International Nucleotide Sequence Database Collaboration.

79% among compared species except strain mt16. When comparing strain mt16 to other species, the AGIOS value ranged from 65% with *P. harei* to 88% with *A. vaginalis* (Table 6). dDDH values estimated between strain mt16^T and the compared genomes ranged from 18.8% to 34.7%. Well under 70%, these values confirm strain mt16^T as a new species [30].

3.6. Comparison with other Anaerococcus species

The characteristics of strain mt16 were compared to the characteristics of closely related *Anaerococcus* species. Differences were observed for several characteristics such as cell diameter, indole production, enzymes production, acid production from carbohydrates, G + C content, and habitat as reported in Table 7.

4. Discussion

Strain mt16^T was isolated from the stool sample of an 8-month old infant afflicted with kwashiorkor. Phenotypic characterization, MALDI-TOF MS, 16S rRNA sequencing, and comparative genomic analyses among close phylogenetic relatives allowed identification of strain mt16^T as a new species within the genus *Anaerococcus*, with closest relative being *A. vaginalis*. However, several discrepancies with *A. vaginalis* including a 16S rRNA similarity, low dDDH and low AGIOS values confirmed it as a new species.

Typical features of the present strain included a susceptibility to all tested antibiotics consistent with the absence of known antibiotic resistance genes (Table 4). This high antibiotic susceptibility is clinically relevant and suggesting that gut *Anaerococcus* species could be altered by oral antibiotics. Further studies are needed to determine whether *Anaerococcus rubiinfantis*, found in this study in a severely malnourished child, is part of the Healthy Mature Anaerobic Gut Microbiota [3], is associated with malnutrition, or has no role on nutritional status. Moreover, future studies should elucidate if gut symbionts resistant to commonly used antibiotics are more likely to be pathogenic, such as *Clostridium difficile*

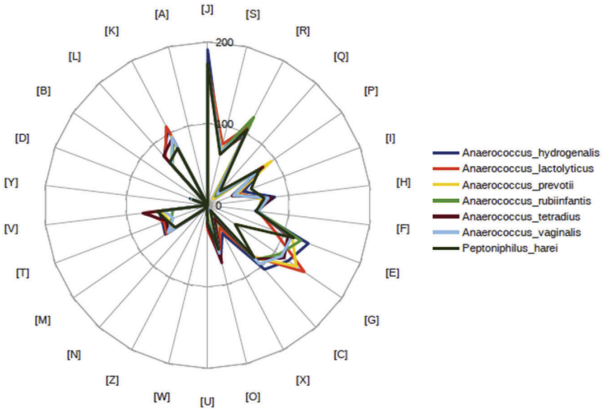


Fig. 7. Distribution of functional classes of predicted genes according to the clusters of orthologous groups of protein. For the codes of COG categories, see Table 4.

Table 6
The number of orthologous proteins shared between genomes (upper right)[a].

	Anaerococcus lactolyticus	*Peptoniphilus harei*	*Anaerococcus tetradius*	*Anaerococcus hydrogenalis*	*Anaerococcus rubiinfantis*	*Anaerococcus vaginalis*	*Anaerococcus prevotii*
A. lactolyticus	**2253**	727	1145	938	1003	1047	1108
P. harei	0.65	**1724**	752	677	738	753	733
A. tetradius	0.73	0.65	**2079**	980	1061	1111	1193
A.hydrogenalis	0.6	0.58	0.60	**2069**	1073	1083	989
A.rubiinfantis	0.70	0.65	0.71	0.68	**1808**	1188	1091
A.vaginalis	0.70	0.66	0.71	0.67	0.88	**1764**	1082
A.prevotii	0.72	0.65	0.79	0.60	0.70	0.70	**1806**

Bold represents the total number of orthologous proteins for this species.
[a] Average percentage similarity of nucleotides corresponding to orthologous proteins shared between genomes (lower left) and numbers of proteins per genome.

(antibiotic resistance by sporulation) and the aerotolerant *Klebsiella oxytoca* previously identified as a causative organism of antibiotic-associated hemorrhagic colitis [33].

A paradoxical feature observed in our strain and several other *Anaerococcus* members (*A. vaginalis, A. senegalensis, A. prevotii and some strains of A. tetradius and A. octavius* – Table 7) was the presence of a catalase enzyme [34]. In fact, a catalase activity has been found in several strictly anaerobic species such as some members of the genera *Bacteroides*, *Clostridium*, *Porphyromonas*, *Desulfovibrio* and even few archaeal species [34]. This suggest a different oxidative stress resistance level among anaerobic prokaryotes. Indeed, while *Methanobrevibacter smithii* and *Faecalibacterium prausnitzii* are prokaryotes among the most sensitive ones to oxygen (*F. prausnitzii* loses its viability within 2 min after exposure to ambient air [35]), we observed a trace but non-significant growth of *A. rubiinfantis* in microaerophilic conditions, as previously reported for *A. senegalensis* [12]. These results suggest that the aerobic/anaerobic bacterial culture dichotomy is no longer relevant [36] and that there is a continuum between species whose growth is not affected by a normal (ambient air) partial pressure of oxygen and/or positive redox potential, and species that only grow in total absence of oxygen and/or when redox potential is very low [37]. Moreover, tolerance to oxygen and to redox potential of the environment seems completely dependent on the presence of diverse antioxidants [36] and electron acceptors in the medium [35]. This opens new perspectives for exploration in the field of the 'anaerobic' prokaryotes.

Finally, we confirmed the usefulness of genomic analysis as an updated tool to identify, classify and describe more accurately new bacterial strains as promoted by the taxonogenomics description concept [10]. Differences between strain mt16 and *A. vaginalis* (its closest neighbor) also included the production of alkaline phosphatase as well as the production of acids from the following carbohydrates: arabinose, mannitol and sucrose [38]. Strain mt16 was classified among the *Anaerococcus* genus that currently contains eight validated species [5] which are Gram-positive, strictly anaerobic, non-motile, and mostly non-sporeforming species [6,12]. The phylogenetic relationship between several members of the family *Peptoniphilaceae* to which the genus *Anaerococcus* belongs, displayed in Fig. 5, showed strain mt16 in the midst of the validated *Anaerococcus* species.

Given the phylogenetic analysis of the 16S rRNA gene sequence of strain mt16^T, its proteomic profile obtained with MALDI-TOF MS, its genomic description and annotation and dDDH values, we thus defined strain mt16^T as a new species called *Anaerococcus rubiinfantis* sp. nov. (=CSUR P2032 = DSM 101186).

Taxonomic and nomenclatural proposals Description of *Anaerococcus rubiinfantis* strain mt16^T sp. nov. (from ru.be.us (adj) infantis (L. gen. n.) meaning of a red infant, which is a reference to the hair discoloration observed in kwashiorkor patients).

Table 7
Differential characteristics of *Anaerococcus rubiinfantis* strain mt16^T, *Anaerococcus vaginalis* JCM 8138^T, *Anaerococcus hydrogenalis* JCM 7635^T, *Anaerococcus senegalensis* DSM 25366^T, *Anaerococcus prevotii* ATCC 9321^T, *Anaerococcus tetradius* JCM 1964^T, *Anaerococcus lactolyticus* JCM 8140^T, *Anaerococcus octavius* NCTC 9810^T, *Anaerococcus murdochii* ATCC-1385^T, *Bacillus sonorensis* DSM 13779^T [12,38–43].

Properties	*Anaerococcus rubiinfantis*[a]	*Anaerococcus vaginalis*[b]	*Anaerococcus hydrogenalis*[b]	*Anaerococcus senegalensis*[a]	*Anaerococcus prevotii*[b]	*Anaerococcus tetradius*[b]	*Anaerococcus lactolyticus*[b]	*Anaerococcus octavius*[b]	*Anaerococcus murdochii*[b]
Cell diameter (μm)	0.9	na	0.7–1.8	0.87	0.7–1.8	<1.8	na	0.7–0.9	≥0.7
Indole	+	+	+	+	−	−	−	−	−
Major fatty acid[a]	16:0	16:0	na	16:0	na	16:0	na	16:0	18:1n9
G + C content (%)	29.5	30–34	30–34	28.56	29–33	30–32	30–34	26–31	na
Production of									
Alkaline phosphatase	+	−	−	+	−	−	−	−	+
Catalase	+	na	−	+	+	±	na	±	−
Nitrate reductase	−	na	−	−	−	−	na	−	−
Urease	−	−	±	+	±	+	+	−	−
β-galactosidase	−	−	−	−	−	−	+	−	+
Acid from									
L-Arabinose	+	−	na	−	−	−	−	−	na
Ribose	−	na	na	+	+	−	−	−	na
Mannose	+	±	+	+	+	+	+	+	+
Mannitol	+	−	+	−	−	−	−	±	na
Sucrose	+	−	−	+	−	+	−	−	na
D-glucose	+	+	+	+	±	±	+	+	+
D-fructose	−	na	−	+	w	+	na	−	na
D-maltose	+	na	−	−	−	−	na	−	na
D-lactose	−	−	+	−	−	−	+	−	−
Habitat	Human gut	Vaginal discharges	Vaginal discharges	Human gut	Clinical specimens	Vaginal discharges	Vaginal discharges	Nasal cavity	Clinical specimens

na: non available data.
[a] Testing was realized in house for these species or this property.
[b] The test results reported in this table are provided by the literature except for the results for the major fatty acid.

The cells are Gram-positive, non-sporeforming, non-motile cocci with a diameter of 0.9 μm that form small white colonies on 5% sheep-blood enriched Colombia agar with a mean diameter of 2 mm. Strain mt16^T optimal growth is at 37° under anaerobic atmosphere. The major fatty acid was hexadecanoic acid. Catalase positive. Oxidase negative. Positive reactions were observed for alkaline phosphatase, leucine arylamidase, valine arylamidase, β-glucosidase, acid phosphatase and Naphthol-AS-BI-phosphohydrolase with an API ZYM strip. Neither urease nor nitrate reduction were observed. Indole was produced. Arabinose, glucose, sucrose, mannose, mannitol, tagatose, potassium-5-ketogluconate, *N*-acetyl-glucosamine, maltose, potassium gluconate, capric acid, adipic acid, trisodium citrate and phenylacetic acid were metabolized according to the API 50CH and API 20A strips. D-lactose, D-salicine, D-xylose, glycerol, D-cellobiose, D-melezitose, D-raffinose, D-sorbitol, L-rhamnose, D-trehalose, glycerol, erythritol, D-adonitol, methyl-βD-xylopyranoside, D-fructose, L-sorbose, L-rhamnose, dulcitol, inositol, D-sorbitol, methyl-αD-mannopyranoside, methyl-αD-glucopyranoside, amygdalin, arbutin, D-trehalose, inulin, starch, glycogen, xylitol, gentiobiose, D-turanose, D-lyxose, D-lyxose, L-lyxose, D-arabitol, L-arabitol, D-fucose, L-fucose, potassium ketogluconate and potassium 2-gluconate were not metabolized. Strain mt16 was susceptible to amoxicillin, amoxicillin/clavulanic acid, benzylpenicillin, clindamycin, imipenem, metronidazole and vancomycin. The genome is 1,929,161 bp long with a 29.5% G + C content and accessible under FAVH00000000 in the GenBank collection. The 16S rRNA sequence is also accessible in the GenBank collection under number LN881592. The type strain mt16^T (= CSUR P2032 = DSM 101186) was isolated from the stool of a child living in Senegal and suffering from kwashiorkor.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

The authors thank the Xegen Company (www.xegen.fr) for automating the genomic annotation process. This study was funded by the Méditerranée Infection Foundation.

We thank Karolina Griffiths and Magdalen Lardiere for English reviewing.

References

[1] H. Wang, C.A. Liddell, M.M. Coates, M.D. Mooney, C.E. Levitz, A.E. Schumacher, et al., Murray, Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013, Lancet 384 (2014) 957–979, http://dx.doi.org/10.1016/S0140-6736(14)60497-9.
[2] M. Million, M. Tidjani Alou, S. Khelaifia, D. Bachar, J.-C. Lagier, N. Dione, et al., Increased gut redox and depletion of anaerobic and methanogenic prokaryotes in severe acute malnutrition, Sci. Rep. 6 (2016) 26051, http://dx.doi.org/10.1038/srep26051.
[3] M. Million, A. Diallo, D. Raoult, Gut microbiota and malnutrition, Microb. Pathog. (2016), http://dx.doi.org/10.1016/j.micpath.2016.02.003.
[4] J.-C. Lagier, F. Armougom, M. Million, P. Hugon, I. Pagnier, C. Robert, et al., Microbial culturomics: paradigm shift in the human gut microbiome study, Clin. Microbiol. Infect. 18 (2012) 1185–1193, http://dx.doi.org/10.1111/1469-0691.12023.
[5] A.C. Parte, LPSN—list of prokaryotic names with standing in nomenclature, Nucleic Acids Res. 42 (2014) D613–D616, http://dx.doi.org/10.1093/nar/gkt1111.
[6] P. De Vos, G.M. Garrity, D. Jones, N.R. Krieg, W. Ludwig, F.A. Rainey, et al., Bergey's Manual of Systematic Bacteriology, in: The *Firmicutes*, second ed., vol. 3, Springer-Verlag, New York, New York, 2009 http://dx.doi.org/10.1007/978-0-387-68489-5.
[7] A.M. Viale, A.K. Arakaki, F.C. Soncini, R.G. Ferreyra, Evolutionary relationships

among eubacterial groups as inferred from GroEL (chaperonin) sequence comparisons, Int. J. Syst. Bacteriol. 44 (1994) 527–533, http://dx.doi.org/10.1099/00207713-44-3-527.
[8] C.R. Woese, O. Kandler, M.L. Wheelis, Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya, Proc. Natl. Acad. Sci. U. S. A. 87 (1990) 4576–4579.
[9] M. Wolf, T. Müller, T. Dandekar, J.D. Pollack, Phylogeny of *Firmicutes* with special reference to *Mycoplasma (Mollicutes)* as inferred from phosphoglycerate kinase amino acid sequence data, Int. J. Syst. Evol. Microbiol. 54 (2004) 871–875, http://dx.doi.org/10.1099/ijs.0.02868-0.
[10] P.-E. Fournier, J.-C. Lagier, G. Dubourg, D. Raoult, From culturomics to taxonomogenomics: a need to change the taxonomy of prokaryotes in clinical microbiology, Anaerobe 36 (2015) 73–78, http://dx.doi.org/10.1016/j.anaerobe.2015.10.011.
[11] P. Seng, M. Drancourt, F. Gouriet, B. La Scola, P.-E. Fournier, J.M. Rolain, 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 (2009) 543–551, http://dx.doi.org/10.1086/600885.
[12] J.-C. Lagier, K. El Karkouri, T.-T. Nguyen, F. Armougom, D. Raoult, P.-E. Fournier, Non-contiguous finished genome sequence and description of *Anaerococcus senegalensis* sp. nov, Stand. Genom. Sci. 6 (2012) 116–125, http://dx.doi.org/10.4056/sigs.2415480.
[13] J.-C. Lagier, P. Hugon, S. Khelaifia, P.-E. Fournier, B. La Scola, D. Raoult, The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota, Clin. Microbiol. Rev. 28 (2015) 237–264, http://dx.doi.org/10.1128/CMR.00014-14.
[14] P.R. Murray, E.J. Baron, J.H. Jorgensen, M.L. Landry, M.A. Pfaller, Manual of Clinical Microbiology, ninth ed., ASM Press, Washington, D.C, 2007.
[15] M. Sasser, Bacterial Identification by Gas Chromatographic Analysis of Fatty Acids Methyl Esters (GC-FAME), http://www.midi-inc.com/pdf/MIS_Technote_101.pdf (accessed 27.05.16).
[16] E. Matuschek, D.F.J. Brown, G. Kahlmeter, Development of the EUCAST disk diffusion antimicrobial susceptibility testing method and its implementation in routine microbiology laboratories, Clin. Microbiol. Infect. 20 (2014) O255–O266, http://dx.doi.org/10.1111/1469-0691.12373.
[17] D.M. Citron, M.I. Ostovari, A. Karlsson, E.J. Goldstein, Evaluation of the E test for susceptibility testing of anaerobic bacteria, J. Clin. Microbiol. 29 (1991) 2197–2203.
[18] The European Committee on Antimicrobial Susceptibility Testing, Breakpoint Tables for Interpretation of MICs and Zone Diameters. Version 6.0, 2016. http://www.eucast.org (accessed 22.05.16).
[19] P. Seng, C. Abat, J.M. Rolain, P. Colson, J.-C. Lagier, F. Gouriet, 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 (2013) 2182–2194, http://dx.doi.org/10.1128/JCM.00492-13.
[20] M. Drancourt, C. Bollet, A. Carlioz, R. Martelin, J.P. Gayral, D. Raoult, 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates, J. Clin. Microbiol. 38 (2000) 3623–3630.
[21] E. Stackebrandt, J. Ebers, Taxonomic parameters revisited: tarnished gold standards, Microbiol. Today 33 (2006) 152.
[22] D. Hyatt, G.-L. Chen, P.F. Locascio, M.L. Land, F.W. Larimer, L.J. Hauser, Prodigal: prokaryotic gene recognition and translation initiation site identification, BMC Bioinforma. 11 (2010) 119, http://dx.doi.org/10.1186/1471-2105-11-119.
[23] T.M. Lowe, S.R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, Nucleic Acids Res. 25 (1997) 955–964.
[24] K. Lagesen, P. Hallin, E.A. Rødland, H.-H. Staerfeldt, T. Rognes, D.W. Ussery, RNAmmer: consistent and rapid annotation of ribosomal RNA genes, Nucleic Acids Res. 35 (2007) 3100–3108, http://dx.doi.org/10.1093/nar/gkm160.
[25] L. Käll, A. Krogh, E.L.L. Sonnhammer, A combined transmembrane topology and signal peptide prediction method, J. Mol. Biol. 338 (2004) 1027–1036, http://dx.doi.org/10.1016/j.jmb.2004.03.016.
[26] P. Gouret, J.D. Thompson, P. Pontarotti, PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees, BMC Bioinforma. 10 (2009) 298, http://dx.doi.org/10.1186/1471-2105-10-298.
[27] M. Lechner, S. Findeiß, L. Steiner, M. Marz, P.F. Stadler, S.J. Prohaska, Proteinortho: detection of (Co-)orthologs in large-scale analysis, BMC Bioinforma. 12 (2011) 124, http://dx.doi.org/10.1186/1471-2105-12-124.
[28] D. Ramasamy, A.K. Mishra, J.-C. Lagier, R. Padhmanabhan, M. Rossi, E. Sentausa, et al., A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species, Int. J. Syst. Evol. Microbiol. 64 (2014) 384–391, http://dx.doi.org/10.1099/ijs.0.057091-0.
[29] A.F. Auch, M. von Jan, H.-P. Klenk, M. Göker, Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison, Stand. Genom. Sci. 2 (2010) 117–134, http://dx.doi.org/10.4056/sigs.531120.
[30] J.P. Meier-Kolthoff, A.F. Auch, H.-P. Klenk, M. Göker, Genome sequence-based species delimitation with confidence intervals and improved distance functions, BMC Bioinforma. 14 (2013) 60, http://dx.doi.org/10.1186/1471-2105-14-60.
[31] P. Gouret, J. Paganini, J. Dainat, D. Louati, E. Darbo, P. Pontarotti, et al., Integration of evolutionary biology concepts for functional annotation and automation of complex research in evolution: the multi-agent software system DAGOBAH, in: P. Pontarotti (Ed.), Evol. Biol. — Concepts Biodivers. Macroevolution Genome Evol., Springer, Berlin Heidelberg, 2011, pp. 71–87. http://link.springer.com/chapter/10.1007/978-3-642-20763-1_5 (accessed 27.05.16).
[32] P. Gouret, V. Vitiello, N. Balandraud, A. Gilles, P. Pontarotti, E.G. Danchin, FIGENIX: intelligent automation of genomic annotation: expertise integration in a new software platform, BMC Bioinform. 6 (2005) 198, http://dx.doi.org/10.1186/1471-2105-6-198.
[33] C. Högenauer, C. Langner, E. Beubler, I.T. Lippe, R. Schicho, G. Gorkiewicz, et al., *Klebsiella oxytoca* as a causative organism of antibiotic-associated hemorrhagic colitis, N. Engl. J. Med. 355 (2006) 2418–2426, http://dx.doi.org/10.1056/NEJMoa054765.
[34] A.L. Brioukhanov, A.I. Netrusov, Catalase and superoxide dismutase: distribution, properties, and physiological role in cells of strict anaerobes, Biochemistry (Mosc) 69 (2004) 949–962.
[35] M.T. Khan, S.H. Duncan, A.J. Stams, J.M. van Dijl, H.J. Flint, H.J. Harmsen, The gut anaerobe *Faecalibacterium prausnitzii* uses an extracellular electron shuttle to grow at oxic-anoxic interphases, ISME J. 6 (2012) 1578–1585, http://dx.doi.org/10.1038/ismej.2012.5.
[36] N. Dione, S. Khelaifia, B. La Scola, J.C. Lagier, D. Raoult, A quasi-universal medium to break the aerobic/anaerobic bacterial culture dichotomy in clinical microbiology, Clin. Microbiol. Infect. 22 (2016) 53–58, http://dx.doi.org/10.1016/j.cmi.2015.10.032.
[37] S. Hirano, N. Matsumoto, M. Morita, K. Sasaki, N. Ohmura, Electrochemical control of redox potential affects methanogenesis of the hydrogenotrophic methanogen *Methanothermobacter thermautotrophicus*, Lett. Appl. Microbiol. 56 (2013) 315–321, http://dx.doi.org/10.1111/lam.12059.
[38] T. Ezaki, Y. Kawamura, N. Li, Z.Y. Li, L. Zhao, S. Shu, Proposal of the genera *Anaerococcus* gen. nov., *Peptoniphilus* gen. nov. and *Gallicola* gen. nov. for members of the genus *Peptostreptococcus*, Int. J. Syst. Evol. Microbiol. 51 (2001) 1521–1528, http://dx.doi.org/10.1099/00207713-51-4-1521.
[39] T. Ezaki, S.L. Liu, Y. Hashimoto, E. Yabuuchi, *Peptostreptococcus hydrogenalis* sp. nov. from human fecal and vaginal flora, Int. J. Syst. Evol. Microbiol. 40 (1990) 305–306, http://dx.doi.org/10.1099/00207713-40-3-305.
[40] Y. Song, C. Liu, S.M. Finegold, *Peptoniphilus gorbachii* sp. nov., *Peptoniphilus olsenii* sp. nov., and *Anaerococcus murdochii* sp. nov. isolated from clinical specimens of human origin, J. Clin. Microbiol. 45 (2007) 1746–1752, http://dx.doi.org/10.1128/JCM.00213-07.
[41] D.A. Murdoch, M.D. Collins, A. Willems, J.M. Hardie, K.A. Young, J.T. Magee, Description of three new species of the genus *Peptostreptococcus* from human clinical specimens: *Peptostreptococcus harei* sp. nov., *Peptostreptococcus ivorii* sp. nov., and *Peptostreptococcus octavius* sp. nov, Int. J. Syst. Bacteriol. 47 (1997) 781–787, http://dx.doi.org/10.1099/00207713-47-3-781.
[42] N. Li, Y. Hashimoto, S. Adnan, H. Miura, H. Yamamoto, T. Ezaki, Three new species of the genus *Peptostreptococcus* isolated from humans: *Pepstreptococcus vaginalis* sp. nov., *Pepstreptococcus lacrimalis* sp. nov., and *Peptostreptococcus lactolyticus* sp. nov, Int. J. Syst. Bacteriol. 42 (1992) 602–605, http://dx.doi.org/10.1099/00207713-42-4-602.
[43] T. Ezaki, N. Yamamoto, K. Ninomiya, S. Suzuki, E. Yabuuchi, Transfer of *Peptococcus indolicus, Peptococcus asaccharolyticus, Peptococcus prevotii* and *Peptococcus magnus* to the genus *Peptostreptococcus* and proposal of *Peptostreptococcus tetradius* sp. nov, Int. J. Syst. Bacteriol. 33 (1983) 683–698.

Article 6: *Bacillus rubiinfantis* sp. nov. strain mt2^{T} a new bacterial species isolated from human gut.

Maryam Tidjani Alou, Jaishriram Rathored, Saber Khelaifia, Caroline Michelle, Souleymane Brah , Bouli Ali Diallo, Didier Raoult, Jean-Christophe Lagier.

Publié dans New Microbes and New Infections

Bacillus rubiinfantis sp. nov. strain mt2^T, a new bacterial species isolated from human gut

M. Tidjiani Alou[1], J. Rathored[1], S. Khelaifia[1], C. Michelle[1], S. Brah[2], B. A. Diallo[3], D. Raoult[1,4] and J.-C. Lagier[1]

1) *Faculté de médecine, Unité des Maladies Infectieuses et Tropicales Emergentes (URMITE), UM63, CNRS7278, IRD198, Inserm 1095, Aix-Marseille Université, Marseille, France,* 2) *Hopital National de Niamey,* 3) *Laboratoire de microbiologie, département de biologie, Université Abdou Moumouni de Niamey, Niamey, Niger and* 4) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Bacillus rubiinfantis sp. nov. strain mt2^T is the type strain of *B. rubiinfantis* sp. nov., isolated from the fecal flora of a child with kwashiorkor in Niger. It is Gram-positive facultative anaerobic rod belonging to the *Bacillaceae* family. We describe the features of this organism alongside the complete genome sequence and annotation. The 4 311 083 bp long genome (one chromosome but no plasmid) contains 4028 protein-coding gene and 121 RNA genes including nine rRNA genes.

Keywords: *Bacillus rubiinfantis*, culturomics, genome, taxonogenomics
Original Submission: 7 July 2015; **Revised Submission:** 1 September 2015; **Accepted:** 7 September 2015
Article published online: 16 September 2015

Corresponding author: J.-C. Lagier, Faculté de Médecine, URMITE, UMR CNRS 7278, IRD 198, INSERM U1095, Aix-Marseille Université, 27 Bd Jean Moulin, 13385 Marseille cedex 5, France
E-mail: jclagier@yahoo.fr

Introduction

Bacterial classification is currently based on phenotypic and genomic characteristics and the building of phylogenetic relationships using a comparison of the 16S ribosomal RNA [1–3]. Culturomics studies are based on the multiplication of culture conditions with a rapid identification method, which has allowed us to recently extend the gut microbiota repertoire [4,5]. In parallel, a new concept of bacterial description, termed taxonogenomics, has recently been developed combining a proteomic description with the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) profile associated with a biochemical and genomic description of the new bacterial species [4]. To date, nine bacterial species have officially been recognized through taxonogenomics [5–15].

The genus *Bacillus* was established in 1872 by Cohn and is composed of strictly aerobic and facultatively anaerobic rod-shaped bacteria that form heat-resisting endospores [16–19]. It includes over 200 described species and subspecies belonging to the *Firmicutes* phylum (http://www.bacterio.net/bacillus.html). These bacteria are widely distributed in the environment (soil, water, air), plants, food and human clinical samples [20]. The most known pathogenic *Bacillus* species is *Bacillus anthracis,* but most *Bacillus* are considered to be harmless to humans [20].

Bacillus rubiinfantis strain mt2^T (= Collection de souches de l'Unité des Rickettsies (CSUR) P1141 = Deutsche Sammlung von Mikroorganismen (DSM) 28615) is the type strain of *B. rubiinfantis* sp. nov. This bacterium is Gram positive, spore forming, facultative anaerobic and motile. It was isolated from the stool of a child living in Niamey, Niger, with severe acute malnutrition (kwashiorkor).

Material and Methods

Organism information: classification and features

A stool sample was collected from a 2-year-old girl living in Niamey, Niger, with severe acute malnutrition. Consent was

http://dx.doi.org/10.1016/j.nmni.2015.09.008

obtained from the child's parents by the National Hospital of Niamey, and the study was approved by the Institut Fédératif de Recherche 48, Faculty of Medicine, Marseille, France (agreement 09-022). The patient did not receive any antibiotics at the time of sample collection; the fecal sample was stored at −80°C.

Strain identification

All the colonies were identified by MALDI-TOF, as described below. In the case of no matching spectra in the database, we further characterized the colonies using 16S rRNA sequencing, as previously described. If the 16S rRNA sequence similarity value was lower than 98.7%, we considered a new species without performing DNA-DNA hybridization as suggested by Stackebrandt and Ebers [21].

Phenotypic characteristics

The main phenotypic characteristics (i.e. Gram staining, motility, sporulation, catalase and oxidase test) were performed as previously described [22]. The chemical characteristics of the newly isolated strain were investigated using API 20NE, API ZYM and API 50CH strips (bioMérieux, Marcy l'Étoile, France). Growth temperatures were tested at 25°C, 30°C, 37°C, 45°C and 55°C. The 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. Regarding electron microscopy, detection formvar coated grids were deposited on a 40 µL bacterial suspension drop and incubated during 30 minutes at 37°C. The grids were incubated for 10 seconds on ammonium molybdate 1%, dried on blotting paper and then observed with a Morgani 268D at an operating voltage of 60 kV.

MALDI-TOF protein analysis was carried out as previously described [23] using a Microflex spectrometer (Brüker Daltonics, Leipzig, Germany). Twelve individual colonies were deposited on a MTP 96 MALDI-TOF target plate (Brüker). The 12 spectra were imported into the MALDI BioTyper software (version 2.0; Brüker) and analysed by standard pattern matching (with default parameter settings) against the main spectra of 6.252 bacteria, including 199 spectra from 104 validly named *Bacillus* species used as reference data in the BioTyper database. A score enabled the presumptive identification and discrimination of the tested species from those in a database: a score of >2 with a validated species enabled the identification at the species level, and a score of <1.7 did not enable any identification [23]. No significant score was obtained for strain mt2, thus suggesting that our isolate was not a member of any known species. The obtained reference spectrum for *B. rubiinfantis* strain mt2^T (Fig. 4) was incremented in our database and compared to other members of the *Bacillus* genus. The observed differences are shown in gel view in Fig. 5.

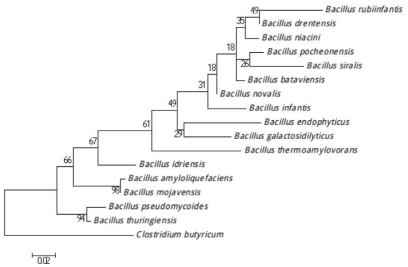


FIG. 1. Phylogenetic tree highlighting position of *Bacillus rubiinfantis* sp. nov. strain mt2^T (= CSUR P1141 = DSM 28615) relative to other type strains within *Bacillus* genus. Strains and their corresponding GenBank accession numbers for 16S rRNA genes are (type = ^{T}): *B. novalis* strain SCTB 113, JN650278; *B. drentensis* DRG 4, HQ436340; *B. niacini* strain TSII-13, JN993716; *B. bataviensis* strain LMG21832, AJ542507; *B. pocheonensis* strain Gsoil 420, AB245377.1; *B. infantis* strain A-49, KC751070; *B. siralis* strain 171544, NR_028709.1; *B. endophyticus* strain 2DT, NR_025122.1; *B. pseudomycoides* strain BIHB 360, FJ859700.1; *B. thuringiensis* strain GMC 108, AB741470; *B. idriensis* strain SMC 4352−2, NR_043268.1; *B. amyloliquefaciens* strain BIHB 35, FJ859496 ; *B. mojavensis* strain BCRC 17124, EF433405.1 ; *B. galactosidilyticus* strain LMG 17892, NR_025580.1 ; *B. thermoamylovorans* strain DKP, NR_029151.1. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using maximum-likelihood method within MEGA6. Numbers at nodes are percentages of bootstrap values obtained by repeating analysis 1000 times to generate majority consensus tree. *Bacteroides thetaiotaomicron* strain ATCC 29148^T (L16489) was used as outgroup. Scale bar = 1% nucleotide sequence divergence.

Growth conditions and genomic DNA preparation

B. rubiinfantis sp. nov., strain mt2^T (= CSUR P1141 = DSM 28615), was grown on 5% sheep's blood−enriched Columbia agar (bioMérieux) at 37°C in aerobic atmosphere. Bacteria

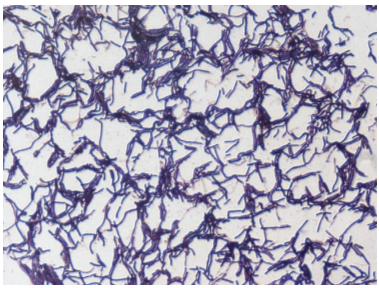


FIG. 2. Gram staining of *B. rubiinfantis strain* mt2^T.

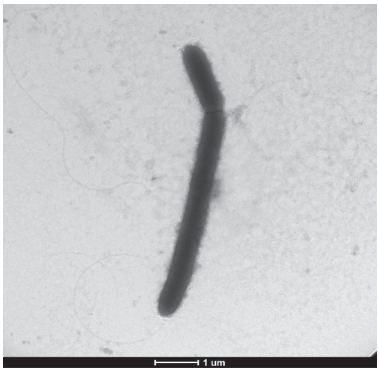


FIG. 3. Transmission electron microscopy of *B. rubiinfantis strain* mt2^T using Morgani 268D (Philips) at operating voltage of 60 kV. Scale bar = 1 µm

grown on three petri dishes were collected and resuspended in 4 × 100 µL of Tris-EDTA (TE) buffer. Then 200 µL of this suspension was diluted in 1 ml TE buffer for lysis treatment that included a 30- minute incubation with 2.5 µg/µL lysozyme at 37°C, followed by an overnight incubation with 20 µg/µL proteinase K at 37°C. Extracted DNA was then purified using 3 successive phenol–chloroform extractions and ethanol precipitations at −20°C overnight. After centrifugation, the DNA was resuspended in 160 µL TE buffer.

Genome sequencing and assembly

Genomic DNA of *Bacillus rubiinfantis* was sequenced on the MiSeq Technology (Illumina, 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). gDNA was quantified by a Qubit assay with the high sensitivity kit (Life Technologies, Carlsbad, CA, USA) to 66.2 ng/µL. The mate-pair library was prepared with 1 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 labchip. The DNA fragments ranged in size from 1 kb up to 11 kb, with an optimal size of 3.927 kb. No size selection was performed, and 505 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal at 597 bp on a

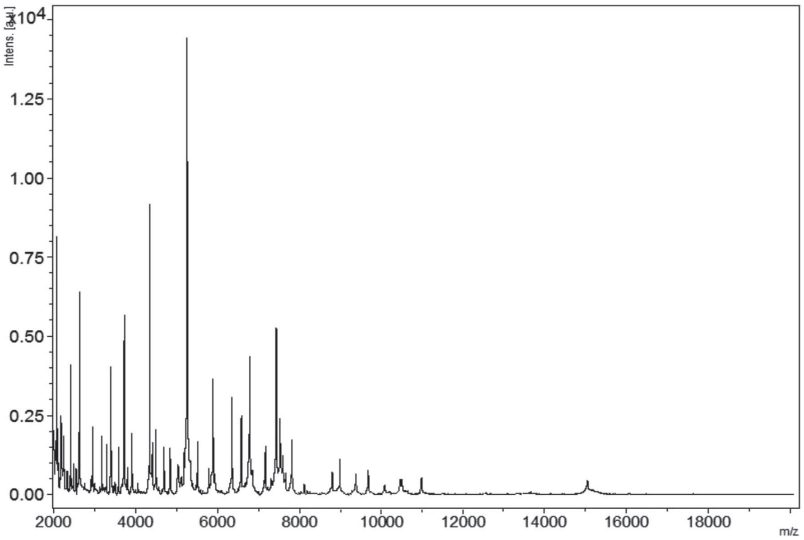


FIG. 4. Reference mass spectrum from *B. rubiinfantis strain* mt2^T. Spectra from 12 individual colonies were compared and reference spectrum was generated.

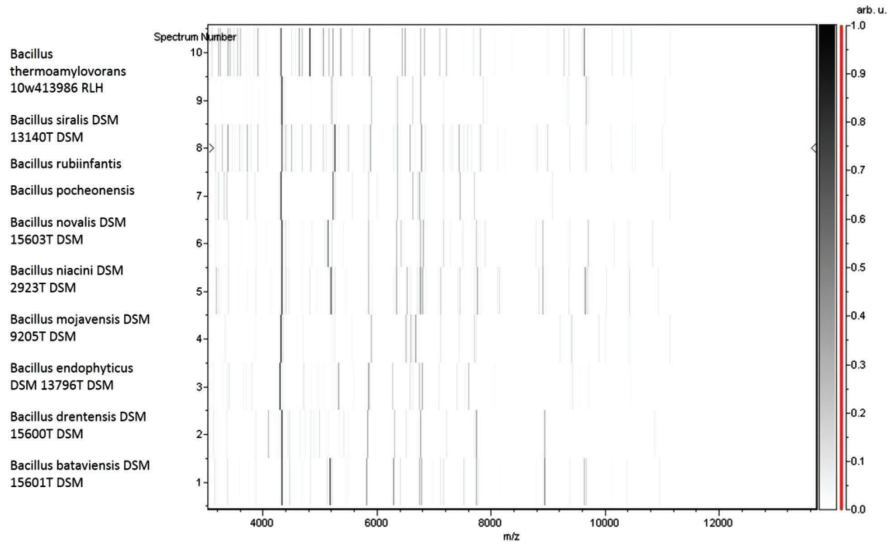


FIG. 5. Gel view comparing *B. rubiinfantis* (= CSUR P1141 = DSM 28615) to other species within genus *Bacillus*. Gel view displays raw spectra of loaded spectrum files arranged in pseudo-gel-like look. X-axis records m/z value. Left y-axis displays running spectrum number originating from subsequent spectra loading. Peak intensity is expressed by greyscale scheme code. Color bar and right y-axis indicate relation between color peak; peak intensity in arbitrary units. Displayed species are indicated on left.

Covaris S2 device in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent), and the final concentration library was measured at 59.2 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 2 × 251 bp. Total information of 10.7 Gb was obtained from a 1283 K/mm^2 cluster density, with a cluster passing quality control filters of 99% (24 072 208 clusters). Within this run, the index representation for *Bacillus rubiinfantis* was determined to 9.71%. The 2 053 904 paired reads were filtered according to the read qualities. These reads were trimmed, then assembled. A genome with 12 scaffolds and a size of 4.32 Mb was generated. The percentage GC was calculated at 40.0%.

Genome annotation and genome analysis

Open reading frames (ORFs) were predicted using Prodigal (http://prodigal.ornl.gov/) with default parameters, but the predicted ORFs were excluded if they spanned a sequencing gap region (contain N). The predicted bacterial protein sequences were searched against the Clusters of Orthologous Groups (COGs) using BLASTP (E-value 1e^{-03}, coverage 0.7, identity 30%). If no match was found, it searched against the NR database using BLASTP with an E-value of 1e^{-03}, coverage 0.7 and identity 30%, If the sequence lengths were smaller than 80 amino acids, we used an E-value of 1e^{-05}. The tRNAScanSE tool [24] was used to find tRNA genes, whereas ribosomal RNAs were found by RNAmmer [25]. Lipoprotein signal peptides and the number of transmembrane helices were predicted using Phobius [26]. ORFans were identified if all the performed BLASTP did not give positive results (E value smaller than 1e^{-03} for ORFs with sequence size upper than 80 aa or E value smaller than 1e^{-05} for ORFs with sequence length smaller 80 aa). Such parameter thresholds have already been used in previous works to define ORFans.

For the genomic comparison, we compared our type strain *Bacillus rubiinfantis* strain mt2^T to *Bacillus bataviensis* (GenBank accession number AJ542508.1), *Bacillus infantis* (AY904032.1), *Bacillus endophyticus* (AF295302.1), *Bacillus pseudomycoides*

(AF013121.1), *Bacillus thuringiensis* (D16281.1), *Bacillus amyloliquefaciens* (AB006920.1), *Bacillus mojavensis* (AB021191.1), *Bacillus thermoamylovorans* (L27478.1) and *Clostridium butyricum* (AJ58420.1). Genomes were automatically retrieved from the 16S RNA tree using Xegen software (Phylopattern) [27]. For each selected genome, the complete genome sequence, proteome genome sequence and Orfeome genome sequence were retrieved from the File Transfer Protocol (FTP) of National Center for Biotechnology Information. All proteomes were analysed with proteinOrtho [28]. For each couple of genomes, a similarity score was then computed. This score is the mean value of nucleotide similarity between all couples of orthologues between the two genomes studied (AGIOS) [6]. An annotation of the entire proteome was performed to define the distribution of the functional classes of predicted genes according to the clusters of orthologous groups of proteins (using the same method as for the genome annotation).

Results

Phenotypic description

Strain mt2^T (Table 1) was first isolated in January 2014 by a 21-day anaerobic preincubation in blood culture with sheep's blood and cultivation 5% sheep's blood–enriched Colombia agar (BioMérieux) in anaerobic atmosphere at 37°C.

This strain showed a 96% nucleotide sequence similarity with *B. bataviensis*, the phylogenetically closest *Bacillus* species with a valid published name (Fig. 1). This similarity value is below the 16S rRNA threshold of 98.7% set by Stackebrandt and Ebers to delineate a new species without carrying out DNA-DNA hybridization [21].

Various growth temperatures (25°C, 30°C, 37°C, 45°C and 56°C) were tested. Growth occurred between 25°C and 45°C on blood-enriched Columbia agar (bioMérieux), with the optimal growth being obtained at 37°C after 48 hours of incubation. 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. Optimal growth was achieved aerobically. Weak cell growth was observed under microaerophilic and anaerobic conditions. The motility test was positive, and the cells were sporulating. Colonies were translucent and 10 mm in diameter on blood-enriched Columbia agar. Cells were Gram-positive rods (Fig. 2). Under electron microscopy, the bacteria grown on agar had a mean diameter and length of 0.54 and 0.62 µm, respectively (Fig. 3).

TABLE 1. Classification and general features of *Bacillus rubiinfantis* strain mt2

Property	Term
Current classification	Domain: *Bacteria* Phylum: *Firmicutes* Class: *Bacilli* Order: *Bacillales* Family: *Bacillaceae* Genus: *Bacillus* Species: *Bacillus rubiinfantis* Type strain: mt2
Gram stain	Positive
Cell shape	Rod
Motility	Motile
Sporulation	sporulating
Temperature range	Mesophilic
Optimum temperature	37°C

Strain mt2^T exhibited catalase activity but was negative for oxidase. Using an API ZYM strip (BioMérieux), positive reactions were observed for esterase (C4), esterase lipase (C8), leucine arylamidase, valine arylamidase, cystine arylamidase, α-chymotrypsin, acid phosphatase, naphtol-AS-BI-phosphohydrolase, α-galactosidase and β-galactosidase. Negative reactions were observed for lipase (C14), trypsin, alkaline phosphatase, β-glucuronidase, α-glucosidase, β-glucosidase, α-fucosidase, α-mannosidase and N-acetyl-β-glucosaminidase. Using an API 20 NE strip, positive reactions were obtained for L-arginine, urea, 4-nitrophenyl- βD-galactopyranoside and capric acid. All other reactions were negative.

Using an API 50 CH strip (BioMérieux), positive reactions were observed for D-galactose, D-glucose, D-fructose, D-mannose, N-acetylglucosamine, esculin ferric citrate, D-maltose, D-lactose, D-melibiose and D-trehalose. Negative reactions were observed glycerol, erythrol, D-arabinose, L-arabinose, D-ribose, L-xylose, D-xylose, D-adonitol, methyl-βD-xylopranoside, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl-αD-mannopyranoside, methyl-αD-glucopyranoside, amygdaline, arbutin, salicin, D-cellobiose, D-maltose, D-saccharose, inulin, D-melezitose, D-raffinose, amidon, glycogen, gentiobiose, D-turanose, D-lyxose, D-tagatose, D-fucose, L-fucose, D-arabitol, L-arabitol, potassium gluconate, potassium 2-ketogluconate and potassium 5-ketogluconate.

Cells are susceptible to vancomycin, rifampicin, doxycycline, erythromycin, amoxicillin, nitroflurantoin, ciprofloxacin, ceftriaxone, penicillin G, gentamicin and imipenem but resistant to trimethoprim/ sulfamethoxazole and metronidazole. The differences exhibited by the comparison with other representatives of the genus *Bacillus* are detailed in Table 2.

Genome properties

The genome of *B. rubiinfantis* strain mt2^T is 4 311 083 bp long with a 40.0% G+C content (Fig. 6). Of the 4149 predicted genes, 4028 were protein-coding genes and 121 were RNAs. Out of these 121 rRNA genes, 11 are 16S rRNA genes, 9 are 23S rRNA, 11 are 5S rRNA and 90 predicted tRNA genes were identified in

TABLE 2. Differential characteristics of *Bacillus rubinfantis* strain mt2^T, *Bacillus novalis* LMG 21837, *Bacillus bataviensis* LMG 21833, *Bacillus drentensis* LMG 21831, *Bacillus niacini* DSM 2923, *Bacillus endophyticus* UCM 5715, *Bacillus siralis* NCIMB 13601, *Bacillus thermoamylovorans* CNCM I 1378, *Bacillus pocheonensis* DSM 18135, *Bacillus mojavensis* NRRL B 14698 [31–39]

Property	*B. rubinfantis*	*B. nolis*	*B. bataviensis*	*B. drentensis*	*B. niacini*	*B. endophyticus*	*B. siralis*	*B. thermo-amylovorans*	*B. pocheonensis*	*B. mojavensis*
Cell diameter (μm)	0.5–0.6	0.6–1.2	0.7–1.2	0.6–1.2	0.9–1.4	0.5–1.5	0.5–0.8	0.45–0.5	1.5–3	1–2
Oxygen requirement	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic
Gram stain	+	+	+	+	+	+	+	+	+	+
Salt requirement	NA	NA	NA	NA	NA	+	+	NA	+	NA
Motility	+	+	+	+	–	+	NA	+	–	+
Endospore formation	+	+	+	+	+	+	+	–	+	+
Indole	–	–	–	–	+	–	NA	–	–	–
Production of:										
Alkaline phosphatase	–	NA	NA	NA	NA	NA	NA	NA	NA	NA
Catalase	+	NA	NA	NA	+	+	+	+	+	+
Oxidase	–	NA	NA	NA	+	+	+	+	+	+
Nitrate reductase	–	+	+	+/–	+	–	+	–	+	–
Urease	+	–	–	–	–	–	NA	NA	–	–
β-Galactosidase	+	+	+	+	+	–	–	+	+	+
N-Acetyl-glucosamine	–	+	+	+	+	NA	NA	+	+	NA
Acid from:										
L-Arabinose	–	–	–	–	–	+	–	–	+	+
Ribose	–	+	+	+	–	+	–	+	–	–
Mannose	+	+	+	+	+	+	–	+	+	+
Mannitol	–	+	+	+	–	+	–	–	–	+
Sucrose	+	–	+	+	+	+	–	+	–	+
D-Glucose	+	+	+	+	+	+	–	+	–	+
D-Fructose	+	+	+	+	+	NA	–	+	+	+
D-Maltose	–	+	+	+	+	+	–	+	+	+
D-Lactose	+	–	+	+	+	+	+	+	–	–
Habitat	Human gut	Soil	Soil	Soil	soil	Cotton plants	Grassland soil	Palm vine	Soil	soil

NA: non-available data

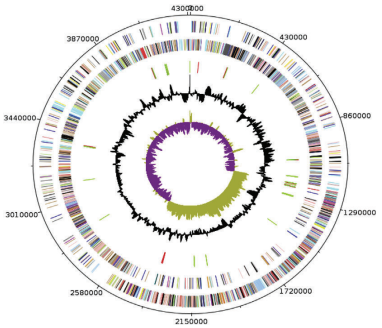


FIG. 6. Graphical circular map of chromosome. From outside to center: genes on forward strain colored by COGs categories (only gene assigned to COGs), RNA genes (Trnas green, rRNAs red), GC content and GC skew. COGs, Clusters of Orthologous Groups.

the genome. A total of 3026 genes (72.93%) were assigned a putative function. Two hundred two genes were identified as ORFans (5.04%). The remaining genes were annotated as hypothetical proteins. The properties and statistics of the genome are summarized in Tables 3 and 4. The distribution of genes into COGs functional categories is presented in Table 4.

