Noncontiguous finished genome sequence and description of *Prevotella phocaeensis* sp. nov., a new anaerobic species isolated from human gut infected by *Clostridium difficile*

P. Afouda[1], S. Ndongo[1], S. Khelaifia[1], N. Labas[1], F. Cadoret[1], F. Di Pinto[1], J. Delerce[1], D. Raoult[1,2] and M. Million[1]

1) *Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM 63, CNRS 7278, IRD 198, INSERM 1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine, Aix-Marseille Université, Marseille, France and* 2) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Prevotella phocaeensis sp. nov. strain SN19^T (= DSM 103364) is a new species isolated from the gut microbiota of patient with colitis due to *Clostridium difficile*. Strain SN19^T is Gram-negative rod-shaped bacteria, strictly anaerobic, nonmotile and non–endospore forming. The predominance fatty acid is hexadecanoic acid. Its 16S rRNA showed a 97.70% sequence identity with its phylogenetically closest species, *Prevotella oralis*. The genome is 2 922 117 bp long and contains 2486 predicted genes including 56 RNA genes.

Keywords: *Clostridium difficile*, culturomics, gut microbiota, *Prevotella phocaeensis*, taxonogenomics
Original Submission: 28 October 2016; **Accepted:** 8 December 2016
Article published online: 13 December 2016

Corresponding author: M. Million, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM 63, CNRS 7278, IRD 198, INSERM 1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine, Aix-Marseille Université, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France
E-mail: matthieumillion@gmail.com

Introduction

The human gut includes approximately 10^{14} microorganisms, known as the intestinal microbiota, and contains four main phyla: *Firmicutes, Bacteroidetes, Actinobacteria* and *Proteobacteria* [1]. This microbial ecosystem contributes to the host's health [1–3] but may also be involved in human disease [1,4,5]. Because intestinal conditions are difficult to reproduce in the laboratory, particularly a strongly reduced environment with a quasi absence of oxygen, most of this microbial universe has not yet been cultivated [1]. However, strict anaerobic microbes should not be neglected because they are the mainstay of the healthy mature anaerobic gut microbiota indispensable for a healthy life [6].

Several studies have already been conducted to identify the microorganisms included in this ecosystem by using both culture methods and culture-independent methods, such as sequencing of the gene encoding 16S ribosomal RNA and metagenomics. The latter sheds light on several unknown microorganisms. However, genomics identifies DNA or RNA sequences but not living microbes.

Recently in our laboratory the culturomics and taxonogenomics approaches were established to unravel gut microbiota diversity while overcoming biases associated with culture-independent techniques. Through this innovative strategy associating several culture conditions, mass spectrometry and molecular biology, more than 200 new species have been isolated from the human gut [7]. Using this approach, we isolated for the first time *Prevotella phocaeensis* strain SN19^T (= CSUR P2259 = DSM 103364) from the gut of a woman infected by *Clostridium difficile*. Here we describe its phenotypic features along with the genome annotation and comparison with the closest species.

http://dx.doi.org/10.1016/j.nmni.2016.12.013

Material and Methods

Sample information

A stool specimen was collected from an 81-year-old white woman with *Clostridium difficile* infection (toxin B, ribotype 027 negative) at La Timone hospital (Marseille, France) in November 2015. The stool sample was collected in sterile plastic containers and stored at −80°C once aliquots were made. Consent was obtained, and the study was approved by the Institut Fédératif de Recherche 48 (Faculty of Medicine, Marseille, France) under agreement number 09-022. The patient was treated with metronidazole; the outcome was a full recovery and no relapse.

Growth conditions

Growth of this isolate was obtained after 5 days' incubation in a 5% sheep's blood– and 5% rumen-enriched medium in anaerobic atmosphere at 37°C. This isolate was then subcultured on 5% sheep's blood–enriched Columbia agar (COS; bioMérieux, Marcy l'Etoile, France). The growth environment of strain SN19^T was determined by testing different temperatures (25, 28, 30, 37 and 56°C) in aerobic and anaerobic (anaeroGEN; Oxoid, Thermo Scientific, Dardilly, France) conditions. Salinity (5–100 g/L) and pH (6–8.5) were also tested.

Strain identification

Matrix-assisted desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) analysis of strains SN19^T was performed on a MicroFlex mass spectrometer (Bruker Daltonics, Leipzig, Germany) as previously described [8]. The acquired spectrum was then loaded into the MALDI Biotyper Software (Bruker) and analyzed as previously described [8] by using the standard pattern-matching algorithm, which compared the spectrum acquired with that present in the library (Bruker database and ours, constantly updated), including 7463 species. Score values of $\geq$1.7 but <2 indicated identification beyond the genus level, and score values of $\geq$2.0 indicated identification at the species level. Scores of <1.7 were interpreted as not relevant.

In this case, we performed the 16S rRNA gene sequencing. For DNA extraction, we used the EZ1 DNA Tissue Kit using Biorobot EZ1 Advanced XL (Qiagen, Courtaboeuf, France). The amplification of the 16s rRNA gene was carried out using PCR technology and universal primers FD1 and RP2 [8] (Eurogentec, Angers, France). Then we realized the purification, sequencing and assembly of the amplified products as previously described [9]. Sequences of 16S rRNA genes were confronted with those which are available in GenBank by BLASTn (Basic Local Alignment Search Tool) (http://www.ncbi.nlm.nih.gov/GenBank/). When the percentage of identity was <98.7%, the studied strain was considered to be a new species [10].

Phylogenetic analysis

A custom Python script was used to automatically retrieve all species from the same family of the new species and download 16S sequences from National Center for Biotechnology Information (NCBI) by parsing NCBI eUtils results and the NCBI taxonomy page. It only keeps sequences from type strains. In case of multiple sequences for one type strain, it selects the sequence obtaining the best identity rate from the BLASTn alignment with our sequence. The script then separates 16S sequences in two groups: one containing the sequences of strains from the same genus (group a), and one containing the others (group b). It only keeps the 15 closest strains from group a and the closest one from group b. If it is impossible to get 15 sequences from group a, the script selects more sequences from group b to get at least nine strains from both groups.

Phenotypic and biochemical characterization

Biochemical characterization and sporulation. For biochemical characterization of strain SN19^T, we used gallery API. ZYM, API Rapid ID 20NE and API 50CH (bioMérieux) according to the manufacturer's instructions. Then we tested catalase (bioMérieux) and oxidase (BD BBL DrySlide; Becton Dickinson, Franklin Lakes, NJ, USA) activities. To determine sporulation, thermal shock was performed on a bacterial suspension of this strain at 80°C for 20 minutes; then this suspension was seeded Columbia blood agar. Incubation was done for 72 hours under anaerobic conditions at different temperatures: 28, 37, 42 and 56°C.

Fatty acid methyl ester (FAME) analysis by gas chromatography/mass spectrometry (GC/MS). Cellular FAME analysis was performed by GC/MS. Two samples were prepared with approximately 30 mg of bacterial biomass per tube collected from several culture plates. FAME were prepared as described by Sasser [11]. GC/MS analyses were carried out as previously described [12]. 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 (National Institute of Standards and Technology, Gaithersburg, USA) and the FAMEs mass spectral database (Wiley, Chichester, UK).

Microscopy. Mobility and Gram staining of strain SN19^T were determined using an optical microscope (Leica, Wetzlar, Germany) with a 40× and 100× oil-immersion objective lens, respectively. In order to observe the cells' morphology, they