Taxonogenomics and description of *Vaginella massiliensis* gen. nov., sp. nov., strain Marseille P2517^T, a new bacterial genus isolated from the human vagina

K. Diop[1], F. Bretelle[2], C. Michelle[1], M. Richez[1], J. Rathored[1], D. Raoult[1,3], P.-E. Fournier[1] and F. Fenollar[1]

1) *Institut hospitalo-universitaire Méditerranée-infection, Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine,* 2) *Department of Gynecology and Obstetrics, Gynépole, Aix-Marseille Université, Hôpital Nord, Assistance Publique-Hôpitaux de Marseille, Marseille, France and* 3) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

An obligate aerobic, Gram-negative, nonmotile and nonsporulating rod designated Marseille P2517 was isolated from the vaginal flora. We describe its features, annotate the genome and compare it to the closest species. The 16S rRNA analysis shows 93.03% sequence similarity with *Weeksella virosa,* the phylogenetically closest species. Its genome is 2 434 475 bp long and presents 38.16% G+C. On the basis of these data, it can be considered as a new genus in the *Flavobacteriaceae* family, for which we proposed the name *Vaginella massiliensis* gen. nov., sp. nov. The type strain is Marseille P2517^T.

Keywords: Bacterial vaginosis, *Flavobacteriaceae*, genome, vaginal flora, *Vaginella massiliensis*
Original Submission: 13 October 2016; **Accepted:** 9 November 2016
Article published online: 25 November 2016

Corresponding author: F. Fenollar, Aix-Marseille Université, Institut hospitalo-universitaire Méditerranée-infection, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France
E-mail: florence.fenollar@univ-amu.fr

Introduction

Bacterial vaginosis is a common yet poorly understood condition affecting women of childbearing age in both industrialized and developing countries. Bacterial vaginosis is characterized simultaneously by an abnormal loss of the normal Doderleïn flora accompanied by an unexplained overgrowth of anaerobic bacteria that were previously minor in the vagina [1,2]. In pregnant women, this vaginal dysbiosis is the consequence of certain complications such as miscarriage, preterm birth or chorioamnionitis [3]. Bacterial vaginosis is mostly treated with antibiotics, mainly metronidazole and clindamycin, but treatment frequently fails; the relapse rate is estimated at 50% at 6 months [4,5].

In order to describe the vaginal flora as fully as possible and to better understand the condition in order to provide better treatment, we studied the vaginal microbiota from healthy women and patients with bacterial vaginosis using the culturomics concept. This is based on the multiplication of culture conditions (atmosphere, media, and temperature) and a rapid bacterial identification using matrix-assisted desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) [6].

A new member of *Flavobacteriaceae* was therefore isolated. Proposed in 1985 by Jooste, it was only in 1992 that the name of this family was validated by Reichenbach. The type genus is *Flavibacterium* (http://www.bacterio.net/flavobacteriaceae.html) [7]. The family currently contains 114 genera (http://www.bacterio.net/-classifgenerafamilies.html). Some species are found in soil and the marine environment, while others are pathogens found in fish and the human urogenital tract [8].

The classical bacterial description presents some limitations. Hence, in order to describe a new bacterium, our laboratory introduced taxonogenomics, a new approach that complements classic features with the proteomic information obtained by MALDI-TOF MS and the description of the annotated whole genome [9,10].

In the following section, we describe the *Vaginella massiliensis* strain Marseille P2517^T (= DSM 102346^T = CSUR P2517), a new genus isolated from a vaginal swab taken from a healthy 22-year-old French woman without bacterial vaginosis.

Materials and Methods

Sample collection

As previously described [11], a vaginal sample was taken from a healthy 22-year-old French woman without bacterial vaginosis at

http://dx.doi.org/10.1016/j.nmni.2016.11.006

La Timone Hospital in Marseille (France) in January 2016 using a Sigma Transwab (Medical Wire, Corsham, United Kingdom). The study was authorized by the local IFR48 ethics committee (Marseille, France) under agreement number 09-022. The patient also provided written informed consent. When the sample was collected, she was not receiving any antibiotic treatment.