It was the 70th organism identified within the *Bacillus* genera. The genome's GenBank accession number is CCFE01000000 and consists of 12 scaffolds and 33 contigs.

Genome comparison

The draft genome sequence of *Bacillus rubiinfantis* is larger than those of *Bacillus amyloliquefaciens, Bacillus mojavensis* and *Bacillus thermoamylovorans* (4.01, 3.96 and 3.71) but smaller than those of *Bacillus pseudomycoides, Bacillus bataviensis, Bacillus*

TABLE 3. Nucleotide content and gene count levels of the genome

Attribute	Genome (total)	
	Value	% of total[a]
Size (bp)	4 311 083	100
G+C content (bp)	1 724 433	40.0
Coding region (bp)	3 672 471	84.88
Total genes	4149	100
RNA genes	121	2.80
Protein-coding genes	4028	97.08
Genes with function prediction	3026	72.93
Genes assigned to COGs	2666	64.25
Genes with peptide signals	244	5.88
Genes with transmembrane helices	1045	25.18
Genes with Pfam domains	1405	34.02
CRISPR repeats	01	0.02

COGs, Clusters of Orthologous Groups.
[a]Total is based on either size of genome (in base pairs) or total number of protein coding genes in annotated genome.

TABLE 4. Number of genes associated with 25 general COGs functional categories

Code	Value	% of total[a]	Description
J	163	3.97	Translation
A	0	0	RNA processing and modification
K	265	5.34	Transcription
L	235	4.62	Replication, recombination, repair
B	1	0.02	Chromatin structure and dynamics
D	40	0.92	Cell cycle control, mitosis, meiosis
Y	0	0	Nuclear structure
V	65	1.46	Defense mechanisms
T	197	3.57	Signal transduction mechanisms
M	184	3.60	Cell wall/membrane biogenesis
N	60	1.24	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	46	0.97	Intracellular trafficking and secretion
O	122	2.73	Posttranslational modification, protein turnover, chaperones
C	211	4079	Energy production and conversion
G	183	3.85	Carbohydrate transport and metabolism
E	397	9.14	Amino acid transport and metabolism
F	91	2.09	Nucleotide transport and metabolism
H	107	2.46	Coenzyme transport and metabolism
I	133	3.15	Lipid transport and metabolism
P	218	4.87	Inorganic ion transport and metabolism
Q	81	1.79	Secondary metabolites biosynthesis, transport, catabolism
R	428	9.41	General function prediction only
S	291	6.41	Function unknown
—	3518	33.47	Not in COGs

COGs, Clusters of Orthologous Groups.
[a]Total is based on total number of protein coding genes in annotated genome.

thuringiensis, Bacillus infantis, Bacillus endophyticus, Bacillus infantis and *Clostridium butyricum* (5.75, 5.37, 5.24, 5.1, 4.88 and 4.62 MB, respectively). The G+C content of *Bacillus rubiinfantis* is smaller than those of *Bacillus amyloliquefaciens, Bacillus infantis* and *Bacillus mojavensis* (46.1, 46.0 and 43.7) but larger than those of *Bacillus bataviensis, Bacillus thermoamylovorans, Bacillus endophyticus, Bacillus pseudomycoides, Bacillus thuringiensis* and *Clostridium butyricum* (39.6%, 37.3%, 36.4%, 35.4%, 35.4% and 28.8%, respectively). The gene content of *Bacillus rubiinfantis* is larger than those of *Bacillus mojavensis, Bacillus amyloliquefaciens* and *Bacillus thermoamylovorans* (4131, 4005 and 3458) but lower than that of *Bacillus pseudomycoides, Bacillus thuringiensis, Bacillus bataviensis, Bacillus endophyticus, Bacillus infantis* and *Clostridium butyricum* (5941, 5263, 5238, 4996, 4837 and 4231, respectively). To evaluate the genomic similarity among studied *Bacillus* strains, we determined two parameters, dDDH, Which exhibits a high correlation with DDH [29,30], and AGIOS [6], which was designed to be independent from DDH (Tables 5 and 6).

however, the distribution of genes into COGs categories was similar in all compared genomes (Fig. 7(. In addition, *B. rubiinfantis* shared 4022, 5265, 5077, 4078, 2152, 5863, 3468, 5271 and 4148 orthologous genes with *B. amyloliquefaciens, B. bataviensis, B. endophyticus, B. mojavensis, B. infantis, B. pseudomycoides, B. thuringiensis, B. thermoamylovorans* and *Clostridium butyricum* (Table 6). Among species with standing in nomenclature, AGIOS values ranged from 57.74 between *Clostridium*

TABLE 5. Pairwise comparison of *Bacillus rubiinfantis* (upper right) with eight other species using GGDC, formula 2 (DDH estimates based on identities/HSP length)[a]

	B. rubiinfantis	*B. amyloliquefaciens*	*B. bataviensis*	*B. endophyticus*	*B. mojavensis*	*B. infantis*	*B. pseudomycoides*	*B. thermoamylovorans*	*B. thuringiensis*	*C. butyricum*
B. rubiinfantis	100% ± 00	2.54% ± 0.14	2.66% ± 0.23	2.54% ± 0.16	2.55% ± 0.14	2.57% ± 0.20	2.54% ± 0.16	2.54% ± 0.23	2.55% ± 0.14	2.53% ± 0.16
B. amyloliquefaciens		100% ± 00	2.54% ± 0.22	2.54% ± 0.20	3.01% ± 0.22	2.54% ± 0.20	2.53% ± 0.19	2.53% ± 0.24	2.54% ± 0.13	2.53% ± 0.18
B. bataveinsis			100% ± 00	2.54% ± 0.21	2.54% ± 0.19	2.61% ± 0.20	2.54% ± 0.21	2.53% ± 0.24	2.55% ± 0.18	2.53% ± 0.17
B. endophyticus				100% ± 00	2.55% ± 0.18	2.54% ± 0.23	2.55% ± 0.22	2.53% ± 0.22	2.56% ± 0.20	2.53% ± 0.19
B. mojavensis					100% ± 00	2.54% ± 0.16	2.54% ± 0.16	2.53% ± 0.23	2.55% ± 0.12	2.53% ± 0.17
B. infantis						100% ± 00	2.53% ± 0.19	2.53% ± 0.22	2.55% ± 0.18	2.52% ± 0.11
B. pseudomycoides							100% ± 00	2.53% ± 0.19	3.00% ± 0.16	2.53% ± 0.18
B. thermoamylovorans								100% ± 00	2.53% ± 0.17	2.52% ± 0.21
B. thuringiensis									100% ± 00	2.54% ± 0.17
C. butyricum										100% ± 00

[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 16S rRNA (Fig. 1) and phylogenomic analyses as well as GGDC results.
GGDC, Genome-to-Genome Distance Calculator; DDH, DNA-DNA Hybridization; HSP, High-Scoring Pair.

TABLE 6. Numbers of orthologous protein shared between genomes (upper right) average percentage similarity of nucleotides corresponding to orthologous protein shared between genomes (lower left)

	B. amyloliquefaciens	*B. bataviensis*	*B. endophyticus*	*B. mojavensis*	*B. infantis*	*B. pseudomycoides*	*B. rubiinfantis*	*B. thermoamylovorans*	*B. thuringiensis*	*C. butyricum*
B. amyloliquefaciens	4022[a]	1674	1864	2028	686	1642	1486	1442	1707	798
B. bataviensis	65.44	5265[a]	1765	1653	878	1785	1850	1599	1807	854
B. endophyticus	65.43	66.90	5077[a]	1849	749	1769	1547	1481	1839	833
B. mojavensis	77.91	65.86	66.25	4078[a]	693	1637	1466	1413	1694	776
B. infantis	65.39	74.32	67.37	65.97	2152[a]	744	740	633	757	391
B. pseudomycoides	64.79	67.00	68.06	65.55	67.10	5863[a]	1556	1416	2212	868
B. rubiinfantis	65.31	73.95	66.82	65.78	72.33	67.19	4028[a]	1532	1571	778
B. thermoamylovorans	64.64	67.30	66.27	65.10	67.67	66.64	67.41	3468[a]	1445	785
B. thuringiensis	64.42	66.65	67.70	65.23	66.92	83.41	66.80	66.33	5271[a]	879
C. butyricum	57.74	60.66	61.81	59.14	61.44	62.07	60.72	61.49	62.11	4148[a]

[a]Number of proteins per genome.

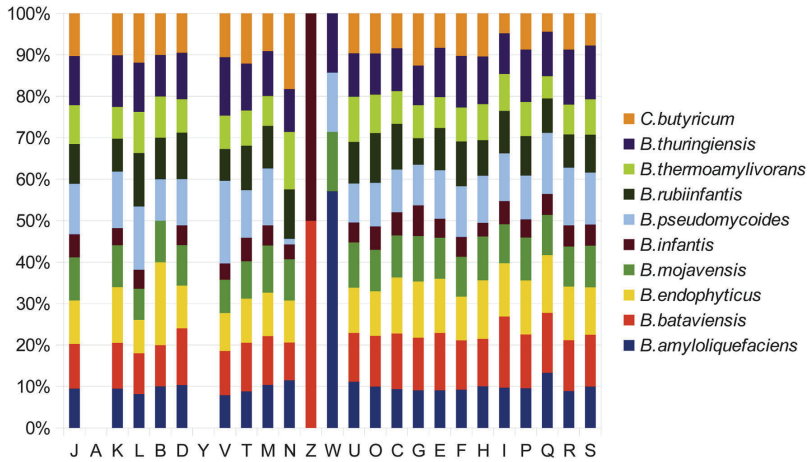


FIG. 7. Distribution of functional classes of predicted genes according to COGs. COGs, Clusters of Orthologous Groups.

butyricum and *B. amyloliquefaciens* to 83.41% between *B. thuringiensis* and *B. pseudomycoides*.

Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Bacillus rubiinfantis* sp. nov. that contains the strain mt2. This bacterial strain has been isolated from the fecal flora of a 2-year-old girl from Niamey, Niger, with kwashiorkor.

Taxonomic and nomenclatural proposals: description of *Bacillus rubiinfantis* strain mt2^T sp. nov.

The cells are Gram-positive, sporulating and rod-shaped bacilli with a diameter of 0.5 µm. Colonies are translucent and 10 mm in diameter on 5% sheep's blood–enriched Columbia agar (bioMérieux) and are catalase positive and oxidase negative. Positive reactions are observed for esterase (C4), esterase lipase (C8), leucine arylamidase, valine arylamidase, cystine arylamidase, α-chymotrypsin, acid phosphatase, naphtol-AS-BI-phosphohydrolase, α-galactosidase and β-galactosidase. Negative reactions are observed for lipase (C14), trypsin, nitrate reduction, alkaline phosphatase, β-glucuronidase, α-glucosidase, β-glucosidase, α-fucosidase, α-mannosidase and N-acetyl-β-glucosaminidase.

Cells are susceptible to vancomycin, rifampicin, doxycycline, erythromycin, amoxicillin, nitrofurantoin, ciprofloxacin, ceftriaxone, penicillin G, gentamicin and imipenem but resistant to trimethoprim/sulfamethoxazole and metronidazole.

The G+C content of the genome is 40.0%. The 16S rRNA gene sequence and whole-genome shotgun sequence of *B. rubiinfantis* strain mt2^T are deposited in GenBank under accession numbers LK021113 and CCFE01000000, respectively. The type strain mt2^T (= CSUR P1141 = DSM 28615) was isolated from the stool of a child living in Niamey, Niger, with kwashiorkor.

Acknowledgements

The authors thank the Xegen Company (http://www.xegen.fr/) for automating the genomic annotation process. This study was funded by the Mediterranée-Infection Foundation. We thank K. Griffiths for her English-language review.

Conflict of Interest

None declared.

References

[1] Viale AM, Arakaki AK, Soncini FC, Ferreyra RG. Evolutionary relationships among eubacterial groups as inferred from GroEL

(Chaperonin) sequences comparison. Int J Syst Bacteriol 1944;44(3): 527–33.
[2] Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposals for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 1990;87(12):4576–9.
[3] Wolf M, Muller T, Dandekar T, Pollack JD. Phylogeny of Firmicutes with special reference to Mycoplasma (Mollicutes) as inferred from phosphoglycerate kinase amino acid sequence data. Inst J Syst Evol Microbiol 2004;54(Pt3):871–5.
[4] 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 Jan;28(1): 237–64.
[5] 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.
[6] Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi-Tamisier M, Sentausa E, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of new bacterial species. Int J Syst Evol Microbiol 2014;64:384–91.
[7] Lagier JC, El Karkouri K, Nguyen TT, Armougom F, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Anaerococcus senegalensis* sp. nov. Stand Genomic Sci 2012;6: 116–25.
[8] Lagier JC, Armougom F, Mishra AK, Nguyen TT, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Alistipes timonensis* sp. nov. Stand Genomic Sci 2012;6:315–24.
[9] Roux V, El Karkouri K, Lagier JC, Robert C, Raoult D. Non-contiguous finished genome sequence and description of *Kurthia massiliensis* sp. nov. Stand Genomic Sci 2012;7:221–32.
[10] Kokcha S, Ramasamy D, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Brevibacterium senegalense* sp. nov. Stand Genomic Sci 2012;7:233–45.
[11] Ramasamy D, Kokcha S, Lagier JC, Nguyen T-T, Raoult D, Fournier PE. Genome sequence and description of *Aeromicrobium massiliense* sp. nov. Stand Genomic Sci 2012;7:246–57.
[12] Lagier JC, Elkarkouri K, Rivet R, Couderc C, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Senegalemassilia anaerobia* gen. nov., sp. nov. Stand Genomic Sci 2013;7: 343–56.
[13] Lagier JC, El Karkouri K, Mishra AK, Robert C, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Enterobacter massiliensis* sp. nov. Stand Genomic Sci 2013;7:399–412.
[14] Hugon P, Mishra AK, Lagier JC, Nguyen TT, Couderc C, Raoult D, et al. Non-contiguous finished genome sequence and description *of Brevibacillus massiliensis* sp. nov. Stand Genomic Sci 2013;8:1–14.
[15] Lagier JC, Gimenez G, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of Herbaspirillum massiliense sp. nov. Stand Genomic Sci 2012 Dec 19;7(2):200–9.
[16] Priest FG, Goodfellow M, Todd C. A numerical classification of the genus Bacillus. J Gen Microbiol 1988;134:1847–82.
[17] Fritze D. Taxonomy of the genus Bacillus and related genera: the aerobic endospore-forming bacteria. Physiopathology 2004;94(11): 1245–8.
[18] Gordon RE, Haynes WC, Pang CH. The genus Bacillus. US Dep Agric Handh 1973;427.
[19] Goto K, Omura T, Hara Y, Sadaie Y. Application of the 16s rDNA sequence as an index for rapid identification of species in the genus Bacillus. J Gen Appl Microbiol 2000;46(1):1–8.
[20] http://www.bacterio.net/bacillus.html.
[21] Prevot AR. Dictionnaire des bactéries pathogens. In: Hauduroy P, Ehringer G, Guillot G, Magrou J, Prevot AR, Rosset, Urbain A, editors. Paris: Masson; 1953. p.1–692.
[22] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152–5.
[23] Lagier JC, Edouard S, Pagnier I, Mediannikov O, Drancourt M, Raoult D. Current and past strategies for bacterial culture in clinical microbiology. Clin Microbiol Rev 2015 Jan;28(1):208–36.
[24] 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.
[25] Prodigal (http://prodigal.ornl.gov/).
[26] Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucl Acids Res 2007;35:3100–8.
[27] 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.
[28] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340:783–95.
[29] Gouret P, Thompson JD, Pontarotti P. PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinformatics 2009;10:298. http://dx.doi.org/10.1186/1471-2105-10-298.
[30] Lechner M, Findeib S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics 2011;12:124.
[31] 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.
[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] Heyram J, Vanparys B, Logan NA, Balcaen A, Rodriguez-Diaz M, Felske A, et al. *Bacillus novalis* sp. nov., *Bacillus vireti* sp. nov.,*Bacillus soli* sp. nov., *Bacillus bataviensis* sp. nov., and *Bacillus drentensis* sp. nov., from the Drentse A grasslands. Int J Syst Evol Microbiol 2004;54(1):47–57.
[34] Nagel M, Andreesen JR. *Bacillus niacini* sp. nov., a nicotinate- metabolizing mesophile isolated from soil. Int J Syst Evol Microbiol 1991;41(1): 134–9.
[35] Reva ON, Smirnov VV, Petterson B, Priest FG. *Bacillus endophyticus* sp. nov., isolated from the inner tissues of cotton plants (Gossypium sp.). Int J Syst Evol Microbiol 2002;52(1):101–7.
[36] Petterson B, De Silva SK, Uhlén M, Priest FG. *Bacillus siralis* sp. nov., a novel species from silage with a higher order structural attribute in the 16S rRNA genes. Int J Syst Evol Microbiol 2000;50(6):2181–7.
[37] Combet-Blanc Y, Olivier B, Streicher C, Patel BK, Dwivedi PP, Pot B, et al. *Bacillus thermoamylovorans sp. nov.*, a moderately thermophilic and amylolytic bacterium. Int J Syst Evol Microbiol 1995;45(1):9–16.
[38] Ten LN, Baek SH, Im WT, Larina LL, Lee JS, Oh HM, et al. *Bacillus pocheonensis* sp. nov., a moderately halotolerant, aerobic bacterium isolated from soil of a ginseng field. Int J Syst Evol Microbiol 2007;57(11):2532–7.
[39] Robert MS, Nakamura LK, Cohan FM. *Bacillus mojavensis* sp. nov., distinguishable from Bacillus subtilis by sexual isolation, divergence in DNA sequence, and differences in fatty acid composition. Int J Syst Evol Microbiol 1994;44(2):256–64.

Article 7: *Bacillus niameyensis* sp. nov. strain SIT3^{T} a new bacterial species isolated from human gut.

Maryam Tidjani Alou, Jaishriram Rathored, Sory Ibrahima Traore, Saber Khelaifia, Caroline Michelle, Souleymane Brah, Bouli Ali Diallo, Didier Raoult, Jean-Christophe Lagier.

Publié dans New Microbes and New Infections

Bacillus niameyensis sp. nov., a new bacterial species isolated from human gut

M. Tidjani Alou[1], J. Rathored[1], S. I. Traore[1], S. Khelaifia[1], C. Michelle[1], S. Brah[2], B. A. Diallo[3], D. Raoult[1,4] and J.-C. Lagier[1]

1) *Aix-Marseille Université, URMITE, UM63, CNRS7278, IRD198, Inserm 1095, Faculté de médecine, Marseille, France,* 2) *Hopital National de Niamey,* 3) *Laboratoire de Microbiologie, Département de Biologie, Université Abdou Moumouni de Niamey, Niamey, Niger and* 4) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Bacillus niameyensis sp. nov. strain SIT3^T (= CSUR P1266 = DSM 29725) is the type strain of *B. niameyensis* sp. nov. This Gram-positive strain was isolated from the digestive flora of a child with kwashiorkor and is a facultative anaerobic rod and a member of the *Bacillaceae* family. This organism is hereby described alongside its complete genome sequence and annotation. The 4 286 116 bp long genome (one chromosome but no plasmid) contains 4130 protein-coding and 66 RNA genes including five rRNA genes.

Keywords: *Bacillus niameyensis*, culturomics, genome, malnutrition, taxonogenomics
Original Submission: 8 July 2015; **Revised Submission:** 13 September 2015; **Accepted:** 14 September 2015
Article published online: 28 September 2015

Corresponding author: J.-C. Lagier, Aix-Marseille Université, URMITE, UMR CNRS 7278, IRD 198, INSERM U1095, Faculté de Médecine, 27 Bd Jean Moulin, 13385 Marseille Cedex 5, France
E-mail: jclagier@yahoo.fr
M. Tidjani Alou, J. Rathored and S.I. Traore contributed equally to this article, and all should be considered first author.

Introduction

Bacillus niameyensis strain SIT3^T (= CSUR P1266 = DSM 29725) is the type strain of *B. niameyensis* sp. nov. This bacterium is a Gram-positive bacillus, spore-forming, facultative anaerobic, and motile. It was isolated from the stool of a child living in Niamey, Niger, afflicted with kwashiorkor. This isolation was part of a culturomics study of the gut microbiota of children with severe acute malnutrition aiming to characterize their microbiota. Culturomics aims to explore as exhaustively as possible a microbial ecosystem using multiple culture conditions [1,2].

Phylogenetic relationships based on the 16S ribosomal RNA gene are currently used for bacterial classification alongside phenotypic and genotypic characteristics [3–5]. However, with the development of genomic sequencing and its decreasing cost, a new concept of bacterial description has been used in our laboratory [2,6–11] combining a proteomics description with the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) profile [12] associated with a biochemical and genomic description of the new bacterial species.

The genus *Bacillus* was established in 1872 by Cohn and encompasses over 200 described species and subspecies belonging to the *Firmicutes* phylum. *Bacillus* species are strictly aerobic and facultative anaerobic rod-shaped bacteria that form heat-resisting endospores [12–15]. They are often found in the environment (water, soil, air), food, plants and human clinical samples [16]. Some *Bacillus* species, such as *Bacillus thuringiensis,* are known to be pathogenic for insects and are used as biological control agents for crops [14].

Materials and methods

Organism information: classification and features

A stool sample was collected from a 2-year-old boy living in Niamey, Niger, with kwashiorkor, a form of severe acute

http://dx.doi.org/10.1016/j.nmni.2015.09.011

malnutrition. Consent was obtained from the child's parents at the National Hospital of Niamey and the study was approved by the Institut Fédératif de Recherche 48, Faculty of Medicine, Marseille, France, under agreement 09-022. The patient did not receive antibiotics at the time of sample collection, and the fecal sample was stored at −80°C.

Strain identification

The stool sample was cultured using the culturomics concept [2]. MALDI-TOF was used for colony identification as described below. In case of a failed identification using this technique, the 16S ribosomal RNA was sequenced. Stackebrandt and Ebers [17] suggested that a similarity level lower than 98.7% defined a new species (Fig. 1) without performing DNA-DNA hybridization (DDH). This similarity value is below the 16S rRNA threshold of 98.7% set by Stackebrandt and Ebers to delineate a new species without carrying out DDH.

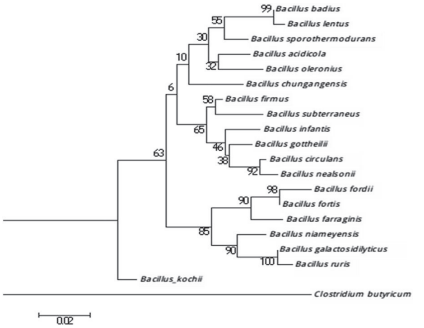


FIG. 1. Phylogenetic tree highlighting position of *Bacillus niameyensis* sp. nov. strain SIT3 (= CSUR P1266 = DSM 29725) relative to other type strains within *Bacillus* genus. Strains and their corresponding GenBank accession numbers for 16S rRNA genes are *Bacillus amyloliquefaciens* AB006920, *Bacillus bataviensis* AJ542508, *Bacillus drentensis* AJ542506, *Bacillus endophyticusv* AF295302, *Bacillus galactosidilyticus* AJ535638, *Bacillus idriensis* AY904033, *Bacillus mojavensis* AB021191, *Bacillus infantis* AY904032, *Bacillus niacini* AB021194, *Bacillus novalis* AJ542512, *Bacillus pseudomycoides* AF13121, *Bacillus pocheonensis* AB245377, *Bacillus thuringiensis* D16281, *Bacillus thermoamylovorans* L27478. Sequences were aligned using Clustal W (http://www.clustal.org/clustal2/), and phylogenetic inferences were obtained using maximum-likelihood method within MEGA 6 (http://www.megasoftware.net/mega.php). Numbers at nodes are percentages of bootstrap values obtained by repeating analysis 1000 times to generate majority consensus tree. *Clostridium butyricum* AJ458420 was used as outgroup. Scale bar = 1% nucleotide sequence divergence.

Phenotypic characteristics

Phenotypic characteristics like Gram staining (Fig. 2), sporulation, motility, catalase and oxidase were highlighted as previously described [18]. Biochemical features of our strain were investigated using API 20NE, ZYM and 50 CH strips according to the manufacturer's instructions (bioMérieux, Marcy l'Étoile, France). Various growth temperatures (25, 30, 37, 45 and 56°C) and atmospheres were tested. Growth under anaerobic and microaerophilic conditions occurrence was tested using GENbag Anaer and GENbag miroaer systems, respectively (bioMérieux). Aerobic growth was achieved with or without 5% CO_2.

In order to perform the electronic microscopy to observe our strain (Fig. 3), detection Formvar-coated grids were deposited on a 40 µL bacterial suspension drop and incubated during 30 minutes at 37°C. The grids were incubated for 10 seconds on ammonium molybdate 1%, dried on blotting paper and then observed with a Morgani 268D transmission electron microscope (Philips/FEI, Hillsboro, OR, USA) at an operating voltage of 60 kV.

MALDI-TOF protein analysis was carried out as previously described [12,19] using a Microflex spectrometer (Bruker Daltonics, Leipzig, Germany). Twelve individual colonies were deposited on a MTP 96 MALDI-TOF target plate (Bruker). The 12 spectra were imported into the MALDI BioTyper software (version 2.0, Bruker) and analysed by standard pattern matching (with default parameter settings) against the main spectra of 6.252 bacteria, including 199 spectra from 104 validly named *Bacillus* species used as reference data in the BioTyper database. A score enabled the presumptive identification and discrimination of the tested species from those in a database: a score of >2 with a validated species enabled the identification at the species level, and a score of <1.7 did not enable any identification. No significant score was obtained for strain SIT3^T, thus

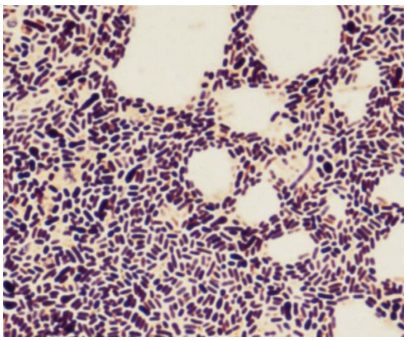


FIG. 2. Gram staining of *B. niameyensis* strain SIT3^T.

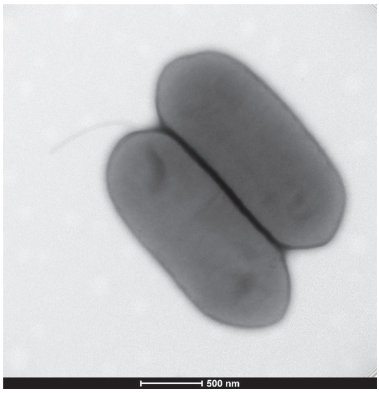


FIG. 3. Transmission electron microscopy of *B. niameyensis* strain SIT3, using Morgani 268D transmission electron microscope (Philips/FEI, Hillsboro, OR, USA) at operating voltage of 60 kV. Scale bar = 500 nm.

suggesting that our isolate was not a member of any known species. The reference spectrum for strain SIT3^T (Fig. 4) was incremented in our database and then compared to other known species of the *Bacillus* genus. The differences are shown in Fig. 5.

Growth conditions and genomic DNA preparation

B. niameyensis strain SIT3^T (= CSUR P1266 = DSM 29725) was grown on 5% sheep's blood–enriched Columbia agar (bioMérieux) at 37°C in aerobic atmosphere. Bacteria grown on three petri dishes were collected and resuspended in 4 × 100 µL of Tris-EDTA (TE) buffer. Then 200 µL of this suspension was diluted in 1 mL TE buffer for lysis treatment that included a 30-minute incubation with 2.5 µg/µL lysozyme at 37°C, followed by an overnight incubation with 20 µg/µL proteinase K at 37°C. Extracted DNA was then purified using three successive phenol–chloroform extractions and ethanol precipitations at −20°C overnight. After centrifugation, the DNA was resuspended in 160 µL TE buffer.

Genome sequencing and assembly

Genomic DNA of *Bacillus niameyensis* was sequenced on the MiSeq Technology (Illumina, San Diego, CA, USA) with the mate pair strategy. The gDNA was bar coded in order to be mixed with 11 other projects with the Nextera Mate Pair sample prep kit (Illumina). gDNA was quantified by a Qubit assay with a high-sensitivity kit (Life Technologies, Carlsbad, CA, USA) to 66.2 ng/µL. The mate pair library was prepared with 1 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 to 11 kb, with an optimal size at 3.927 kb. No size selection was performed, and 505 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal at 597 bp on a Covaris device S2 (Covaris, Woburn, MA, USA) in microtubes. The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies), and the final concentration library was measured at 59.2 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

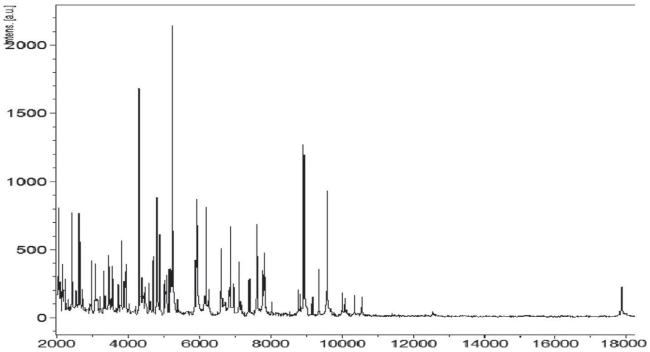


FIG. 4. Reference mass spectrum from *B. niameyensis* strain SIT3^T. Spectra from 12 individual colonies were compared and reference spectrum generated.

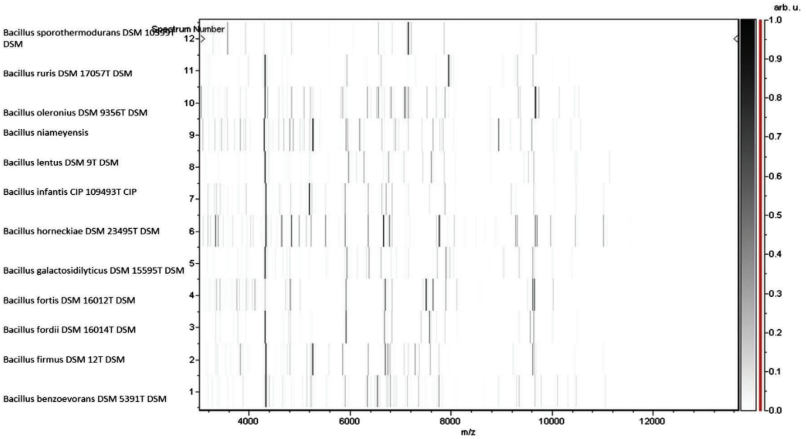


FIG. 5. Gel view comparing *B. niameyensis* (= CSUR P1266 = DSM 29725) to other species within genus *Bacillus*. Gel view displays raw spectra of loaded spectrum files arranged in pseudo-gel-like look. X-axis records m/z value. Left y-axis displays running spectrum number originating from subsequent spectra loading. Peak intensity expressed by greyscale scheme code. Color bar and right y-axis indicate relation between color peak, with peak intensity in arbitrary units. Displayed species are indicated at left.

instrument along with the flow cell. An automated cluster generation and sequencing run was performed in a single 39-hour run in a 2 × 251 bp.

Genome annotation and genome analysis

Open reading frames (ORFs) were predicted using Prodigal 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 (COGs) database using BLASTP (E value $1e^{-03}$, coverage 0.7 and identity percentage 30%). If no hit was found, it was searched against the NR database using BLASTP with an E value of $1e^{-03}$, coverage 0.7 and identity percentage 30%. If the sequence length was smaller than 80 aa, we used an E value of $1e^{-05}$. The tRNAScanSE tool [20] was used to find tRNA genes, whereas ribosomal RNAs were found by using RNAmmer [21]. Lipoprotein signal peptides and the number of transmembrane helices were predicted using Phobius [22]. ORFans were identified if all the BLASTP analyses performed did not provide 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 than 80 aa). Such parameter thresholds have already been used in previous studies to define ORFans.

Genomes were automatically retrieved from the 16S RNA tree using Xegen software (Phylopattern) [23]. For each selected genome, complete genome sequence, proteome genome sequence and Orfeome genome sequence were retrieved via File Transfer Protocol (FTP) from the National Center for Biotechnology Information. All proteomes were analysed with proteinOrtho [24]. Then for each couple of genomes, a similarity score was computed. This score is the mean value of nucleotide similarity between all couple of orthologues between the two genomes studied (average genomic identity of orthologous gene sequences, AGIOS) [25]. An annotation of the entire proteome was performed to define the distribution of functional classes of predicted genes according to the

TABLE 1. **Classification and general features of *Bacillus niameyensis* strain SIT3^T**

Property	Term
Current classification	Domain: *Bacteria* Phylum: *Firmicutes* Class: *Bacilli* Order: *Bacillales* Family: *Bacillaceae* Genus: *Bacillus* Species: *Bacillus niameyensis* Type strain: SIT3
Gram stain	Positive
Cell shape	Rod
Motility	Motile
Sporulation	sporulating
Temperature range	Mesophilic
Optimum temperature	37°C

TABLE 2. Differential characteristics of *Bacillus niameyensis* strain SIT3^T, *Bacillus hackensackii, Bacillus galactosidilyticus* DSM 15595, *Bacillus fordii* DSM 16014, *Bacillus fortis* DSM 16012, *Bacillus oceanisediminis* JCM 16506, *Bacillus sporothermodurans* DSM 10599, *Bacills infantis* JCM 13438, *Bacillus horneckiae* MTCC 9535, *Bacillus lentus* ATCC 10840 [29–36]

Property	*B. niameyensis*	*B. hackensackii*	*B. galactosidilyticus*	*B. fordii*	*B. fortis*	*B. oceanisediminis*	*B. sporothermodurans*	*B. infantis*	*B. horneckiae*	*Bacillus lentus*
Cell diameter (µm)	0.7–1.0	2.0–8.0	0.7–0.9	0.6–0.8	0.6–0.8	0.6–0.8	0.7	NA	1.0–1.5	0.7–1.2
Oxygen requirement	+	+	+	+	+	+	+	+	+	+
Gram stain	+	+	+	–	–	+	+	+	+	+
Salt requirement	NA	NA	NA	NA	NA	NA	NA	NA	–	–
Motility	+	+	+	+	+	NA	+	NA	+	+
Endospore formation	+	+	+	+	+	+	+	NA	+	+
Indole	–	–	–	–	–	+	NA	NA	–	NA
Production of										
Alkaline phosphatase	–	NA	NA	NA	NA	NA	NA	NA	+	NA
Catalase	+	+	+	+	+	+	+	+	+	+
Oxidase	–	–	NA	+	+	+	+	–	–	NA
Nitrate reductase	+	–	+	NA	NA	+	–	NA	+	–
Urease	–	+	+/–	NA	NA	–	–	NA	–	+
β-Galactosidase	+	NA	+	NA	NA	+	NA	NA	–	NA
N-acetyl-glucosamine	+	NA	+	+/–	+	NA	NA	+	+	NA
Acid from:										
L-Arabinose	–	–	+/–	w	–	–	NA	NA	+	+
Ribose	–	–	+/–	w	+	–	NA	+	–	NA
Mannose	–	–	+	–	–	–	–	–	+	+
Mannitol	–	–	–	–	–	+	–	+	+	+
Sucrose	–	–	+/–	–	–	+	NA	+	–	+
D-Glucose	–	–	w	–	+	+	–	+	–	+
D-Fructose	–	–	w	–	–	–	–	+	–	NA
D-Maltose	+	–	w	–	–	–	NA	+	+	NA
D-Lactose		–	+/–	–	–	–	–	+	–	NA
Habitat	Human gut	Blood	Raw milk	Raw milk	Milking apparatus	Marine sediment	Milk	Blood	Spacecraft assembly	soil

NA, data not available; w, weak reaction.

clusters of orthologous groups of proteins (using the same method than for the genome annotation). To evaluate the genomic similarity among studied *Bacillus* strains, we determined two parameters, dDDH – digital DDH (DNA-DNA Hybridization), which exhibits a high correlation with DDH [24,25], and AGIOS [23], which was designed to be independent from DDH.

Results

Phenotypic description

Strain SIT3^T (Table 1) was first isolated in May 2014 by a 15-day preincubation in blood culture with sheep blood and cultivation on 5% sheep blood–enriched Colombia agar (bioMérieux) in aerobic atmosphere at 37°C. This strain showed a 95.1% nucleotide sequence similarity with *Bacillus lentus,* the phylogenetically closest *Bacillus* species with a valid published name (Fig. 1).

Growth occurred between 28 and 45°C on blood-enriched Columbia agar (bioMérieux), with the optimal growth being obtained at 37°C after 48 hours of incubation. Optimal growth was achieved aerobically. Weak cell growth was observed under microaerophilic and anaerobic conditions. The motility test was positive, and the cells were sporulating. Colonies were translucent and 1 to 2 mm in diameter on sheep's blood–enriched Columbia agar. Cells were Gram-positive rods (Fig. 2). Under electron microscopy, the bacteria grown on agar had a mean diameter and length of 0.87 µm and 2.6 µm, respectively (Fig. 3).

Strain SIT3^T exhibited catalase activity but was negative for oxidase.

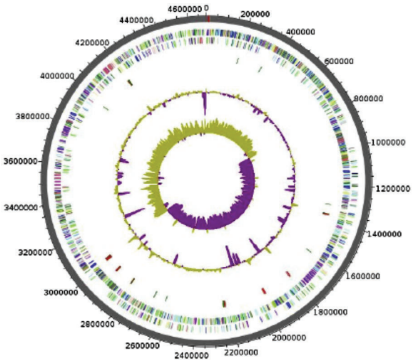


FIG. 6. Graphical circular map of chromosome. From outside to center: genes on forward strain colored by COGs categories (only gene assigned to COGs), RNA genes (tRNAs green, rRNAs red), GC content and GC skew. COGs, Clusters of Orthologous Groups database.

Using an API ZYM strip (bioMérieux), positive reactions were observed for esterase (C4), esterase lipase (C8), leucine arylamidase, α-chymotrypsin, naphtol-AS-BI-phosphohydrolase, β-galactosidase and N-acetyl-β-glucosaminidase. Negative reactions were observed for acid phosphatase, lipase (C14), valine arylamidase, cysteine arylamidase, trypsin, alkaline phosphatase, α-galactosidase, β-glucuronidase, α-glucosidase, β-glucosidase, α-fucosidase and α-mannosidase.

Using an API 20 NE strip, positive reactions were obtained for nitrate reduction, β-glucosidase and β-galactosidase. All other reactions were negative.

Using an API 50 CH strip (bioMérieux), positive reactions were recorded for methyl-αD-glucopyranoside, N-acetylglucosamine, esculin ferric citrate, D-maltose, D-trehalose and starch. Negative reactions were recorded for glycerol, erythritol, D-arabinose, L-arabinose, D-ribose, D-xylose, L-xylose, D-adonitol, methyl-βD-xylopranoside, D-galactose, D-glucose, D-fructose, D-mannose, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl-αD-mannopyranoside, amygdaline, arbutin, salicin, D-cellobiose, D-lactose, D-melibiose, D-saccharose, inulin, D-raffinose, D-melezitose, glycogen, xylitol, gentiobiose, D-turanose, D-lyxose, D-tagatose, D-fucose, L-fucose, D-arabitol, L-arabitol, potassium gluconate, potassium 2-ketogluconate and potassium 5-ketogluconate.

Cells are susceptible to imipenem, metronidazole, doxycycline, rifampicin, vancomycin, amoxicillin, ceftriaxone and gentamicin (500 µg) and are resistant to trimethoprim/sulfamethoxazole, erythromycin, ciprofloxacin and gentamicin (15 µg).

TABLE 3. Nucleotide content and gene count levels of the genome

Attribute	Genome (total)	
	Value	% of total[a]
Size (bp)	4 286 116	100
G + C content (bp)	1 603 007	37.4
Coding region (bp)	3 665 154	85.51
Total genes	4196	100
RNA genes	66	1.57
Protein-coding genes	4130	98.42
Genes with function prediction	3058	72.88
CRISPRs	7	0.16
Genes assigned to COGs	2694	64.20
Genes with peptide signals	294	7.0
Genes with transmembrane helices	1180	28.12
Genes associated to PKS or NRPS	10	0.23
Genes associated to mobilome	2198	51.80
Genes associated to toxin/antitoxin	111	2.61
Genes associated to resistance genes	0	0
Genes with paralogues (E value $1e^{-10}$)	1317	31.03
Genes larger than 5000 nt	2	0.05

COGs, Clusters of Orthologous Groups database; CRISPR, clustered regularly interspaced short palindromic repeat; NRPS, nonribosomal peptide synthase; PKS, polyketide synthase.
[a]Total is based on either size of genome (bp) or total number of protein coding genes in annotated genome.

TABLE 4. Number of genes associated with the 25 general COGs functional categories

Code	Value	% of total[a]	Description
J	180	4.36	Translation
A	0	0.00	RNA processing and modification
K	297	7.19	Transcription
L	185	4.48	Replication, recombination and repair
B	1	0.02	Chromatin structure and dynamics
D	36	0.87	Cell cycle control, mitosis and meiosis
Y	0	0.00	Nuclear structure
V	111	2.69	Defense mechanisms
T	186	4.50	Signal transduction mechanisms
M	173	4.19	Cell wall/membrane biogenesis
N	64	1.55	Cell motility
Z	0	0.00	Cytoskeleton
W	0	0.00	Extracellular structures
U	44	1.07	Intracellular trafficking and secretion
O	99	2.40	Posttranslational modification, protein turnover, chaperones
C	156	3.78	Energy production and conversion
G	351	8.50	Carbohydrate transport and metabolism
E	279	6.76	Amino acid transport and metabolism
F	86	2.08	Nucleotide transport and metabolism
H	97	2.35	Coenzyme transport and metabolism
I	85	2.06	Lipid transport and metabolism
P	218	5.28	Inorganic ion transport and metabolism
Q	73	1.77	Secondary metabolites biosynthesis, transport and catabolism
R	520	12.59	General function prediction only
S	326	7.89	Function unknown
—	374	8.91	Not in COGs

COGs, Clusters of Orthologous Groups database.
[a]Total is based on either size of genome (bp) or total number of protein coding genes in annotated genome.

The differences exhibited by the comparison with other representatives of the genus *Bacillus* are detailed in Table 2.

Genome properties

The genome of *B. niameyensis* strain SIT3^T is 4 286 116 bp long with a 37.40% G + C content (Fig. 6, Table 3) and 46 generated contigs. Of the 4196 predicted genes, 4130 were protein-coding genes and 66 were RNAs. Of these 66 rRNA genes, one is a 16S rRNA gene, three are 5S RNA genes, one is a 23S rRNA and 61 predicted tRNA genes were identified in the genome. A total of 3068 genes (73.11%) were assigned a putative function. A total of 230 genes were identified as ORFans (5.42%). Using ARG-ANNOT [26], no resistance genes were found. Nevertheless, ten genes associated to polyketide synthase (PKS) or nonribosomal peptide synthase (NRPS) were discovered in the analysis of the genome. The remaining genes were annotated as hypothetical proteins. The properties and

TABLE 5. Closely related species with *Bacillus niameyensis*

Name of species	Similarity (%)	Accession No.
Bacillus niameyensis	100	LK985389
Bacillus oceanisediminis	94.88	KP120947.1
Bacillus nealsonii	94.45	KC433938.1
Bacillus firmus	93.50	KF478238.1
Bacillus infantis	94.60	KP696715.1
Bacillus fordii	95.56	NR_025786.1

TABLE 6. Number of orthologous proteins shared between genomes (upper right)[a]

	Bacillus niameyensis	*Bacillus badius*	*Bacillus firmus*	*Bacillus fordii*	*Bacillus infantis*	*Bacillus nealsonii*	*Bacillus oceanisediminis*	*Clostridium butyricum*
Bacillus niameyensis	4,130	1,511	1,769	1,648	1,800	1,658	1,846	832
Bacillus badius	61.24	4,486	1,940	1,730	1,886	1,601	2,040	838
Bacillus firmus	61.8	62.60	4,142	1,956	2,634	2,062	3,039	971
Bacillus fordii	65.06	61.82	62.01	4,688	1,842	1,608	2,069	824
Bacillus infantis	59.99	62.25	67.18	61.3	4,142	2,074	2,728	946
Bacillus nealsonii	61.44	60.15	62.84	59.95	61.08	4,789	2,149	957
Bacillus oceanisediminis	61.82	62.67	88.49	62.03	67.17	63.04	5,578	1,012
Clostridium butyricum	54.21	51.46	52.78	52.44	50.8	54.49	52.8	4,152

[a]Average percentage similarity of nucleotides corresponding to orthologous proteins shared between genomes (lower left) and numbers of proteins per genome.

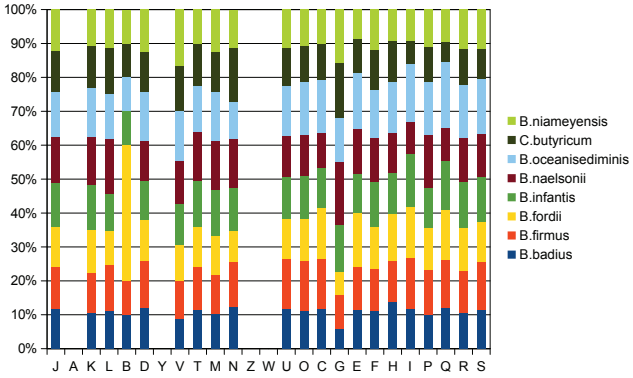


FIG. 7. Distribution of functional classes of predicted genes according to clusters of orthologous groups of proteins.

statistics of the genome are summarized in Table 3. The distribution of genes into COGs functional categories is presented in Table 4.

Genome comparison

Bacillus oceanideminis, *Bacillus nealsonii*, *Bacillus firmus*, *Bacillus infantis* and *Bacillus fordii* are closely related species to *B. niameyensis* with available genomes (Table 5) and were thus chosen for this comparative analysis. The draft genome sequence of *Bacillus niameyensis* is smaller than those of *Bacillus oceanisediminis*, *Bacillus nealsonii*, *Bacillus firmus*, *Bacillus infantis* and *Bacillus fordii* (5.76, 4.98, 4.97, 4.88, 4.62 and 4.51 MB, respectively) but larger than those of *Bacillus badius* (4.04 MB). It should be noted that the size of our genome is an estimation

TABLE 7. Pairwise comparison of *Bacillus niameyensis* with eight other species using GGDC, formula 2 (DDH estimates based on identities/HSP length)[a] **[27,28]**

	Bacillus niameyensis	*Bacillus badius*	*Bacillus firmus*	*Bacillus fordii*	*Bacillus infantis*	*Bacillus nealsonii*	*Bacillus oceanisediminis*	*Clostridium butyricum*
Bacillus niameyensis	100% ± 00	2.53% ± 0.18	2.53% ± 0.20	2.54% ± 0.22	2.54% ± 0.13	2.54% ± 0.20	2.53% ± 0.22	2.53% ± 0.17
Bacillus badius		100% ± 00	2.54% ± 0.19	2.53% ± 0.17	2.55% ± 0.18	2.53% ± 0.17	2.54% ± 0.21	2.53% ± 0.19
Bacillus firmus			100% ± 00	2.53% ± 0.19	2.60% ± 0.22	2.55% ± 0.20	3.04% ± 0.11	2.53% ± 0.18
Bacillus fordii				100% ± 00	2.53% ± 0.18	2.53% ± 0.17	2.53% ± 0.19	2.52% ± 0.18
Bacillus infantis					100% ± 00	2.54% ± 0.15	2.60% ± 0.21	2.53% ± 0.18
Bacillus nealsonii						100% ± 00	2.56% ± 0.16	2.53% ± 0.18
Bacillus. oceanisediminis							100% ± 00	2.53% ± 0.18
Clostridium butyricum								100% ± 00

DDH, DNA-DNA hybridization; GGDC, genome-to-genome distance; HSP, high-scoring pair.
[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 (Fig. 1) and phylogenetic analyses as well as GGDC results.

because 46 contigs were obtained in the sequencing of the genome.

The G + C content of *Bacillus niameyensis* is smaller than those of *Bacillus infantis, Bacillus badius, Bacillus firmus, Bacillus fordii* and *Bacillus oceanisediminis* (46.0, 43.90, 41.40, 41.20 and 40.8%, respectively) but larger than those of *Bacillus nealsonii* and *Clostridium butyricum* (35.1 and 28.8%).

The gene content of *Bacillus niameyensis* is smaller than those of *Bacillus oceanisediminis Bacillus nealsonii, Bacillus infantis, Bacillus badius, Bacillus fordii, Bacillus firmus* and *Clostridium butyricum* (5722, 4864, 4837, 4486, 4229, 4229 and 4231, respectively). However the distribution of genes into COGs categories was similar in all compared genomes (Table 6, Fig. 7). In addition, *Bacillus niameyensis* shared 4130, 4486, 4142, 4688, 4142, 4789, 5578 and 4152 orthologous genes with *Bacillus badius, Bacillus firmus, Bacillus fordii, Bacillus infantis, Bacillus nealsonii, Bacillus oceanisediminis* and *Clostridium butyricum* (Table 7). Among species with standing in nomenclature, AGIOS values ranged from 61.24 between *Bacillus niameyensis* and *Bacillus badius* to 63.04% between *Bacillus oceanisediminis* and *Bacillus nealsonii.*

Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Bacillus niameyensis* sp. nov. that contains the strain SIT3^T. This bacterial strain was isolated from the fecal flora of a 2-year-old boy from Niamey, Niger, with kwashiorkor.

Taxonomic and nomenclatural proposals

Description of *Bacillus niameyensis* strain SIT3^T sp. nov. Cells are Gram-positive, sporulating, rod-shaped bacilli with a diameter of 0.1 µm. Colonies are translucent and 1 to 2 mm in diameter on 5% sheep's blood–enriched Columbia agar (bioMérieux). Cells are catalase positive and oxidase negative. Positive reactions were observed for esterase (C4), esterase lipase (C8), leucine arylamidase, α-chymotrypsin, naphtol-AS-BI-phosphohydrolase, β-galactosidase and N-acetyl-β-glucosaminidase. Negative reactions were observed for acid phosphatase, lipase (C14), valine arylamidase, cysteine arylamidase, trypsin, alkaline phosphatase, α-galactosidase, β-glucuronidase, α-glucosidase, β-glucosidase, α-fucosidase and α-mannosidase.

Cells are susceptible to imipenem, metronidazole, doxycycline, rifampicin, vancomycin, amoxicillin, ceftriaxone and gentamycin (500 µg) and are resistant to trimethoprim/sulfamethoxazole, erythromycin, ciprofloxacin and gentamicin (15 µg).

The G + C content of the genome is 37.40%. The 16S rRNA gene sequence and whole-genome shotgun sequence of *B. niameyensis* strain SIT3^T are deposited in GenBank under accession numbers LK955389 and CTDY01000000, respectively. The type strain SIT3^T (= CSUR P1266 = DSM 29725) was isolated from the stool of a child living in Niamey, Niger, with kwashiorkor.

Conflict of interest

None declared.

Acknowledgements

The authors thank the Xegen Company (http://www.xegen.fr/) for automating the genomic annotation process. This study was funded by the Méditerranée Infection Foundation. We thank Karolina Griffiths for English-language review.

References

[1] Dubourg G, Lagier JC, Armougom F, et al. The gut microbiota of a patient with resistant tuberculosis is more comprehensively studied by culturomics than by metagenomics. Eur J Clin Microbiol Infect 2013;32:637–45.
[2] Lagier JC, Armougom F, Million M, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 2012;18:1185–93.
[3] Viale AM, Arakaki AK, Soncini FC, et al. Evolutionnary relationships among eubacterial groups as inferred from GroEL (Chaperonin) sequences comparison. Int J Syst Bacteriol 1944;44:527–33.
[4] Woese CR, Kandler O, Wheelis ML, et al. Towards a natural system of organisms: proposals for the domains *Archaea, Bacteria*, and *Eucarya*. Proc Natl Acad Sci U S A 1990;87:4576–9.
[5] Wolf M, Muller T, Dandekar T, et al. Phylogeny of firmicutes with special reference to *Mycoplasma (Mollicutes)* as inferred from phosphoglycerate kinase amino acid sequence data. Inst J Syst Evol Microbiol 2004;54(pt 3):871–5.
[6] Hugon P, Mishra AK, Lagier JC, et al. Non-contiguous finished genome sequence and description of *Brevibacillus massiliensis* sp. nov. Stand Genomic Sci 2013;8:1–14.
[7] Kokcha S, Ramasamy D, Lagier JC, et al. Non-contiguous finished genome sequence and description of *Brevibacterium senegalense* sp. nov. Stand Genomic Sci 2012;7:233–45.
[8] Lagier JC, El Karkouri K, Nguyen TT, et al. Non-contiguous finished genome sequence and description of *Anaerococcus senegalensis* sp. nov. Stand Genomic Sci 2012;6:116–25.
[9] Lagier JC, Armougom F, Mishra AK, et al. Non-contiguous finished genome sequence and description of *Alistipes timonensis* sp. nov. Stand Genomic Sci 2012;6:315–24.
[10] Lagier JC, Elkarkouri K, Rivet R, et al. Non contiguous-finished genome sequence and description of *Senegalemassilia anaerobia* gen. nov., sp. nov. Stand Genomic Sci 2013;7:343–56.

[11] Lagier JC, El Karkouri K, Mishra AK, et al. Non contiguous-finished genome sequence and description of *Enterobacter massiliensis* sp. nov. Stand Genomic Sci 2013;7:399–412.
[12] Seng P, Drancourt M, Gouriet F, 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.
[13] Fritze D. Taxonomy of the genus *Bacillus* and related genera: the aerobic endospore-forming bacteria. Physiopathology 2004;94: 1245–8.
[14] Gordon RE, Haynes WC, Pang CH, et al. The genus *Bacillus*. US Dep Agric Handh 1973;427.
[15] Priest FG, Goodfellow M, Todd C, et al. A numerical classification of the genus *Bacillus*. J Gen Microbiol 1988;134:1847–82.
[16] Goto K, Omura T, Hara Y, Sadaie Y, et al. Application of the 16S rDNA sequence as an index for rapid identification of species in the genus *Bacillus*. J Gen Appl Microbiol 2000;46:1–8.
[17] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152–5.
[18] Lagier JC, Edouard S, Pagnier I, Mediannikov O, Drancourt M, Raoult D. Current and past strategies for bacterial culture in clinical microbiology. Clin Microbiol Rev 2015;28:208–36.
[19] Seng P, Abat C, Rolain JM, 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.
[20] Bendtsen JD, Nielsen H, von Heijne G, et al. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340:783–95.
[21] 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.
[22] Käll L, Krogh A, Sonnhammer EL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol 2004;338: 1027–36.
[23] Gouret P, Thompson JD, Pontarotti P. PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinform 2009;10:298.
[24] Lechner M, Findeib S, Steiner L, et al. Proteinortho: detection of (co-) orthologs in large-scale analysis. BMC Bioinform 2011;12:124.
[25] Ramasamy D, Mishra AK, Lagier JC, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of new bacterial species. Int J Syst Evol Microbiol 2014;64:384–91.
[26] Gupka SK, Padmanabhan BR, Diene SM, et al. ARG-ANNOT, a new bioinformatics tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 2014;58:212–20.
[27] Auch AF, Von Jan M, Klenk HP, et al. DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci 2010;2:117–34.
[28] Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinform 2013;14:60.
[29] Hong T, Heibler N, Tang Y, et al. *"Bacillus hackensackii"* sp. nov., a novel carbone dioxide sensitive bacterium isolated from blood culture. Diagn Microbiol Infect Dis 2003;45:143–7.
[30] Heyndrickx M, Logan NA, Lebbe L, et al. *Bacillus galactosidilyticus* sp. nov., an alkali-tolerant beta-galactosidase producer. Int J Syst Evol Microbiol 2004;54:617–21.
[31] Scheldeman P, Rodriguez-Diaz M, Goris J, et al. Heyndrickx M. *Bacillus farraginis* sp. nov., *Bacillus fortis* sp. nov., *Bacillus fordii* sp. nov., isolated from dairy farms. Int J Syst Evol Microbiol 2004;54:1355–64.
[32] Zhang J, Wang J, Fang C, et al. *Bacillus oceanisediminis* sp. nov., isolated from marine sediment. Int J Syst Evol Microbiol 2010;60:2924–9.
[33] Pettersson B, Lembke F, Hammer P, et al. *Bacillus sporothermodurans*, a new species producing highly heat-resistant endospores. Int J Syst Evol Microbiol 1996;46:759–63.
[34] Ko KS, Oh WS, Lee MY, Lee, et al. *Bacillus infantis* sp. nov. and *Bacillus idriensis* sp. nov., isolated from a patient with a neonatal sepsis. Int J Syst Evol Microbiol 2006;56:2541–4.
[35] Vaishampayan P, Probst A, Krishnamurthi S, et al. *Bacillus horneckiae* sp. nov., isolated from a spacecraft-assembly clean room. Int J Syst Evol Microbiol 2010;60:1031–7.
[36] Gordon RE, Hyde JL, Moore Jr JA. *Bacillus firmus–Bacillus lentus*: a series or one species? Int J Syst Bacteriol 1977;27:256–62.

Article 8: *Massilibacterium senegalense* gen. nov., sp. nov., a new bacterial genus isolated from the human gut.

Maryam Tidjani Alou, Jaishriram Rathored, Jean-Christophe Lagier, Saber Khelaifia, Noémie Labas, Cheikh Sokhna, Aldiouma Diallo, Didier Raoult, Gregory Dubourg.

Publié dans New Microbes and New Infections

Massilibacterium senegalense gen. nov., sp. nov., a new bacterial genus isolated from the human gut

M. Tidjani Alou[1], J. Rathored[1], J.-C. Lagier[1], S. Khelaifia[1], N. Labas[1], C. Sokhna[1,2], A. Diallo[1,2], D. Raoult[1,3] and G. Dubourg[1]

1) *Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes IRD 198, CNRS 7278, Aix-Marseille Université, Marseille, France,* 2) *Campus Commun UCAD-IRD of Hann, Dakar, Senegal and* 3) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Massilibacterium senegalense gen. nov., sp. nov., strain mt8^T, is the type strain of *Massilibacterium* gen. nov., a new genus within the *Bacillaceae* family. This Gram-negative facultative anaerobic rod was isolated from the gut microbiota of a severely malnourished boy. Its phenotypic description is hereby presented with a complete annotation of its genome sequence. This genome is 5 697 950 bp long and contains 5615 protein-coding genes and 178 RNA genes, among which are 40 rRNA genes.

Keywords: Culturomics, genome, gut microbiota, *Massilibacterium senegalense*, taxonogenomics
Original Submission: 1 December 2015; **Revised Submission:** 11 January 2016; **Accepted:** 14 January 2016
Article published online: 22 January 2016

Corresponding author: G. Dubourg, Aix-Marseille Université, URMITE, UM63, CNRS7278, IRD198, INSERM 1095, Faculté de médecine, 27 Boulevard jean Moulin, 13385 Marseille cedex 05, France
E-mail: greg.dubourg@gmail.com
The first two authors contributed equally to this article, and both should be considered first author.

Introduction

The human gut microbiota is a complex and vast ecosystem harbouring eukaryotes, viruses, archaea and bacteria, these being by far the most abundant [1]. Its cell count is estimated to approximately 10^{14}, representing ten times the human somatic cell count, and its collective bacterial genome size is 150 times the size of the human genome [1–4]. The development of metagenomics has allowed a better exploration of gut microbiota by bypassing the noncultivable bacteria problem and unveiling links between altered gut microbiota and several diseases such as obesity, inflammatory bowel disease and irritable bowel syndrome [2]. It has also been demonstrated that the microbiota plays key roles in digestion and in immunologic and metabolic functions [2–4]. Nevertheless, a cultivation approach would be a complementary way to explore the gut microbiome in order to have a better representation of the viable population. In addition, it would allow further knowledge about the gut bacterial repertoire.

A new approach was developed in our laboratory in order to explore as exhaustively as possible the human gut microbiota by multiplying culture conditions with different atmospheres, media and temperatures [5]. This approach, known as culturomics, allowed us to isolate a new member of the *Bacillaceae* family. This family was created by Cohn in 1872 and consists of 52 validated genera (http://www.bacterio.net/). *Bacillus* is the type genus of this family, containing genera that are mostly aerobic or facultative anaerobic, rod-shaped, spore-forming, Gram-positive bacteria. These ubiquitous species are found in many ecosystems—mainly soil but also other environmental and clinical samples. Most *Bacillaceae* species are harmless, but some can be opportunistic pathogens, and *Bacillus anthracis,* the agent of anthrax, is well known to be pathogenic for humans [6].

Bacterial classification is currently based on phylogenetic relationships built on the 16S ribosomal RNA gene, phenotypic and genotypic characteristics including G+C content and DNA-DNA hybridization [7–9]. However, a great breakthrough has been

http://dx.doi.org/10.1016/j.nmni.2016.01.010

made in the last years in the area of genome sequencing, partly due to its decreasing cost. In fact, to this day, almost 70 000 genomes have been sequenced (https://gold.jgi.doe.gov/). With the development of this innovation, we proposed a new concept of bacterial description, including a proteomic description with a matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) profile [10,11] alongside a biochemical and genomic description of the new species [12–17].

We describe here a new member of this family, the genus *Massilibacterium*, isolated in the faeces of a patient with kwashiorkor. *Massilibacterium senegalense* is the type species (= CSUR P1510 = DSM 100455) of this new genus.

Materials and Methods

Organism information

As part of a culturomics study of the gut microbiota of children with severe acute malnutrition, a stool sample was collected from a 2-month-old Senegalese boy with kwashiorkor (body mass index, 14 kg/m^2) in April 2014. The patient was not treated with antibiotics at the time of sample collection; the sample was stored at −80°C. This study was authorized by the child's parents and was approved by the Institut Fédératif de Recherche 48 (Faculty of Medicine, Marseille, France) under agreement 09-022.

Strain identification by MALDI-TOF and 16S rRNA sequencing

Using the 18 culture conditions of the culturomics concept, the fecal sample was cultivated, and the obtained colonies were identified by MALDI-TOF as described below [5]. Proteomic analysis of our strain was carried out with MALDI-TOF as previously described [10,11]. A Microflex spectrometer (Bruker Daltonics, Leipzig, Germany) was used with a MTP 96 MALDI-TOF target plate (Bruker) on which 12 individual colonies were deposited. Twelve spectra were thus obtained, imported into MALDI BioTyper 2.0 software (Bruker) and analysed by standard pattern matching (with default parameter settings) against the main spectra of 7567 bacteria. Comparison with the BioTyper database spectra enabled the identification and discrimination of the analysed species from those in the database in accordance with the obtained score: a score >2 with a validated species enabled the identification at the species level, and a score <1.7 did not enable any identification. After a failed identification of the colony with a clean spectrum, it was identified by sequencing the 16S ribosomal RNA as previously described [18]. A threshold of 98.7% similarity level was determined to define a new species without performing DNA-DNA hybridization [19].

Growth conditions

In order to determine the ideal growth condition of *M. senegalense*, different growth temperatures (28, 30, 37, 45 and 56°C) were tested under anaerobic and microaerophilic atmospheres using GENbag anaer and GENbag miroaer systems respectively (bioMérieux, Marcy l'Étoile, France). The strain growth was also tested aerobically with or without 5% CO_2 supplementation.

Morphologic, biochemical and antibiotics susceptibility tests

The phenotypic characteristics (Gram staining, sporulation, motility, catalase, oxidase) were analysed as previously described [20]. Antibiotic susceptibility testing was performed using the disk diffusion method according to EUCAST 2015 recommendations (http://www.eucast.org/). Using API 20NE, API ZYM and API 50CH strips, we investigated the biochemical characteristics of the strain according to the manufacturer's instructions (bioMérieux). Electronic microscopy was performed with detection Formvar-coated grids which were deposited on a 40 µL bacterial suspension drop and incubated at 37°C for 30 minutes. Then followed a 10-second incubation on ammonium molybdate 1%. The grids were dried on blotting paper and finally observed with a Tecnai G20 transmission electron microscope (FEI Company, Limeil-Brevannes, France).

Genomic DNA preparation

M. senegalense strain mt8^T was cultured on 5% sheep's blood–enriched Columbia agar (bioMérieux) at 37°C aerobically. Bacteria grown on three petri dishes were resuspended in 4 × 100 µL of Tris-EDTA (TE) buffer. Then 200 µL of this suspension was diluted in 1 mL TE buffer for lysis treatment that included a 30-minute incubation with 2.5 µg/µL lysozyme at 37°C, followed by an overnight incubation with 20 µg/µL proteinase K at 37°C. Extracted DNA was then purified using three successive phenol–chloroform extractions and ethanol precipitations at −20°C overnight. After centrifugation, the DNA was resuspended in 160 µL TE buffer.

Genome sequencing and assembly

Genomic DNA (gDNA) of *M. senegalense* was sequenced on the MiSeq Technology (Illumina, 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). gDNA was quantified by a Qubit assay with a high sensitivity kit (Thermo Fisher Scientific, Waltham, MA, USA) to 66.2 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, USAUSA) with a DNA 7500 lab chip. The DNA fragments ranged in size from 1 to 11 kb, with an optimal size at 3.927 kb. No size selection was performed, and 505 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal at 597 bp on the Covaris device S2 in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies, Santa Clara, CA, USA), and the final concentration library was measured at 59.2 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. An automated cluster generation and sequencing run was performed in a single 39-hour run in a 2 × 251 bp read length.

Genome annotation and analysis

Open reading frames (ORFs) were predicted using Prodigal [21] 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 against the Clusters of Orthologous Groups (COGs) database using BLASTP (*E* value 1e-03, coverage 0.7 and identity percentage of 30%). If no hit was found, then it was searched against the NR database using BLASTP with an *E* value of 1e-03, coverage 0.7 and identity percentage of 30%, and if the sequence length was smaller than 80 amino acids, we used an *E* value of 1e-05. The tRNAScanSE tool [22] was used to find tRNA genes, whereas ribosomal RNAs were found by using RNAmmer [23]. Lipoprotein signal peptides and the number of transmembrane helices were predicted using Phobius [24]. 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 amino acids or *E* value smaller than 1e-05 for ORFs with sequence length smaller 80 amino acids). Such parameter thresholds have already been used in previous studies to define ORFans.

Genomes were automatically retrieved from the 16S RNA tree using Xegen software (Phylopattern) [25]. For each selected genome, the complete genome sequence, proteome genome sequence and Orfeome genome sequence were retrieved from the FTP site of the National Center for Biotechnology Information (NCBI). All proteomes were analysed with proteinOrtho [26]. Then for each couple of genomes a similarity score was computed. This score is the mean value of nucleotide similarity between all couple of orthologues between the two genomes studied (average genomic identity of orthologous gene sequences, AGIOS) [27]. 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). To evaluate the genomic similarity among studied *Bacillus* strains, we determined two parameters: digital DNA-DNA hybridization (dDDH), which exhibits a high correlation with DNA-DNA hybridization (DDH) [28,29], and AGIOS [27], which was designed to be independent from DDH. Annotation and comparison processes were performed in the Multi-Agent software system DAGOBAH [30] that included Figenix [31] libraries that provide pipeline analysis.

Results

Strain identification and phylogenetic analyses

Strain mt8^T (Table 1) was first isolated in January 2015 after 30-day preincubation in a blood culture bottle with sheep's blood and cultivation on 5% sheep's blood–enriched Colombia agar in an aerobic atmosphere at 37°C. MALDI-TOF displayed an identification score under 1.7 for strain mt8^T, suggesting that the obtained spectra was not matched to any spectra in our database. The 16S ribosomal RNA sequence (accession no. LN828943) of strain mt8^T showed a 93% nucleotide sequence similarity with *Bacillus halodurans,* which is the phylogenetically closest species with a validly published name (Fig. 1). Consequently, as this 16S rRNA nucleotide sequence similarity was lower than the threshold of 95% recommended by Stackebrandt and Ebers [19] to delineate a new genus, it was classified as a new genus called *Massilibacterium,* type species *Massilibacterium senegalense* strain mt8^T. The reference spectrum for strain mt8^T (Fig. 2) was thus incremented in our database and then compared to other known species of the genus *Bacillus.* The differences exhibited are shown in the obtained gel view (Fig. 3).

TABLE 1. Classification and general features of *Massilibacterium senegalense* strain mt8^T

Property	Term
Current classification	Domain: *Bacteria* Phylum: *Firmicutes* Class: *Bacilli* Order: *Bacillales* Family: *Bacillaceae* Genus: *Massilibacterium* Species: *Massilibacterium senegalense* Type strain: mt8
Gram stain	Negative
Cell shape	Rod
Motility	Motile
Sporulation	sporulating
Temperature range	Mesophilic
Optimum temperature	37°C

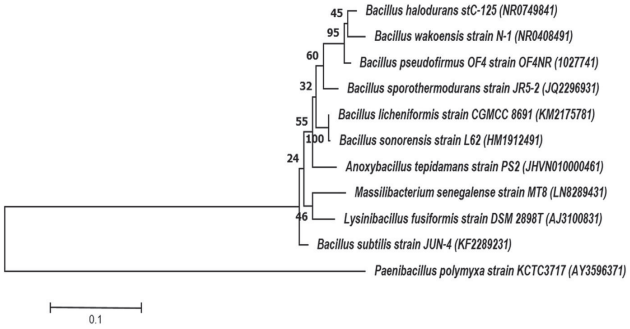


FIG. 1. Phylogenetic tree highlighting the position of *Massilibacterium senegalense* strain mt8^T relative to other close strains. The respective GenBank accession numbers for 16S rRNA genes are indicated in parenthesis. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using the maximum-likelihood method within the MEGA6. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 1,000 times to generate a majority consensus tree. *Paenibacillus polymyxa* strain KCTC3717 was used as an outgroup. The scale bar represents a 1% nucleotide sequence divergence.

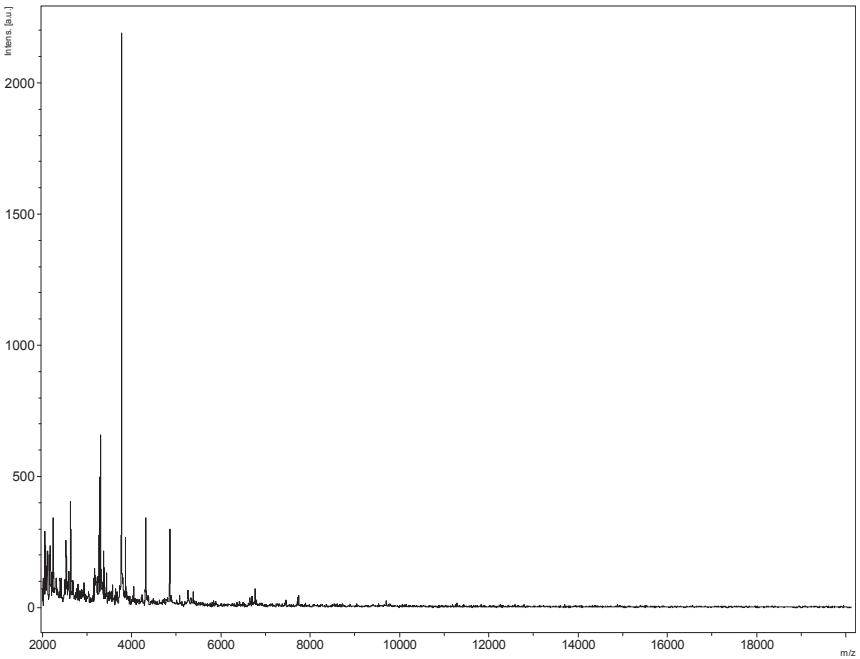


FIG. 2. Reference mass spectrum from *Massilibacterium senegalense* strain mt8^T. Spectra from 12 individual colonies were compared and a reference spectrum was generated.

Phenotypic description

Growth of strain mt8^{T}was observed between 28 and 45°C on 5% sheep's blood–enriched Columbia agar, and optimal growth was achieved at 37°C after 24 hours' incubation in aerobic conditions. Poor growth occurred under microaerophilic and anaerobic conditions. Cells were motile and sporulating. Colonies were irregular white colonies with a mean diameter of 5 mm on blood-enriched Colombia agar. The Gram staining (Fig. 4) showed Gram-negative rods. Using electron microscopy, the rods had a mean diameter of 1.8 µm and a length of 5.9 µm (Fig. 5).

Catalase and oxidase activities were negative for strain mt8^T. Using API ZYM, positive reactions were observed for esterase (C4) and acid phosphatase. Reactions for alkaline phosphatase, esterase lipase (C8), lipase (C14), leucine arylamidase, valine arylamidase, trypsin, α-chymotrypsin, naphthol-AS-BI-phosphohydrolase, β-galactosidase, N-acetyl-β-glucosaminidase, α-galactosidase, β-glucuronidase, α-glucosidase, β-glucosidase, α-fucosidase and α-mannosidase were negative. Using API 20NE, nitrate reduction and esculin hydrolysis were observed. All other reactions were negative, including indole formation and urease. An API 50CH strip showed positive reactions for N-acetyl-glucosamine, amygdalin, arbutin, salicin, D-maltose, D-fructose, inulin, D-mannose, D-sucrose and D-raffinose. Negative reactions were recorded for glycerol, erythritol, D-arabinose, L-arabinose, 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, esculin ferric citrate, D-cellobiose, D-melibiose, D-trehalose, D-melezitose, starch, glycogen, xylitol, gentiobiose, D-turanose, D-lyxose, D-tagatose, D-fucose, L-fucose, D-

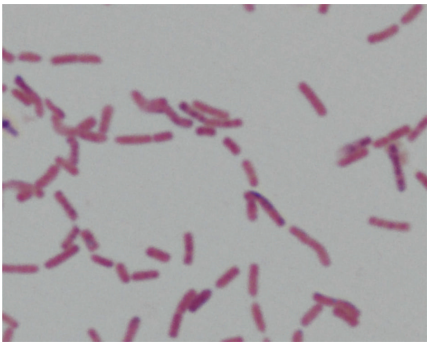


FIG. 4. Gram staining of *Massilibacterium senegalense* strain mt8^T.

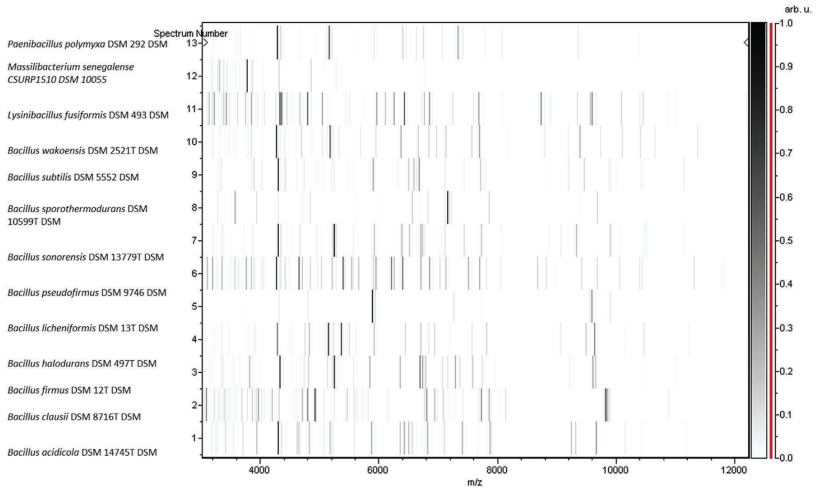


FIG. 3. Gel view comparing *Massilibacterium senegalense* strain mt8^T to other species within the Bacillaceae family. The gel view displays the raw spectra of loaded spectrum files arranged in a pseudo-gel like look. The x-axis records the m/z value. The left y-axis displays the running spectrum number originating from subsequent spectra loading. The peak intensity is expressed by a Gray scale scheme code. The color bar and the right y-axis indicate the link between the color in which a peak is displayed and the peak intensity in arbitrary units. Displayed species are indicated on the left.

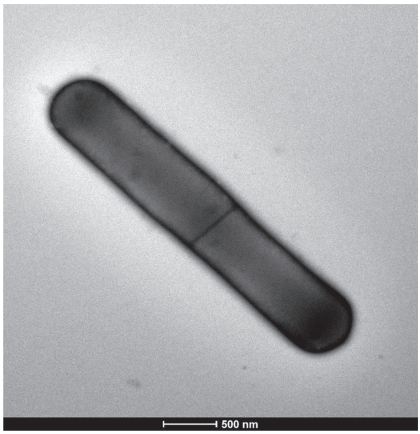


FIG. 5. Transmission electron microscopy of *Massilibacterium senegalense* strain mt8^T using Tecnai G20 transmission electron microscope (FEI Company). Scale bar = 500 nm.

arabitol, L-arabitol, potassium gluconate, potassium 2-ketogluconate and potassium 5-ketogluconate.

Cells were susceptible to imipenem, doxycycline, rifampicin, vancomycin, amoxicillin, ceftriaxone, gentamicin (500 µg), trimethoprim/sulfamethoxazole, erythromycin, ciprofloxacin and gentamicin (15 µg) but were resistant to metronidazole.

Table 2 shows the differences between the biochemical features of close relatives of *Massilibacterium senegalense* strain mt8^T, members of the *Bacillaceae* family.

Genome properties

The genome of *M. senegalense* strain mt8^T genome (accession no. CTRN00000000) was 5 697 950 bp long with a 35.67% G+C content (Fig. 6, Table 3). It was composed of ten scaffolds and 12 contigs. There were 5793 predicted genes, among which 5615 were protein-coding genes and 178 RNAs (14 5S rRNA genes, 16 16S rRNA genes, ten 23S rRNA genes and 138 tRNA genes). A total of 4262 genes were assigned a putative function, and 208 genes were identified as ORFans. The remaining 386 genes were annotated as hypothetical proteins. Using ARG-ANNOT [32], no resistance genes were found. The properties and statistics of the genome are summarized in Table 3, and the gene distribution into COGs functional categories is presented in Table 4.

Genome comparison

The genomic characteristics (size, percentage of G+C content, protein-coding genes and total number of genes) were used to compare strain mt8^T with the genome of closely related species (Table 5). The size of *Massilibacterium senegalense* strain mt8^T (5.69 Mb) is larger than *Bacillus wakoensis* strain N_1, *Lysinibacillus fusiformis* strain DSM 2898^T, *Bacillus halodurans* strain C-125, *Bacillus pseudofirmus* strain OF4, *Anoxybacillus tepidamans* strain PS2 and *Bacillus smithii* strain 7_3_47FAA (5.53, 4.84, 4.2, 3.86, 3.36 and 3.24 Mb respectively). The G+C content of *M. senegalense* (35.6%) is smaller than those of *B. halodurans, A. tepidamans, B. smithii, B. pseudofirmus, B. wakoensis* and *L. fusiformis* (43.7, 43.0, 40.7, 40.3, 38.3 and 37.6%). The gene content of *M. senegalense* (5793) is bigger than the gene content of *L. fusiformis, B. wakoensis, B. halodurans, B. pseudofirmus, A. tepidamans* and *B. smithii* (4764, 4460, 4076, 3841, 3400 and 3235 respectively) (Table 5). There are more protein-coding genes (5615) in the genome of *M. senegalense* than in the genomes of *L. fusiformis, B. wakoensis, B. halodurans, B. pseudofirmus, A. tepidamans,* and *B. smithii* (4548, 3912, 3903, 3704, 3245 and 2832 respectively). The distribution of genes into COGs categories was similar in all compared genomes (Fig. 7, Table 4). *M. senegalense* also shared 1368, 1244, 1263, 1318, 1321 and 1231 orthologous genes with *B. pseudofirmus, L. fusiformis, B. wakoensis, A. tepidamans, B. halodurans,* and *B. smithii* respectively (Table 6). Among species with standing in nomenclature, AGIOS values ranged from 63.43 to 70.67% among compared species except *M. senegalense*. When compared to other species, the AGIOS values ranged from 65.40% with *B. halodurans* to 66.37% with *A. tepidamans* (Table 6). To evaluate the genomic similarity among studied *Bacillaceae* strains, we determined two parameters, dDDH, which exhibits a high correlation with DDH [28,29], and AGIOS [27], which was designed to be independent from DDH (Table 7).

Conclusion

Given the 93% similarity level to *Bacillus halodurans* for the 16S rRNA sequence of strain mt8^T, its MALDI-TOF spectrum and the analysis of its annotated genome, we created a new genus, *Massilibacterium. Massilibacterium senegalense* is the type strain.

Description of *Massilibacterium* gen. nov.

Massilibacterium (from Massilia, Marseille's old Roman and Greek name; Marseille is the city in which the strain was isolated).

Facultative anaerobic rod-shaped bacteria. Gram negative. Optimal growth in aerobic conditions at 37°C. Catalase and oxidase negative. Nitrates were reduced into nitrites. Negative for indole formation. β-Glucosidase positive. Urease negative. The type species is *Massilibacterium senegalense* strain mt8^T. Habitat is human gut.

TABLE 2. Differential characteristics of *Massilibacterium senegalense* strain mt8^T, *Bacillus halodurans* **DSM 497**, *Bacillus acidicola* **DSM 14745**^{T}, *Bacillus wakoensis* **DSM 2521**^{T}, *Bacillus hemicellulosilyticus* **DSM 16731**^{T}, *Bacillus cellulosilyticus* **DSM 2522**^{T}, *Bacillus akibai* **ATCC 43226**^{T}, *Bacillus mannanilyticus* **DSM 16130**^{T}, *Bacillus okuhidensis* **DSM 13666**^{T}, *Bacillus sonorensis* **DSM 13779**^{T} [33–37]

Property	*Massilibacterium senegalense*	*Bacillus halodurans*	*Bacillus acidicola*	*Bacillus wakoensis*	*Bacillus hemicellulosilyticus*	*Bacillus cellulosilyticus*	*Bacillus akibai*	*Bacillus mannanilyticus*	*Bacillus okuhidensis*	*Bacillus sonorensis*
Cell diameter (µm)	1.7–1.9	0.5–0.6	1.0–1.6	0.5–0.8	0.3–0.5	0.6–0.8	0.6–0.8	0.6–0.8	0.5–1.0	1.0
Oxygen requirement	+	+	+	+	+	+	+	+	+	+
Gram stain	–	NA	+	+	+/–	+	+	+/–	+/–	+
Salt requirement	–	+	–	–	–	–	–	–	–	–
Motility	+	NA	–	+	–	+	+	–	+	+
Endospore formation	+	+	+	+	+	+	+	+	+	+
Indole	–	NA	–	–	–	–	–	–	NA	NA
Production of:										
Alkaline phosphatase	–	NA	NA	NA	NA	NA	NA	NA	NA	NA
Catalase	–	NA	+	+	+	+	+	+	+	+
Oxidase	–	NA	–	–	–	–	–	–	+	NA
Nitrate reductase	+	–	–	+	–	+	+	–	+	+
Urease	–	NA	NA	NA	NA	NA	NA	NA	NA	NA
β-Galactosidase	–	NA	NA	NA	NA	NA	NA	NA	NA	NA
N-acetyl-glucosamine	+	+	NA	NA	NA	NA	NA	NA	+	NA
Acid from:										
L-Arabinose	–	+	–	–	–	–	+	–	+	+
Ribose	–	–	+	–	–	–	–	–	NA	NA
Mannose	+	+	+	+	+	+	+	+	–	NA
Mannitol	+	+	+	+	+	–	+	+	+	+
Sucrose	+	+	+	+	+	+	+	+	NA	NA
D-Glucose	–	+	+	+	+	+	+	+	NA	+
D-Fructose	+	+	+	+	+	+	+	+	NA	NA
D-Maltose	+	+	+	+	+	+	+	+	NA	NA
D-Lactose	–	+	+/–	–	+	+	–	–	+	NA
Habitat	Human gut	Soil	Acidic peat bogs	Industry	Industry	Industry	Industry	Industry	Water	Desert

NA, not available.

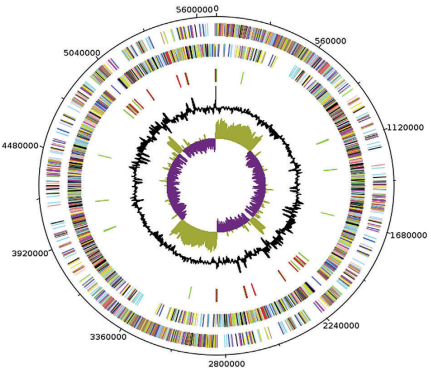


FIG. 6. Graphical circular map of chromosome. From outside to center: genes on forward strain coloured by COGs categories (only gene assigned to COGs), RNA genes (tRNAs green, rRNAs red), G+C content and G+C skew.

TABLE 4. Number of genes associated with 25 general COGs functional categories

Code	Value	% of total[a]	Description
J	298	10.63	Translation
A	0	0.0	RNA processing and modification
K	334	11.91	Transcription
L	406	14.48	Replication, recombination and repair
B	2	0.07	Chromatin structure and dynamics
D	68	2.43	Cell cycle control, mitosis and meiosis
Y	0	0.0	Nuclear structure
V	80	2.85	Defense mechanisms
T	252	8.99	Signal transduction mechanisms
M	224	7.99	Cell wall/membrane biogenesis
N	132	4.71	Cell motility
Z	0	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	98	3.50	Intracellular trafficking and secretion
O	190	6.78	Posttranslational modification, protein turnover, chaperones
C	248	8.84	Energy production and conversion
G	140	4.99	Carbohydrate transport and metabolism
E	370	13.20	Amino acid transport and metabolism
F	138	4.92	Nucleotide transport and metabolism
H	226	8.06	Coenzyme transport and metabolism
I	190	6.78	Lipid transport and metabolism
P	286	10.20	Inorganic ion transport and metabolism
Q	92	3.28	Secondary metabolites biosynthesis, transport and catabolism
R	602	21.47	General function prediction only
S	432	15.41	Function unknown
—	424	7.31	Not in COGs

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

Description of *Massilibacterium senegalense* strain mt8^T gen. nov., sp. nov.

Massilibacterium senegalense (se.ne.gal.e′n.sis. L. gen. masc., meaning originating from Senegal, the country from which the stool sample was collected).

Cells are sporulating, motile and facultative anaerobic, Gram-negative, rod-shaped bacilli with a mean diameter of 1.8 µm and a length of 5.9 µm. Colonies were 5 mm diameter white irregular colonies on 5% sheep's blood–enriched Colombia agar. Catalase and oxidase negative.

Positive reactions were observed for esterase (C4) and acid phosphatase. Nitrate reduction and aesculin hydrolysis were positive. N-acetyl-glucosamine, amygdalin, arbutin, salicin, D-maltose, D-fructose, inulin, D-mannose, D-sucrose and D-raffinose were metabolized. Cells were susceptible to imipenem, doxycycline, rifampicin, vancomycin, amoxicillin, ceftriaxone, gentamicin (500 µg), trimethoprim/sulfamethoxazole, erythromycin, ciprofloxacin and gentamicin (15 µg) but were resistant to metronidazole.

The G+C content of the genome is 35.67%. The 16S rRNA gene sequence and whole genome shotgun sequence of *M. senegalense* strain mt8^T are deposited in GenBank under accession numbers LN828943 and CTRN01000000 respectively.

TABLE 3. Nucleotide content and gene count levels of the genome

Attribute	Genome (total)	
	Value	% of total[a]
Size (bp)	5 697 950	100
G+C content (%)	2 034 168	35.7
Coding region (bp)	4 442 019	77.95
Total genes	5793	100
RNA genes	178	3.07
Protein-coding genes	5615	96.92
Genes with function prediction	4262	73.57
Genes assigned to COGs	3838	66.25
Genes with peptide signals	210	3.62
Genes with transmembrane helices	504	8.70
CRISPR repeats	0	0
ORFan genes	208	3.59
Genes associated with PKS or NRPS	13	0.22
No. of antibiotic resistance genes	0	0

COGs, Clusters of Orthologous Groups database; CRISPR, clustered regularly interspaced short palindromic repeat.
[a]Total is based on either size of genome (bp) or total number of protein-coding genes in annotated genome.

TABLE 5. Genome comparison of species closely related to *Massilibacterium senegalense* strain mt8^T

Organism	INSDC	Size (Mb)	G+C%	Protein-coding genes	Total genes
Massilibacterium senegalense strain mt8^T	CTRN00000000.1	5.69	35.6	5615	5793
Bacillus pseudofirmus strain OF4	CP001878.2	3.86	40.3	3704	3841
Lysinibacillus fusiformis strain DSM 2898^T	CP010820.1	4.84	37.6	4548	4764
Bacillus wakoensis strain N_1	BAUT00000000.1	5.53	38.3	3912	4460
Anoxybacillus tepidamans strain PS2	JHVN00000000.1	3.36	43.0	3245	3400
Bacillus halodurans strain C-125	BA000004.3	4.2	43.7	3903	4076
Bacillus smithii strain *7_3_47FAA*	ACWF00000000.1	3.24	40.7	2832	3235

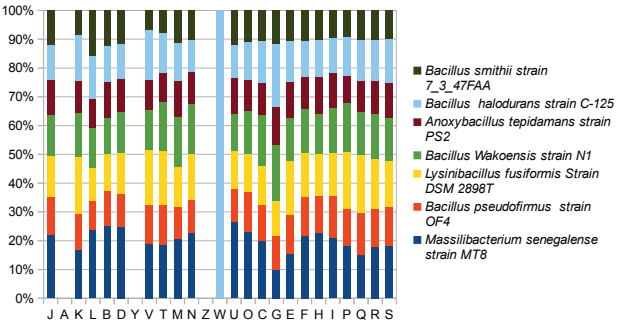


FIG. 7. Distribution of functional classes of predicted genes according to clusters of orthologous groups of protein.

TABLE 6. Number of orthologous proteins shared between genomes (upper right)[a]

	Bacillus pseudofirmus	*Lysinibacillus fusiformis*	*Massilibacterium senegalense*	*Bacillus wakoensis*	*Anoxybacillus tepidamans*	*Bacillus halodurans*	*Bacillus smithii*
B. pseudofirmus	4335	1496	1368	1868	1615	1959	1457
L. fusiformis	64.28	4767	1244	1356	1456	1464	1325
M. senegalense	65.99	65.59	5615	1263	1318	1321	1231
B. wakoensis	70.67	64.46	66.35	4576	1495	1819	1358
A. tepidamans	65.38	64.71	66.37	65.28	3463	1611	1456
B. halodurans	68.48	63.43	65.40	68.03	65.66	4066	1425
B. smithii	65.06	65.14	65.95	64.98	68.03	65.00	3294

[a]Average percentage similarity of nucleotides corresponding to orthologous proteins shared between genomes (lower left) and numbers of proteins per genome.

TABLE 7. Pairwise comparison of *Massilibacterium senegalense* strain mt8^T with other species using GGDC, formula 2 (DDH estimates based on identities/HSP length)[a] (upper right)

	Bacillus pseudofirmus	*Lysinibacillus fusiformis*	*Massilibacterium senegalense*	*Bacillus wakoensis*	*Anoxybacillus tepidamans*	*Bacillus halodurans*	*Bacillus smithii*
B. pseudofirmus	100%	29.1% ± 2.54	26.3% ± 2.55	21.2% ± 2.59	21.9% ± 2.53	27.5% ± 2.56	25.7% ± 2.54
L. fusiformis		100%	29.5% ± 2.54	26.1% ± 2.54	23.2% ± 2.53	30.8% ± 2.54	31.9% ± 2.54
M. senegalense			100%	25% ± 2.54	21% ± 2.54	24.4% ± 2.55	30.4% ± 2.54
B. wakoensis				100%	19.5% ± 2.53	23.3% ± 2.56	28.3% ± 2.53
A. tepidamans					100%	22.2% ± 2.53	21.6% ± 2.55
B. halodurans						100%	25.9% ± 2.54
B. smithii							100%

DDH, DNA-DNA hybridization; GGDC, Genome-to-Genome Distance Calculator; HSP, high-scoring segment pairs.
[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 (Fig. 3) and phylogenomic analyses as well as GGDC results.

The type strain mt8^T (= CSUR P1510 = DSM 100455) was isolated from the stool of a young Senegalese boy with kwashiorkor.

Conflict of Interest

None declared.

Acknowledgements

The authors thank the Xegen Company (http://www.xegen.fr/) for automating the genomic annotation process. This study was funded by the Fondation Méditerranée Infection. We thank K. Griffiths for English-language review and C. Andrieu for administrative assistance.

References

[1] Xu Z, Knight R. Dietary effects on human gut microbiome diversity. Br J Nutr 2015;113:S1–5.
[2] Salazar N, Arboleya S, Valdès L, et al. The human intestinal microbiome at extreme ages of life. Dietary intervention as a way to counteract alterations. Front Genet 2014;5:406.
[3] Sankar SA, Lagier JC, Pontarotti P, Raoult D, Fournier PE. The human gut microbiome, a taxonomic conundrum. Syst Appl Microbiol 2015;38:276–86.
[4] Simpson HL, Campbell PJ. Review article: dietary fibre–microbiota interactions. Aliment Pharmacol Ther 2015;42:158–78.
[5] Lagier JC, Armougom F, Million M, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 2012;18:1185–93.
[6] De Vos P, Garrity GM, Jones D, et al. Bergey's manual of systematic bacteriology. The firmicutes. 2nd ed., Vol. 3. New York: Springer; 2009.
[7] Viale AM, Arakaki AK, Soncini FC, Ferreyra RG. Evolutionary relationships among eubacterial groups as inferred from GroEL (Chaperonin) sequences comparison. Int J Syst Bacteriol 1944;44:527–33.
[8] Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposals for the domains *Archaea, Bacteria*, and *Eucarya*. Proc Natl Acad Sci U S A 1990;87:4576–9.
[9] Wolf M, Muller T, Dandekar T, Pollack JD. Phylogeny of *Firmicutes* with special reference to *Mycoplasma (Mollicutes)* as inferred from phosphoglycerate kinase amino acid sequence data. Int J Syst Evol Microbiol 2004;54:871–5.
[10] Seng P, Drancourt M, Gouriet F, 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.
[11] Seng P, Abat C, Rolain JM, 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.
[12] Hugon P, Mishra AK, Lagier JC, et al. Non-contiguous finished genome sequence and description of *Brevibacillus massiliensis* sp. nov. Stand Genomic Sci 2013;8:1–14.
[13] Kokcha S, Ramasamy D, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Brevibacterium senegalense* sp. nov. Stand Genomic Sci 2012;7:233–45.
[14] Lagier JC, Armougom F, Mishra AK, Nguyen TT, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Alistipes timonensis* sp. nov. Stand Genomic Sci 2012;6:315–24.
[15] Lagier JC, Elkarkouri K, Nguyen TT, Armougom F, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Anaerococcus senegalensis* sp. nov. Stand Genomic Sci 2012;6:116–25.
[16] Lagier JC, Elkarkouri K, Rivet R, Couderc C, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Senegalemassilia anaerobia* gen. nov., sp. nov. Stand Genomic Sci 2013;7:343–56.
[17] Lagier JC, Elkarkouri K, Mishra AK, Robert C, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Enterobacter massiliensis* sp. nov. Stand Genomic Sci 2013;7:399–412.
[18] 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(10):3623–30.
[19] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152–5.
[20] Lagier JC, Edouard S, Pagnier I, Mediannikov O, Drancourt M, Raoult D. Current and past strategies for bacterial culture in clinical microbiology. Clin Microbiol Rev 2015;28:208–36.
[21] Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:119–30.
[22] 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.
[23] 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.
[24] Käll L, Krogh A, Sonnhammer EL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol 2004;338:1027–36.
[25] Gouret P, Thompson JD, Pontarotti P. PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinformatics 2009;10:298.
[26] 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.
[27] Ramasamy D, Mishra AK, Lagier JC, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of new bacterial species. Int J Syst Evol Microbiol 2014;64:384–91.
[28] 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.
[29] 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.
[30] Gouret P, Paganini J, Dainat J, 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.
[31] Gouret P, Vitiello V, Balandraud N, Gilles A, Pontarotti P, Danchin EGJ. FIGENIX: intelligent automation of genomic annotation: expertise integration in a new software platform. BMC Bioinformatics 2005;6:198.
[32] Gupka SK, Padmanabhan BR, Diene SM, et al. ARG-ANNOT, a new bioinformatics tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 2014;58:212–20.

[33] Li Z, Kawamura Y, Shida O, Yamagata S, Deguchi T, Ezaki T. *Bacillus okuhidensis* sp. nov. isolated from the Okuhida spa area of Japan. Int J Syst Evol Microbiol 2002;52:1205–9.
[34] Nielsen P, Fritze D, Priest FG. Phenetic diversity of alkaliphilic *Bacillus* strains: proposal of nine new species. Microbiology 1995;141:1745–61.
[35] Noqi Y, Takami H, Horikoshi K. Characterization of alkaliphilic *Bacillus* strains used in industry: proposal of five novel species. Int J Syst Evol Microbiol 2005;55:2309–15.
[36] Palmisano MM, Nakamura LK, Duncan KE, Istock CA, Cohan FM. *Bacillus sonorensis* sp. nov., a close relative of *Bacillus licheniformis*, isolated from soil in the Sonoran desert, Arizona. Int J Syst Evol Microbiol 2001;51:1671–9.
[37] Albert RA, Archambault J, Rosselò-Mora R, Tindall BJ, Matheny M. *Bacillus acidicola* sp. nov., a novel mesophilic, acidophilic species isolated from acidic sphagnum peat bogs in Wisconsin. Int J Syst Evol Microbiol 2005;55:2125–30.

Article 9: Genome sequence and description of *Desnuesiella massiliensis* gen. nov., sp. nov. a new member of family Clostridiaceae.

Linda Hadjadj, Maryam Tidjani Alou, Cheikh Sokhna,

Jean-Christophe Lagier, Didier Raoult, Jean-Marc Rolain.

Publié dans New Microbes and New Infections

Genome sequence and description of *Desnuesiella massiliensis* gen. nov., sp. nov. a new member of family *Clostridiaceae*

L. Hadjadj[1], M. Tidjani Alou[1], C. Sokhna[2], J.-C. Lagier[1], D. Raoult[1,3] and J.-M. Rolain[1]

1) *Unité de recherche sur les maladies infectieuses et tropicales émergentes (URMITE), UMR CNRS, IHU Méditerranée Infection, Faculté de Médecine et de Pharmacie, Aix-Marseille-Université, Marseille, France,* 2) *Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM63, CNRS7278, IRD198, InsermU1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Aix-Marseille Université, Marseille, France and Dakar, Senegal and* 3) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Desnuesiella massiliensis, strain MT10^T gen. nov., sp. nov. is a newly proposed genus within the family *Clostridiaceae*, isolated from the digestive microbiota of a child suffering from kwashiorkor. *Desnuesiella massiliensis* is a facultatively anaerobic, Gram-positive rod. Here we describe the features of this organism, together with the complete genome sequence and annotation. The 5 503 196-bp long genome (one chromosome but no plasmid) contains 5227 protein-coding and 81 RNA genes, including 14 rRNA genes.

New Microbes and New Infections © 2016 The Authors. Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases.

Keywords: Culturomics, *Desnuesiella massiliensis*, genome, kwashiorkor, taxono-genomics
Original Submission: 4 February 2016; **Revised Submission:** 12 March 2016; **Accepted:** 14 March 2016
Article published online: 19 March 2016

Corresponding author: J.-M. Rolain, Unité de recherche sur les maladies infectieuses et tropicales émergentes (URMITE), UMR CNRS, IHU Méditerranée Infection, Faculté de Médecine et de Pharmacie, Aix-Marseille-Université, Marseille, France
E-mail: jean-marc.rolain@univ-amu.fr

Introduction

Strain MT10^T (= CSUR P1918 = DSM 101500) is the type strain of *Desnuesiella massiliensis* gen. nov., sp. nov. This bacterium, which is proposed to belong to the family *Clostridiaceae*, is a Gram-positive, flagellated, facultative anaerobic bacillus. It was isolated from the stool sample of a 1-year-old boy with kwashiorkor living in Senegal, through a culturomics study of the bacterial diversity of the faeces of children with kwashiorkor disease [1].

The newly proposed strategy of applying high throughput genome sequencing, matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) spectral analysis of cellular proteins, coupled with more traditional methods of phenotypic characterization has been demonstrated as a useful approach for the description of new bacterial taxa [2–5].

The family *Clostridiaceae* [6] belongs to the phylum Firmicutes and includes 40 genera. Members belonging to this family were isolated mainly from the environment and from commensal digestive microbiota of mammals. Some are major human pathogens, including *Clostridium botulinum*, *Clostridium difficile*, *Clostridium tetani* and *Clostridium perfringens* [7].

Here we present a summary classification and a set of features for *D. massiliensis* gen. nov., sp. nov. strain MT10^T together with the description of the complete genome sequence and annotation. These characteristics support the circumscription of a novel genus, *Desnuesiella* gen. nov. within the family *Clostridiaceae*, with *Desnuesiella massiliensis* gen. nov., sp. nov. as the type species.

Organism Information

In March 2015, a faecal sample was collected from a 1-year-old boy living in Dakar, Senegal who had kwashiorkor, which is a form of severe malnutrition. Consent was obtained from the

http://dx.doi.org/10.1016/j.nmni.2016.03.003

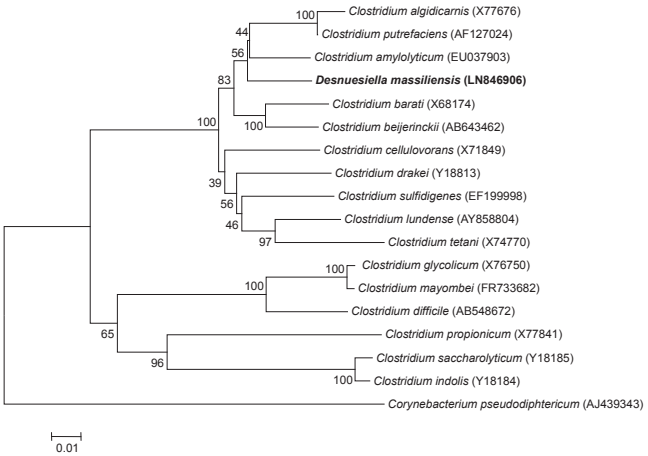


FIG. 1. Phylogenetic tree highlighting the position of *Desnuesiella massiliensis* strain MT10^T relative to other type strains within the *Clostridiaceae* family. GenBank accession numbers are indicated in parentheses. Sequences were aligned using CLUSTAL X, and phylogenetic interferences were obtained using the maximum-likelihood method within the MEGA 5 software. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 1000 times to generate a majority consensus tree. *Corynebacterium pseudodiphtericum* was used as out-group. The scale bar represents a 1% nucleotide sequence divergence.

child's parents. The study was approved by the Institut Fédératif de Recherche 48 under agreement 09-022. The boy had not received antibiotics at the time of sample collection. The faecal specimen was preserved at −80°C after collection and sent to Marseille. Strain MT10^T was isolated in liquid Columbia broth after 21 days of aerobic incubation at 37°C using the 'culturomics' concept [1]. After a failed identification using MALDI-TOF mass spectrometry, the 16s rRNA was sequenced.

When BLAST was performed with to NCBI database, the 16S rRNA gene sequence of *D. massiliensis* strain MT10^T (GenBank Accession number LN846906) exhibited an identity of 94.60% with *Clostridium amylolyticum* (Fig. 1). This value was the highest similarity observed, but was lower than the 95% 16s rRNA gene sequence threshold recommended by Stackebrandt and Ebers to delineate a new genus without carrying out DNA–DNA hybridization [8] and by Tindall *et al.* [9].

Different growth temperatures (25, 30, 37, 45°C) were tested. Growth occurred between 25°C and 37°C, but optimal growth was observed in anaerobic conditions at 37°C, 24 h after inoculation. Colonies were smooth, opaque and approximately 1 mm in diameter on 5% sheep blood-enriched agar (BioMérieux, Marcy l'Étoile, France). Growth of the strain was tested in anaerobic and microaerophilic atmospheres using GasPak EZ Anaerobe Pouch (Becton Dickinson Co., Franklin Lakes, NJ, USA) and CampyGen Compact (Oxoid, Basingstoke, UK) systems, respectively, and in aerobic atmosphere, with or without 5% CO_2. Growth was observed under aerobic (with and without CO_2), microaerophilic and anaerobic conditions. Gram staining showed Gram-positive rods unable to form spores (Fig. 2a). A motility test produced a positive result. Cells grown on agar did not sporulate and the rods exhibited monotrichous flagella. The size of cells were determined by negative staining transmission electron microscopy on a Technai G^{20} Cryo (FEI, Hillsboro, OR, USA) at an operating voltage of 200 kV, the rods have a length ranging from 2.2 to 2.6 µm (mean 2.4 µm) and a width ranging from 0.4 to 0.5 µm (mean 0.43 m) (Fig. 2b).

Desnuesiella massiliensis is catalase-, oxidase-, urease- and indole-negative but alkaline-phosphatase-positive. Fermentation of sucrose, D-ribose, D-glucose, D-lactose, D-mannose, D-maltose, D-fructose were positive but not that of L-arabinose, D-sorbitol and D-xylose. Differential phenotypic characteristics using API 50CH and API Zym system (BioMérieux) between *D. massiliensis* gen. nov., sp. nov. strain MT10^T and others species from the family *Clostridiaceae* [10–14] are detailed in Table 1.

Susceptibility testing was performed by E-test method (BioMérieux) and MIC was expressed in mg/L. *Desnuesiella massiliensis* was susceptible to amoxicillin (0.032), ceftriaxone (1), imipenem (0.016), erythromycin (0.25), doxycycline (0.032)

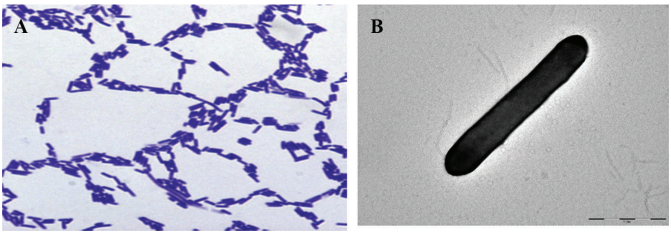


FIG. 2. (a) Gram staining of *Desnuesiella massiliensis* strain MT10^T. (b) Transmission electron microscopy of *D. massiliensis* strain MT10^T using a Technai G^{20} Cryo (FEI) at an operating voltage of 200 kV. The scale bar represents 1 μm.

and rifampicin (0.125) but resistant to colistin (>256), gentamicin (>256) and metronidazole (>256).

Extended Features Descriptions

MALDI-TOF mass spectrometry protein analysis was performed as previously described [15] using a Microflex spectrometer (Bruker Daltonics, Leipzig, Germany). Twelve distinct deposits were made for strain MT10^T from 12 isolated colonies. The 12 MT10^T spectra were imported into the MALDI BioTyper software (version 2.0, Bruker) and analysed by standard pattern matching (with default parameter settings) against the main spectra of 4108 bacteria, including 120 spectra from the family *Clostridiaceae*, used as reference data, in the BioTyper database. A score enabled the identification, or not, from the tested species: a score >2 with a validated species enabled identification at species level, a score >1.7 but <2 enabled identification at genus level; and a score <1.7 did not enable any identification. No significant

TABLE 1. **Differential phenotypic characteristics between *Desnuesiella massiliensis* gen. nov., sp. nov., strain MT10^T and phylogenetically close *Clostridiaceae* species**

Characteristic	1	2	3	4	5	6
Cell diameter (μm)	2.4	2-7.5	3-4	2-5	1.7	0.5
Oxygen requirement	AAF	AS	AS	na	AS	AS
Gram stain	+	+	–	+	v	+
Motility	+	+	+	–	+	+
Endospore formation	–	–	+	na	+	+
Production of:						
Alkaline phosphatase	+	na	na	na	na	na
Catalase	–	–	na	na	–	na
Oxidase	–	–	na	na	na	na
Urease	–	na	na	–	–	na
β-galactosidase	w	na	na	na	na	na
Indole	–	–	–	+	na	na
Leucine arylamidase	w	na	na	+	na	+
Cystine arylamidase	–	na	na	na	na	na
Valine arylamidase	–	na	na	na	na	+
Utilization of:						
Mannitol	–	+	+	–	+	na
Threhalose	+	+	–	–	na	na
Sucrose	+	+	+	–	+	+
L-arabinose	–	–	+	–	+	+
D-sorbitol	–	–	–	–	na	na
D-xylose	–	–	+	+	na	+
D-ribose	+	+	+	+	–	na
D-glucose	+	+	+	+	+	+
D-lactose	+	+	–	–	+	+
D-mannose	+	+	+	+	+	+
D-maltose	+	+	–	–	+	+
D-fructose	+	+	+	+	+	+
Glycerol	+	–	+	–	na	na
N-Acetylglucosamine	+	na	na	+	na	na
G+C content (mol%)	32.1	33.1	32	na	28	32.3
Habitat	Human gut	Anaerobic reactor	Sediment	Pork meat	Human gut	Sediment

Strains: (1) *Desnuesiella massiliensis*; (2) *Clostridium amylolyticum*; (3) *Clostridium drakei*; (4) *Clostridium algidicarnis*; (5) *Clostridium beijerinckii*; (6) *Clostridium sulfidigenes*.
+, positive result; –, negative result; w, weak positive result; na, data not available; AS, strictly anaerobic; AAF, facultatively anaerobic.

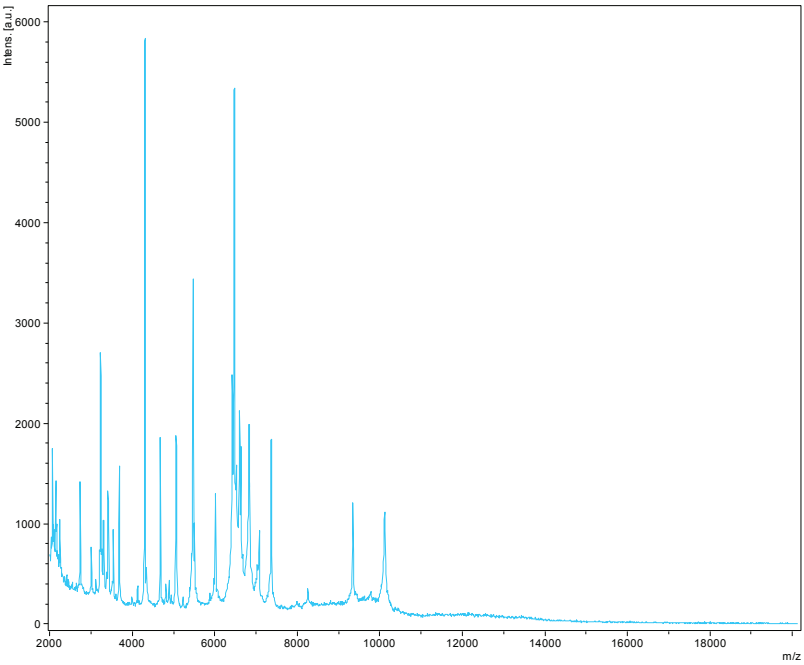


FIG. 3. Reference mass spectrum from *Desnuesiella massiliensis* strain MT10^T. Spectra from 12 individual colonies were compared and a reference spectrum was generated.

MALDI-TOF score was obtained for strain MT10^T against the Bruker database, suggesting that our isolate was a new species. We incremented our database with the spectrum from strain MT10^T (Fig. 3).

Genome Sequencing Information

The organism was selected for sequencing on the basis of its phylogenetic position and 16S rRNA similarity to other members of the family *Clostridiaceae*. It was the first genome of *Desnuesiella massiliensis* gen. nov., sp. nov. (CYSK00000000).

After DNA extraction by the phenol–chloroform method, genomic DNA of *Desnuesiella massiliensis* was sequenced using MiSeq Technology (Illumina Inc., San Diego, CA, USA) with the mate pair strategy.

For genome annotation, open reading frames (ORFs) were predicted using PRODIGAL (http://prodigal.ornl.gov) with default parameters but the predicted ORFs were excluded if they spanned a sequencing gap. The predicted bacterial protein sequences were searched against the Clusters of Orthologous Groups (COG) databases and the GenBank database [16] using BLASTP. The tRNAScanSE tool [17] was used to find tRNA genes whereas ribosomal RNAs were found by using RNAmmer [18] and BLASTn against the GenBank database. Transmembrane helices and lipoprotein signal peptides were predicted using the Phobius web server [19]. ORFans were identified if their BLASTP *E*-value was lower than 1e-03 for alignment length >80 amino acids. If alignment lengths were <80 amino acids, we used an *E*-value of 1e-05.

The genome is 5 503 196 bp long with 32.09% GC content (Fig. 4 and Table 2). It is composed of 14 scaffolds (composed of 16 contigs). Of the 5308 predicted genes, 5227 were protein-coding genes, and 81 were RNAs (eight genes are 5S rRNA, five genes are 16S rRNA, one gene is 23S rRNA, 67 genes are tRNA genes). A total of 3890 genes (74.42%) were assigned as putative function (by COGs or by NR blast). In all, 276 genes were identified as ORFans (5.28%). The remaining genes were annotated as

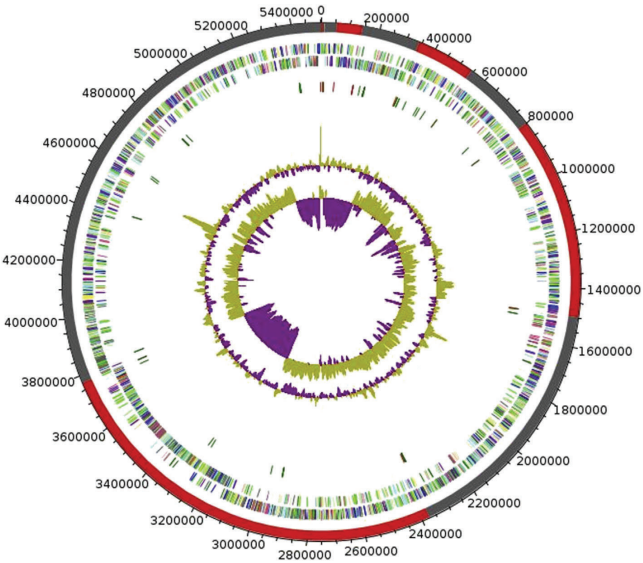


FIG. 4. Graphical circular map of the chromosome. From outside to the centre: genes on forward strand (coloured by COG categories), genes on reverse strand (coloured by COG categories), RNA genes (tRNAs green, rRNAs red), GC content, GC skew.

hypothetical proteins (843 genes, 16.13%). The distribution of genes into COGs functional categories is presented in Table 3.

The draft genome sequence of *D. massiliensis* is smaller than those of *Clostridium drakei* and *Clostridium beijerinckii* (5.50, 5.64 and 6.00 Mb, respectively), but larger than those of *Clostridium algidicarnis* and *Clostridium sulfidigenes* (3.06 and 3.72 Mb, respectively).

TABLE 2. Nucleotide content and gene count levels of the genome

Attribute	Value	% of total [a]
Genome size (bp)	5 503 196	100
DNA coding (bp)	4 723 283	85.8
DNA G+C (bp)	1 765 744	32.1
DNA scaffolds	14	—
Total genes	5308	100
Protein-coding genes	5227	98.5
RNA genes	81	1.5
Pseudo genes	13	—
Genes in internal clusters	2212	—
Genes with function prediction	3890	74.4
Genes assigned to COGs	3873	74.1
Genes with Pfam domains	4814	90
Genes with signal peptides	489	9.3
Genes with transmembrane helices	1309	25

[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.

TABLE 3. Number of genes associated with the 25 general COG functional categories

Code	Value	% of total[a]	Description
A	0	0	RNA processing and modification
B	1	0.02	Chromatin structure and dynamics
C	232	4.44	Energy production and conversion
D	60	1.15	Cell cycle control, mitosis and meiosis
E	286	5.47	Amino acid transport and metabolism
F	115	2.2	Nucleotide transport and metabolism
G	379	7.25	Carbohydrate transport and metabolism
H	154	2.95	Coenzyme transport and metabolism
I	122	2.33	Lipid transport and metabolism
J	321	6.14	Translation, ribosomal structure and biogenesis
K	475	9.1	Transcription
L	182	3.5	Replication, recombination and repair
M	243	4.65	Cell wall/membrane biogenesis
N	87	1.66	Cell motility
O	182	3.48	Post-translational modification, protein turnover, chaperones
P	183	3.5	Inorganic ion transport and metabolism
Q	54	1.03	Secondary metabolites biosynthesis, transport and catabolism
R	436	8.34	General function prediction only
S	218	4.17	Function unknown
T	294	5.62	Signal transduction mechanisms
U	47	0.9	Intracellular trafficking and secretion
V	188	3.6	Defence mechanisms
W	14	0.27	Extracellular structures
Y	0	0	Nuclear structure
Z	2	0.04	Cytoskeleton
—	1354	25.9	Not in COGs

[a]The total is based on the total number of protein-coding genes in the annotated genome

TABLE 4. **Genomic comparison of *Desnuesiella massiliensis* with five others members of the family *Clostridiaceae*[a]**

	C. drakei	*C. algidicarnis*	*C. baratii*	*C. beijerinckii*	*C. sulfidigenes*	*D. massiliensis*
Clostridium drakei	**5748**	747	818	911	761	910
Clostridium algidicarnis	61.97	**2787**	1062	1113	1114	1358
Clostridium baratii	57.32	62.01	**2839**	1288	1038	1229
Clostridium beijerinckii	62.06	69.21	63.51	**5020**	1142	1428
Clostridium sulfidigenes	62.25	69.86	61.57	68.84	**3148**	1313
Desnuesiella massiliensis	61.78	72.16	61.60	68.69	69.39	**5227**

[a]Numbers of orthologous proteins shared between genomes (above diagonal), AGIOS values (below diagonal) and numbers of proteins per genome (bold numbers).

The G+C content of *D. massiliensis* is smaller than those of *C. drakei* (32.09, 35.02%, respectively), but larger than those of *C. algidicarnis*, *C. beijerinckii* and *C. sulfidigenes* (30.26%, 29.86% and 30.00%, respectively).

The gene content of *D. massiliensis* is smaller than that of *C. drakei* (5227, 5748, respectively), but larger than those of *C. algidicarnis*, *C. beijerinckii* and *C. sulfidigenes* (2787, 5020 and 3148, respectively).

Table 4 summarizes the number of orthologous genes and the average percentage of nucleotide sequence identity between the different genomes studied. The nucleotide sequence identity of orthologous genes ranges from 57.3% to 69.9% among previously published genomes.

Conclusions

On the basis of phenotypic, phylogenetic and genomic analysis (taxonogenomics), we formally propose the creation of *Desnuesiella massiliensis* gen. nov., sp. nov., which contains the strain MT10^T. This bacterium has been isolated from the digestive flora of a child living in Senegal suffering from kwashiorkor.

Taxonomic and Nomenclatural Proposals

Description of *Desnuesiella massiliensis* gen. nov., sp. nov.

Desnuesiella (Des.nue.siel'la. ML. dim. suffix tella; M.L. fem. n. Desnuesiella named after the French bacteriologist Christelle Desnues, Aix-Marseille University, Marseille, France), *massiliensis* (mas.si.li.en'sis. L. gen. fem. n. massiliensis of *Massilia*, the Roman name of Marseille, France, where the type strain was isolated).

Desnuesiella massiliensis are Gram-positive rods, flagellated, motile, facultative anaerobic, mesophilic. Optimal growth is achieved at 37°C. Colonies are moderately opaque and approximately 1 mm in diameter on 5% sheep blood-enriched agar. Cells have a mean length of 2.4 µm and a mean width of 0.43 µm.

Desnuesiella massiliensis is catalase-, oxidase-, urease- and indole-negative but alkaline-phosphatase-positive. A positive reaction was obtained for the fermentation of sucrose, D-ribose, D-glucose, D-lactose, D-mannose, D-maltose, D-fructose, but not for L-arabinose, D-sorbitol and D-xylose.

Desnuesiella massiliensis was susceptible to amoxicillin, ceftriaxone, imipenem, erythromycin, doxycycline and rifampicin but resistant to colistin, gentamicin and metronidazole.

The G+C content of the genome is 32.09%. The 16S rRNA and genome sequences are deposited in GenBank under accession numbers LN846906 and CYSK00000000, respectively. The type strain MT10^T (= CSUR P1918 = DSM 101500) was isolated from the stool of a child living in Dakar, Senegal with kwashiorkor disease.

Acknowledgements

The authors thank Xegen company for automating the genome annotation process.

Transparency Declaration

The authors declare that they have no competing interests.

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] 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;64:384–91.
[3] Bendjama E, Loucif L, Diene SM, Michelle C, Gacemi-Kirane D, Rolain JM. Non-contiguous finished genome sequence and description of "*Bacillus massilioalgeriensis*" sp. nov. Stand Genomic Sci 2014;9:1046–61.
[4] Hadjadj L, Bentorki AA, Michelle C, Amoura K, Djahoudi A, Rolain JM. Genome sequence and description of "*Mannheimia massiliguelmaensis*" sp. nov. New Microbes New Infect 2015;8:131–6.

[5] Keita MB, Padhmanabhan R, Caputo A, Robert C, Delaporte E, Raoult D, et al. Non-contiguous finished genome sequence and description of "*Gorillibacterium massiliense*" gen. nov, sp. nov., a new member of the family *Paenibacillaceae*. Stand Genomic Sci 2014;9:807–20.
[6] Skerman VBD, McGowan, Sneath PHA. Approved lists of bacterial names. Int J Syst Bacteriol 1980;30:225–420.
[7] Wells CL, Wilkins TD. *Clostridia*: spore forming anaerobic bacilli. In: Baron S, et al., editors. Baron's Medical Microbiology. 4th ed. University of Texas Medical Branch; 1996.
[8] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2015;33:152–5.
[9] Tindall BJ, Rossello-Mora R, Busse HJ, Ludwig W, Kampfer P. Notes on the characterization of prokaryote strains for taxonomic purposes. Int J Syst Evol Microbiol 2010;60:249–66.
[10] Lawson P, Dainty RH, Kristiansen N, Berg J, Collins MD. Characterization of a psychrotrophic *Clostridium* causing spoilage in vacuum-packed cooked pork: description of *Clostridium algidicarnis* sp. nov. Lett Appl Microbiol 1994;19:153–7.
[11] Liou JS, Balkwill DL, Drake GR, Tanner RS. *Clostridium carboxidivorans* sp. nov., a solvent-producing clostridium isolated from an agricultural settling lagoon, and reclassification of the acetogen *Clostridium scatologenes* strain SL1 as *Clostridium drakei* sp. nov. Int J Syst Evol Microbiol 2005;55:2085–91.
[12] Sallam A, Steinbuchel A. *Clostridium sulfidigenes* sp. nov., a mesophilic, proteolytic, thiosulfate- and sulfur-reducing bacterium isolated from pond sediment. Int J Syst Evol Microbiol 2009;59:1661–5.
[13] Song L, Dong X. *Clostridium amylolyticum* sp. nov., isolated from H2-producing UASB granules. Int J Syst Evol Microbiol 2008;58:2132–5.
[14] Suresh K, Prakash D, Rastogi N, Jain RK. *Clostridium nitrophenolicum* sp. nov., a novel anaerobic p-nitrophenol-degrading bacterium, isolated from a subsurface soil sample. Int J Syst Evol Microbiol 2007;57:1886–90.
[15] Seng P, Drancourt M, Gouriet F, La SB, 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] Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2012;40(Database issue):D48–53.
[17] 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.
[18] 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;35:3100–18.
[19] Kall L, Krogh A, Sonnhammer EL. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. Nucleic Acids Res 2007;35(Web Server issue):W429–32.

Article 10: *Numidum massiliense* gen. nov., sp. nov., a new member of the *Bacillaceae family* isolated from the human gut.

Maryam Tidjani Alou, Thi-Thien Nguyen, Nicholas Armstrong, Jaishriram Rathored, Saber Khelaifia, Didier Raoult, Pierre-Edouard Fournier, Jean-Christophe Lagier.

Publié dans New Microbes and New Infections

Numidum massiliense gen. nov., sp. nov., a new member of the *Bacillaceae* family isolated from the human gut

M. Tidjani Alou[1], T.-T. Nguyen[1], N. Armstrong[1], J. Rathored[1], S. Khelaifia[1], D. Raoult[1,2], P.-E. Fournier[1] and J.-C. Lagier[1]

1) *Aix-Marseille Université, URMITE, UM63, CNRS7278, IRD198, Inserm 1095, Faculté de médecine, Marseille, France and* 2) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Numidum massiliense gen. nov., sp. nov., strain mt3^T is the type strain of *Numidum* gen. nov., a new genus within the family *Bacillaceae*. This strain was isolated from the faecal flora of a Tuareg boy from Algeria. We describe this Gram-positive facultative anaerobic rod and provide its complete annotated genome sequence according to the taxonogenomics concept. Its genome is 3 755 739 bp long and contains 3453 protein-coding genes and 64 RNA genes, including eight rRNA genes.

Keywords: Bacillaceae, Culturomics, gut microbiota, *Numidum massiliense* genome, taxonogenomics
Original Submission: 15 April 2016; **Revised Submission:** 10 May 2016; **Accepted:** 12 May 2016
Article published online: 17 May 2016

Corresponding author: J.-C. Lagier, Aix-Marseille Université, URMITE, UM63, CNRS7278, IRD198, Inserm 1095, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille Cedex 05, France
E-mail: jclagier@yahoo.fr

Introduction

Several microbial ecosystems are harboured by the human body, among which is the human gut microbiota. This particular ecosystem is so vast that its cell count (10^{14} cells) is evaluated at ten times the number of human cells in the human body, and its collective bacterial genome size is 150 times larger than the human genome [1–4]. Over the years, with the evolution of exploratory techniques of microbial ecosystems from culture to metagenomics, the gut microbiota has been shown to be involved in many conditions such as obesity, inflammatory bowel disease and irritable bowel disease [1]. It has also been shown to play key roles in digestion as well as metabolic and immunologic functions [1–3]. A better knowledge of the gut microbiota's composition is thus required for an improved understanding of its functions.

In order to extend the gut microbiota repertoire and bypass the noncultivable bacteria issue, the culturomics concept was developed in order to cultivate as exhaustively as possible the viable population of a bacterial ecosystem; it consists in the multiplication of culture conditions, as well as varying of media, temperature and atmosphere [5]. Using this technique, strain mt3^T was isolated and identified as a previously unknown member of the *Bacillaceae* family. Currently there are 53 validated genera in the *Bacillaceae* family. This family was created by Fisher in 1895 (http://www.bacterio.net/*Bacillaceae*.html). The genus *Bacillus* was described as its type genus. The genera that belong to this family are rod shaped, mostly aerobic and facultative anaerobic bacteria. They are found in various ecosystems like the human body, soil, water, air and other environmental ecosystems [6].

Bacterial classification is currently based on a polyphasic approach with phenotypic and genotypic characteristics such as DNA-DNA hybridization, G+C content and 16S rRNA sequence similarity [7–9]. Nevertheless, this classification system has its limits, among which is the high cost of the DNA-DNA hybridization technique and its low reproducibility [7,10]. With the recent development of genome sequencing technology [11], a new concept of bacterial description was developed in our laboratory [12–16]. This taxonogenomics

http://dx.doi.org/10.1016/j.nmni.2016.05.009

concept [17] combines a proteomic description with the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) profile [18] associated with a phenotypic description and the sequencing, annotation and comparison of the complete genome of the new bacterial species [19].

We describe strain mt3^T, a new genus *Numidum massiliense* gen. nov., sp. nov. (= CSUR P1305 = DSM 29571), a new member of the *Bacillaceae* family using the concept of taxonogenomics.

Materials and Methods

Organism information

A stool sample was collected from a healthy Tuareg boy living in Algeria. Verbal consent was obtained from the patient, and the study was approved by the Institut Fédératif de Recherche 48, Faculty of Medicine, Marseille, France, under agreement 09-022.

Strain identification by MALDI-TOF MS and 16S rRNA sequencing

The sample was cultured using the 18 culture conditions of culturomics [20]. The colonies were obtained by seeding on solid medium, purified by subculture and identified using MALDI-TOF MS [18,21]. Colonies were deposited in duplicate on a MTP 96 MALDI-TOF MS target plate (Bruker Daltonics, Leipzig, Germany), which was analysed with a Microflex spectrometer (Bruker). The 12 spectra obtained were matched against the references of the 7567 bacteria contained in the database by standard pattern matching (with default parameter settings), with MALDI BioTyper database software 2.0 (Bruker). An identification score over 1.9 with a validated species allows identification at the species level, and a score under 1.7 does not enable any identification. When identification by MALDI-TOF MS failed, the 16S rRNA was sequenced [22]. Stackebrandt and Ebers [23] suggest similarity levels of 98.7% and 95% of the 16s rRNA sequence as a threshold to define, respectively, a new species and a new genus without performing DNA-DNA hybridization.

Growth conditions

In order to determine our strain's ideal growth conditions, different temperatures (25, 28, 37, 45 and 56°C) and atmospheres (aerobic, microerophilic and anaerobic) were tested. GENbag anaer and GENbag miroaer systems (bioMérieux, Marcy l'Étoile, France) were used to respectively test anaerobic and microaerophilic growth. Aerobic growth was achieved with and without 5% CO_2.

Morphologic, biochemical and antibiotic susceptibility testing

Gram staining, motility, catalase, oxidase and sporulation were tested as previously described [20]. Biochemical description was performed using API 20 NE, ZYM and 50CH (bioMérieux) according to the manufacturer's instructions. Cellular fatty acid methyl ester (FAME) analysis was performed by gas chromatography/mass spectrometry (GC/MS). Two samples were prepared with approximately 70 mg of bacterial biomass per tube collected from several culture plates. FAMEs were prepared as previously described (http://www.midi-inc.com/pdf/MIS_Technote_101.pdf). GC/MS analyses were carried out as described before [24]. Briefly, fatty acid methyl esters 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, MD, USA) and the FAMEs mass spectral database (Wiley, Chichester, UK).

Antibiotic susceptibility testing was performed using the disk diffusion method according to European Committee on Antimicrobial Susceptibility Testing (EUCAST) 2015 recommendations (http://www.eucast.org/). To perform the negative staining of strain mt3^T, detection Formvar-coated grids were deposited on a 40 µL bacterial suspension drop, then incubated at 37°C for 30 minutes and on ammonium molybdate 1% for 10 seconds. The dried grids on blotted paper were observed with a Tecnai G20 transmission electron microscope (FEI Company, Limeil-Brevannes, France).

Growth conditions and genomic DNA preparation

N. massiliense strain mt3^T (= CSUR P1305 = DSM 29571) was grown on 5% sheep's blood–enriched Columbia agar (bioMérieux) at 37°C in aerobic atmosphere. Bacteria grown on three petri dishes were collected and resuspended in 4 × 100 µL of Tris-EDTA (TE) buffer. Then 200 µL of this suspension was diluted in 1 mL TE buffer for lysis treatment that included a 30- minute incubation with 2.5 µg/µL lysozyme at 37°C, followed by an overnight incubation with 20 µg/µL proteinase K at 37°C. Extracted DNA was then purified using three successive phenol–chloroform extractions and ethanol precipitations at −20°C overnight. After centrifugation, the DNA was resuspended in 160 µL TE buffer.

Genome sequencing and assembly

Genomic DNA of *N. massiliense* was sequenced on the MiSeq Technology (Illumina, 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). gDNA was quantified by a Qubit assay with a high

sensitivity kit (Life Technologies, Carlsbad, CA, USA) to 66.2 ng/μL. The mate pair library was prepared with 1 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 labchip. The DNA fragments ranged in size from 1 to 11 kb, with an optimal size at 3.927 kb. No size selection, was performed and 505 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal at 597 bp on a Covaris device S2 in microtubes (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 59.2 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. An automated cluster generation and sequencing run were performed in a single 39-hour run in a 2 × 251 bp read length.

Genome annotation and comparison

Open reading frames (ORFs) were predicted using Prodigal [25] with default parameters, but the predicted ORFs were excluded if they were spanning a sequencing gap region (contains N). The predicted bacterial protein sequences were searched against the Clusters of Orthologous Groups (COGs) using BLASTP (*E* value 1e-03, coverage 70%, identity percent 30%). If no hit was found, it searched against the NR database using BLASTP with an *E* value of 1e-03 coverage 70% and identity percent of 30%. If sequence lengths were smaller than 80 amino acids, we used an *E* value of 1e-05. The tRNAScanSE tool [26] was used to find tRNA genes, whereas rRNAs were found by using RNAmmer [27]. Lipoprotein signal peptides and the number of transmembrane helices were predicted using Phobius [28]. ORFans were identified if all the performed BLASTP procedures did not give positive results (*E* value smaller than 1e-03 for ORFs with sequence size upper than 80 aa or *E* value smaller than 1e-05 for ORFs with sequence length smaller 80 aa). Such parameter thresholds have already been used in previous works to define ORFans.

Genomes were automatically retrieved from the 16s RNA tree using Xegen software (Phylopattern [29]). For each selected genome, complete genome sequence, proteome and ORFeome genome sequence were retrieved from the National Center for Biotechnology Information FTP site. All proteomes were analysed with proteinOrtho [30]. Then for each couple of genomes, a similarity score was computed. This score is the mean value of nucleotide similarity between all couples of orthologues between the two genomes studied (AGIOS) [19]. 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). To evaluate the genomic similarity among the compared strains, we determined two parameters: digital DNA-DNA hybridization (dDDH), which exhibits a high correlation with DNA-DNA hybridization (DDH) [31,32], and AGIOS [19], which was designed to be independent from DDH.

Results

Strain identification and phylogenetic analyses

Strain mt3^T (Table 1) was first isolated in April 2014 by a preincubation of 21 days in brain–heart infusion supplemented with 5% sheep's blood and cultivated on 5% sheep's blood–enriched Colombia agar (bioMérieux) in an aerobic atmosphere at 37°C.

No significant score was obtained for strain mt3^T using MALDI-TOF MS, thus suggesting that our isolate's spectrum did not match any spectra in our database. The nucleotide sequence of the 16S r RNA of strain mt3^T (GenBank accession no. LK985385) showed a 90.5% similarity level with *Bacillus firmus*, the phylogenetically closest species with a validly published name (Fig. 1), therefore defining it as a new genus within the *Bacillaceae* family named *Numidum massiliense* (= CSUR P1305 = DSM29571). *N. massiliense* spectra (Fig. 2) were added as reference spectra to our database. The reference spectrum for *N. massiliense* was then compared to the spectra of phylogenetically close species, and the differences were exhibited in a gel view (Fig. 3).

Phenotypic description

Growth was observed from 25 to 56°C on blood-enriched Columbia agar (bioMérieux), with optimal growth being

TABLE 1. Classification and general features of *Numidum massiliense* strain mt3^T

Property	Term
Current classification	Domain: *Bacteria* Phylum: *Firmicutes* Class: *Bacilli* Order: *Bacillales* Family: *Bacillaceae* Genus: *Numidum* Species: *Numidum massiliense* Type strain: mt3
Gram stain	Positive
Cell shape	Rod
Motility	Nonmotile
Sporulation	Sporulating
Temperature range	Mesophilic
Optimum temperature	37°C

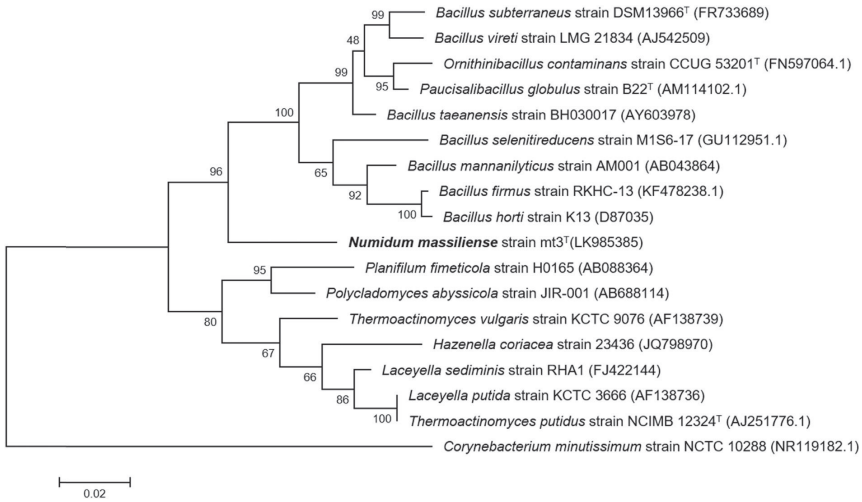


FIG. 1. Phylogenetic tree highlighting position of *Numidum massiliense* gen. nov., sp. nov. strain mt3^T (= CSUR P1305 = DSM 29571) relative to other strains within family *Bacillaceae*. Scale bar represents 1% nucleotide sequence divergence.

obtained aerobically at 37°C after 48 hours of incubation. Weak cell growth was observed under microaerophilic and anaerobic conditions. The cells were nonmotile and sporulating. Cells were Gram-positive rods (Fig. 4) and formed greyish colonies with a mean diameter of 10 mm on blood-enriched Columbia agar. Under electron microscopy, the bacteria had a mean diameter of 0.5 µm and length of 2.7 µm (Fig. 5).

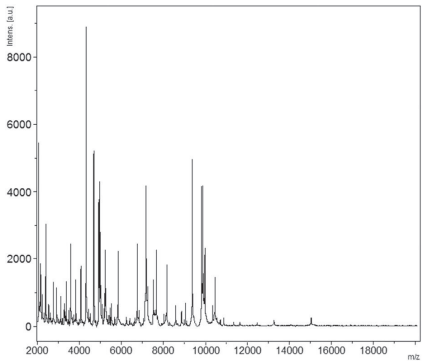


FIG. 2. Reference mass spectrum from *Numidum massiliense* strain mt3^T. Spectra from 12 individual colonies were compared and reference spectrum was generated.

The major fatty acid by far is the branched 13-methyl-tetradecanoic acid (88%). Other fatty acids are described with low abundances (below 6%). The majority of them were branched fatty acids (Table 2).

Strain mt3^T was positive for catalase and negative for oxidase. Alkaline phosphatase, esterase (C4), esterase lipase (C8), leucine arylamidase, trypsin, α-chymotrypsin, acid phosphatase, β-galactosidase, β-glucuronidase, α-glucosidase, β-glucosidase, protease and N-acetyl-β-glucosaminidase activities were exhibited. Nitrates were reduced into nitrites. D-Ribose, D-xylose, D-mannose, D-galactose, D-fructose, D-glucose, D-mannitol, N-acetylglucosamine, amygdalin, esculin ferric citrate, D-maltose, D-lactose, D-trehalose and D-tagatose and adipic acid were metabolized.

Cells were susceptible to doxycycline, ceftriaxone, gentamicin 500 µg, ticarcillin/clavulanic acid, rifampicin, teicoplanin, metronidazole and imipenem. Resistance was exhibited against erythromycin, colistin/polymyxin B, ciprofloxacin, penicillin, trimethoprim/sulfamethoxazole, nitrofurantoin and gentamicin 15 µg.

The biochemical and phenotypic features of strain mt3^T were compared to the corresponding features of other close representatives of the *Bacillaceae* family (Table 3).

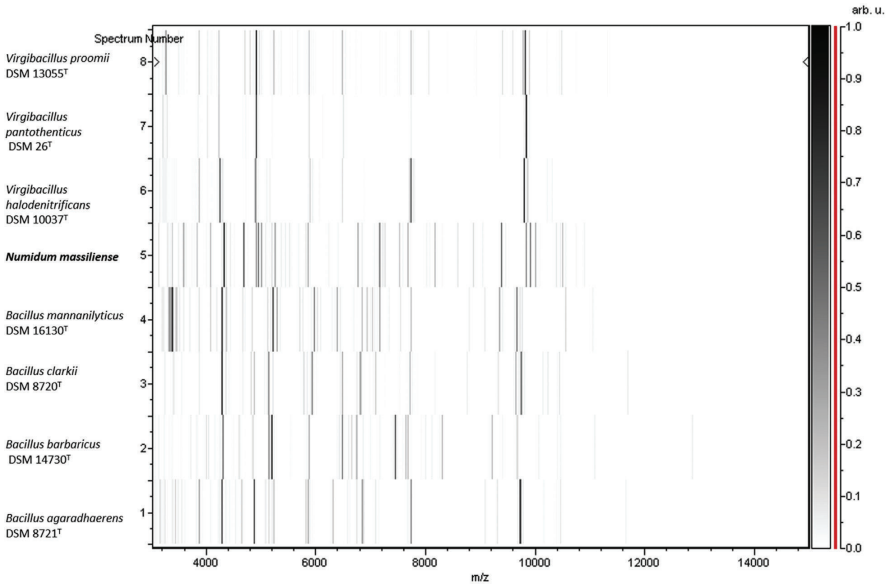


FIG. 3. Gel view comparing *Numidum massiliense* strain mt3^T (= CSUR P1305 = DSM 29571) to other species within *Bacillaceae* family. Gel view displays raw spectra of loaded spectrum files arranged in pseudo-gel-like look. X-axis records m/z value. Left y-axis displays running spectrum number originating from subsequent spectra loading. Peak intensity is expressed by greyscale scheme code. Colour bar and right y-axis indicate relation between colour peak is displayed with and peak intensity in arbitrary units. Displayed species are indicated on left.

Genome properties

The genome of *N. massiliense* strain mt3^T is 3 757 266 bp long with a 52.05% G+C content (Table 4, Fig. 6). Of the 3513 predicted genes, 3448 were protein-coding genes and 65 were RNAs (three genes are 5S rRNA, four genes are 16S rRNA,

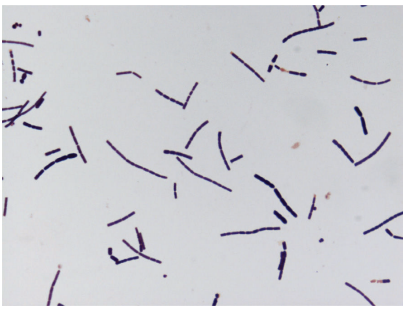


FIG. 4. Gram staining of *Numidum massiliense* strain mt3^T.

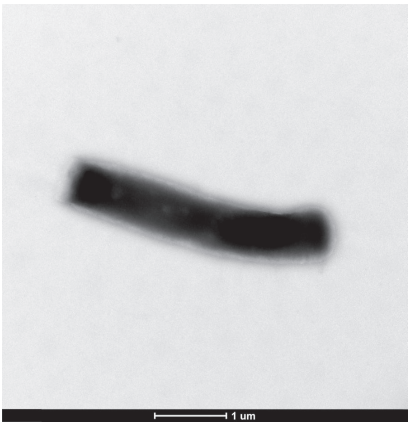


FIG. 5. Transmission electron microscopy of *Numidum massiliense* strain mt3^T, using Morgani 268D (Philips, Amsterdam, The Netherlands) at operating voltage of 60 kV. Scale bar represents 1 µm.

TABLE 2. Cellular fatty acid composition (%)

Fatty acid	IUPAC name	Mean relative %[a]
15:0 iso	13-methyl-tetradecanoic acid	87.6 ± 1.6
15:0 anteiso	12-methyl-tetradecanoic acid	5.5 ± 0.4
17:0 iso	15-methyl-Hexadecanoic acid	3.0 ± 0.9
16:0	Hexadecanoic acid	0.9 ± 0.1
16:0 iso	14-methyl-Pentadecanoic acid	0.6 ± 0.1
18:1n9	9-Octadecenoic acid	0.6 ± 0.2
13:0 anteiso	10-methyl-Dodecanoic acid	TR
5:0 iso	3-methyl-butanoic acid	TR
17:0 anteiso	14-methyl-Hexadecanoic acid	TR
18:0	Octadecanoic acid	TR
18:2n6	9,12-Octadecadienoic acid	TR
14:0 iso	12-methyl-Tridecanoic acid	TR
14:0	Tetradecanoic acid	TR

IUPAC, International Union of Pure and Applied Chemistry; TR, trace amounts (<1%).
[a]Mean peak area percentage ± standard deviation.

TABLE 4. Nucleotide content and gene count levels of genome

	Genome (total)	
Attribute	Value	% of total[a]
Size (bp)	3 757 266	100
G+C content (bp)	1 955 657	52.05
Coding region (bp)	3 1815 69	84.67
Total genes	3513	100
RNA genes	65	1.85
Protein-coding genes	3448	98.14
Genes with function prediction	2570	73.15
Genes assigned to COGs	2314	65.86
Genes with peptide signals	229	6.51
Genes with transmembrane helices	977	27.81

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.

two genes are 23S rRNA and 56 genes are tRNA genes). A total of 2570 genes (73.15%) were assigned as putative function (by COGs or by NR blast). Four hundred twelve genes were identified as ORFans (11.93%). The remaining 503 genes were annotated as hypothetical proteins (14.57%). The National Center for Biotechnology Information ID project is PRJEB8811, and the genome is deposited under accession number CTDZ01000000. The distribution of genes into COGs functional categories is presented in Table 5.

Genome comparison

N. massiliense genomic characteristics were compared to other close species (Table 6).

The draft genome sequence of *N. massiliense* strain mt3^T (3.76 MB) is smaller than the draft genome sequences of *Bacillus vireti* LMG 21834, *Bacillus mannanilyticus* JCM 10596, *Paucisalibacillus globulus* DSM 18846 *and Bacillus subterraneus* DSM 13966^T (5.29, 4.53, 4.24 and 3.9 MB respectively) and larger than those of *Bacillus selenitireducens* MLS10 and *Laceyella sacchari* 1-1 (3.59 and 3.32 MB respectively). The G+C content of *N. massiliense* (52.05%) is larger than the G+C contents of *L. sacchari* 1-1, *B. selenitireducens* MLS10, *B. subterraneus* DSM 13966^T, *B. vireti* LMG 21834, *B. mannanilyticus* JCM 10596 and *P. globulus* DSM 18846 (48.9, 48.7, 42.1, 39.7, 39.6 and 35.8% respectively).

TABLE 3. Differential characteristics of *Numidum massiliense* strain mt3^T, *Bacillus mannanilyticus* strain DSM 16130, *Virgibacillus pantothenticus* strain ATCC 14576, *Virgibacillus dokdonensis* DSM 16826, *Ornithinibacillus contaminans* DSM 22953, *Bacillus polygoni* strain NCIMB 14282^T, *Bacillus agaradhaerens* strain DSM 8721, *Paucisalibacillus globulus* strain LMG 23148^T, *Bacillus barbaricus* strain DSM 14730^T and *Virgibacillus koreensis* strain JCM 12387^T [33–41]

Property	*N. massiliense*	*B. mannanilyticus*	*V. pantothenticus*	*V. dokdonensis*	*O. contaminans*	*B. polygoni*	*B. agaradhaerens*	*P. globulus*	*B. barbaricus*	*V. koreensis*
Cell diameter (μm)	0.5–0.6	0.6–0.8	0.5–0.7	0.6–0.8	0.8–1	0.4–0.5	0.5–0.6	0.5	0.5	0.5–0.7
Oxygen requirement	+	+	+	−	+	+	+	+	+	−
Gram stain	+	+/−	+	+/−	+	+	NA	+	+	+
Salt requirement	−	−	+	−	−	+	+	−	−	+
Motility	−	+	+	+	−	−	NA	−	−	+
Endospore formation	+	+	+	+	+	+	+	+	+	+
Indole	−	−	−	−	NA	−	NA	NA	−	−
Production of:										
Alkaline phosphatase	+	NA	NA	−	NA	NA	NA	NA	NA	NA
Catalase	+	+	+	+	+	+	−	+	+	+
Oxidase	−	−	NA	+	+	−	NA	−	−	+
Nitrate reductase	+	−	+/−	−	NA	+	+	−	−	−
Urease	−	NA	NA	−	NA	NA	−	−	−	−
β-Galactosidase	+	NA	NA	−	NA	NA	NA	NA	NA	+
N-acetyl-glucosamine	+	NA	+	−	NA	NA	+	+	+	+
Acid from:										
L-Arabinose	−	−	−	−	NA	NA	+	−	−	+
Ribose	+	NA	+	+	−	−	NA	−	+	NA
Mannose	+	+	+	+	−	+	+	+	+	+
Mannitol	+	+	−	−	w	+	+	+	−	+
Sucrose	−	+	+/−	+	NA	+	+	+	−	NA
D-Glucose	+	+	+	+	+	+	+	+	+	−
D-Fructose	+	+	+	+	NA	+	+	+	+	+
D-Maltose	+	+	+	+	NA	+	+	+	+	+
D-Lactose	+	+	+/−	+	NA	−	NA	+	+/−	NA
Habitat	Human gut	Industry	Soil	Seawater	Blood	Indigo balls	Industry	Soil	Paint	Salt

+, positive result; −, negative result; w, weakly positive result; NA, data not available.

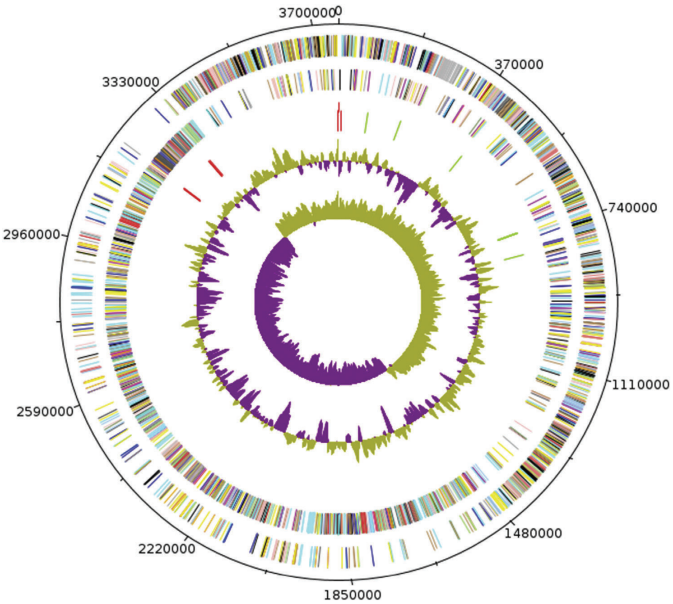


FIG. 6. Graphical circular map of chromosome. From outside to center: Genes on forward strain coloured by COGs categories (only gene assigned to COGs), RNA genes (tRNAs green, rRNAs red), GC content and GC skew.

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

Code	Value	% of total[a]	Description
J	150	4.35	Translation
A	0	0	RNA processing and modification
K	247	7.16	Transcription
L	169	4.90	Replication, recombination and repair
B	1	0.03	Chromatin structure and dynamics
D	30	0.87	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	85	2.47	Defense mechanisms
T	123	3.57	Signal transduction mechanisms
M	143	4.15	Cell wall/membrane biogenesis
N	8	0.23	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	34	0.99	Intracellular trafficking and secretion
O	93	2.70	Posttranslational modification, protein turnover, chaperones
C	156	4.52	Energy production and conversion
G	234	6.79	Carbohydrate transport and metabolism
E	278	8.06	Amino acid transport and metabolism
F	63	1.83	Nucleotide transport and metabolism
H	95	2.76	Coenzyme transport and metabolism
I	117	3.39	Lipid transport and metabolism
P	163	4.73	Inorganic ion transport and metabolism
Q	75	2.18	Secondary metabolites biosynthesis, transport and catabolism
R	370	10.73	General function prediction only
S	269	4.80	Function unknown
—	2903	84.19	Not in COGs

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

The gene content of *N. massiliense* (3513) is smaller than the gene contents of *B. vireti* LMG 21834, *B. mannanilyticus* JCM 10596, *P. globulus* DSM 18846 and *B. subterraneus* DSM 13966^T (5050, 4369, 4127 and 3772 respectively) but larger than those of *B. selenitireducens* MLS10 and *L. sacchari* 1-1 (3368 and 3256 respectively).

TABLE 6. Genome comparison of closely related species to *Numidum massiliense* strain mt3^T

Organism	INSDC	Size (Mb)	G+C (%)	Total genes
Numidum massiliense strain mt3^T	CTDZ00000000.1	3.76	52.05	3513
Bacillus vireti strain LMG 21834	ALAN00000000.1	5.29	39.7	5050
Bacillus mannanilyticus JCM 10596	BAMO00000000.1	4.53	39.6	4369
Paucisalibacillus globulus DSM 18846	AXVK00000000.1	4.24	35.8	4127
Bacillus subterraneus DSM 13966^T	JXIQ00000000.1	3.9	42.1	3772
Bacillus selenitireducens strain MLS10	CP001791.1	3.59	48.7	3368
Laceyella sacchari strain 1-1	ASZU00000000.1	3.32	48.9	3256

INSDC, International Nucleotide Sequence Database Collaboration.

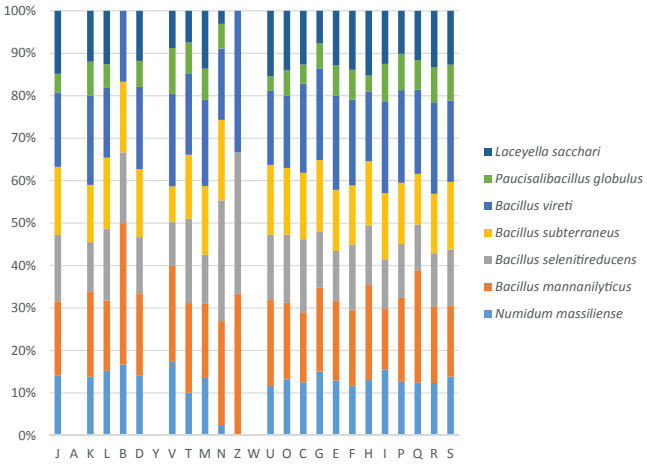


FIG. 7. Distribution of functional classes of predicted genes according to clusters of orthologous groups of proteins.

TABLE 7. Numbers of orthologous protein shared between genomes (upper right)[a]

	Numidum massiliense	*Bacillus mannanilyticus*	*Bacillus selenitireducens*	*Bacillus subterraneus*	*Bacillus vireti*	*Laceyella sacchari*	*Paucisalibacillus globulus*
N. massiliense	**3453**	1162	1028	1191	1294	1121	456
B. mannanilyticus	53.11	**3710**	1194	1369	1471	1174	511
B. selenitireducens	55.11	54.58	**3212**	1301	1318	972	461
B. subterraneus	54.81	56.4	58.15	**3648**	1632	1141	558
B. vireti	54.39	56.91	57.56	66.1	**4963**	1244	656
L. sacchari	57.96	54.52	55.66	55.94	55.59	**3152**	412
P. globulus	52.26	55.47	54.19	58.18	58.89	52.72	**4000**

[a]Average percentage similarity of nucleotides corresponding to orthologous protein shared between genomes (lower left) and numbers of proteins per genome (bold).

However, the distribution of genes into COGs categories was similar in all compared genomes except for those corresponding to the cytoskeleton category, which were only present in *B. vireti, B. selenitireducens* and *B. mannanilyticus* (Fig. 7). *N. massiliense* strain mt3^T shared 1162, 1028, 1191, 1294, 1121 and 456 orthologous genes with *B. mannanilyticus, B. selenitireducens, B. subterraneus, B. vireti, L. sacchari* 1-1 and *P. globulus* respectively (Table 7). Among species with standing in nomenclature, AGIOS values ranged from 52.26% between *N. massiliense* and *P. globulus* to 66.1% between *B. vireti* and *B. subterraneus*. When *N. massiliense* was compared to the other species, AGIOS values ranged from 52.26% with *P. globulus* to

TABLE 8. Pairwise comparison of *Bacillus niameyensis* with eight other species using GGDC, formula 2 (DDH estimates based on identities/HSP length)[a]

	Numidum massiliense	*Bacillus mannanilyticus*	*Bacillus selenitireducens*	*Bacillus subterraneus*	*Bacillus vireti*	*Laceyella sacchari*	*Paucisalibacillus globulus*	*Corynebacterium minutissimum*
N. massiliense	100% ± 00	2.52% ± 0.13	2.53% ± 0.15	2.52% ± 0.14	2.52% ± 0.13	2.52% ± 0.13	2.52% ± 0.16	2.52% ± 0.21
B. mannanilyticus		100% ± 00	2.53% ± 0.10	2.52% ± 0.18	2.52% ± 0.13	2.52% ± 0.12	2.52% ± 0.19	2.52% ± 0.20
B. selenitireducens			100% ± 00	2.53% ± 0.13	2.52% ± 0.12	2.53% ± 0.16	2.52% ± 0.11	2.52% ± 0.21
B. subterraneus				100% ± 00	2.55% ± 0.23	2.52% ± 0.18	2.52% ± 0.21	2.52% ± 0.21
B. vireti					100% ± 00	2.52% ± 0.17	2.52% ± 0.23	2.52% ± 0.20
L. sacchari						100% ± 00	2.52% ± 0.06	2.52% ± 0.21
P. globulus							100% ± 00	2.52% ± 00
C. minutissimum								100% ± 00

DDH, DNA-DNA hybridization; GGDC, Genome-to-Genome Distance Calculator; HSP, high-scoring segment pairs.
[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 16S rRNA (Fig. 1) and phylogenomic analyses as well as GGDC results.

57.96% with *L. sacchari.* To evaluate the genomic similarity among the compared strains, dDDH was also determined (Table 8).

Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Numidum massiliense* which contains the type strain mt3^T. This bacterial strain has been isolated from the faecal flora of a Tuareg boy living in Algeria.

Description of *Numidum* gen. nov.

Numidum (nu.mi'dum, from Numidum, which relates to a nomad people from Africa), is a Gram-positive, sporulating, facultative anaerobic bacilli. Optimal growth in aerobic condition at 37°C. Catalase positive and oxidase negative. Nitrates were reduced into nitrites. It is urease negative. The type strain is *Numidum massiliense* strain mt3^T.

Description of *Numidum massiliense* strain mt3^T gen. nov., sp. nov.

Numidum massiliense (mas.il'ien'se. L. gen. masc., massiliense, of Massilia, the Latin name of Marseille, where strain mt3^T was isolated) cells have a mean diameter of 0.5 µm. Colonies are greyish and 10 mm in diameter on 5% sheep's blood–enriched Columbia agar (bioMérieux). Positive reactions are observed for alkaline phosphatase, esterase (C4), esterase lipase (C8), leucine arylamidase, trypsin, α-chymotrypsin, acid phosphatase, β-galactosidase, β-glucuronidase, α-glucosidase and N-acetyl-β-glucosaminidase. D-Ribose, D-xylose, D-mannose, D-galactose, D-fructose, D-glucose, D-mannitol, N-acetylglucosamin, amygdalin, esculin ferric citrate, D-maltose, D-lactose, D-trehalose and D-tagatose and adipic acid were metabolized.

Cells were susceptible to doxycycline, ceftriaxone, gentamicin 500 µg, ticarcillin/clavulanic acid, rifampicin, teicoplanin, metronidazole and imipenem.

The G+C content of the genome is 52.05%. The 16S rRNA gene sequence and whole-genome shotgun sequence of *N. massiliense* strain mt3^T are deposited in GenBank under accession numbers LK985385 and CTDZ01000000, respectively. The type strain mt3^T (= CSUR P1305 = DSM 29571) was isolated from the stool of a Tuareg boy living in Algeria.

Acknowledgements

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

Conflict of Interest

None declared.

References

[1] Salazar N, Arboleya S, Valdès L, Stanton C, Ross P, Ruiz L, et al. The human intestinal microbiome at extreme ages of life. Dietary intervention as a way to counteract alterations. Front Genet 2014;5:406.
[2] Sankar SA, Lagier JC, Pontarotti P, Raoult D, Fournier PE. The human gut microbiome, a taxonomic conundrum. Syst Appl Microbiol 2015;38:276–86.
[3] Simpson HL, Campbell BJ. Review article: dietary fibre-microbiota interactions. Aliment Pharmacol Ther 2015;42:158–76.
[4] Xu Z, Knight R. Dietary effects on human gut microbiome diversity. Br J Nutr 2015;113(Suppl.):S1–5.
[5] 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.
[6] Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, et al. Bergey's manual of systematic bacteriology. The firmicutes. 2nd ed., vol. 3. New York: Springer; 2009.
[7] Rosselló-Móra 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; 2006. p. 23–50.
[8] Wayne LG, Brenner DJ, Colwell PR, Grimont PAD, Kandler O, Krichevsky MI, et al. Report of the ad hoc committee on reconciliation of approaches to bacterial systematic. Int J Syst Bacteriol 1987;37:463–4.
[9] Woese R, Kandler O, Wheelis ML. Towards a natural system of organisms: proposals for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 1990;87:4576–9.
[10] Viale AM, Arakaki AK, Soncini FC, Ferreyra RG. Evolutionnary relationships among eubacterial groups as inferred from GroEL (Chaperonin) sequences comparison. Int J Syst Bacteriol 1944;44:527–33.
[11] Reddy TB, Thomas AD, Stamatis D, Bertsch J, Isbandi M, Jansson J, et al. The Genome OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. Nucleic Acids Res 2015;43(Database issue):D1099–106.
[12] Caputo A, Lagier JC, Azza S, Robert C, Mouelhi D, Fournier PE, et al. *Microvirga massiliensis* sp. nov., the human commensal with the largest genome. Microbiologyopen 2016;5:307–22.
[13] Tidjani Alou M, Rathored J, Khelaifia S, Michelle C, Brah S, Diallo BA, et al. *Bacillus rubiinfantis* sp. nov. strain mt2^T, a new bacterial species isolated from human gut. New Microbes New Infect 2015;8:51–60.
[14] Seck E, Rathored J, Khelaifia S, Croce O, Robert C, Couderc C, et al. *Virgibacillus senegalensis* sp. nov., a new moderately halophilic bacterium isolated from human gut. New Microbes New Infect 2015;8:116–26.
[15] Hugon P, Mishra AK, Lagier JC, Nguyen TT, Couderc C, Raoult D, et al. Non-contiguous finished genome sequence and description of *Brevibacillus massiliensis* sp. nov. Stand Genomic Sci 2013;8:1–14.
[16] Lagier JC, Armougom F, Mishra AK, Nguyen TT, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Alistipes timonensis* sp. nov. Stand Genomic Sci 2012;6:315–24.

[17] Fournier PE, Lagier JC, Dubourg G, Raoult D. From culturomics to taxonogenomics: a need to change the taxonomy of prokaryotes in clinical microbiology. Anaerobe 2015;36:73–8.
[18] 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.
[19] 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.
[20] 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.
[21] 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.
[22] 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.
[23] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152–5.
[24] 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.
[25] 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.
[26] 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.
[27] 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.
[28] Käll L, Krogh A, Sonnhammer EL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol 2004;338:1027–36.
[29] Gouret P, Thompson JD, Pontarotti P. PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinformatics 2009;10:298.
[30] Lechner M, Findeib S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics 2011;12:124.
[31] 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;2:117–37.
[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] Aino K, Hirato K, Matsuno T, Morita N, Nodosaka Y, Fujiwara T, et al. *Bacillus polygoni* sp. nov., a moderately halophilic, non-motile obligate alkaliphile isolated from indigo balls. Int J Syst Evol Microbiol 2008;58:120–4.
[34] Heyndrickx M, Lebbe L, Kersters K, De Vos P, Forsyth G, Logan NA. Virgibacillus: a new genus to accommodate *Bacillus pantothenticus* (Prom and Knight 1950). Emended description of *Virgibacillus pantothenticus*. Int J Syst Evol Microbiol 1998;48:99–106.
[35] Lee JS, Lim JM, Lee KC, Lee JC, Park YH, Kim CJ. *Virgibacillus koreensis* sp. nov., a novel bacterium from a salt field, and transfer of *Virgibacillus picturae* to the genus *Oceanobacillus picturae* comb. nov. with emended descriptions. Int J Syst Evol Microbiol 2006;56:251–7.
[36] Nielsen P, Fritze D, Priest FG. Phenetic diversity of alkaliphilic *Bacillus* strains: proposal for nine species. Microbiology 1995;141:1745–61.
[37] Noqi Y, Takami H, Horikoshi K. Characterization of alkaliphilic *Bacillus* strains used in industry: proposal of five novel species. Int J Syst Evol Microbiol 2005;55:2309–15.
[38] Nunes I, Tiago I, Pires AL, Da Costa MS, Verissimo A. *Paucisalibacillus globulus* gen. nov., sp. nov., a Gram-positive bacterium isolated from potting soil. Int J Syst Evol Microbiol 2006;56:1841–5.
[39] Täubel M, Kämpfer P, Buczolits S, Lubitz W, Busse HJ. *Bacillus barbaricus* sp. nov. isolated from an experimental wall painting. Int J Syst Evol Microbiol 2003;53:725–30.
[40] Yoon JH, Kang SJ, Lee SY, Lee MH, Oh TK. *Virgibacillus dokdonensis* sp. nov., isolated from a Korean island, Dokdo, located at the edge of the East Sea in Korea. Int J Syst Evol Microbiol 2005;55:1833–7.
[41] Kämpfer P, Falsen E, Lodders N, Langer S, Busse HJ, Schumann P. *Ornithinibacillus contaminans* sp. nov., an endospore-forming species. Int J Syst Evol Microbiol 2010;60:2930–4.

Article 11: *Rubeoparvulum massiliense* gen. nov., sp. nov., a new bacterial genus isolated from the human gut from a Senegalese infant with severe acute malnutrition.

Maryam Tidjani Alou, Jaishriram Rathored, Jean-Christophe Lagier, Saber Khelaifia, Caroline Michelle, Cheikh Sokhna, Aldiouma Diallo, Bouli Ali Diallo, Pierre-Edouard Fournier, Didier Raoult and Sophie Edouard

Soumis à New Microbes New Infections

***Rubeoparvulum massiliense* gen. nov., sp. nov., a new bacterial genus isolated from the human gut from a Senegalese infant with severe acute malnutrition.**

Maryam Tidjani Alou[1], Jaishriram Rathored[1], Jean-Christophe Lagier[1,2], Saber Khelaifia[1], Caroline Michelle[1], Cheikh Sokhna[2], Aldiouma Diallo[2], Bouli Ali Diallo[3], Pierre-Edouard Fournier, Didier Raoult[1,4] and Sophie Edouard[1]*

[1]Aix-Marseille Université, URMITE, UM63, CNRS7278, IRD198, INSERM 1095, Faculté de médecine, 27 Boulevard jean Moulin, 13385 Marseille cedex 05, France.

[2] Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes IRD 198, CNRS 7278, Aix-Marseille Université, Marseille, France; Campus Commun UCAD-IRD of Hann, Dakar, Senegal.

[3]Laboratoire de microbiologie, département de biologie, Université Abdou Moumouni de Niamey

[4]Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

Running title: *Rubeoparvulum massiliense* gen. nov., sp. nov

Word count: 2,916

*Corresponding author: Sophie Edouard (sophie.edouard@univ-amu.fr) 27 boulevard Jean Moulin 13385 Marseille cedex 05

Tel: (33) 4 91 38 55 17 -Fax: (33) 4 91 38 77 72

Abstract

Rubeoparvulum massiliense strain mt6^{T} was isolated from the gut microbiota of a severely malnourished boy from Senegal and consisted of facultative anaerobic, sporeforming, non-motile and Gram-stain negative rods. *R. massiliense* showed a 92% similarity with the 16S rRNA of *Bacillus mannanilyticus.* The genome of strain mt6^{T} is 2,843,796 bp long genome with a 43.75% G+C content. It contains 2,735 protein-coding genes and 76 RNA genes among which 9 rRNA genes.

Keywords: *Rubeoparvulum massiliense*; gut microbiota; genome; culturomics; taxonogenomics.

Abbreviations

aa: aminoacids

AGIOS: Average of Genomic Identity of Orthologous gene Sequences

bp: base pairs

COG: Clusters of Orthologous Groups

CSUR: Collection de souches de l'Unité des Rickettsies

DDH: DNA-DNA Hybridization

DSM: Deutsche Sammlung von Mikroorganismen

FAME: Fatty Acid Methyl Ester

GC/MS: Gaz Chromatography/Mass Spectrometry

gDNA: genomic DNA

kb: kilobases

MALDI-TOF MS: Matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry

ORF: Open Reading Frame

TE buffer: Tris-EDTA buffer

URMITE: Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes

1. Introduction

The human microbiome is defined as the sum of all microbes colonizing the human body [1]. The gut microbiota is one of the largest microbial ecosystems of the human body consisting of 10^{14} microbial cells with a microbiome 150 times larger than the human genome [2]. The gastro-intestinal microbiota colonization starts before birth with the maternal microbiota and its early composition is influenced by the mode of birth. Its composition matures rapidly for the first year and reaches adult form by three years old [2,3]. A disruption of its equilibrium has been proven to be implicated in a growing number of pathologies like inflammatory bowel disease, irritable bowel syndrome, obesity [3,4], and severe acute malnutrition [5–7].

A new cultural approach called “microbial culturomics” based on the multiplication of culture conditions with a variation of temperature, media, atmosphere, was developed in our laboratory in order to explore as exhaustively as possible a microbial ecosystem [8,9]. Using this new approach, we isolated a new member of the *Bacillaceae* family. At this time, 52 validated genera are part of the *Bacillaceae* family created in 1895 by Fisher and *Bacillus* is its type genus described by Cohn

in 1872 [10]. Most species of this family were found in environment (soil, water, plant) and are opportunistic pathogens in human except *Bacillus anthracis* which is very pathogenic as well known. The *Bacillaceae* family includes gram-positive, rod-shaped, mostly aerobic and facultative anaerobic genera [11]. By adding the description of the assembled and annotated genome of the species and the proteomic description of the strain with the MALDI-TOF profile to the classical description principles (phylogenetic relationships based on the 16S rRNA sequence, phenotypic and genotypic characteristics), a new concept of description called "taxonogenomics" was developed in our laboratory [12]. We hereby describe the genus *Rubeoparvulum* which type species is *Rubeoparvulum massiliense* strain mt6 (=CSUR P1473 =DSM 100479) from a stool sample collected in a 2 months-old infant living in Senegal and presenting with kwashiorkor, a type of severe acute malnutrition.

2. Materials and methods

2.1. Sample information

The strain mt6 was isolated from a stool taken from a severely malnourished 2-month old boy with a height-for-age score of -5.87 and presenting with nutritional edema collection

from Senegal in April 2014. This sampling was realized as part of an exploratory study of the human gut microbiota in African children presenting malnutrition. The study was approved by the local IFR 48 ethics committee, under agreement number 09-022 and the boy's parents provided informed oral consent. The sample was stored at -80°C after collection.

2.2. Strain identification by MALDI-TOF MS and 16S rRNA sequencing

In order to explore as exhaustively as possible the bacterial diversity of the fecal sample, the "culturomics" concept was used to culture this sample using eighteen culture conditions [8]. The purified colonies obtained were identified using MALDI-TOF MS as described previously [13,14]. Colonies were deposited on a MTP 96 MALDI-TOF target plate (Brüker Daltonics, Leipzig, Germany) which was analyzed with a Microflex spectrometer (Brüker Daltonics). The spectra obtained were matched against the references of the 7,567 bacteria contained in the database, by standard pattern matching (with default parameter settings), with the MALDI BioTyper database software (version 2.0, Brüker). An identification score over 1.9 with a validated species allowed the identification at the species level and a score under 1.7 did not enable any identification. In the latter case, if a clean spectrum was

obtained, the 16S rRNA was sequenced and the obtained nucleotide sequence was matched against the NCBI database using the BLAST algorithm [15]. A new species or genus was defined without performing DNA-DNA hybridization by a similarity level of the 16S rRNA sequence under 98.65% or 95% respectively [16].

2.3. Growth conditions

The ideal growth conditions of strain mt6^T were determined by testing different culture conditions. Five growth temperatures (25, 30, 37, 45 and 56°C) were tested under anaerobic and microaerophilic atmospheres using GENbag Anaer and GENbag microer systems respectively (bioMerieux, Marcy l'Etoile, France). Aerobic growth was tested with and without 5% CO_2. Growth was also tested at various pH (6, 6.5, 7, 7.5, 8 and 8.5) using a pH adjusted Colombia agar (BioMerieux). Salt tolerance was also tested with 0.5, 1, 5, 7.5 and 10% (w/v) NaCl.

2.4. Morphologic, biochemical and antibiotics susceptibility tests

Phenotypic characteristics such as Gram staining, sporulation, motility, catalase and oxidase were determined as previously described [8]. To perform the negative staining, detection formvar coated grids were deposited on a 40 µL bacterial

suspension drop, incubated at 37°C for 30 minutes and on ammonium molybdate 1% for ten seconds. The dried grids on blotted paper were observed with a Tecnai G20 transmission electron microscope (FEI Company, Limeil-Brevannes, France). The biochemical features of strain mt6^{T} were investigated with API 50CH, API ZYM and API 20A strips (bioMérieux) according to the manufacturer's instructions. Cellular fatty acid methyl ester (FAME) analysis was performed by GC/MS. Strain mt6 was grown on 5% sheep blood enriched Colombia agar (bioMerieux) for the fatty acids analysis which was carried out by GC/MS. Approximately 67 mg of bacterial biomass each harvested from several culture plates. Cellular fatty acid methyl esters (FAME) were prepared as described by Sasser [17]. Briefly, GC/MS analyses were realized by using a Clarus 500 gas chromatograph equipped with a SQ8S MS detector (Perkin Elmer, Courtaboeuf, France). 2 µL of FAME extracts were volatized at 250°C (split 20 mL/min) in a Focus liner with wool and separated on an Elite-5MS column (30 m, 0.25 mm i.d., 0.25 mm film thickness) using a linear temperature gradient (70 to 290°C at 6°C/min), allowing the detection of C4 to C24 fatty acid methyl esters. Helium flowing at 1.2 mL/min was used as carrier gas. The MS inlet line was set at 250°C and EI source at 200°C. Full scan monitoring was performed from 45 to 500 m/z.

All data were collected and processed using Turbomass 6.1 (Perkin Elmer). FAMEs were identified by a spectral database search using MS Search 2.0 operated with the Standard Reference Database 1A (National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA) and the FAMEs mass spectral database (Wiley, Chichester, UK). Retention time correlations with estimated nonpolar retention indexes from the NIST database were obtained using a 37-component FAME mix (Supelco; Sigma-Aldrich, Saint-Quentin Fallavier, France); FAME identifications were confirmed using this index. Antibiotic susceptibility testing was performed using a disk diffusion method according to EUCAST 2015 recommendations [18].

2.5. Genomic DNA preparation

For genomic DNA preparation, *R. massiliense* strain mt6^{T} was cultured on 5 % sheep blood-enriched Columbia agar (BioMerieux) at 37°C aerobically. Bacteria grown on three Petri dishes were resuspended in 4x100µL of TE buffer. Then, 200 µL of this suspension was diluted in 1ml TE buffer for lysis treatment that included a 30- minute incubation with 2.5 µg/µL lysozyme at 37°C, followed by an overnight incubation with 20 µg/µL proteinase K at 37°C. Extracted DNA was then purified

using three successive phenol-chloroform extractions and ethanol precipitations at -20°C overnight. After centrifugation, the DNA was resuspended in 160 µL TE buffer.

2.6. Genome sequencing and assembly

Using the Mate-Pair strategy, the genomic DNA (gDNA) of *R. massiliense* strain mt6^T was sequenced on the MiSeq sequencer (Illumina, San Diego, CA, USA). 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 1µg of gDNA using the Nextera Mate-Pair Illumina guide and the gDNA sample was simultaneously fragmented and tagged with a Mate-Pair junction adapter. The pattern of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) with a DNA 7500 labchip. The DNA fragments ranged in size from 1 kb up to 10 kb with an optimal size at 4.08 kb. No size selection was performed and only 464 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal at 569 bp on the Covaris S2 device in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies)

and the final library concentration was measured at 24.4 nmol/L. The libraries were normalized at 2nM 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 in a 2x251-bp. Total information of 10.1Gb was obtained from a 1,189 K/mm2 cluster density with a cluster passing quality control filters of 99.1% (22,579,000 clusters). The reads obtained were trimmed, then assembled was performed using the CLC genomicsWB4 software.

2.7. Genome annotation and comparison

Open Reading Frames (ORFs) were predicted using Prodigal [19] with default parameters but the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched against the GenBank [20] and the Clusters of Orthologous Groups (COG) databases 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 an E-value of $1e^{-03}$, a coverage of 0.7 and an identity percentage of 30%, and if the sequence length was smaller than 80 amino acids (aa), we

used an E-value of $1e^{-05}$. The tRNAScanSE tool [21] was used to find tRNA genes, while ribosomal RNAs were found using RNAmmer [22]. Lipoprotein signal peptides and the number of transmembrane helices were predicted using Phobius [23]. Mobile genetic elements were predicted using PHAST [24] and RAST [25]. ORFans were identified if all the BLASTP performed did not give positive results (*E*-value smaller than $1e^{-03}$ for ORFs with sequence size larger than 80 aa or *E*-value smaller than $1e^{-05}$ for ORFs with sequence length smaller 80 aa). Such parameter thresholds have already been used in previous studies to define ORFans. Artemis [26] and DNA Plotter [27] were used for data management and the visualization of genomic features, respectively. The Mauve alignment tool (version 2.3.1) was used for multiple genomic sequence alignment [28].

Comparator species for genomic comparison were identified in the 16S RNA tree using the Phylopattern software [29]. The genome of strain $mt6^T$ was compared to those of *Alkaliphilus metalliredigens* strain QYMF, *Clostridium aceticum* strain DSM 1496, *Alkaliphilus transvaalensis* strain SAGM1 and *Alkaliphilus oremlandii* strain OhILAs.

For each selected genome, the complete genome sequence, proteome genome sequence and Orfeome genome sequence

were retrieved from the FTP of NCBI. 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). Annotation and comparison processes were performed in the Multi-Agent software system DAGOBAH [30] which includes Figenix [31] libraries which provide pipeline analysis. To evaluate the genomic similarity between studied genomes, we determined two parameters, digital DDH (dDDH) which exhibits a high correlation with DDH [32,33] and AGIOS [34], which was designed to be independent from DDH [34]. The AGIOS score is the mean value of nucleotide similarity between all couples of orthologous proteins between the two studied genomes [34].

3. Results

3.1. Strain identification and phylogenetic analyses

The mt6 strain was isolated after a 30-day pre-incubation at 37°C in an anaerobic blood culture bottle supplemented with 5 ml of rumen fluid filter-sterilized through a 0.2 μm pore filter (Thermo Fisher Scientific, Villebon sur Yvette, France). Strain mt6^{T} was subcultured on 5% blood-enriched Colombia agar (bioMerieux) in anaerobic atmosphere at 37°C. The bacterium could not be identified by matrix-assisted laser desorption-

ionization time-of-flight screening (score under 1.7) but sequencing of the 16S rRNA revealed that strain mt6^T's nucleotide sequence (accession number LN828926) had a 92% similarity level with *Bacillus mannanilyticus*, the phylogenetically closest validated species (Figure 1). According to Kim *et al* [16], a new genus can be defined by a similarity level threshold lower than 95% without carrying out DNA-DNA hybridization. Consequently, strain mt6^T was classified as a new genus called *Rubeoparvulum*, its type species being *Rubeoparvulum massiliense* (Table 1).

3.2. Phenotypic description

The growth of the mt6 stains occurred between 25 and 56°C on 5% sheep blood Colombia agar. Optimal growth was achieved at 37°C after 48 hours of incubation in both anaerobic and microaerophilic conditions. The cell growth was weaker in aerobic conditions. Strain mt6 was able to grow at pH values ranging from 6 to 8.5 and 0.5 to 5% NaCl concentrations. Cells were sporeforming, motile and form translucent colonies with a mean diameter of 0.5 mm on blood-enriched Colombia agar. Microscopic observations showed Gram-stain negative rod-shaped cells (Figure 2) and electron microscopy showed rods with a mean diameter of 1 µm and a mean length of 6.8 µm (Figure 3). Our MALDI-TOF database was incremented with

the reference spectrum obtained for strain mt6^T (Figure 4). Comparisons of the aforementioned spectrum to that of other known species of the *Bacillaceae* family are represented in the gel view (Figure 5).

Strain mt6^T was negative for catalase activity and positive for oxidase activity. Using an API ZYM strip, positive reactions were recorded for alkaline phosphatase, esterase (C4), esterase lipase (C8), leucine arylamidase, α-chymotrypsin, acid phosphatase and naphtol-AS-BI-phosphohydrolase. Nitrate reduction was observed; urease, β-glucosidase and protease activities were positive using an API 20 NE strip. All other reactions were negative on both strips. An API 50CH was used to test the carbohydrates metabolization. The following carbohydrates were metabolized by strain mt6^T: glycerol, D-lactose, D-fucose, D-mannose, D-cellobiose, D-maltose, salicin, D-arabitol, N-acetyl-glucosamine and potassium-5-ketogluconate. Amygdalin, arbutin, D-fructose, inulin, D-sucrose, D-raffinose, erythritol, D-arabinose, L-arabinose, D-ribose, D-xylose, L-xylose, D-adonitol, methyl-βD-xylopyranoside, D-glucose, D-galactose, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl-αD-mannopyranoside, methyl-αD-glucopyranoside, esculin ferric citrate, D-melibiose, D-trehalose, D-melezitose, starch,

glycogen, xylitol, gentiobiose, D-turanose, D-lyxose, D-tagatose, L-fucose, L-arabitol, potassium gluconate and potassium 2-ketogluconate showed negative reactions. Differences between the biochemical features of close members of the family *Bacillaceae* are exhibited in Table 2.

The major cellular fatty acids found for strain mt6 were 12-methyl-tetradecanoic acid (25 %), 10-methyldodecanoic acid (15 %) and hexadecanoic acid (12 %). This strain was composed of many branched structures (iso/anteiso). A specific 3-OH fatty acid was also described (< 1 %, Table 3).

Cells were susceptible to amoxicillin, gentamicin, ceftriaxone, ciprofloxacin, penicillin, imipenem, tobramycin and oxacillin but were resistant to metronidazole, trimethoprim/ sulfamethoxazole, rifampicin, doxycycline, vancomycin, nitrofurantoin and erythromycin.

3.3. Genome properties

With an estimated size of 2,843,796 bp, the *R. massiliense* genome had a G+C content of 43.75% (Table 4, Figure 6). It was composed of six scaffolds composed of six contigs. Out of 2,811 predicted genes, 2,735 were protein-coding genes, and 76 were RNAs (seven 5S rRNA, one 16S rRNA gene, one 23S rRNA gene, 67 tRNA genes). A putative function was assigned to 1,873genes (66.63%) by cogs or NR

blast. 233 genes (8.28%) were identified as ORFans. The remaining 402 genes (14.70%) were annotated as hypothetical proteins. Table 4 shows the statistics of the genome while Table 5 presents the distribution of genes into COG functional categories.

3.4. Genome comparison

The genome of strain mt6^T was compared to those of closely related species (Table 6) by comparing their main genomic characteristics (size, G+C content, protein coding genes, total number of genes). The genome size of strain mt6^T (2.84 Mb) is smaller than *B. agri* (5.51 Mb), *B. borstelensis* (5.16 Mb), *B. mannanilyticus* (4.53 Mb), *B. thermoruber* (4.43 Mb) and *C. thermarum* (2.9Mb). Strain mt6^T had a higher G+C content (43.75%) than *B. mannanilyticus* (39.6%) but lower than *B. thermoruber* (58.4%), *B. agri* (54.2%), *B. borstelensis* (52%) and *C. thermarum* (47.6%). Strain mt6^T has the smallest number of protein coding genes as well as the smallest number of total genes than all of the other compared genomes, as summarized in Table 6.

Among species with standing in nomenclature, AGIOS values ranged from 75.55 between *B. borstelensis* and *B. thermoruber* to 59.20 between *B. thermoruber* and *B.*

mannanilyticus. The comparison of the AGIOS value of strain mt6^T with the other species gave AGIOS values ranging from 60.30 with *B. thermoruber* to 63.12 with *B. mannanilyticus* (Table 7). In addition, strain mt6^T shared 1,296, 1,316, 1,039, 1,079 and 1,605 orthologous genes with *B. thermoruber, B. borstelensis, C. thermarum, B. mannanilyticus,* and *B. agri*, respectively. Finally, we observed that in each COG categories, all compared genomes have nearly the same number of genes (Figure 7).

4. Conclusion

The proteomic analysis of strain mt6^T with its MALDI-TOF spectrum, the 92% similarity level of the 16S rRNA nucleotide sequence to *Bacillus mannanilyticus* and the analysis of its complete assembled and annotated genome allowed us to classify it as a new genus called *Rubeoparvulum*, its type species being *Rubeoparvulum massiliense.*

Description of *Rubeoparvulum* gen. nov.

Rubeoparvulum (ru.be.o (adj) meaning red; par.vu.lum meaning infant. Strain mt6^T was isolated from the stool of a patient with kwashiorkor ; Red infant refers to the hair discoloration observed in kwashiorkor patients).

Cells are rod-shaped Gram-stain negative bacteria. Optimal growth in anaerobic and microaerophilic conditions at 37°C. Catalase negative. Oxidase positive. Nitrate reduction, urease, β-glucosidase and alkaline phosphatase were positive. The type species is *Rubeoparvulum massiliense* strain mt6^{T}.

Description of *Rubeoparvulum massiliense* strain mt6^{T} gen. nov., sp. nov.

Rubeoparvulum massiliense (mas.si.li.en'se. L. adj. *massiliense* of Massilia, the old Greek and Roman name of Marseille (France) where the strain was isolated).

Cells are sporeforming, motile and facultative anaerobe, Gram-stain negative, rod-shaped bacilli with a mean diameter of 1 μm and a mean length of 6.8 μm. Colonies were small (mean diameter of 0.5 mm) and translucent on 5% sheep blood-enriched Colombia agar. Catalase negative and oxidase positive. Positive reactions were recorded for alkaline phosphatase, esterase (C4), esterase lipase (C8), leucine arylamidase, α-chymotrypsin, acid phosphatase and naphtol-AS-BI-phosphohydrolase. Urease, β-glucosidase, protease activities and nitrate reduction were also positive. Glycerol, D-lactose, D-fucose, D-mannose, D-cellobiose, D-maltose, salicin, N-acetyl-glucosamine, potassium-5-ketogluconate and D-arabitol were metabolized. Cells were susceptible to amoxicillin, gentamicin,

ceftriaxone, ciprofloxacin, penicillin, imipenem, tobramycin and oxacillin but were resistant to metronidazole, trimethoprim/ sulfamethoxazole, rifampicin, doxycycline, vancomycin, nitrofurantoin and erythromycin.

The G+C content of the genome is 43.75%. The 16S rRNA gene sequence and whole-genome shotgun sequence of *R. massiliense* strain mt6^{T} are deposited in EMBL/EBI under accession numbers LN828926 and CVPE00000000 respectively. The type strain mt6^{T} (=CSUR P1473 =DSM 100479) was isolated from the fecal matter of a 2-month old boy from Senegal suffering from kwashiorkor.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

The authors thank the Xegen Company (www.xegen.fr) for automating the genomic annotation process. This study was funded by the"Fondation Méditerranée Infection".

We thank Karolina Griffiths for English reviewing and Claudia Andrieu for administrative assistance.

Reference list

[1] Fricke WF. The More the Merrier? Reduced Fecal Microbiota Diversity in Preterm Infants Treated with Antibiotics. J Pediatr 2014; 165:8–10. doi:10.1016/j.jpeds.2014.03.022.

[2] Simpson HL, Campbell BJ. Review article: dietary fibre-microbiota interactions. Aliment Pharmacol Ther 2015; 42:158–79. doi:10.1111/apt.13248.

[3] Gritz EC, Bhandari V. The Human Neonatal Gut Microbiome: A Brief Review. Front Pediatr 2015;3. doi:10.3389/fped.2015.00017.

[4] Salazar N, Arboleya S, ValdÃ©s L, Stanton C, Ross P, Ruiz L, et al. The human intestinal microbiome at extreme ages of life. Dietary intervention as a way to counteract alterations. Front Genet 2014;5. doi:10.3389/fgene.2014.00406.

[5] Million M, Diallo A, Raoult D. Gut microbiota and malnutrition. Microb Pathog 2016. doi:10.1016/j.micpath.2016.02.003.

[6] Million M, Tidjani Alou M, Khelaifia S, Bachar D, Lagier J-C, Dione N, et al. Increased Gut Redox and Depletion of Anaerobic and Methanogenic Prokaryotes in Severe Acute Malnutrition. Sci Rep 2016; 6:26051. doi:10.1038/srep26051.

[7] Tidjani Alou M, Lagier J-C, Raoult D. Diet influence on the gut microbiota and dysbiosis related to nutritional disorders. Hum Microbiome J 2016; 1:3–11. doi:10.1016/j.humic.2016.09.001.

[8] 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; 28:237–64. doi:10.1128/CMR.00014-14.

[9] 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. doi:10.1111/1469-0691.12023.

[10] Parte AC. LPSN—list of prokaryotic names with standing in nomenclature. Nucleic Acids Res 2014; 42:D613–6. doi:10.1093/nar/gkt1111.

[11] De Vos P, Garrity GM, Jones D, Krieg NR, Ludwig W, Rainey FA, et al., editors. Systematic Bacteriology. New York, NY: Springer New York; 2009.

[12] 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. doi:10.1016/j.anaerobe. 2015.10.011.

[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; 49:543–51. doi:10.1086/600885.

[14] Seng P, Abat C, Rolain JM, Colson P, Lagier J-C, 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. doi:10.1128/JCM.00492-13.

[15] 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.

[16] 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; 64:346–51. doi:10.1099/ijs.0.059774-0.

[17] Sasser M. Bacterial identification by gas chromatographic analysis of fatty acids methyl esters (GC-FAME). Microbial ID Inc. Newark, NY; 2006.

[18] 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 Off Publ Eur Soc Clin Microbiol Infect Dis 2014; 20:O255-266. doi:10.1111/1469-0691.12373.

[19] Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010; 11:119. doi:10.1186/1471-2105-11-119.

[20] Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2012; 40: D48-53. doi:10.1093/nar/gkr1202.

[21] 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.
[22] Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 2007; 35:3100–8. doi:10.1093/nar/gkm160.
[23] Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol 2004; 338:1027–36. doi:10.1016/j.jmb.2004.03.016.
[24] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res 2011; 39: W347-352. doi:10.1093/nar/gkr485.
[25] Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 2008; 9:75. doi:10.1186/1471-2164-9-75.
[26] Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. Bioinforma Oxf Engl 2000; 16:944–5.
[27] Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinforma Oxf Engl 2009; 25:119–20. doi:10.1093/bioinformatics/btn578.
[28] Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004; 14:1394–403. doi:10.1101/gr.2289704.

[29] Gouret P, Thompson JD, Pontarotti P. PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinformatics 2009; 10:298. doi:10.1186/1471-2105-10-298.

[30] 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. Evol. Biol. – Concepts Biodivers. Macroevolution Genome Evol., Springer Berlin Heidelberg; 2011, p. 71–87.

[31] 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. doi:10.1186/1471-2105-6-198.

[32] 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; 2:117–34. doi:10.4056/sigs.531120.

[33] Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 2013; 14:60. doi:10.1186/1471-2105-14-60.

[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; 64:384–91. doi:10.1099/ijs.0.057091-0.

[35] Gao X-H, Gao S, Zhou Y, Guan H-L, Zhang Y-J, Jia M, et al. Tenuibacillus halotolerans sp. nov., a novel bacterium isolated from a soil sample from a salt lake in Xinjiang, China and emended description of the genus Tenuibacillus. Antonie Van Leeuwenhoek 2013; 103:207–15. doi:10.1007/s10482-012-9802-3.

[36] Garcia MT. Thalassobacillus devorans gen. nov., sp. nov., a moderately halophilic, phenol-degrading, Gram-positive bacterium. Int J Syst Evol Microbiol 2005; 55:1789–95. doi:10.1099/ijs.0.63560-0.

[37] Jeon CO. Reclassification of Bacillus haloalkaliphilus Fritze 1996 as Alkalibacillus haloalkaliphilus gen. nov., comb. nov. and the description of Alkalibacillus salilacus sp. nov., a novel halophilic bacterium isolated from a salt lake in China. Int J Syst Evol Microbiol 2005; 55:1891–6. doi:10.1099/ijs.0.63456-0.

[38] Mayr R. Ornithinibacillus gen. nov., with the species Ornithinibacillus bavariensis sp. nov. and Ornithinibacillus californiensis sp. nov. Int J Syst Evol Microbiol 2006; 56:1383–9. doi:10.1099/ijs.0.64038-0.

[39] Nogi Y. Characterization of alkaliphilic Bacillus strains used in industry: proposal of five novel species. Int J Syst Evol Microbiol 2005; 55:2309–15. doi:10.1099/ijs.0.63649-0.

[40] Pereira SG, Albuquerque L, Nobre MF, Tiago I, Verissimo A, Pereira A, et al. Pullulanibacillus uraniitolerans sp. nov., an acidophilic, U(VI)-resistant species isolated from an acid uranium mill tailing effluent and emended description of the genus Pullulanibacillus. Int J Syst Evol Microbiol 2013; 63:158–62. doi:10.1099/ijs.0.040923-0.

[41] Ren P-G. Salinibacillus aidingensis gen. nov., sp. nov. and Salinibacillus kushneri sp. nov., moderately halophilic bacteria isolated from a neutral saline lake in Xin-Jiang, China. Int J Syst Evol Microbiol 2005; 55:949–53. doi:10.1099/ijs.0.63287-0.

[42] Slobodkina GB, Panteleeva AN, Kostrikina NA, Kopitsyn DS, Bonch-Osmolovskaya EA, Slobodkin AI. Tepidibacillus fermentans gen. nov., sp. nov.: a moderately thermophilic anaerobic and microaerophilic bacterium from an underground gas storage. Extremophiles 2013; 17:833–9. doi:10.1007/s00792-013-0564-2.

Figure legends

Figure 1. Phylogenetic tree highlighting the position of *Rubeoparvulum massiliense* strain mt6^T relative to other close strains. The respective GenBank accession numbers for 16S rRNA genes are indicated in parenthesis. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using the maximum-likelihood method within the MEGA6 software. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 1,000 times to generate a majority consensus tree. *Salinimicrobium catena* strain HY1 was used as an outgroup. The scale bar represents a 5% nucleotide sequence divergence.

Figure 2. Gram staining of *Rubeoparvulum massiliense* strain mt6^T (=CSUR P1473 =DSM 100479).

Figure 3. Transmission electron microscopy of *Rubeoparvulum massiliense* strain mt6^T (=CSUR P1473 =DSM 100479) using a Tecnai G20 transmission electron microscope (FEI Company) at operating voltage of 60 kV. The scale bar represents 500 nm.

Figure 4. Reference mass spectrum from *Rubeoparvulum massiliense* strain mt6^T (=CSUR P1473 =DSM 100479). Spectra from 12 individual colonies were compared and a reference spectrum was generated.

Figure 5. Gel view comparing *Rubeoparvulum massiliense* strain mt6^{T} (=CSUR P1473 =DSM 100479) to other species within the *Bacillaceae* family. 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-axis displays the running spectrum number originating from subsequent spectra loading. The peak intensity is expressed by a Gray scale scheme code. The color bar and the right y-axis indicate the relation between the color a peak is displayed with and the peak intensity in arbitrary units. Displayed species are indicated on the left.

Figure 6. Graphical circular map of the chromosome. From outside to the center: Genes on the forward strain colored by COG categories (only gene assigned to COG), RNA genes (tRNAs green, rRNAs red), G+C content and G+C skew. COGs, Clusters of Orthologous Groups database.

Figure 7. Distribution of functional classes of predicted genes according to the clusters of orthologous groups of protein.

Table 1. Classification and general features of *Rubeoparvulum massiliense* strain mt6^{T}

Property	Term
Current classification	Domain: *Bacteria*
	Phylum: *Firmicutes*
	Class: *Bacilli*
	Order: *Bacillales*
	Family: *Bacillaceae*
	Genus: *Rubeoparvulum*
	Species: *Rubeoparvulum massiliense*
	Type strain: mt6
Gram stain	Negative
Cell shape	Rod
Motility	Motile
Sporulation	Sporulating
Temperature range	Mesophilic
Optimum temperature	37°C

Table 2. Differential characteristics of *Rubeoparvulum massiliense* strain mt6^{T} CSUR P1473 =DSM 100479, *Bacillus mannanilyticus* strain AM-001^{T} DSM 16130^{T}, *Tepidibacillus fermentans* strain STGH^{T} DSM 23802^{T}, *Pullulanibacillus uraniitolerans* strain UG-2^{T} DSM 19429^{T}, *Alkalibacillus haloalkaliphilus* DSM 5271^{T}, *Tenuibacillus halotolerans* strain YIM 94025^{T} KCTC 33046^{T}, *Thalassobacillus devorans* strain G-19.1^{T} DSM 16966^{T}, *Salinibacillus aidingensis* strain 25-7^{T} JCM 12389^{T}, *Salinibacillus kushneri* strain 8-2^{T} JCM 12390^{T}, *Ornithinibacillus bavariensis* strain WSBC 24001^{T} DSM 15681^{T} [35–42].

Properties	*Rubeoparvulum massiliense*	*Bacillus mannanilyticus*	*Tepidibacillus fermentans*	*Pullulanibacillus uraniitolerans*	*Alkalibacillus haloalkaliphilus*	*Tenuibacillus halotolerans*	*Thalassobacillus devorans*	*Salinibacillus kushneri*	*Salinibacillus aidingensis*	*Ornithinibacillus bavariensis*
Cell diameter (µm)	1.0	0.6-0.8	0.3	1.0	0.3-0.5	0.2-0.3	1.0-1.2	0.4-0.6	0.3-0.5	0.4
Oxygen requirement	+/-	+	-	+	+	+	+	+	+	+
Gram stain	-	+/-	+	+	-	+	+	+	+	+
Salt requirement	-	-	+	+	+	+	+	+	+	+
Indole	-	-	na	na	na	-	-	na	na	-
Production of										
Catalase	-	+	-	+	+	+	+	+	+	+
Oxidase	+	-	+	-	-	+	-	+/-	+/-	+
Nitrate reductase	+	-	+	-	-	-	+	-	-	-
Urease	+	na	na	na	-	na	-	-	-	-

Acid from										
L-Arabinose	-	-	-	+	na	-	-	+	+	-
Ribose	-	na	na	+	+	na	-	na	na	+
Mannose	+	+	na	+	+	na	+	+	+	-
Mannitol	-	+	na	na	-	-	+	+	+	-
Sucrose	-	+	-	+	na	+	+	+	+	-
D-glucose	-	+	+	+	-	+	+	+	+	+
D-fructose	-	+	+	+	+	+	+	+	+	-
D-maltose	+	+	+	+	-	-	+	+	+	-
D-lactose	+	-	-	+	na	-	-	+	**+**	-
Habitat	Human stool	Industry	Gas storage	Mill tailing effluent	Salt lake	Salt lake	Hypersaline environments	Neutral saline lake	Neutral saline lake	Pasteurized milk

na: data not available

Table 3. Cellular fatty acid composition (%) of *Rubeoparvulum massiliense* strain mt6^{T}.

Fatty acids	Name	Mean relative %*
15:0 anteiso	12-methyl-tetradecanoic acid	25.2 ± 0.3
13:0 anteiso	10-methyl-Dodecanoic acid	15.2 ± 0.2
16:0	Hexadecanoic acid	12.0 ± 0.6
18:1n9	9-Octadecenoic acid	9.3 ± 0.4
13:0 iso	11-methyl-Dodecanoic acid	7.1 ± 0.1
18:0	Octadecanoic acid	6.3 ± 0.1
15:0 iso	13-methyl-tetradecanoic acid	5.8 ± 0.1
14:0 iso	12-methyl-Tridecanoic acid	5.6 ± 0.3
18:2n6	9,12-Octadecadienoic acid	5.3 ± 0.2
5:0 iso	3-methyl-butanoic acid	1.7 ± 0.1
14:0	Tetradecanoic acid	1.6 ± 0.1
11:0 anteiso	8-methyl-decanoic acid	1.1 ± 0.1
18:1n6	12-Octadecenoic acid	TR
15:0	Pentadecanoic acid	TR
12:0 iso	10-methyl-Undecanoic acid	TR
17:0	Heptadecanoic acid	TR
18:1n7	11-Octadecenoic acid	TR
13:0	Tridecanoic acid	TR
15:0 3-OH anteiso	3-hydroxy-12-methyl-Tetradecanoic acid	TR
11:0 iso	9-methyl-decanoic acid	TR
16:0 iso	14-methyl-Pentadecanoic acid	TR
10:0	Decanoic acid	TR
16:1n7	9-Hexadecenoic acid	TR
20:4n6	5,8,11,14-Eicosatetraenoic acid	TR
17:0 anteiso	14-methyl-Hexadecanoic acid	TR
12:0	Dodecanoic acid	TR
17:0 iso	15-methyl-Hexadecanoic acid	TR

* Mean peak area percentage; TR = trace amounts < 1 %

Table 4. Nucleotide content and gene count levels of the genome

Attribute	Genome (total)	
	Value	**% of total***
Size (bp)	2,843,796	100
G+C content (%)	1,244,365	43.75
Coding region (bp)	2,498,460	87.85
Total genes	2,811	100
RNA genes	76	2.70
Protein-coding genes	2,735	97.29
Genes with function prediction	1,873	66.63
Genes assigned to COGs	862	30.66
Genes with peptide signals	319	11.34
Number of pseudogenes	13	0.46
CRISPR repeats	0	0
Nº of genes with Pfam-A domains	2,557	90.96
ORFans genes	233	8.28
Genes associated with PKS or NRPS	12	0.42
Nº of antibiotic resistance genes	0	0

* 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.

Table 5. Number of genes associated with the 25 general COG functional categories

Code	Value	% of total*	Description
J	150	5.48	Translation
A	0	0	RNA processing and modification
K	141	5.15	Transcription
L	139	5.08	Replication, recombination and repair
B	1	0.03	Chromatin structure and dynamics
D	28	1.02	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	50	1.82	Defense mechanisms
T	90	3.29	Signal transduction mechanisms
M	82	2.99	Cell wall/membrane biogenesis
N	49	1.79	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	39	1.42	Intracellular trafficking and secretion
O	82	2.99	Post-translational modification, protein turnover, chaperones
C	129	4.71	Energy production and conversion
G	81	2.96	Carbohydrate transport and metabolism
E	227	8.29	Amino acid transport and metabolism

F	62	2.26	Nucleotide transport and metabolism
H	76	2.77	Coenzyme transport and metabolism
I	66	2.41	Lipid transport and metabolism
P	155	5.66	Inorganic ion transport and metabolism
Q	30	1.09	Secondary metabolites biosynthesis, transport catabolism
R	273	9.98	General function prediction only
S	189	6.91	Function unknown
-	862	31.51	Not in COGs

*The total is based on the total number of protein coding genes in the annotated genome.

Table 6. Genome comparison of closely related species to *Rubeoparvulum massiliense* strain mt6^{T}

Name of Organisms	**INSDC**	**Size (Mb)**	**G+C (%)**	**Protein coding genes**	**Total Genes**
Rubeoparvulum massiliense strain mt6^{T}	CVPE00000000	2.84	43.75	2,735	2,811
Bacillus mannanilyticus strain AM-001	BAMO000000000.1	4.53	39.6	3,846	4,454
Brevibacillus agri strain DSM 6348^{T}	JATL00000000.1	5.51	54.2	5,047	5,297
Brevibacillus borstelensis strain DSM 6347^{T}	APBN00000000.1	5.16	52.0	4,817	5,039
Brevibacillus thermoruber strain DSM 7064	ATNE00000000.1	4.43	58.4	4,072	4,269
Caldalkalibacillus thermarum strain HA6	AFCE00000000.1	2.9	47.6	2,741	2,969

Table 7. The numbers of orthologous protein shared between genomes (upper right)*

	Bacillus mannanilyticus strain AM-001	*Brevibacillus agri* strain DSM 6348^T	*Brevibacillus borstelensis* strain DSM 6347^T	*Brevibacillus thermoruber* strain DSM 7064	*Caldalkalibacillus thermarum* strain HA6	*Rubeoparvulum massiliense* strain mt6
Bacillus mannanilyticus strain AM-001	**4,842**	1,605	1,606	1,533	1,178	1,079
Brevibacillus agri strain DSM6348^T	60.62	**5,273**	2,713	2,625	1,368	1,286
Brevibacillus borstelensis strain DSM 6347T	61.19	73.07	**5,019**	2,645	1,376	1,316
Brevibacillus thermoruber strain DSM 7064	59.20	75.04	75.55	**4,253**	1,35	1,269
Caldalkalibacillus thermarum strain HA6	66.10	62.97	63.15	63.13	**2,986**	1,039
Rubeoparvulum massiliense strain mt6	*63.12*	61.17	61.57	60.30	62.87	**2,733**

* average percentage similarity of nucleotides corresponding to orthologous protein shared between genomes (lower left) and numbers of proteins per genome (bold).

Table 8. Pairwise comparison of *Rubeoparvulum massiliense* strain mt6^T with other species using GGDC, formula 2 (DDH estimates based on identities / HSP length)* upper right.

	Bacillus mannanilyticus	*Brevibacillus agri*	*Brevibacillus borstelensis*	*Brevibacillus thermoruber*	*Caldalkalibacillus thermarum*	*Rubeoparvulum massiliense*
Bacillus mannanilyticus	100% ± 00	32.9% ±2.52	32.2% ±2.52	31.6% ±2.52	26.2% ±2.52	26.3% ±2.52
Brevibacillus agri		100% ± 00	18.8% ±2.70	19.4% ±2.73	30% ±2.52	35.4% ±2.52
Brevibacillus borstelensis			100% ± 00	20.2% ±2.80	29.9% ±2.52	30.7% ±2.52
Brevibacillus thermoruber				100% ± 00	35.1% ±2.52	23.2% ±2.52
Caldalkalibacillus thermarum					100% ± 00	27.2% ±2.52
Rubeoparvulum massiliense						100% ± 00

*The confidence intervals indicate the inherent uncertainty in estimating DDH values from intergenomic distances based on models derived from empirical test datasets (which are always limited in size). These results are in accordance with the 16S rRNA (Figure 1) and phylogenomic analyses as well as the GGDC results.

Figure 1.

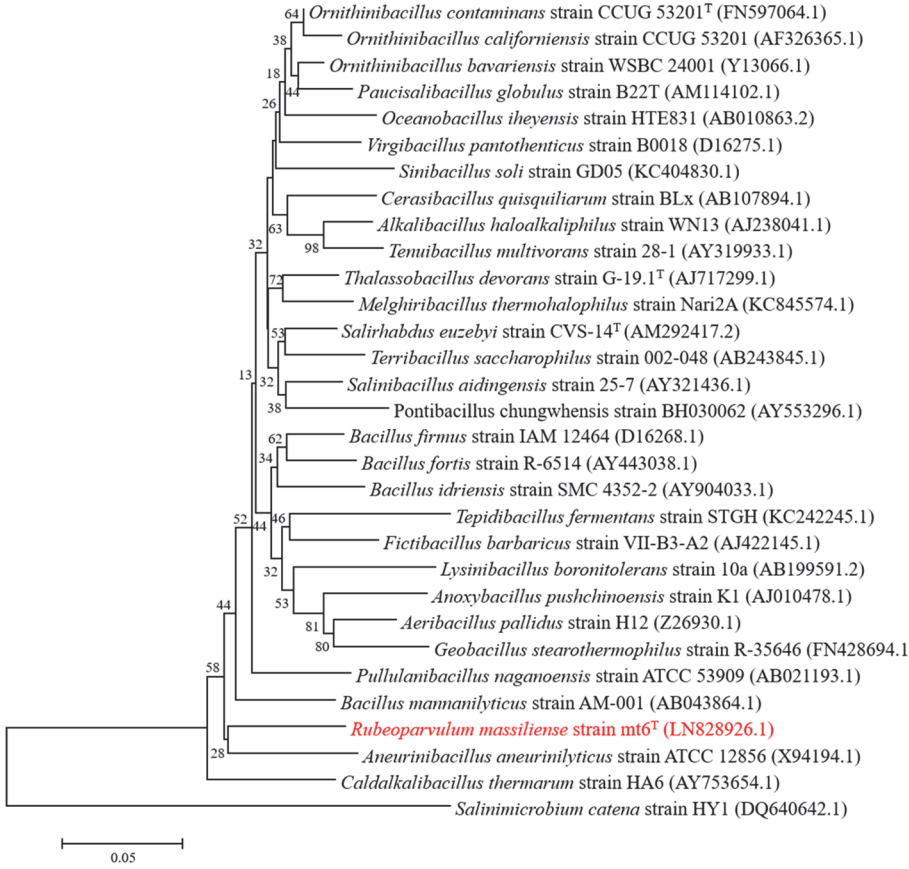

64
38
18
44
26
63
98
32
72
53
13
32
38
62
34
52
44
46
32
53
81
80
58
28
Ornithinibacillus contaminans strain CCUG 53201T (FN597064.1)
Ornithinibacillus californiensis strain CCUG 53201 (AF326365.1)
Ornithinibacillus bavariensis strain WSBC 24001 (Y13066.1)
Paucisalibacillus globulus strain B22T (AM114102.1)
Oceanobacillus iheyensis strain HTE831 (AB010863.2)
Virgibacillus pantothenticus strain B0018 (D16275.1)
Sinibacillus soli strain GD05 (KC404830.1)
Cerasibacillus quisquiliarum strain BLx (AB107894.1)
Alkalibacillus haloalkaliphilus strain WN13 (AJ238041.1)
Tenuibacillus multivorans strain 28-1 (AY319933.1)
Thalassobacillus devorans strain G-19.1T (AJ717299.1)
Melghiribacillus thermohalophilus strain Nari2A (KC845574.1)
Salirhabdus euzebyi strain CVS-14T (AM292417.2)
Terribacillus saccharophilus strain 002-048 (AB243845.1)
Salinibacillus aidingensis strain 25-7 (AY321436.1)
Pontibacillus chungwhensis strain BH030062 (AY553296.1)
Bacillus firmus strain IAM 12464 (D16268.1)
Bacillus fortis strain R-6514 (AY443038.1)
Bacillus idriensis strain SMC 4352-2 (AY904033.1)
Tepidibacillus fermentans strain STGH (KC242245.1)
Fictibacillus barbaricus strain VII-B3-A2 (AJ422145.1)
Lysinibacillus boronitolerans strain 10a (AB199591.2)
Anoxybacillus pushchinoensis strain K1 (AJ010478.1)
Aeribacillus pallidus strain H12 (Z26930.1)
Geobacillus stearothermophilus strain R-35646 (FN428694.1)
Pullulanibacillus naganoensis strain ATCC 53909 (AB021193.1)
Bacillus mannanilyticus strain AM-001 (AB043864.1)
Rubeoparvulum massiliense strain mt6T (LN828926.1)
Aneurinibacillus aneurinilyticus strain ATCC 12856 (X94194.1)
Caldalkalibacillus thermarum strain HA6 (AY753654.1)
Salinimicrobium catena strain HY1 (DQ640642.1)
0.05

Figure 2.

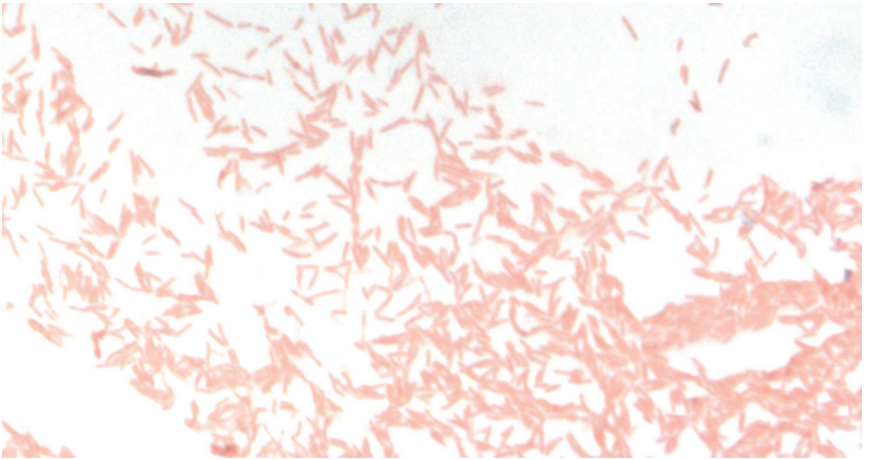


Figure 3.

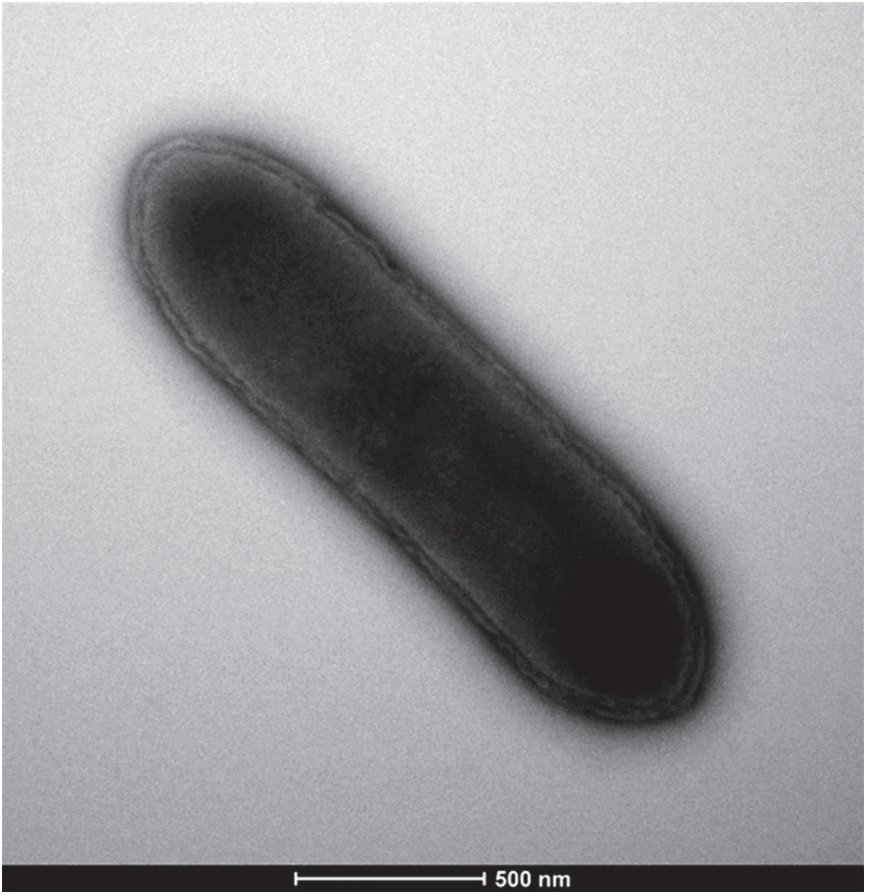


Figure 4.

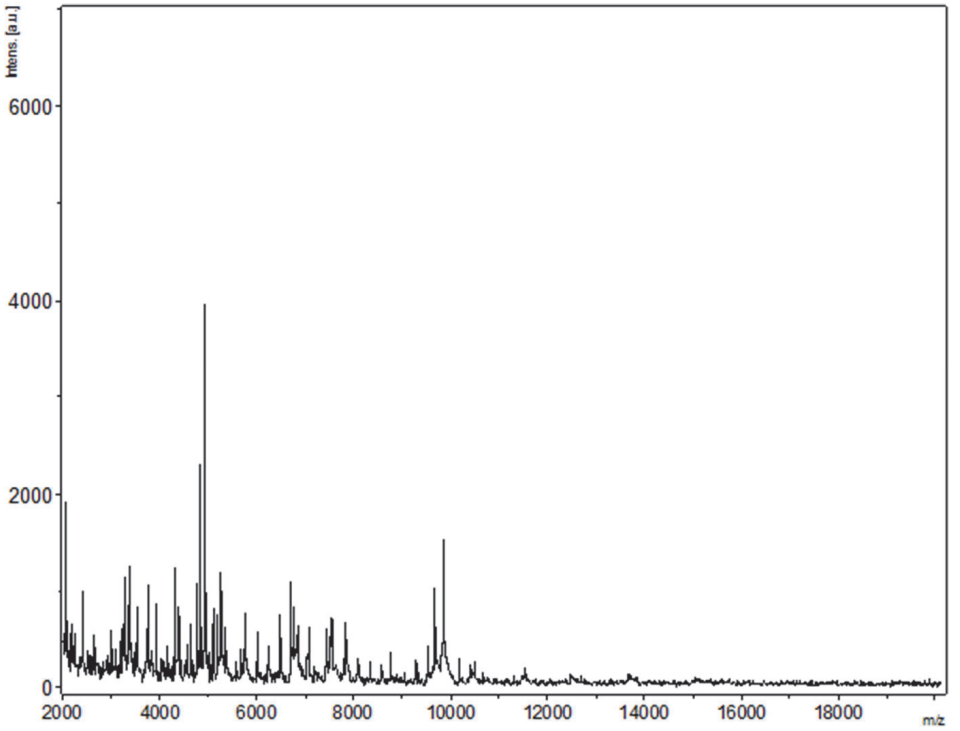

Intens. [a.u.]
6000
4000
2000
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
m/z

Figure 5.

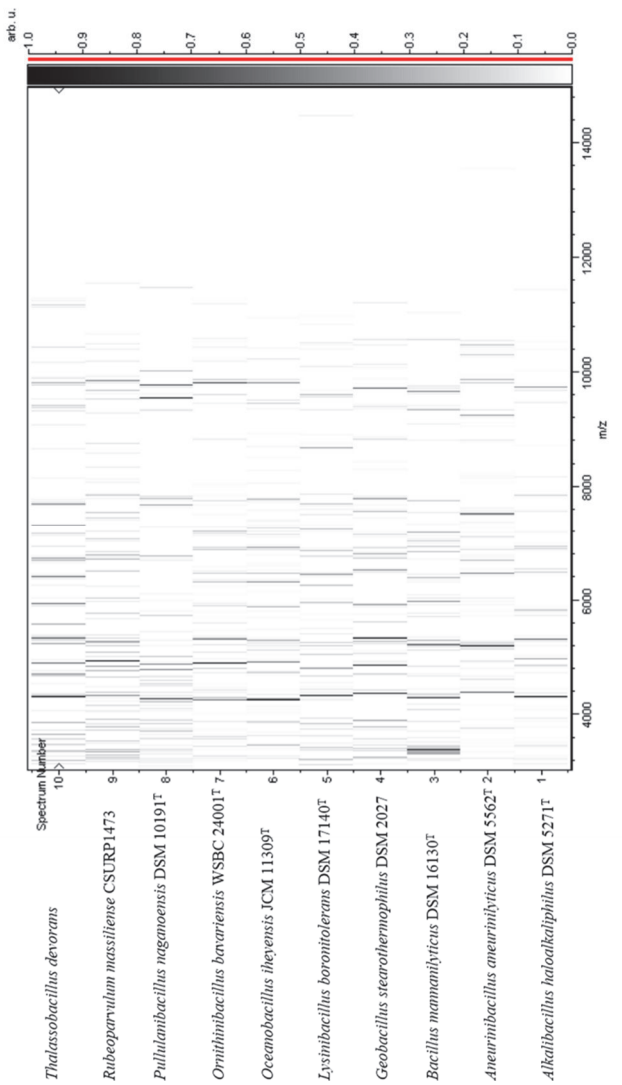

arb. u.
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Spectrum Number
10
9
8
7
6
5
4
3
2
1
4000
6000
8000
10000
12000
14000
m/z
Thalassobacillus devorans
Rubeoparvulum massiliense CSURP1473
Pullulanibacillus nagamoensis DSM 10191T
Ornithinibacillus bavariensis WSBC 24001T
Oceanobacillus iheyensis JCM 11309T
Lysinibacillus boronitolerans DSM 17140T
Geobacillus stearothermophilus DSM 2027
Bacillus mannanilyticus DSM 16130T
Aneurinibacillus aneurinilyticus DSM 5562T
Alkalibacillus haloalkaliphilus DSM 5271T

Figure 6.

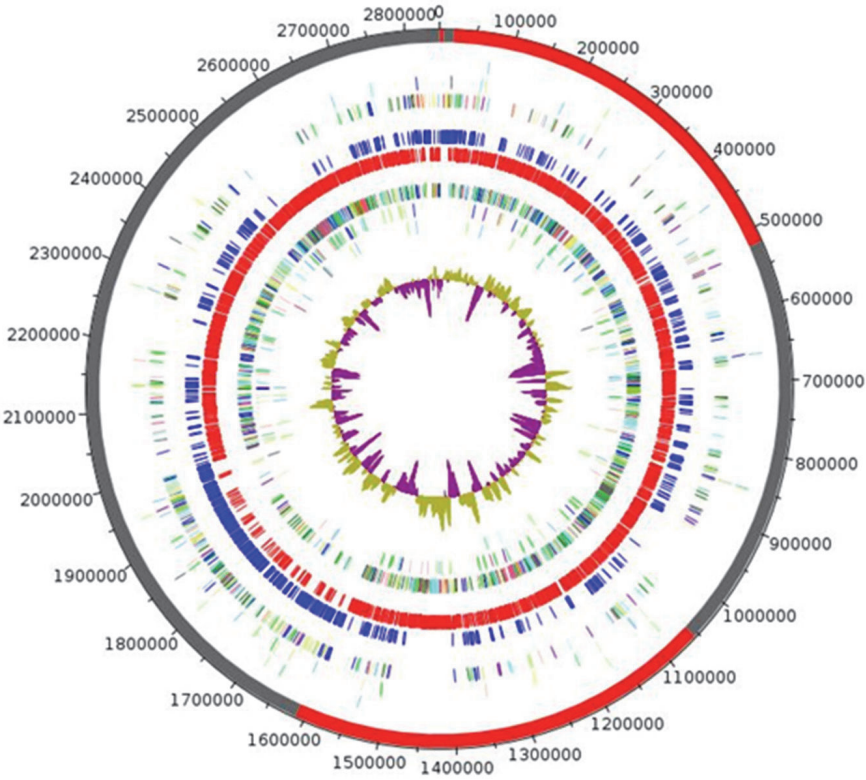
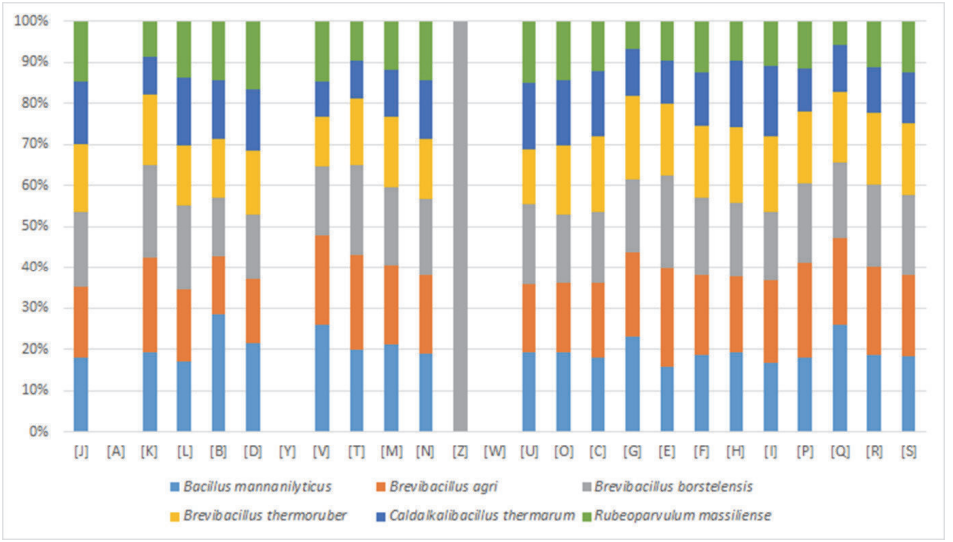

2800000
100000
200000
300000
400000
500000
600000
700000
800000
900000
1000000
1100000
1200000
1300000
1400000
1500000
1600000
1700000
1800000
1900000
2000000
2100000
2200000
2300000
2400000
2500000
2600000
2700000

Figure 7.


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
[J] [A] [K] [L] [B] [D] [Y] [V] [T] [M] [N] [Z] [W] [U] [O] [C] [G] [E] [F] [H] [I] [P] [Q] [R] [S]
Bacillus mannanilyticus
Brevibacillus agri
Brevibacillus borstelensis
Brevibacillus thermoruber
Caldalkalibacillus thermarum
Rubeoparvulum massiliense

Article 12: *Malnutritionisia massiliensis* strain mt12^{T} gen. nov., sp. nov., a new bacterial genus isolated from the gut microbiota of a severely malnourished infant.

Maryam Tidjani Alou, Jaishriram Rathored, Gregory Dubourg, Claudia Andrieu, Nicholas Armstrong, Cheikh Sokhna, Aldiouma Diallo, Didier Raoult and Pierre-Edouard Fournier.

Soumission prochaine à Antonie van Leeuwenhoek

Malnutritionisia massiliensis **gen. nov., sp. nov., a new bacterial species isolated from the gut microbiota of a severely malnourished infant.**

Maryam Tidjani Alou[1], Jaishriram Rathored[1], Caroline Michelle[1], Gregory Dubourg[1], Claudia Andrieu[1], Nicholas Armstrong[1], Cheikh Sokhna[2], Aldiouma Diallo[2], Didier Raoult[1,3] and Pierre-Edouard Fournier[1*].

[1] Aix-Marseille Université, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM63, CNRS 7278, IRD 198, INSERM 1095, 27 Boulevard Jean Moulin, 13385 Marseille, Provence-Alpes-Côte d'Azur, France.

[2] Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes IRD 198, CNRS 7278, Aix-Marseille Université, Marseille, France; Campus Commun UCAD-IRD of Hann, Dakar, Senegal

[3] Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

*Corresponding author: Pierre-Edouard Fournier

(pierre-edouard.fournier@univ-amu.fr)

tel: (33) 4 91 38 55 17

fax: (33) 4 91 38 77 72

Abstract

Malnutritionisia massiliensis gen. nov., sp. nov., strain Mt12^{T} (=CSUR P1907 = DSM 100590) was isolated from the fecal sample of a seven-month-old girl from Senegal. This bacterium is a strictly anaerobic, spore-forming Gram-stain-negative bacilli. The major cellular fatty acid was tetradecanoic acid. Its 16S rRNA exhibited 94.9% similarity with *Crassaminicella profunda* strain Ra1766H^{T}, the closest species with a validly published name. The genome of strain Mt12^{T} is 3,497,275-bp long with a 30.45 % of G+C content. 3,397 genes were predicted, including 3,268 protein-coding genes and 129 RNAs, including eight 16S rRNAs.

Keywords: Culturomics; genome; gut microbiota; *Malnutritionisia massiliensis*; Taxonogenomics.

Abbreviations

AGIOS: Average of Genomic Identity of Orthologous gene Sequences

bp: base pairs

COG: Clusters of Orthologous Groups

CSUR: Collection de souches de l'Unité des Rickettsies

DDH: DNA-DNA Hybridization

DSM: Deutsche Sammlung von Mikroorganismen

FAME: Fatty Acid Methyl Ester

GC/MS: Gas Chromatography/Mass Spectrometry

kb: kilobases

MALDI-TOF MS: Matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry

ORF: Open Reading Frame

TE buffer: Tris-EDTA buffer

SDS: sodium dodecyl sulfate

URMITE: Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergent

Introduction

The gut microbiota is a fairly complex ecosystem involved in many aspects of human health. The gut microbiota is heavily implicated in digestion, mainly through the metabolism of non-digestible carbohydrates, in antimicrobial protection, and in immunomodulation in the gut. It also contributes to the integrity and maintenance of the intestinal barrier (Jandhyala *et al.* 2015). The gut microbiota consists essentially of bacteria but also of viruses, archaea, fungi and protozoans (Lagier *et al.* 2012b). Colonization of the gastrointestinal tract starts at birth and its composition reaches a mature and stable state between the ages of two to three years old (Palmer *et al.* 2007). The gut microbiota composition is dependent on various factors such as age, diet, environment, genetics and gut wall structure (Graf *et al.* 2015; Sankar *et al.* 2015). The major bacterial phyla include the *Firmicutes*, *Bacteroidetes*, *Actinobacteria* and *Proteobacteria* (Jandhyala *et al.* 2015).

Dysbiosis of the gut microbiota is instrumental in several diseases such as inflammatory bowel disease, obesity, type 2 diabetes (Sankar *et al.* 2015) and severe acute malnutrition (Million *et al.* 2016). In order to investigate the implication of the gut microbiota in severe acute malnutrition, a large-scale study was conducted using “culturomics”, a high throughput

culture method which consists in varying physico-chemical parameters of culture to explore the gut microbiota as thoroughly as possible (Lagier *et al.* 2012a). In the process, we isolated a new genus, a member of the family *Clostridiaceae*. The family *Clostridiaceae* is currently made of 30 genera which contain strictly anaerobic species, most of which are Gram-stain-positive and spore-forming (Parte 2014). To describe this new genus, we used the taxonogenomics approach (Fournier *et al.* 2015). Taxonogenomics mixes next generation sequencing, phylogenetic and phenotypic characteristics including a MALDI-TOF mass spectrometry protein profile to describe new bacterial species. We herein describe *Malnutritionisia massiliensis* strain Mt12^{T} (=CSUR P1907 =DSM 100590).

Material and methods

Sample informations

A stool sample was collected from a seven-month-old girl from Senegal who suffered from Kwashiorkor, a severe form of acute malnutrition. She had a height-for-age z-score of -0.17 and presented with a nutritional edema. The patient was not receiving antibiotics at the time of stool collection. The specimen was preserved at +4 °C for seven days until it was transferred to our laboratory in Marseille where it was stored at -

80 °C until further use. This child's parents gave their oral informed consent to participation in this culturomics study of the gut microbiota of malnourished children. The study was approved by the Institut Fédératif de Recherche 48 (Faculty of Medicine, Marseille, France) under agreement number 09-022.

Strain identification by MALDI-TOF MS and 16S rRNA sequencing

The stool sample was cultured using 18 culture conditions as previously described (Lagier *et al.* 2015). Colonies were purified through subculture and identified by MALDI-TOF MS using a Microflex spectrometer and a MTP 96 MALDI-TOF target plate (Bruker Daltonics, Leipzig, Germany), as described previously (Seng *et al.* 2009; Seng *et al.* 2013). The spectra obtained for each colony were matched against the MALDI Biotyper software version 3.0 (Bruker) and URMITE databases using standard pattern matching (with default parameter settings). Identification scores used were as follows: a score over 1.9 allowed identification at the species level while a score under 1.9 did not allow any identification. In the latter case, the colony was identified by sequencing its 16S rRNA using the fD1 and rP2 primers, as previously described (Drancourt *et al.* 2000). According to Kim *et al.* (Kim *et al.* 2014), a similarity threshold of 98.65% was used to define a new species whereas a

threshold of 95% was used to define a new genus without performing DNA-DNA hybridization. Upon identification, twelve MALDI-TOF MS spectra for strain Mt12^T were incremented in URMITE database.

Growth conditions

Ideal growth conditions for strain Mt12^T were determined by testing five growth temperatures (25, 28, 37, 45 and 56 °C) in an aerobic atmosphere with or without 5% CO_2, and under anaerobic and microaerophilic conditions using the GENbag Anaer and GENbag microaer systems, respectively (BioMerieux, Marcy l'Etoile, France). Different pH values (6, 6.5, 7, 8.5) and NaCl concentrations (0.5, 1, 5, 7.5 and 10%) were also tested.

Morphological, biochemical and antibiotic susceptibility tests

Phenotypic characteristics such as Gram staining, motility, sporulation, catalase and oxidase activities were tested as previously described (Lagier *et al.* 2015). Biochemical analysis of strain Mt12^T was carried out using API 50CH, API 20A, API ZYM strips (BioMérieux) in an anaerobic atmosphere. Antibiotic susceptibility was tested using the disk diffusion method (Matuschek *et al.* 2014) and according to EUCAST 2015 recommendations.

Transmission electron microscopy was performed as follows: cells were fixed with 2.5 % glutaraldehyde in 0.1M cacodylate buffer for at least 1h at 4°C. A drop of cell suspension was deposited for approximately 5 minutes on glow-discharged formvar carbon film on 400 mesh nickel grids (FCF400-Ni, 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 G20 Cryo (FEI) transmission electron microscope operated at 200 keV.

Cellular fatty acid methyl ester (FAME) analysis was performed by GC/MS. Two samples were prepared with approximately 30 mg of bacterial biomass per tube harvested from several culture plates. Fatty acid methyl esters were prepared as described by Sasser (Sasser 2006). GC/MS analyses were carried out as described before (Dione *et al.* 2016). Briefly, fatty acid methyl esters were separated using an Elite 5-MS column and monitored by mass spectrometry (Clarus 500 - SQ 8 S, Perkin Elmer, Courtaboeuf, France). A 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).

Genomic DNA preparation

After lysis pretreatements by a lysozyme incubation at 37 °C for 2 hours followed by a proteinase K action respectively, DNA was extracted on the EZ1 biorobot (Qiagen) with EZ1 DNA tissues kit. The elution volume is 50µL. gDNA was quantified by a Qubit assay with the high sensitivity kit (Life technologies, Carlsbad, CA, USA) to 123.7 ng/µl.

Genome sequencing

The genomic DNA (gDNA) of strain Mt12^{T} was sequenced using a MiSeq sequencer (Illumina Inc, San Diego, CA, USA) and the mate pair strategy. The gDNA was barcoded in order to be mixed with 11 other projects using 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 fragmentation pattern 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 11kb, with an optimal size at 4.031 kb. No size selection was performed and 385.5ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal size of 1070 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 2.40 nmol/l. The libraries were normalized at 2nM 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. The automated cluster generation and sequencing run were performed in a single 2x301-bp run. Total information of 7.3 Gb was obtained from a 511 K/mm2 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 strain Mt12^{T} was determined as 9.51%. The 1,149,152 paired reads were trimmed then assembled to 14 scaffolds.

Genome annotation and analysis

Open Reading Frames (ORFs) were predicted using Prodigal (Hyatt *et al.* 2010) with default parameters but the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched against the GenBank (Benson *et al.* 2012) and the Clusters of Orthologous Groups (COG) databases 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 an E-value of $1e^{-03}$, a coverage of 0.7 and an identity percentage of 30%, and if the sequence length was smaller than 80 amino acids (aa), we used an E-value of $1e^{-05}$. The tRNAScanSE tool (Lowe and Eddy 1997) was used to find tRNA genes, while ribosomal RNAs were found using RNAmmer (Lagesen *et al.* 2007). Lipoprotein signal peptides and the number of transmembrane helices were predicted using Phobius (Käll *et al.* 2004). Mobile genetic elements were predicted using PHAST (Zhou *et al.* 2011) and RAST (Aziz *et al.* 2008). ORFans were identified if all the BLASTP performed did not give positive results (*E*-value smaller than $1e^{-03}$ for ORFs with sequence size larger than 80 aa or *E*-value smaller than $1e^{-05}$ for ORFs with sequence length smaller 80 aa). Such parameter thresholds have already been used in previous studies to define ORFans. Artemis (Rutherford *et al.* 2000) and DNA Plotter (Carver *et al.* 2009) were used for data management and the visualization of genomic features, respectively. The Mauve alignment tool (version 2.3.1) was used for multiple genomic sequence alignment (Darling *et al.* 2004).

Comparator species for genomic comparison were identified in the 16S RNA tree using the Phylopattern software (Gouret *et al.* 2009). The genome of strain $Mt12^T$ was compared to those of

Alkaliphilus metalliredigens strain QYMF, *Clostridium aceticum* strain DSM 1496, *Alkaliphilus transvaalensis* strain SAGM1 and *Alkaliphilus oremlandii* strain OhILAs.

For each selected genome, the complete genome sequence, proteome genome sequence and Orfeome genome sequence were retrieved from the FTP of NCBI. 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). Annotation and comparison processes were performed in the Multi-Agent software system DAGOBAH (Gouret *et al.* 2011) which includes Figenix (Gouret *et al.* 2005) libraries which provide pipeline analysis. To evaluate the genomic similarity between studied genomes, we determined two parameters, digital DDH (dDDH) which exhibits a high correlation with DDH (Auch *et al.* 2010; Meier-Kolthoff *et al.* 2013) and AGIOS (Ramasamy *et al.* 2014), which was designed to be independent from DDH (Ramasamy *et al.* 2014). The AGIOS score (Ramasamy *et al.* 2014) is the mean value of nucleotide similarity between all couples of orthologous proteins between the two studied genomes (Ramasamy *et al.* 2014).

Results

Strain identification and phylogenetic analyses

Primo-isolation of strain Mt12^T occurred after a 10-day pre-incubation of the stool sample in an anaerobic blood culture bottle supplemented with 0.2µ filter-sterilized rumen and sheep blood and seeding on 5% sheep blood-enriched Colombia agar in anaerobic atmosphere conditions at 37 °C. Strain Mt12^T could not be identified using MALDI-TOF MS and, therefore, the 16S rRNA was sequenced. The resulting sequence (Accession number LN850734) revealed a 94.9% similarity level with the 16S rRNA of *Crassaminicella profunda* strain Ra1766H^T, the closest species with a validly published name (Figure 1). According to the current recommended thresholds (Kim *et al.* 2014), strain Mt12^T may, therefore, be a representative strain of a new genus within the family *Clostridiaceae* for which we suggest the name *Malnutritionisia*, type species *Malnutritionisia massiliensis*, type strain Mt12^T (=CSUR P1907 = DSM 100590) (Table 1). The reference protein spectra for *Malnutritionisia massiliensis* (Figure 2) were incremented in URMITE database [http://www.mediterranee-infection.com/article.php?laref=256&titre=urms-database].

Phenotypic description

Growth was observed between 25 and 37 °C in anaerobic conditions. No growth occurred in aerobic and microaerophilic conditions. The ideal growth condition was observed after 48 hours in anaerobic conditions. Growth was observed at all tested pH values (6, 6.5, 7, 8.5) and only at the minimal concentration of NaCl (0.5%). Cells were motile, spore-forming Gram-stain-negative rods (Figure 3) which formed irregular translucent colonies with a white center and a mean diameter of 3 mm. Spores were terminal and deforming (Figure 4). Negative staining visualized with electron microscopy revealed bacilli with a mean length and width of 6.9 µm and 0.5 µm, respectively (Figure 4). The major cellular fatty acids were tetradecanoic acid (46 %), 9-Hexadecenoic acid (21 % and Hexadecanoic acid (18 %). Butanoic acid, a short chain fatty acid, was also detected as shown in Table 2. *Crassaminicella profunda,* the closest related strain, presented a similar cellular fatty acid profile as described previously (Lakhal *et al.* 2015).

Catalase and oxidase activities were absent. Using an API ZYM strip, other enzymatic activities were present, such as esterase C4, esterase lipase C8, leucine arylamidase, valine arylamidase, naphtol-AS-BI-phosphohydrolase and β-

galactosidase. No activity was detected for the following enzymes: alkaline phosphatase, lipase C14, cysteine arylamidase, trypsin, α-chymotrypsin, acid phosphatase, α-galactosidase, β-glucuronidase, α-glucosidase, β-glucosidase, N-acetyl-β-glucosaminidase, α-mannosidase and α-fucosidase. Using an API 20A strip, indole production was present as well as β-glucosidase activity as revealed by the hydrolysis of ferric citrate esculin. Protease and urease activities were negative. Acid was produced from D-glucose, D-lactose, D-sucrose, D-maltose, salicin, D-cellobiose, D-mannose, D-raffinose, D-sorbitol and D-trehalose. No acid production was observed from D-mannitol, D-xylose, L-arabinose, glycerol, D-melezitose and L-rhamnose. Only a few carbohydrates were metabolized: D-ribose, D-melibiose, glycogen, D-tagatose and potassium 5-ketogluconate as revealed by an API 50 CH strip. The other tested carbohydrates (glycerol, erythritol, D-arabinose, L-arabinose, D-xylose, L-xylose, D-adonitol, methyl-βD-xylopyranoside, D-galactose, D-glucose, D-fructose, D-mannose, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl-αD-mannopyranoside, methyl-αD-glucopyranoside, N-acetylglucosamin, amygdalin, arbutin, esculin ferric citrate, salicin, D-cellobiose, D-maltose, D-lactose, D-sucrose, D-trehalose, inulin, D-melezitose, D-raffinose,

starch, xylitol, gentiobiose, D-turanose, D-lyxose, D-fucose, L-fucose, D-arabitol, L-arabitol, potassium gluconate, and potassium 2-ketogluconate) were not utilized.

Strain Mt12^T was susceptible to rifampicin, gentamicin 500 µg, tobramycin, penicillin G, oxacillin, ceftriaxone, doxycycline, and ciprofloxacin but was resistant to gentamicin 30 µg, imipenem, trimethoprim/sulfamethoxazole, teicoplanin, metronidazole, colistin, and erythromycin. Differential characteristics between strain Mt12^T and close relatives are presented in Table 3. Strain Mt12^T differed from other compared species in terms of several phenotypic characteristics including Gram staining, salt tolerance and habitat.

Genome properties

The genome of strain Mt12^T is 3,497,275-bp long with a 30.45% G+C content (Figure 5 and Table 4). It is composed of 14 scaffolds and 15 contigs. Of the 3,397 predicted genes, 3,268 were protein-coding genes, and 129 were RNAs (10 5S rRNAs, 8 16S rRNAs, 4 23S rRNAs, 107 tRNAs). A total of 2,400 genes (73.44 %) were assigned a putative function (through comparison with the COGs or NR databases). A total of 272 genes were identified as ORFans (8.32%). The 448 remaining genes were annotated as hypothetical proteins (13.71%). The properties and statistics of the genome are summarized in Table

4 while the distribution of genes into COG functional categories is presented in Table 5.

Genome comparison

Genomic characteristics of strain Mt12^T were compared to those of closely related species with an available genome (Table 6). The genome size of strain Mt12^T (3.49 Mb) is larger than that of *Alkaliphilus oremlandii* (3.12Mb) but smaller than those of *Alkaliphilus metalliredigens*, *Clostridium aceticum* and *Alkaliphilus transvaalensis* (4.93, 4.2 and 4.02 respectively). The G+C content of strain Mt12^T (30.44%) is lower than those of all compared species. Similarly, he total number of genes (3,397) and the number of protein-coding genes (3,268) of strain Mt12^T were smaller than that of all compared species excepted *A. oremlandii* (3,016). The distribution of genes into COG categories is similar for all compared species with the exception of the extracellular structures category which is only present in strain Mt12^T (Figure 6). Among species with standing in nomenclature, AGIOS values ranged from 68.30 between *A. transvaalensis* and *A. oremlandii* to 69.74 between *C. aceticum* and *A. metalliredigens*. When compared to strain Mt12^T, AGIOS values ranged from 66.56 with *A. metalliredigens* to 67.69 with *C. aceticum* (Table 7). Among species with standing in nomenclature, dDDH values ranged from 12.5% between *A.*

oremlandii and *A. transvaalensis* to 26.8% between *A. metalliredigens* and *A. oremlandii*. dDDH values between strain Mt12^T and compared species ranged from 16.2% with *A. oremlandii* to 29% with *A. metalliredigens* (Table 8).

Conclusion

Considering the specific phenotypic properties of strain Mt12^T, including its low matching MALDI-TOF MS score, the 94.9% 16S rRNA similarity level with *Crassaminicella profunda*, and its genomic analysis, we propose the creation of a new genus within the family *Clostridiaceae*, named *Malnutritionisia*, with *Malnutritionisia massiliensis* as type species and strain Mt12^T as type strain.

Description of *Malnutritionisia* gen. nov.

Malnutritionisia (Mal.nu.tri.tio.ni'sia, N. L. fem. Malnutritionisia from malnutrition, the latin word for undernutrition).

Strictly anaerobic, spore-forming, Gram-stain-negative rod-shaped bacteria. Oxidase and catalase negative. Urease negative. β-glucosidase positive. Indole is produced. The major cellular fatty acid was tetradecanoic acid. Habitat is the human gut. The type species is *Malnutritionisia massiliensis.*

Description of *Malnutritionisia massiliensis* sp. nov.

Malnutritionisia massiliensis (mas.il'ien'sis. L. gen. masc., massiliensis, of Massilia, the Latin name of Marseille, where strain Mt12^T was isolated).

Cells are anaerobic, Gram-stain-negative and motile bacilli with a mean length of 6.9 µm and a mean diameter of 0.5 µm. Strain Mt12^T formed irregular translucent colonies with a white center and a mean diameter of 3 mm. The spore is terminal and deforming. Mesophilic. Positive enzymatic activities included esterase C4, esterase lipase C8, leucine arylamidase, valine arylamidase, naphtol-AS-BI-phosphohydrolase and β-galactosidase. No activity was found for the following enzymes: protease, urease, alkaline phosphatase, lipase C14, cysteine arylamidase, trypsin, α-chymotrypsin, acid phosphatase, α-galactosidase, β-glucuronidase, α-glucosidase, β-glucosidase, N-acetyl-β-glucosaminidase, α-mannosidase and α-fucosidase. Acid was produced from D-glucose, D-lactose, D-sucrose, D-maltose, Salicin, D-cellobiose, D-mannose, D-raffinose, D-sorbitol and D-trehalose. No acid production was observed from D-mannitol, D-xylose, L-arabinose, glycerol, D-melezitose and L-rhamnose. Only a few carbohydrates were metabolized: D-ribose, D-melibiose, glycogen, D-tagatose and potassium 5-ketogluconate as revealed by an API 50 CH strip. The other

tested carbohydrates (glycerol, erythritol, D-arabinose, L-arabinose, D-xylose, L-xylose, D-adonitol, methyl-βD-xylopyranoside, D-galactose, D-glucose, D-fructose, D-mannose, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl-αD-mannopyranoside, methyl-αD-glucopyranoside, N-acetylglucosamin, amygdalin, arbutin, aesculin ferric citrate, salicin, D-cellobiose, D-maltose, D-lactose, D-sucrose, D-trehalose, inulin, D-melezitose, D-raffinose, starch, xylitol, gentiobiose, D-turanose, D-lyxose, D-fucose, L-fucose, D-arabitol, L-arabitol, potassium gluconate, and potassium 2-ketogluconate) were not utilized. Strain Mt12^T was susceptible to rifampicin, gentamicin 500 µg, tobramycin, penicillin G, oxacillin, ceftriaxone, doxycycline, and ciprofloxacin but resistant to gentamicin 30 µg, imipenem, trimethoprim/ sulfamethoxazole, teicoplanin, metronidazole, colistin, and erythromycin.

The genome of strain Mt12^T is 3,497,275 bp long with 30.45% G+C content. The 16S rRNA and genome sequences are available in the EBI/EMBL database under accession numbers LN850734 and CXYX00000000, respectively. The type strain Mt12^T (=CSUR P1907 =DSM 100590) was isolated from the stool sample of a seven-month-old girl from Senegal with Kwashiorkor.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

The authors thank the Xegen Company (www.xegen.fr) for automating the genomic annotation process. This study was funded by the “Fondation Méditerranée Infection”.

We thank TradOnline for English reviewing.

List of references.

Alain K, Pignet P, Zbinden M, *et al* (2002) *Caminicella sporogenes* gen. nov., sp. nov., a novel thermophilic spore-forming bacterium isolated from an East-Pacific Rise hydrothermal vent. Int J Syst Evol Microbiol 52:1621–1628. doi: 10.1099/00207713-52-5-1621

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. doi: 10.4056/sigs.531120

Aziz RK, Bartels D, Best AA, *et al* (2008) The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75. doi: 10.1186/1471-2164-9-75

Benson DA, Karsch-Mizrachi I, Clark K, *et al* (2012) GenBank. Nucleic Acids Res 40:D48-53. doi: 10.1093/nar/gkr1202

Brisbarre N, Fardeau M-L, Cueff V, *et al* (2003) *Clostridium caminithermale* sp. nov., a slightly halophilic and moderately thermophilic bacterium isolated from an Atlantic deep-sea hydrothermal chimney. Int J Syst Evol Microbiol 53:1043–1049. doi: 10.1099/ijs.0.02471-0

Carver T, Thomson N, Bleasby A, *et al* (2009) DNAPlotter: circular and linear interactive genome visualization. Bioinforma Oxf Engl 25:119–120. doi: 10.1093/bioinformatics/btn578

Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. doi: 10.1101/gr.2289704

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. doi: 10.1016/j.nmni.2016.01.002

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.

Fendrich C, Hippe H, Gottschalk G (1990) *Clostridium halophilium* sp. nov. and *C. litorale* sp. nov., an obligate halophilic and a marine species degrading betaine in the Stickland reaction. Arch Microbiol 154:127–132. doi: 10.1007/BF00423321

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. doi: 10.1016/j.anaerobe.2015.10.011

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

Gouret P, Thompson JD, Pontarotti P (2009) PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinformatics 10:298. doi: 10.1186/1471-2105-10-298

Gouret P, Vitiello V, Balandraud N, *et al* (2005) FIGENIX: Intelligent automation of genomic annotation: expertise integration in a new software platform. BMC Bioinformatics 6:198. doi: 10.1186/1471-2105-6-198

Graf D, Di Cagno R, Fåk F, *et al* (2015) Contribution of diet to the composition of the human gut microbiota. Microb Ecol Health Dis 26:26164.

Hong H, Kim S-J, Min U-G, *et al* (2015) *Anaerosolibacter carboniphilus* gen. nov., sp. nov., a strictly anaerobic iron-reducing bacterium isolated from coal-contaminated soil. Int J Syst Evol Microbiol 65:1480–1485. doi: 10.1099/ijs.0.000124

Hyatt D, Chen G-L, Locascio PF, *et al* (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. doi: 10.1186/1471-2105-11-119

Jandhyala SM, Talukdar R, Subramanyam C, *et al* (2015) Role of the normal gut microbiota. World J Gastroenterol 21:8787–8803. doi: 10.3748/wjg.v21.i29.8787

Käll L, Krogh A, Sonnhammer ELL (2004) A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338:1027–1036. doi: 10.1016/j.jmb.2004.03.016

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. doi: 10.1099/ijs.0.059774-0

Klouche N, Fardeau M-L, Lascourreges J-F, *et al* (2007) *Geosporobacter subterraneus* gen. nov., sp. nov., a spore-forming bacterium isolated from a deep subsurface aquifer. Int J Syst Evol Microbiol 57:1757–1761. doi: 10.1099/ijs.0.64642-0

Lagesen K, Hallin P, Rødland EA, *et al* (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108. doi: 10.1093/nar/gkm160

Lagier J-C, Armougom F, Million M, *et al* (2012a) Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 18:1185–1193. doi: 10.1111/1469-0691.12023

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. doi: 10.1128/CMR.00014-14

Lagier J-C, Million M, Hugon P, *et al* (2012b) Human Gut Microbiota: Repertoire and Variations. Front Cell Infect Microbiol. doi: 10.3389/fcimb.2012.00136

Lakhal R, Pradel N, Postec A, *et al* (2015) *Crassaminicella profunda* gen. nov., sp. nov., an anaerobic marine bacterium isolated from deep-sea sediments. Int J Syst

Evol Microbiol 65:3097–3102. doi: 10.1099/ijsem.0.000386

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.

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 Off Publ Eur Soc Clin Microbiol Infect Dis 20:O255-266. doi: 10.1111/1469-0691.12373

Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14:60. doi: 10.1186/1471-2105-14-60

Million M, Diallo A, Raoult D (2016) Gut microbiota and malnutrition. Microb Pathog. doi: 10.1016/j.micpath.2016.02.003

Ogg CD, Patel BKC (2009) *Thermotalea metallivorans* gen. nov., sp. nov., a thermophilic, anaerobic bacterium from the Great Artesian Basin of Australia aquifer. Int J Syst Evol Microbiol 59:964–971. doi: 10.1099/ijs.0.004218-0

Palmer C, Bik EM, DiGiulio DB, *et al* (2007) Development of the human infant intestinal microbiota. PLoS Biol 5:e177. doi: 10.1371/journal.pbio.0050177

Parte AC (2014) LPSN—list of prokaryotic names with standing in nomenclature. Nucleic Acids Res 42:D613–D616. doi: 10.1093/nar/gkt1111

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. doi: 10.1099/ijs.0.057091-0

Rutherford K, Parkhill J, Crook J, *et al* (2000) Artemis: sequence visualization and annotation. Bioinforma Oxf Engl 16:944–945.

Sankar SA, Lagier J-C, Pontarotti P, *et al* (2015) The human gut microbiome, a taxonomic conundrum. Syst Appl Microbiol 38:276–286. doi: 10.1016/j.syapm.2015.03.004

Sasser M (2006) Bacterial identification by gas chromatographic analysis of fatty acids methyl esters (GC-FAME). Microbial ID Inc. Newark, NY

Seng P, Abat C, Rolain JM, *et al* (2013) 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. doi: 10.1128/JCM.00492-13

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. doi: 10.1086/600885

Tan H-Q, Wu X-Y, Zhang X-Q, *et al* (2012) *Tepidibacter mesophilus* sp. nov., a mesophilic fermentative anaerobe isolated from soil polluted by crude oil, and emended description of the genus *Tepidibacter*. Int J Syst Evol Microbiol 62:66–70. doi: 10.1099/ijs.0.027409-0

Zhang Y-Z, Fang M-X, Zhang W-W, *et al* (2013) *Salimesophilobacter vulgaris* gen. nov., sp. nov., an anaerobic bacterium isolated from paper-mill wastewater. Int J Syst Evol Microbiol 63:1317–1322. doi: 10.1099/ijs.0.040915-0

Zhou Y, Liang Y, Lynch KH, *et al* (2011) PHAST: a fast phage search tool. Nucleic Acids Res 39:W347-352. doi: 10.1093/nar/gkr485

Figure legends

Fig. 1 Phylogenetic tree highlighting the position of *Malnutritionisia massiliensis* strain Mt12^T relative to other close strains. The respective GenBank accession numbers for 16S rRNA genes are indicated in parenthesis. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using the maximum-likelihood method within the MEGA6 software. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 1,000 times to generate a majority consensus tree. *Clostridium felsineum* strain NCIMB 10690 was used as an outgroup. The scale bar represents a 2% nucleotide sequence divergence.

Fig. 2. Reference mass spectrum from *Malnutritionisia massiliensis* strain Mt12^T. Spectra from 12 individual colonies were compared and a reference spectrum was generated.

Fig. 3. Gram staining of *Malnutritionisia massiliensis* strain Mt12^T

Fig. 4. Transmission electron microscopy of *Malnutritionisia massiliensis* strain Mt12^T using a Tecnai G20 transmission electron microscope (FEI Company) at operating voltage of 200 keV. The scale bar represents 500 nm.

Fig. 5. Graphical circular map of the chromosome. From outside to the center: Genes on the forward strain colored by COG categories (only gene assigned to COG), RNA genes (tRNAs green, rRNAs red), G+C content and G+C skew. COGs, Clusters of Orthologous Groups database.

Fig. 6. Distribution of functional classes of predicted genes according to the clusters of orthologous groups of protein.

Table 1: Classification and general features of *Malnutritionisia massiliensis* strain Mt12^{T}

Property	Term
Current classification	Domain: *Bacteria*
	Phylum: *Firmicutes*
	Class: *Clostridia*
	Order: *Clostridiales*
	Family: *Clostridiaceae*
	Genus: *Malnutritionisia*
	Species: *Malnutritionisia massiliensis*
	Type strain: Mt12
Gram stain	Negative
Cell shape	Rod
Motility	Motile
Sporulation	Sporulating
Temperature range	Mesophilic
Optimum temperature	37°C

Table 2. Cellular fatty acid composition (%)

Fatty acids	Name	Mean relative % (a)
14:0	Tetradecanoic acid	46.4 ± 3.4
16:1n7	9-Hexadecenoic acid	20.5 ± 2.4
16:0	Hexadecanoic acid	18.1 ± 0.6
16:1n9	7-Hexadecenoic acid	6.9 ± 1.0
4:0	Butanoic acid	3.1 ± 0.6
18:1n9	9-Octadecenoic acid	1.8 ± 0.3
18:0	Octadecanoic acid	1.7 ± 0.3
18:2n6	9,12-Octadecadienoic acid	TR
14:1	Tetradecenoic acid	TR

[a] Mean peak area percentage; TR = trace amounts < 1 %

Table 3: Differential characteristics of *Malnutritionisia massiliensis* strain Mt12^T DSM 100590, *Crassaminicella profunda* strain Ra1766H^T DSM 27501, *Salimesophilobacter vulgaris* Zn2^T DSM 24770, *Anaerosolibacter carboniphilus* strain IRF19^T JCM 19988, *Geosporobacter subterraneus* strain VNs68^T DSM 17957, *Thermotalea metallivorans* strain B2-1^T DSM 21119, *Caminicella sporogenes* strain AM1114^T DSM 14501, *Clostridium caminithermale* strain DVird3^T DSM 15212, *Clostridium halophilium* strain W6 DSM 5388, *Tepidibacter mesophilus* strain B1^T JCM 16806 (Fendrich *et al.* 1990; Alain *et al.* 2002; Brisbarre *et al.* 2003; Klouche *et al.* 2007; Ogg and Patel 2009; Tan *et al.* 2012; Zhang *et al.* 2013; Hong *et al.* 2015; Lakhal *et al.* 2015).

Properties	*Malnutritionisia massiliensis*	*Crassaminicella profunda*	*Salimesophilobacter vulgaris*	*Anaerosalibacter carboniphilus*	*Geosporobacter subterraneus*	*Thermotalea metallivorans*	*Caminicella sporogenes*	*Clostridium caminithermale*	*Clostridium halophilium*	*Tepidibacter mesophilus*
Cell diameter (µm)	0.5-0.7	1.0-2.0	0.5-0.8	0.4-0.6	0.5	0.6-0.7	0.5-0.7	0.4-0.5	0.8-1.0	1.1-1.6
Gram stain	-	+	+	-	+	-	-	+	+	+
Salt tolerance	-	+	+	+	+	+	+	+	+	+
Motility	+	+	+	+	-	+	+	+	+	+
Endospore formation	+	+	-	-	+	-	+	+	+	+
Major cellular fatty acid	14:0	14:0	15:0	iso 15:0	na	na	na	na	na	iso 15:0

Properties	*Malnutritionisia massiliensis*	*Crassaminicella profunda*	*Salimesophilobacter vulgaris*	*Anaerosalibacter carboniphilus*	*Geosporobacter subterraneus*	*Thermotalea metallivorans*	*Caminicella sporogenes*	*Clostridium caminithermale*	*Clostridium halophilium*	*Tepidibacter mesophilus*
Acid from										
L-Arabinose	-	-	-	na	na	+	na	-	na	-
Ribose	+	-	-	+	+	+	na	+	+	+
Mannose	-	+	-	+	-	+	na	+	+	-
Mannitol	-	-	-	+	-	-	na	-	+	-
Sucrose	-	-	-	na	na	+	na	-	na	-
D-glucose	-	+	-	+	+	+	+	+	-	+
D-fructose	-	-	-	+	+	+	-	+	+	+
D-maltose	-	na	-	na	na	+	+	+	na	+
Habitat	Human gut	Sediments	Wastewater	Soil	Water	Water	Chimney rocks	Chimney rocks	Sediments	Soil

340

468 na: Non available data. w: weak reaction.

Table 4. Nucleotide content and gene count levels of the genome

Attribute	Genome (total)	
	Value	**% of total[a]**
Size (bp)	3,497,275	100
G+C content (%)	1,064,697	30.45
Coding region (bp)	3,070,894	87.80
Total genes	3,397	100
RNA genes	129	3.79
Protein-coding genes	3,268	96.20
Genes with function prediction	2,322	70.65
Genes assigned to COGs	2,069	71.44
Genes with peptide signals	346	10.18
CRISPR repeats	3	0.08
ORFans genes	272	8.0
Genes with transmembrane helices	831	24.46
Genes associated with PKS or NRPS	04	0.11
No of antibiotic resistance genes	0	0
No. of genes associated with Pfam-A domains	3,042	89.54

[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.

Table 5. Number of genes associated with the 25 general COG functional categories

Code	Value	% of total[a]	Description
J	173	5.29	Translation
A	0	0.0	RNA processing and modification
K	219	6.70	Transcription
L	189	5.78	Replication, recombination and repair
B	1	0.03	Chromatin structure and dynamics
D	30	0.92	Cell cycle control, mitosis and meiosis
Y	0	0.0	Nuclear structure
V	74	2.26	Defense mechanisms
T	199	6.09	Signal transduction mechanisms
M	139	4.25	Cell wall/membrane biogenesis
N	87	2.66	Cell motility
Z	0	0.0	Cytoskeleton
W	1	0.03	Extracellular structures
U	40	1.22	Intracellular trafficking and secretion
O	77	2.36	Post-translational modification, protein turnover, chaperones

C	154	4.71	Energy production and conversion
G	98	3.00	Carbohydrate transport and metabolism
E	205	6.27	Amino acid transport and metabolism
F	60	1.84	Nucleotide transport and metabolism
H	101	3.09	Coenzyme transport and metabolism
I	54	1.65	Lipid transport and metabolism
P	145	4.44	Inorganic ion transport and metabolism
Q	41	1.25	Secondary metabolites biosynthesis, transport and catabolism
R	298	9.12	General function prediction only
S	228	6.98	Function unknown
-	253	7.44	Not in COGs

[a]The total is based on the total number of protein coding genes in the annotated genome.

Table 6. Genome comparison of closely related species to *Malnutritionisia massiliensis* strain Mt12^{T}

Organisms	**INSDC**	**Size (Mb)**	**G+C (%)**	**Protein coding genes**	**Total Genes**
Malnutritionisia massiliensis strain Mt12^{T}	CXYX00000000.1	3.49	30.44	3,268	3,397
Alkaliphilus metalliredigens strain QYMF	CP000724.1	4.93	36.8	4,576	4,801
Clostridium aceticum strain DSM 1496	CP009687.1	4.2	35.3	3,705	3,847
Alkaliphilus transvaalensis strain SAGM1	JHYF00000000.1	4.02	34.0	3,604	3,725
Alkaliphilus oremlandii strain OhILAs	CP000853.1	3.12	36.3	2,878	3,016

Table 7. The numbers of orthologous proteins shared between g enomes (upper right), average percentage similarity of nucleotides corresponding to orthologous proteins shared between genomes (lower left) and numbers of proteins per genome (bold).

	Alkaliphilus metalliredigens	*Alkaliphilus oremlandii*	*Alkaliphilus transvaalensis*	*Clostridium aceticum*	*Malnutritionisia massiliensis*
Alkaliphilus metalliredigens	**4,823**	1,395	294	1,512	1,357
Alkaliphilus oremlandii	68.38	**2,980**	266	1,342	1,198
Alkaliphilus transvaalensis	68.82	68.30	**433**	288	261

Clostridium aceticum	69.74	68.52	69.69	**3,882**	1,305
Malnutritionisia massiliensis	66.56	66.90	67.23	67.69	**3,271**

Table 8. Pairwise comparison of *Malnutritionisia massiliensis* strain Mt12^{T} with other species using GGDC, formula 2 (DDH estimates based on identities / HSP length)* upper right.

	Alkaliphilus metalliredigens	*Alkaliphilus oremlandii*	*Alkaliphilus transvaalensis*	*Clostridium aceticum*	*Malnutritionisia massiliensis*
Alkaliphilus metalliredigens	100% ± 00	26.8% ±2.55	15.3% ±2.52	25.3% ±2.57	29.0% ±2.54
Alkaliphilus oremlandii		100% ± 00	12.5% ±2.52	14.8% ±2.53	16.2% ±2.52
Alkaliphilus transvaalensis			100% ± 00	26.0% ±2.55	25.3% ±2.54
Clostridium aceticum				100% ± 00	24.1% ±2.54
Malnutritionisia massiliensis					100% ± 00

*The confidence intervals indicate the inherent uncertainty in estimating DDH values from intergenomic distances based on models derived from empirical test data sets (which are always limited in size).

Figure 1.

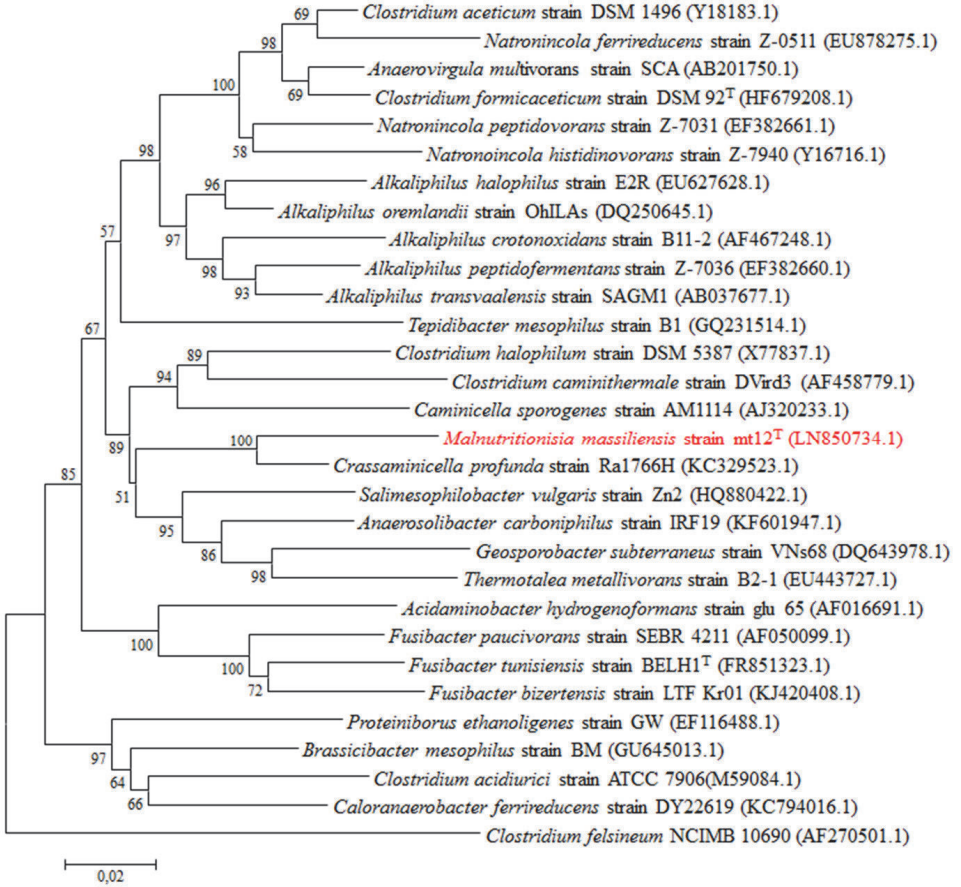

Clostridium aceticum strain DSM 1496 (Y18183.1)
Natronincola ferrireducens strain Z-0511 (EU878275.1)
Anaerovirgula multivorans strain SCA (AB201750.1)
Clostridium formicaceticum strain DSM 92T (HF679208.1)
Natronincola peptidovorans strain Z-7031 (EF382661.1)
Natronoincola histidinovorans strain Z-7940 (Y16716.1)
Alkaliphilus halophilus strain E2R (EU627628.1)
Alkaliphilus oremlandii strain OhILAs (DQ250645.1)
Alkaliphilus crotonoxidans strain B11-2 (AF467248.1)
Alkaliphilus peptidofermentans strain Z-7036 (EF382660.1)
Alkaliphilus transvaalensis strain SAGM1 (AB037677.1)
Tepidibacter mesophilus strain B1 (GQ231514.1)
Clostridium halophilum strain DSM 5387 (X77837.1)
Clostridium caminithermale strain DVird3 (AF458779.1)
Caminicella sporogenes strain AM1114 (AJ320233.1)
Malnutritionisia massiliensis strain mt12T (LN850734.1)
Crassaminicella profunda strain Ra1766H (KC329523.1)
Salimesophilobacter vulgaris strain Zn2 (HQ880422.1)
Anaerosolibacter carboniphilus strain IRF19 (KF601947.1)
Geosporobacter subterraneus strain VNs68 (DQ643978.1)
Thermotalea metallivorans strain B2-1 (EU443727.1)
Acidaminobacter hydrogenoformans strain glu 65 (AF016691.1)
Fusibacter paucivorans strain SEBR 4211 (AF050099.1)
Fusibacter tunisiensis strain BELH1T (FR851323.1)
Fusibacter bizertensis strain LTF Kr01 (KJ420408.1)
Proteiniborus ethanoligenes strain GW (EF116488.1)
Brassicibacter mesophilus strain BM (GU645013.1)
Clostridium acidiurici strain ATCC 7906(M59084.1)
Caloranaerobacter ferrireducens strain DY22619 (KC794016.1)
Clostridium felsineum NCIMB 10690 (AF270501.1)
0,02

Figure 2.

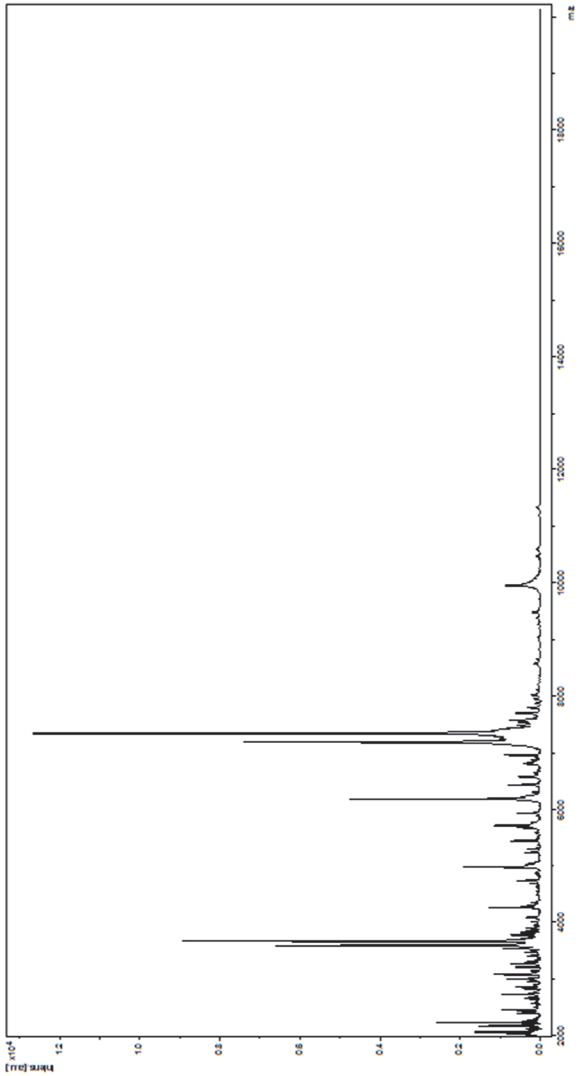

Intens.[au]
x10^4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
2000
4000
6000
8000
10000
12000
14000
16000
18000
m/z

Figure 3.

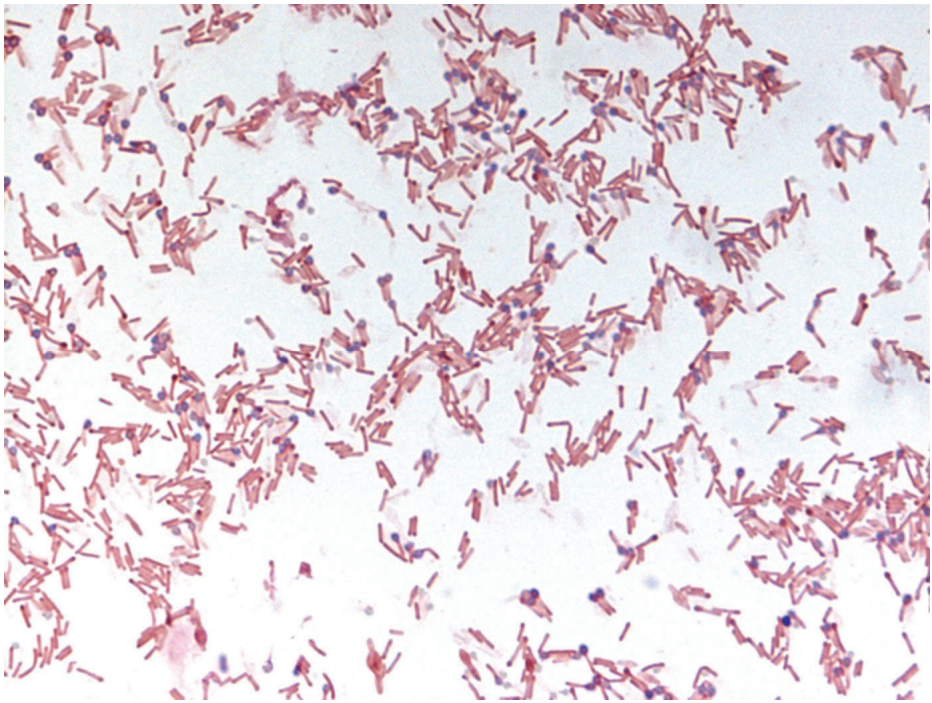


Figure 4.

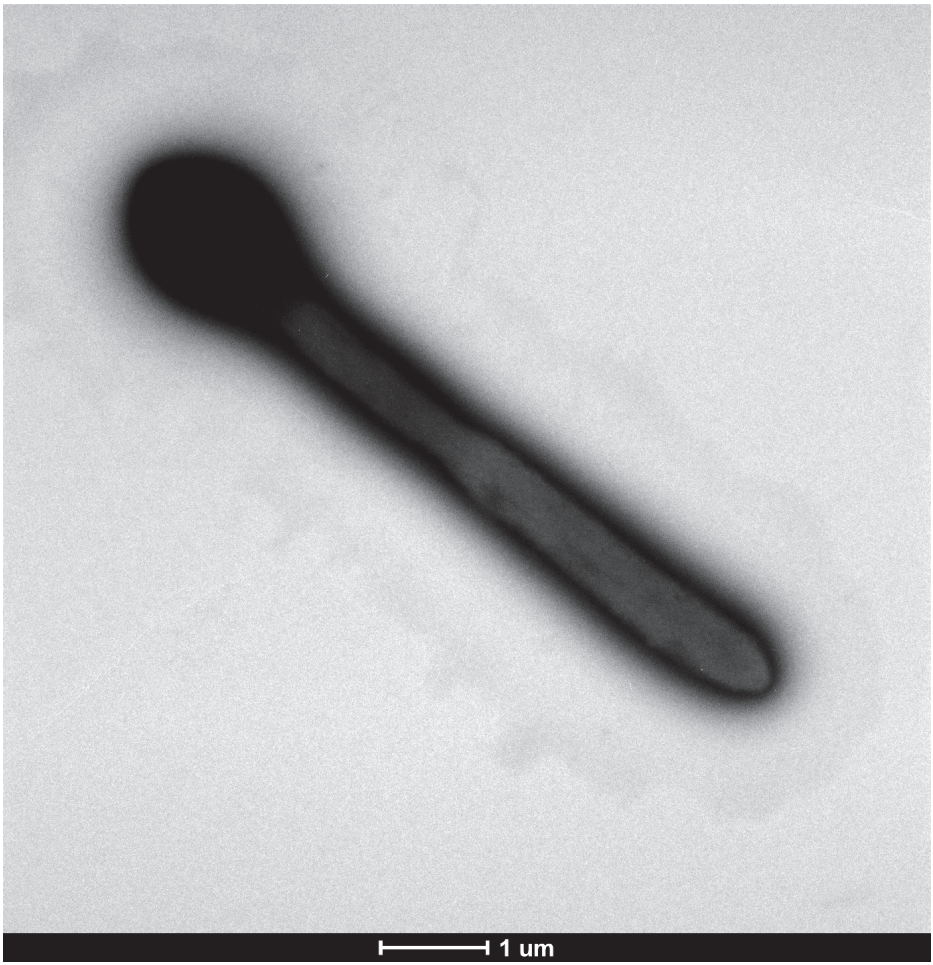


Figure 5.

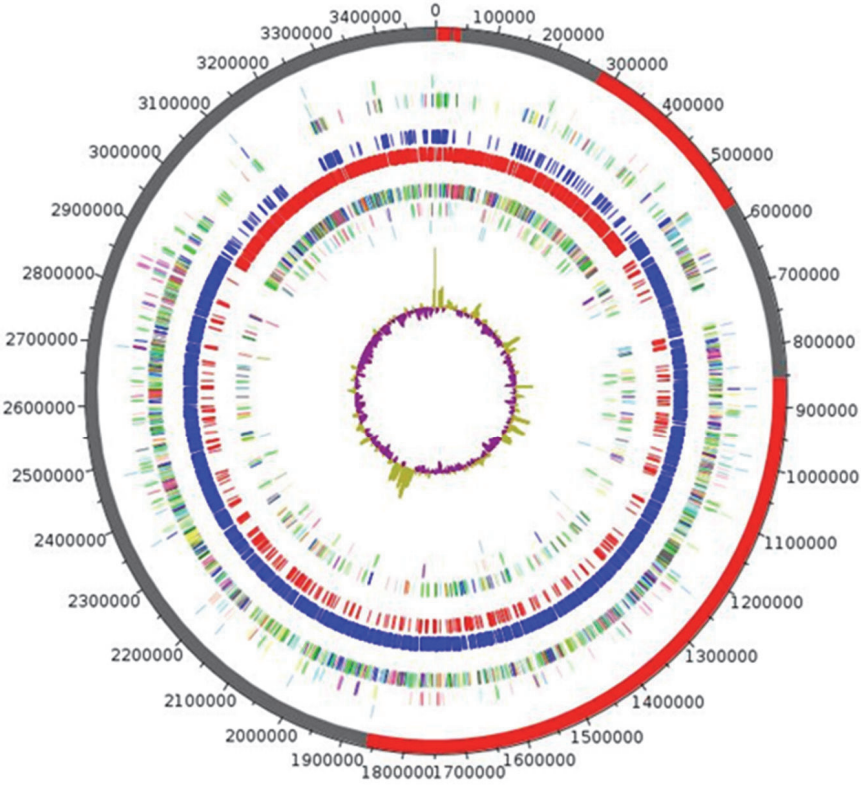
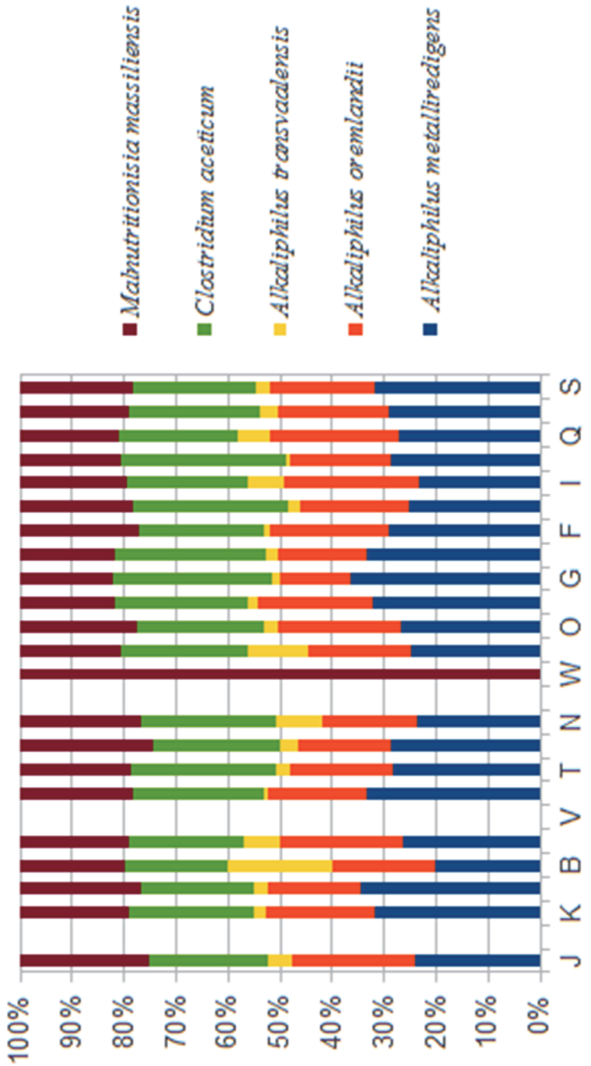

0
100000
200000
300000
400000
500000
600000
700000
800000
900000
1000000
1100000
1200000
1300000
1400000
1500000
1600000
1700000
1800000
1900000
2000000
2100000
2200000
2300000
2400000
2500000
2600000
2700000
2800000
2900000
3000000
3100000
3200000
3300000
3400000

Figure 6.


Mahnutrittionisia massiliensis
Clostridium aceticum
Alkaliphilus transvaalensis
Alkaliphilus oremlandii
Alkaliphilus metalliredigens
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
J K B V T N W O G F I Q S

Article 13: *Anaeromassilibacillus senegalensis* gen. nov., sp. nov., isolated from the gut of a child with kwashiorkor

Elodie Guilhot, Maryam Tidjani Alou, Aldiouma Diallo, Didier Raoult, Saber Khelaifia.

Publié dans New Microbes New Infections

Anaeromassilibacillus senegalensis gen. nov., sp. nov., isolated from the gut of a child with kwashiorkor

E. Guilhot[1], M. Tidjani Alou[1], A. Diallo[3], D. Raoult[1,2,3] and S. Khelaifia[1,2]

1) *Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, CNRS (UMR 7278), IRD (198), INSERM (U1095), AMU (UM63),* 2) *Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine, Aix-Marseille Université, Marseille, France and* 3) *Institut de Recherche pour le Développement, UMR 198 (URMITE), Dakar, Senegal*

Abstract

We report the main characteristics of *Anaeromassilibacillus senegalensis* strain mt9^T (= CSUR P1511) isolated from the stool of a 1-year-old kwashiorkor patient from Senegal.

Keywords: *Anaeromassilibacillus senegalensis*, culturomics, genomics, taxonogenomics, taxonomy
Original Submission: 12 April 2016; **Revised Submission:** 18 April 2016; **Accepted:** 19 April 2016
Article published online: 26 April 2016

Corresponding author: S. Khelaifia, URMITE, CNRS (UMR 7278), IRD (198), INSERM (U1095), AMU (UM63), Faculté de Médecine, Aix-Marseille Université, 27 Boulevard Jean Moulin, 13385 Marseille Cedex 5, France
E-mail: khelaifia_saber@yahoo.fr

Using culturomics [1], a culture-based technique that aims to characterize the microbial diversity of stool samples collected from patients with kwashiorkor, we report the isolation in January 2015 of the strain mt9^T from the stool specimen of a 1-year-old Senegalese kwashiorkor patient. The patient's parents provided signed informed consent, and the agreements of the National Ethics Committee of Senegal and the local ethics committee of the IFR48 (Marseille, France) were obtained under numbers 11-017 and 09-022, respectively. Using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) (http://www.mediterranee-infection.com/article.php?laref=256&titre=urms-database) analysis on a MicroFlex spectrometer (Bruker Daltonics, Leipzig, Germany), we were not able to obtain an identification score of >1.9, which would have allowed a correct identification for this strain [2]. The isolate of the strain mt9^T was obtained by anaerobic culture at 37°C on a 5% sheep's blood–enriched Columbia agar (COS) (bioMérieux, Marcy l'Etoile, France) after a thermic shock and 15-day preincubation in a blood culture bottle. The pure culture of this bacterium grew anaerobically after 72-hour incubation at 37°C. Strain mt9^T was catalase and oxidase negative. Colonies grown on COS medium were white, opaque and 1 to 2 mm in diameter. Bacterial cells were Gram-negative, motile, spore-forming, rod-shaped bacilli with a length of 2.4 µm and a diameter of 0.9 µm. The 16S rRNA gene was sequenced using the fD1–rP2 primers as previously described [3], using a 3130-XL sequencer (Applied Biosciences, Saint Aubin, France). Strain mt9^T exhibited 93.6% sequence similarity with *Clostridium leptum* DSM 753 (GenBank accession no. NR114789), the phylogenetically closest valid species with standing in nomenclature (Fig. 1), which putatively classifies it as a member of a new genus within the family Clostridiaceae in the Firmicutes phylum [4]. In contrast with *C. leptum*, described in 1973 [5] as a Gram-positive bacterium, strain mt9^T is Gram negative.

Because the strain mt9^T exhibited a 16S rRNA sequence divergence of >5% with its phylogenetically closest validated species [6], we propose the creation of the new genus *Anaeromassilibacillus* gen. nov. with the strain mt9^T as the type strain of *Anaeromassilibacillus senegalensis* gen. nov., sp. nov., as the first cultivated representative of this new genera.

http://dx.doi.org/10.1016/j.nmni.2016.04.007

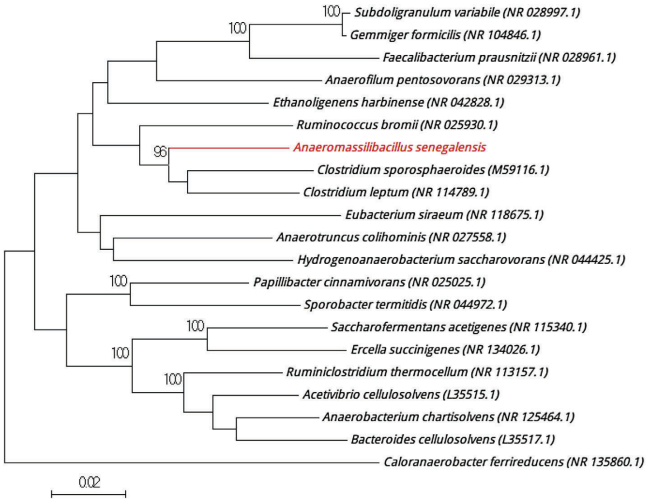


FIG. 1. Phylogenetic tree showing position of *Anaeromassilibacillus senegalensis* strain mt9^T relative to other phylogenetically close members of family Clostridiaceae. GenBank accession numbers are indicated in parentheses. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using maximum-likelihood method within MEGA software. Numbers at nodes are percentages of bootstrap values (≥95%) obtained by repeating analysis 500 times to generate majority consensus tree. Scale bar = 2% nucleotide sequence divergence.

Nucleotide sequence accession number

The 16S rRNA gene sequence was deposited in GenBank under accession number LN866991.

Deposit in a culture collection

Strain mt9^T was deposited in the Collection de Souches de l'Unité des Rickettsies (CSUR, WDCM 875) under accession number P1511.

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, 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.
[3] Mourembou G, Yasir M, Azhar EI, Lagier JC, Bibi F, Jiman-Fatani AA, et al. Rise of microbial culturomics: non-contiguous finished genome sequence and description of *Beduini massiliensis* gen. nov., sp. nov. OMICS 2015;19:766–76.
[4] Euzeby J. Validation list N° 143. Int J Syst Evol Microbiol 2012;62:1–4.
[5] 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 Bacteriol 1976;226:238–52.
[6] Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome Res 2007;17:377–86.

Article 14: "*Bacillus mediterraneensis*" a new bacterial species isolated from the human gut microbiota.

Maryam Tidjani Alou, Pierre-Edouard Fournier and Didier Raoult.

Publié dans New Microbes New Infections

"Bacillus mediterraneensis," a new bacterial species isolated from human gut microbiota

M. T. Alou[1], **P.-E. Fournier**[1] **and D. Raoult**[1,2]

1) *Aix-Marseille Université, URMITE, UM63, CNRS7278, IRD198, Inserm 1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine, Marseille, France and* 2) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

We present a brief description of *"Bacillus mediterraneensis"* strain Marseille-P2366^T (= CSUR P2366 = DSM 102091), a new species isolated from the gastrointestinal tract of a healthy 13-month-old boy from Senegal.

Keywords: *Bacillus mediterraneensis*, culturomics, emerging bacterium, gut microbiota, taxonomy
Original Submission: 6 May 2016; **Revised Submission:** 10 May 2016; **Accepted:** 10 May 2016
Article published online: 15 May 2016

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

While exploring the human gut microbiome using the culturomics approach [1,2], we isolated a bacterial strain from the stool sample of a healthy 13-month-old boy from Senegal. The study was conducted after oral consent was given by the boy's parents, and the ethics committee of the Institut Federatif de Recherche 48 validated the study under agreement 09-022. Strain Marseille-P2366 was first isolated after a 7-day preincubation in an aerobic blood culture bottle (Becton Dickinson, Le Pont de Claix, France) supplemented with sheep's blood and filter-sterilized rumen and subculture on 5% sheep's blood–enriched Columbia agar (bioMérieux, Marcy l'Etoile, France). Light brown and smooth colonies with a diameter of 5 mm were obtained. The bacterial cells were Gram-positive rod-shaped bacilli exhibiting a mean diameter of 0.54 µm and a mean length of 2.25 µm. Oxidase activity was positive but catalase activity was negative. Strain Marseille-P2366^T could not be identified by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) using a Microflex spectrometer (Bruker Daltonics, Leipzig, Germany) [3,4]. We thus relied on 16S rRNA gene sequencing using the fD1-rP2 primers as previously described [5] and a 3130-XL sequencer (Applied Biosciences, Saint Aubin, France) for identification. The obtained sequence was 98.4% similar to *"Bacillus foraminis"* strain CV53 [6] (GenBank accession no. NR_042274.1), the phylogenetically closest species with standing in nomenclature (Fig. 1).

Because the degree of similarity level was lower than the 98.65% threshold to define a new species [7], we propose strain Marseille-P2366^T to be the representative strain of a new species within the genus *Bacillus* that we named *Bacillus mediterraneensis* (me.di.te.ra.ne.en'sis L. masc. adj., mediterraneensis, of Mediterraneum, the Latin name of the Mediterranean Sea by which Marseille, where strain Marseille-P2366^T was isolated, is located).

MALDI-TOF MS spectrum

The MALDI-TOF MS spectrum of *"B. mediterraneensis"* is available at http://www.mediterraneeinfection.com/article.php?laref=256&titre=urms-database.

Nucleotide sequence accession number

The 16S rRNA gene sequence was deposited in GenBank under accession number LT161887.

http://dx.doi.org/10.1016/j.nmni.2016.05.006

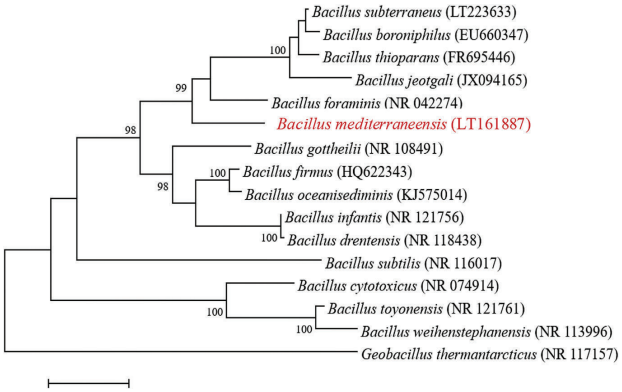


FIG. 1. Phylogenetic tree showing position of *"Bacillus mediterraneensis"* strain Marseille-P2366^T relative to other phylogenetically close neighbours. Sequences were aligned using CLUSTALW and phylogenetic inferences obtained using maximum-likelihood method within MEGA software. Number at nodes is percentage of bootstrap values obtained by repeating analysis 500 times to generate majority consensus tree. Only values ≥95% were indicated. *Geobacillus thermantarcticus* was used as outgroup. Scale bar indicates 1% nucleotide sequence divergence.

Deposit in a culture collection

Strain Marseille-P2366^T was deposited in the Collection de Souches de l'Unité des Rickettsies (CSUR, WDCM 875) under number P2366.

Acknowledgement

This study was funded by the Fondation Méditerranée Infection.

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] Tiago I, Pires C, Mendes V, Morais PV, da Costa MS, Verissimo A. *Bacillus foraminis* sp. nov., isolated from a non-saline alkaline groundwater. Int J Syst Evol Microbiol 2006;56:2571–4.

[7] 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;64:346–51.

Article 15: "*Africanella massiliensis*", a new bacterial genus isolated from the human gut microbiota

Maryam Tidjani Alou., Pierre-Edouard Fournier, Didier Raoult.

Publié dans New Microbes New Infections

"*Africanella massiliensis,*" a new bacterial genus isolated from human gut microbiota

M. T. Alou[1], **P.-E. Fournier**[1] **and D. Raoult**[1,2]

1) *Aix-Marseille Université, URMITE, UM63, CNRS7278, IRD198, Inserm 1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine, Marseille, France and* 2) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

We report the main characteristics of *"Africanella massiliensis"* strain Marseille-P2538 (= CSUR P2538), isolated from the gut microbiota of a healthy 44-month-old girl from Niger.

Keywords: *Africanella massiliensis*, culturomics, emerging bacteria, gut microbiota, taxonomy
Original Submission: 2 May 2016; **Revised Submission:** 9 May 2016; **Accepted:** 10 May 2016
Article published online: 15 May 2016

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 order to describe the bacterial flora of the gastrointestinal tract, a stool sample was collected from a 44-month-old girl from Niger. The patient, for whom her parents gave an informed oral consent, had a weight-for-height z score of −0.65. The study was approved by the ethics committee of the Institut Federatif de Recherche 48 under number 09-022. The stool was cultivated using the culturomics approach [1,2]. Primoculture of strain Marseille-P2538 was achieved following a 3-day preincubation in a liquid medium containing 37 g of Difco Marine Broth (Becton Dickinson, Le Pont de Claix, France) per litre of sterile water at 37°C in an anaerobic atmosphere. After subculture on 5% sheep's blood–enriched Columbia agar (bioMérieux, Marcy l'Etoile, France), colonies were white, circular and smooth, with a mean diameter of 1 mm. Bacterial cells were Gram positive and rod shaped with a mean diameter and length of 0.4 µm and 1.5 µm, respectively. Strain Marseille-P2538^T exhibited 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 of strain Marseille-P2538 using the fD1-rP2 primers as previously described [5] and a 3130-XL sequencer (Applied Biosciences, Saint Aubin, France). The obtained sequence was 94.6% similar to the 16S rRNA of *Ruminococcus gnavus* (GenBank accession no. AB910745) (Fig. 1) [6]. According to the 95% threshold to define a new genus [7], we propose that strain Marseille-P2538 is representative of a new genus within the family *Lachnospiraceae*, for which we propose the name *Africanella* gen. nov. (af.ri.ca.ne'la N.L. fem. n., africanella, of Africa, where the patient from whom strain Marseille-P2538 was isolated lived). Strain Marseille-P2538^T is the type strain of *Africanella massiliensis* gen. nov., sp. nov. (mas.si.li.en'sis, L., fem. adj., massiliensis, for Massilia, the Roman name of Marseille, where strain Marseille-P2538^T was isolated).

MALDI-TOF MS spectrum

The MALDI-TOF MS spectrum of *A. massiliensis* is available at http://www.mediterraneeinfection.com/article.php?laref=256&titre=urms-database.

Nucleotide sequence accession number

The 16S rRNA sequence was deposited in GenBank under accession number LT223700.

http://dx.doi.org/10.1016/j.nmni.2016.05.008

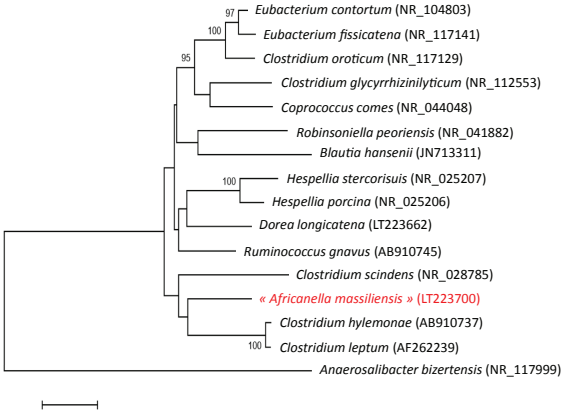


FIG. 1. Phylogenetic tree showing position of *"Africanella massiliensis"* strain Marseille-P2538^T relative to other phylogenetically close species with validly published name. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using maximum-likelihood method within MEGA software. Numbers at nodes are percentages of bootstrap values (≥95%) obtained by repeating analysis 500 times to generate majority consensus tree. *Anaerosalibacter bizertensis* was used as outgroup. Scale bar indicates 2% nucleotide sequence divergence.

Deposit in a culture collection

Strain Marseille-P2538^T was deposited in the Collection de Souches de l'Unité des Rickettsies (CSUR, WDCM 875) under number P2538.

Acknowledgement

This study was funded by the Fondation Méditerranée Infection.

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] 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.

Article 16: "*Clostridium massiliodielmoense*", a new species isolated from the human gut microbiota

Maryam Tidjani Alou, Cheikh Sokhna, Fadi Bittar, Saber Khelaifia, Pierre-Edouard Fournier, Didier Raoult.

Publié dans New Microbes New Infections

'*Clostridium massiliodielmoense*', a new species isolated from the human gut microbiota

Maryam Tidjani Alou[1,2,3], Cheikh Sokhna[2], Fadi Bittar[1,2,3], Saber Khelaifia[1,2,3], Pierre-Edouard Fournier[1,2,3] and Didier Raoult[1,2,3,4]

1) *Aix-Marseille Université, URMITE, UM63, CNRS7278, IRD198, Inserm 1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine,* 2) *Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes IRD 198, CNRS 7278, Aix-Marseille Université, Marseille, France,* 3) *Campus Commun UCAD-IRD of Hann, Dakar, Senegal and* 4) *Special Infectious Agents Unit, King Fahd Medical Research Centre, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Herein, we report the main characteristics of '*Clostridium massiliodielmoense*' strain mt26 (= CSUR P2255), representative of a new species isolated from the gastrointestinal tract of a healthy 28-month-old Senegalese boy.

Keywords: "*Clostridium massiliodielmoense*', culturomics, gut microbiota, taxono-genomics
Original Submission: 11 May 2016; **Revised Submission:** 20 May 2016; **Accepted:** 23 May 2016
Article published online: 2 June 2016

Corresponding author: P.-E. Fournier, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM 63, CNRS7278, IRD198, Inserm1095, Aix-Marseille Université, Marseille, France
E-mail: pierre-edouard.fournier@univ-amu.fr

The culturomics approach is the method of choice to explore as exhaustively as possible the viable population of a microbial ecosystem [1,2]. We used culturomics to study the stool sample of a healthy 28-month-old Senegalese boy with a weight-for-height z-score of −0.99. Oral informed consent was given by the boy's parents and the study was approved by the ethics committee of the Institut Federatif de Recherche IFR48 under number 09-022. Primo-culture of strain mt26 was performed after a 7-day pre-incubation of the stool filtered on a 5-µm filter in an anaerobic blood-culture bottle supplemented with 5% sterile sheep blood. Subculture was performed on 5% sheep blood-enriched Columbia agar (bioMérieux, Marcy-l'Etoile, France) at 37°C in anaerobic atmosphere. Strain mt26 formed a translucent biofilm covering the Petri dish. Cells were Gram-positive, rod-shaped and had mean diameter and length of 0.54 and 2.11 µm, respectively. Strain mt26 exhibited catalase and oxidase activities. Our routine identification method, matrix-assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF-MS) using a Microflex spectrometer (Bruker Daltonics, Bremen, Germany) [3,4], failed to identify strain mt26 at the species level. Following 16S rRNA gene amplification and sequencing using the fD1-rP2 primers, as previously described [5], and using a 3130-XL sequencer (Applied Biosciences, Saint Aubin, France), strain mt26 was 98.4% similar to *Clostridium novyi* strain ATCC 17861 [6] (GenBank Accession number AB045606), the phylogenetically closest species with a validly published name (Fig. 1). *Clostridium novyi* strains have previously been isolated in soil, marine sediments and in animal and human wounds, and sometimes grow forming a biofilm covering the plate [7]. As the 16S rRNA gene similarity level was lower than the 98.65% threshold to define a new species [8,9], we propose the creation of a new species within the genus *Clostridium* that we named '*Clostridium massiliodielmoense*' (mas.si.li.o.di.el.mo.en'se, L. neutr. adj., a combination of *massilio*, of Massilia, the Roman name of Marseille, where strain mt26 was isolated, and *dielmoense*, of Dielmo, the Senegalese village where the boy from whom the strain was cultivated lived). Strain mt26^T is the type strain of '*C. massiliodielmoense*' sp. nov.

http://dx.doi.org/10.1016/j.nmni.2016.05.014

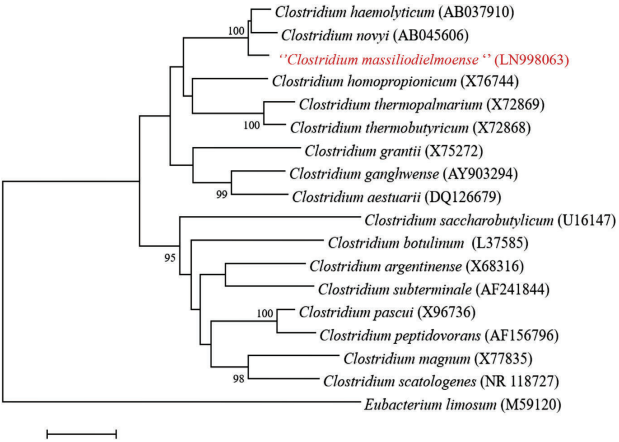


FIG. 1. Phylogenetic tree showing the position of '*Clostridium massiliodielmoense*' strain mt26^T relative to phylogenetically close species with standing in nomenclature. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using the maximum-likelihood method within the 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 bootstraps values >95% were displayed. *Eubacterium limosum* was used as an outgroup. The scale bar indicates a 2% nucleotide sequence.

MALDI-TOF-MS Spectrum

The MALDI-TOF-MS spectrum of '*C. massiliodielmoens*' is available at 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 LN998063.

Deposit in a Culture Collection

Strain mt26^T was deposited in the Collection de Souches de l'Unité des Rickettsies (CSUR, WDCM 875) under number P2255.

Conflicts of Interest

The authors certify that they have no conflict of interest in relation to this research.

Funding

This work was funded by Mediterrannée-Infection Foundation.

References

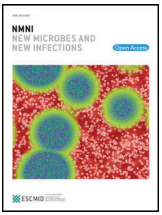
[1] Lagier J-C, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 2012;18:1185–93.
[2] Lagier J-C, Hugon P, Khelaifia S, Fournier P-E, La Scola B, Raoult D. The Rebirth of Culture in microbiology through the example of culturomics to study human gut microbiota. Clin Microbiol Rev 2015;28:237–64.
[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;49:543–51.
[4] Seng P, Abat C, Rolain JM, Colson P, Lagier J-C, 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] Sasaki Y, Takikawa N, Kojima A, Norimatsu M, Suzuki S, Tamura Y. Phylogenetic positions of *Clostridium novyi* and *Clostridium haemolyticum* based on 16S rDNA sequences. Int J Syst Evol Microbiol 2001;51:901–4.
[7] Whitman WB, editor. Systematic Bacteriology. New York, NY: Springer New York; 2009.
[8] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152.
[9] 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;64:346–51.

Article 17: "*Lachnoclostridium touaregense*", a new bacterial species isolated from the human gut microbiota

Maryam Tidjani Alou, Saber Khelaifia, Bernard La Scola, Nadim Cassir.

Publié dans New Microbes New Infections

Accepted Manuscript



"Lachnoclostridium touaregense", a new bacterial species isolated from the human gut microbiota

Maryam Tidjani Alou, Saber Khelaifia, Bernard La Scola, Nadim Cassir

PII: S2052-2975(16)30078-6

DOI: 10.1016/j.nmni.2016.07.007

Reference: NMNI 209

To appear in: *New Microbes and New Infections*

Received Date: 18 July 2016

Accepted Date: 25 July 2016

Please cite this article as: Alou MT, Khelaifia S, La Scola B, Cassir N, *"Lachnoclostridium touaregense",* a new bacterial species isolated from the human gut microbiota, *New Microbes and New Infections* (2016), doi: 10.1016/j.nmni.2016.07.007.

"Lachnoclostridium touaregense", a new bacterial species isolated from the human gut microbiota

Maryam Tidjani Alou[1], Saber Khelaifia[1], Bernard La Scola[1], Nadim Cassir[1*]

[1]Aix-Marseille Univ, URMITE, UM63, CNRS7278, IRD198, Inserm 1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine, 27 Boulevard Jean Moulin, 13385, Marseille cedex 05, France

[*] Corresponding author: Nadim Cassir (cassirnadim@gmail.com)

Running title: "*Lachnoclostridium touaregense"*

Keywords: "*Lachnoclostridium touaregense"*; gut microbiota; culturomics; taxono-genomics.

"Lachnoclostridium touaregense", a new bacterial species isolated from the human gut microbiota

Running title: "*Lachnoclostridium touaregense*"

Keywords: "*Lachnoclostridium touaregense*"; gut microbiota; culturomics; taxono-genomics.

Abstract

We report the main characteristics of "*Lachnoclostridium touaregense*" strain Marseille-P2415^{T} (= CSUR P2415 = DSM 102219), a new bacterial species isolated from the gut microbiota of a healthy young girl from Niger.

Using the 'culturomics' approach [1,2], strain Marseille-P2415^T was isolated in January 2016 from the stool sample of a 44-month-old healthy girl from Niger. Her weight-for-height z-score was of -0.65. This study was validated by the ethics committee of the Institut Fédératif de Recherche IFR48 and oral consent was obtained from the parents. We isolated the strain Marseille-P2415^T through 30-day pre-incubation in anaerobic Colombia like-broth supplemented with sheep blood and seeding on 5% sheep blood enriched Colombia agar (bioMérieux, Marcy L'Etoile, France) in anaerobic atmosphere. Strain Marseille-P2415^T forms on this medium a translucent biofilm formed by Gram-positive rods. Cells have a mean diameter of 0.54 µm and a mean length of 3.35 µm. Oxidase and catalase activities were absent. Protein spectra were obtained for strain Marseille-P2415^T using matrix-assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF MS) on a Microflex spectrometer (Bruker Daltonics, Brenen, Germany) [3,4]. Since these spectra did not match with any species from our database, the 16S rRNA gene was sequenced with fD1-rP2 primers as previously described [5] using a 3130-XL sequencer (Applied Biosciences, Saint Aubin, France). The sequence obtained had a 97.9% similarity with the 16S rRNA gene of *Lachnoclostridium saccharolyticum* strain WM1 (GenBank accession number NR_102852.1), the closest species with a validly published name (Figure 1). According to the 16S rRNA gene sequence similarity for species delineation of prokaryotes [6,7], we propose that strain AT5^T is representative of a new species within the recently described *Lachnoclostridium* genus [8] for which we propose the name *Lachnoclostridium touaregense* (twa.reg'ense touaregense meaning from Touareg; the stool sample was isolated from a young Touareg girl from Niger).

MALDI-TOF-MS spectrum. The MALDI-TOF-MS spectrum of "*Lachnoclostridium touaregense*" is available at 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 LT161895.

Deposit in a culture collection. Strain Marseille-P2415^{T} was deposited in the Collection de Souches de l'Unité des Rickettsies (CSUR, WDCM 875) under number P2415 and in the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ) under number DSM 102219.

Conflict of interest:

The authors certify that they do not have any conflict of interest in relation to this research.

Funding sources:

This work was funded by the Mediterrannée-Infection foundation.

Reference list

[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;18:1185–93.

[2] 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;28:237–64.

[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;49:543–51.

[4] Seng P, Abat C, Rolain JM, Colson P, Lagier J-C, 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.

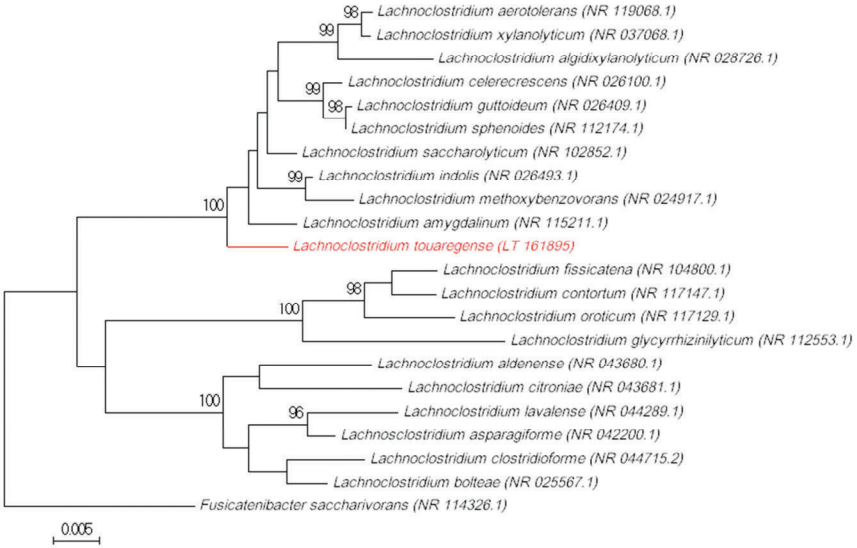
[5] Drancourt M, Bollet C, Carlioz A, Martelin R, Gayral JP, Raoult D. 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. J Clin Microbiol 2000;38:3623–30.

[6] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152.

[7] 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;64:346–51.

[8] Yutin N, Galperin MY. A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. Environ Microbiol 2013;15:2631–41.

Figure 1. Phylogenetic tree showing the position of *Lachnoclostridium touaregense* sp. nov. strain Marseille-P2415^{T} relative to other phylogenetically close neighbors. 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 (> 95%) obtained by repeating the analysis 500 times to generate a majority consensus tree. *Fusicatenibacter saccharivorans* was used as an outgroup. The scale bar indicates a 5% nucleotide sequence divergence.



Article 18: *"Bacillus massilionigeriensis"*, a new bacterial species isolated from the human gut microbiota.

Maryam Tidjani Alou, Souleymane Brah, Jeremy Delerce, Frédéric Cadoret, Pierre-Edouard Fournier, Jean-Christophe Lagier.

Publié dans Human Microbiome Journal

Human Microbiome Journal 1 (2016) 1–2

Contents lists available at ScienceDirect

Human Microbiome Journal

journal homepage: www.elsevier.com/locate/humic

"*Bacillus massilionigeriensis*", a new bacterial species isolated from the human gut microbiota

Maryam Tidjani Alou [a], Souleymane Brah [b], Jeremy Delerce [a], Frédéric Cadoret [a], Pierre-Edouard Fournier [a,*], Jean-Christophe Lagier [a]

[a] Aix-Marseille Université, URMITE, UM63, CNRS7278, IRD198, Inserm1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France
[b] Hopital National de Niamey, BP 238, Niamey, Niger

ARTICLE INFO

Article history:
Received 13 May 2016
Revised 3 June 2016
Accepted 30 June 2016
Available online 7 July 2016

Keywords:
Bacillus massilionigeriensis
Gut microbiota
Culturomics
Emerging bacteria
Niger

ABSTRACT

We present a summary of the main characteristics of "*Bacillus massilionigeriensis*" strain Marseille-P2384^T (=CSUR P2384), isolated from the gut microbiota of a 44-month-old healthy girl from Niger.

Since 2012, we diversified culture conditions to isolate all bacterial members of the human gastro-intestinal tract [1,2]. We applied this strategy, which we named culturomics, to a stool sample from a healthy 44-month-old girl from Niger with a weight-for-height *z*-score of −0.65. Oral consent was given by her parents and the study was validated by the ethics committee of the Institut Federatif de Recherche 48 under number 09-022. Another new bacterium, "*Africanella massiliensis*", was also isolated from this specimen. Strain Marseille-P2384 was first isolated in January 2016 after a 7-day pre-incubation in an aerobic blood culture bottle (Becton Dickinson, Le Pont de Claix, France) supplemented with filter-sterilized rumen followed by subculture on 5% sheep blood-enriched Columbia agar (bioMérieux, Marcy l'Etoile, France) at 37 °C in aerobic atmosphere with 5% CO_2. Colonies were translucent with a mean diameter of 2 mm. Cells were Gram-positive and rod-shaped bacilli, and had a mean diameter and length of 0.48 µm and 2.9 µm, respectively. Strain Marseille-P2384 exhibited catalase activity but no oxidase production.

Routine identification by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF MS) using a Microflex spectrometer (Bruker Daltonics, Bremen, Germany) [3,4] did not allow the identification of strain Marseille-P2384 at the species level. Consequently, we sequenced its 16S rRNA gene using the fD1-rP2 primers as previously described [5], and a 3130-XL sequencer (Applied Biosciences, Saint Aubin, France). Strain Marseille-P2384 exhibited a 96.75% 16S rRNA gene similarity with *Bacillus crescens* strain JC247^T [6] (GenBank accession number LN625239), the phylogenetically-closest species with a validly published name (Fig. 1). As this similarity value was lower than the 98.65% threshold recommended to define a new species [7], we suggest the creation of a new species within the genus *Bacillus* named "*Bacillus massilionigeriensis*" (mas.si.li.o.ni.ge.ri.en's is, L., masc. adj., composed of *massilio*, of Massilia, the Roman name of Marseille where strain Marseille-P2384 was isolated, and *nigeriensis*, of Niger, the country where the patient from whom the sample was obtained, lived). Strain Marseille-P2384^T is the type strain of "*B. massilionigeriensis*".

MALDI-TOF-MS spectrum

The MALDI-TOF-MS spectrum of "*B. massilionigeriensis*" is available at http://www.mediterraneeinfection.com/article.php?laref=256&titre=urms-database.

* Corresponding author.
E-mail address: pierre-edouard.fournier@univ-amu.fr (P.-E. Fournier).

http://dx.doi.org/10.1016/j.humic.2016.06.001
2452-2317/

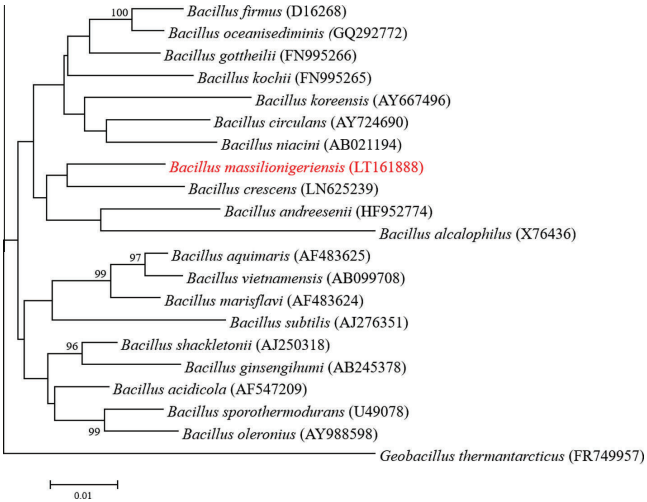


Fig. 1. Phylogenetic tree showing the position of "*Bacillus massilionigeriensis*" strain Marseille-P2384^T relative to other phylogenetically close species with standing in nomenclature. Sequences were aligned using CLUSTALW, and phylogenetic inferences obtained using the maximum-likelihood method within the MEGA software. Numbers at the nodes are percentages of bootstrap values (⩾95%) obtained by repeating the analysis 500 times to generate a majority consensus tree. *Geobacillus thermantarcticus* was used as an outgroup. The scale bar indicates a 1% nucleotide sequence divergence.

Nucleotide sequence accession number

The 16S r RNA gene sequence was deposited in Genbank under accession number LT161888.

Deposit in a culture collection

Strain Marseille-P2384^T was deposited in the Collection de Souches de l'Unité des Rickettsies (CSUR, WDCM 875) under number P2384.

Conflict of interest

The authors certify that they do not have any conflict of interest in relation to this research

Funding sources

This work was funded by Mediterrannée-Infection Foundation.

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. http://dx.doi.org/10.1111/1469-0691.12023.
[2] Lagier JC, 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;28:237–64. http://dx.doi.org/10.1128/CMR.00014-14.
[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;49:543–51. http://dx.doi.org/10.1086/600885.
[4] Seng P, Abat C, Rolain JM, Colson P, Lagier J-C, 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. http://dx.doi.org/10.1128/JCM.00492-13.
[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] Shivani Y, Subhash Y, Subhash Y, Dave Bharti P, Sasikala C, Ramana CV. *Bacillus crescens* sp. nov., isolated from soil. Int J Syst Evol Microbiol 2015;65:2531–6. http://dx.doi.org/10.1099/ijs.0.000296.
[7] 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;64:346–51. http://dx.doi.org/10.1099/ijs.0.059774-0.

Article 19: *"Lachnoclostridium massiliosenegalense"*, a new bacterial species isolated from the human gut microbiota.

Maryam Tidjani Alou, Jean-Christophe Lagier, Bernard La Scola, Nadim Cassir.

Publié dans New Microbes New Infections

'*Lachnoclostridium massiliosenegalense*', a new bacterial species isolated from the human gut microbiota

M. Tidjani Alou, J.-C. Lagier, B. La Scola and N. Cassir

Aix-Marseille Université, URMITE, UM63, CNRS7278, IRD198, Inserm 1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine, Marseille, France

Abstract

We report the main characteristics of '*Lachnoclostridium massiliosenegalense*' strain mt23^T (=CSUR P299 =DSM 102084), a new bacterial species isolated from the gut microbiota of a healthy young girl from Senegal.

Keywords: anaerobe, culturomics, gut microbiota, *Lachnoclostridium massiliosenegalense*, taxono-genomics
Original Submission: 18 July 2016; **Accepted:** 26 July 2016
Article published online: 1 August 2016

Corresponding author: N. Cassir, 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: cassirnadim@gmail.com

As part of the 'culturomics' exploration of the gut microbiota [1,2], a stool sample was collected from a 38-month-old healthy girl from Senegal. Her weight-for-height z-score was −0.12. Oral consent was obtained from the parents, and the ethics committee of the Institut Federatif de Recherche IFR48 validated this study. After a 7-day anaerobic pre-incubation of the stool sample in a marine broth and seeding on 5% sheep-blood-enriched Colombia agar (bioMérieux, La Balme-les-Grottes, France), strain mt23^T was isolated in October 2015. Colonies were translucent with a mean diameter of 2 mm on 5% sheep-blood-enriched Colombia agar. Cells were Gram-positive rods and showed a mean length of 2.5 µm and a mean diameter of 0.5 µm on electron microscopy. Strain mt23^T was catalase and oxidase negative. Routine identification of colonies was carried out using matrix-assisted laser desorption–ionization time-of-flight mass spectrometry (MALDI-TOF MS) on a Microflex spectrometer (Bruker Daltonics, Bremen, Germany) [3,4], which failed to identify strain mt23^T. Therefore, the 16S rRNA gene was sequenced with fD1-rP2 primers as previously described [5] using a 3130-XL sequencer (Applied Biosciences, Saint Aubin, France). The resulting sequence showed a 95.1% similarity with the 16S rRNA gene of *Lachnoclostridium phytofermentans* strain ISDg (GenBank Accession number NR_074652.1), the phylogenetically closest species (Fig. 1) [6]. According to the 16S rRNA gene sequence similarity for species delineation of prokaryotes [7,8], strain mt23^T is representative of a new species within the recently described *Lachnoclostridium* genus. We then propose the name of '*Lachnoclostridium massiliosenegalense*' (mas.si.li.o.se.ne.gal.en′sis. L. gen. masc. n. *massiliosenegalense* contraction of the Latin name of Marseille where strain mt23^T was isolated and Senegal where the stool sample was collected) for which strain mt23 is the type strain.

MALDI-TOF-MS spectrum. The MALDI-TOF-MS spectrum of '*L. massiliosenegalense*' is available at 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 LT161890.

Deposit in a culture collection. Strain mt23^T was deposited in the Collection de Souches de l'Unité des Rickettsies (CSUR, WDCM 875) under number P299 and in the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ) under number DSM 102084.

http://dx.doi.org/10.1016/j.nmni.2016.07.014

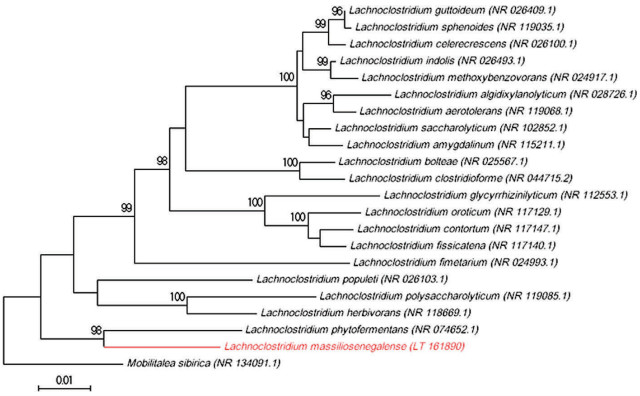


FIG. 1. Phylogenetic tree showing the position of '*Lachnoclostridium massiliosenegalense*' sp. nov. strain mt23^T relative to other phylogenetically close neighbours. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using the maximum-likelihood method within the MEGA software. Numbers at the nodes are percentages of bootstrap values (>95%) obtained by repeating the analysis 500 times to generate a majority consensus tree. *Mobilitalea sibirica* was used as an outgroup. The scale bar indicates a 1% nucleotide sequence divergence.

Conflict of interest

None to declare.

Funding

This work was funded by the Mediterrannée-Infection Foundation.

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;18:1185–93.

[2] 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;28:237–64.

[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;49:543–51.

[4] Seng P, Abat C, Rolain JM, Colson P, Lagier J-C, 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] Yutin N, Galperin MY. A genomic update on clostridial phylogeny: gram-negative spore formers and other misplaced clostridia. Environ Microbiol 2013;15:2631–41.

[7] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152.

[8] 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;64:346–51.

Conclusions et perspectives.

Au cours de mes trois années de recherche doctorales, nous avons pu confirmer l'existence d'une dysbiose chez les enfants atteints de malnutrition sévère aiguë. La caractérisation de cette dysbiose a permis de mettre en évidence une perte de diversité globale chez les malnutris qui s'accompagne à la perte de la diversité anaérobie associée à une perte de la capacité antioxydante du microbiote digestif, estimée par l'absence de *Methanobrevibacter smithii* et un potentiel redox accru, et d'un enrichissement en bactéries aérobies, en *Proteobacteria* et certaines espèces potentiellement pathogènes. Cette caractérisation a été rendue possible par l'association de deux méthodes congruentes et complémentaires, la métagénomique ciblée et la culturomics. Nous avons également pu déterminer l'existence d'un microbiote manquant chez les patients malnutris parmi lesquelles des espèces essentielles à un microbiote digestif sain.

L'analyse du microbiote fécal des enfants malnutris et sains a permis grâce à culturomics l'extension du répertoire du microbiote digestif en isolant et en identifiant 91 espèces non précédemment connues du tractus gastro-intestinal parmi lesquelles 44 nouvelles espèces.

Ce travail nous a ouvert la perspective d'un essai préclinique sur un modèle de souris axénique sur lesquelles notre complexe multi-microbien composé de microbes manquants chez les malnutris sera testé. Nous sommes dans un tournant excitant dans le domaine du microbiote digestif où les connaissances acquises grâce aux études observationnelles permettent le passage à des études interventionnelles où ces connaissances seront testées afin d'ouvrir la voie à la microbiothérapeutique.

List of References.

1. Rajilić-Stojanović M, de Vos WM. The first 1000 cultured species of the human gastrointestinal microbiota. Fems Microbiol Rev. 2014 Sep;38(5):996–1047.
2. Lagier J-C, Million M, Hugon P, Armougom F, Raoult D. Human Gut Microbiota: Repertoire and Variations. Front Cell Infect Microbiol [Internet]. 2012 Nov 2 [cited 2015 Dec 10];2. Available from: http://www.ncbi.nlm.nih.gov /pmc/articles/PMC3487222/
3. Warinner C, Speller C, Collins MJ, Lewis CM. Ancient human microbiomes. J Hum Evol. 2015 Feb; 79:125–36.
4. 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.
5. 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.
6. Salazar N, Arboleya S, ValdÃ©s L, Stanton C, Ross P, Ruiz L, et al. The human intestinal microbiome at extreme ages of life. Dietary intervention as a way to counteract alterations. Front Genet [Internet]. 2014 Nov 21 [cited 2015 Sep 7];5. Available from: http://journal.frontiersin .org/article/10.3389/fgene.2014.00406/abstract
7. Greenhalgh K, Meyer KM, Aagaard KM, Wilmes P. The human gut microbiome in health: establishment and

resilience of microbiota over a lifetime. Environ Microbiol. 2016 Jul;18(7):2103–16.

8. Maukonen J, Saarela M. Human gut microbiota: does diet matter? Proc Nutr Soc. 2015 Feb;74(1):23–36.

9. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. PLoS Biol. 2007 Jul;5(7):e177.

10. Jandhyala SM, Talukdar R, Subramanyam C, Vuyyuru H, Sasikala M, Nageshwar Reddy D. Role of the normal gut microbiota. World J Gastroenterol. 2015 Aug 7;21(29):8787–803.

11. Graf D, Di Cagno R, Fåk F, Flint HJ, Nyman M, Saarela M, et al. Contribution of diet to the composition of the human gut microbiota. Microb Ecol Health Dis. 2015; 26:26164.

12. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. 2012 Jun 14;486(7402):207–14.

13. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, et al. An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol. 2014 Aug;32(8):834–41.

14. Million M, Diallo A, Raoult D. Gut microbiota and malnutrition. Microb Pathog. 2016 Feb 4;

15. Gérard P. Gut microbiota and obesity. Cell Mol Life Sci CMLS. 2016 Jan;73(1):147–62.

16. Manary MJ, Heikens GT, Golden M. Kwashiorkor: more hypothesis testing is needed to understand the aetiology of

oedema. Malawi Medical Journal. Volume 21 numéro 3. 2009 Sep;106–7.

17. OMS | Prise en charge communautaire de la malnutrition aigüe sévère [Internet]. WHO. [cited 2016 Aug 29]. Available from: http://www.who.int/maternal_child_adolescent/documents/a91065/fr/

18. OMS | Normes de croissance OMS et identification de la malnutrition aiguë sévère chez l'enfant [Internet]. WHO. [cited 2016 Aug 29]. Available from: http://www.who.int/nutrition/publications/severemalnutrition/9789241598163/fr/

19. Williams CD. A nutritional disease of childhood associated with a maize diet. Arch Dis Child. 1933 Dec;8(48):423–33.

20. Mukherjee KL. Classification of protein-calorie undernutrition in children. Arch Dis Child. 1967 Dec;42(226):647–51.

21. Smith MI, Yatsunenko T, Manary MJ, Trehan I, Mkakosya R, Cheng J, et al. Gut Microbiomes of Malawian Twin Pairs Discordant for Kwashiorkor. Science. 2013 Feb 1;339(6119):548–54.

22. 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.

23. Dubourg G, Lagier JC, Armougom F, Robert C, Hamad I, Brouqui P, et al. The gut microbiota of a patient with resistant tuberculosis is more comprehensively studied by

culturomics than by metagenomics. Eur J Clin Microbiol Infect Dis. 2013 May;32(5):637–45.

24. 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.

25. Fechner A, Böhme C, Gromer S, Funk M, Schirmer R, Becker K. Antioxidant status and nitric oxide in the malnutrition syndrome kwashiorkor. Pediatr Res. 2001 Feb;49(2):237–43.

26. Raoult A. [Aspects of malnutrition in big child in French East Africa: biochemical data; proteinemia & blood electrophoresis. II]. Bull Société Pathol Exot Ses Fil. 1959 Feb;52(1):114–25.

27. Golden M, Ramdath D. Free radicals in the pathogenesis of kwashiorkor. Proceedings of the nutrition society. Volume 46. 1987;53–68.

28. Subramanian S, Huq S, Yatsunenko T, Haque R, Mahfuz M, Alam MA, et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. Nature [Internet]. 2014 Jun 4 [cited 2015 Sep 7]; Available from: http://www.nature.com/doifinder/10.1038/nature13421

29. Monira S, Nakamura S, Gotoh K, Izutsu K, Watanabe H, Alam NH, et al. Gut microbiota of healthy and malnourished children in bangladesh. Front Microbiol. 2011;2:228.

30. Mata LJ, Jiménez F, Cordón M, Rosales R, Prera E, Schneider RE, et al. Gastrointestinal flora of children with protein—calorie malnutrition. Am J Clin Nutr. 1972 Oct 1;25(10):1118–26.

31. Million M, Tidjani Alou M, Khelaifia S, Bachar D, Lagier J-C, Dione N, et al. Increased Gut Redox and Depletion of Anaerobic and Methanogenic Prokaryotes in Severe Acute Malnutrition. Sci Rep. 2016; 6:26051.

32. UNICEF, Nutrition Section, United Nations System Standing Committee on Nutrition, World Food Programme, World Health Organization. Community-based management of severe acute malnutrition [Internet]. Geneva, Switzerland]; [Rome, Italy]; [Geneva, Switzerland]; ; [New York: World Health Organization] ; [World Food Programme] ; [United Nations System Standing Committee on Nutrition] ; [UNICEF Nutrition Section; 2007 [cited 2016 Jun 29]. Available from: http://www.unicef.org/publications/files/Community_Based_Management_of_Sever_Acute__Malnutirtion.pdf

33. Trehan I, Goldbach HS, LaGrone LN, Meuli GJ, Wang RJ, Maleta KM, et al. Antibiotics as part of the management of severe acute malnutrition. N Engl J Med. 2013 Jan 31;368(5):425–35.

34. Khoruts A, Sadowsky MJ, Hamilton MJ. Development of fecal microbiota transplantation suitable for mainstream medicine. Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc. 2015 Feb;13(2):246–50.

35. Viale AM, Arakaki AK, Soncini FC, Ferreyra RG. Evolutionary relationships among eubacterial groups as inferred from GroEL (chaperonin) sequence comparisons. Int J Syst Bacteriol. 1994 Jul;44(3):527–33.

36. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990 Jun;87(12):4576–9.

37. Wolf M, Müller T, Dandekar T, Pollack JD. Phylogeny of Firmicutes with special reference to Mycoplasma (Mollicutes) as inferred from phosphoglycerate kinase amino acid sequence data. Int J Syst Evol Microbiol. 2004 May;54(Pt 3):871–5.

38. 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 Aug 15;49(4):543–51.

39. Seng P, Abat C, Rolain JM, Colson P, Lagier J-C, 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 Jul;51(7):2182–94.

40. 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 Dec; 36:73–8.

41. Fournier P-E, Raoult D, Drancourt M. « New species announcement », a new format to prompt the description of new human microbial species. New Microbes New Infect [Internet]. 2016 Apr 23 [cited 2016 Sep 3];0(0). Available from: /article/S2052-2975(16)30018-X/abstract