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

The vaginal sample was first preincubated in a blood culture bottle (BD Diagnostics, Le Pont-de-Claix, France) supplemented with 4 mL rumen and filtered at 0.2 µm using a pore filter (Thermo Fisher Scientific, Villebon-sur-Yvette, France) and 3 mL sheep's blood (bioMérieux, Marcy l'Etoile, France). After 7 days of preincubation, 50 µL of the supernatant was inoculated on Chocolat PolyViteX (PVX) agar (BD Diagnostics). After 2 days of incubation at 37°C in aerobic conditions, purified colonies were deposited in duplicate on a MSP 96 MALDI-TOF MS target plate (Bruker Daltonics, Leipzig, Germany), and, as previously described, 1.5 µL of matrix solution was added to each spot. Identification was carried out using a Microflex spectrometer (Bruker) [12], which compares the protein spectra found to those in the Bruker database (constantly updated with spectra of new species discovered in our laboratory). If the score is greater than 1.9, the bacterium is correctly identified. In contrast, if no spectra match the database, and for unidentified bacteria with a clear spectrum, 16S rRNA gene sequencing is performed [13]. As Stackebrandt and Ebers suggested [14], if the 16S rRNA sequence similarity value is lower than 95% or 98.7%, the strain is defined as a new genus or species, respectively.

Phylogenetic tree

A custom Python script was used to automatically retrieve all species from the same order of the new genus and to download 16S sequences from National Center for Biotechnology Information (NCBI) by parsing NCBI eUtils results and the NCBI taxonomy page. This only retains sequences from type strains. In the event 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 into two groups: one containing the sequences of strains from the same family (group A) and one containing the others (group B). Finally, it only retains the 15 closest strains from group A and the closest one from group B. If it is impossible to retrieve 15 sequences from group A, the script selects more sequences from group B to achieve at least nine strains from both groups.

Growth conditions

The ideal growth was tested by cultivating the strain Marseille P2517^T on Colombia agar with 5% sheep's blood incubated at different temperatures (25, 28, 37, 45 and 56°C) and different atmospheres (anaerobic, microaerophilic and aerobic). The anaerobic and microaerophilic atmospheres were generated using, respectively, GENbag anaer and GENbag microaer systems (bioMérieux). Salinity and pH conditions were also tested at different concentrations of NaCl (0, 5, 15 and 45%) and different pH (5, 6, 6.5, 7 and 8.5).

Morphological, biochemical and antibiotic susceptibility tests

Sporulation, motility, Gram stain, catalase and oxidase tests were performed using standard test procedures (https://www.gov.uk/government/collections/standards-for-microbiology-investigations-smi#test-procedures). In order to observe cell morphology, samples were fixed with 2.5% glutaraldehyde in 0.1 M cacodylate buffer for at least an hour at 4°C. One 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 seconds with 1% ammonium molybdate solution in filtered water at room temperature. Electron micrographs were acquired using a Tecnai G^{20} Cryo (FEI Company, Limeil-Brevannes, France) transmission electron microscope operated at 200 keV. Biochemical characteristics were studied using API ZYM, API 20NE and API 50CH strips (bioMérieux) according to the manufacturer's instructions.

A cellular fatty acid methyl ester (FAME) analysis was performed by gas chromatography mass spectrometry (GC/MS). Two samples were prepared with approximately 90 mg of bacterial biomass per tube collected from several culture plates. FAMEs were prepared as described by Sasser (http://www.midi-inc.com/pdf/MIS_Technote_101.pdf). GC/MS analyses were carried out using a Clarus 500 gas chromatograph equipped with a SQ8S MS detector (Perkin Elmer, Courtaboeuf, France). FAME extracts (2 mL) 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), enabling the detection of C4 to C24 fatty acid methyl esters. Helium flowing at 1.2 mL/min was used as a 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 through 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

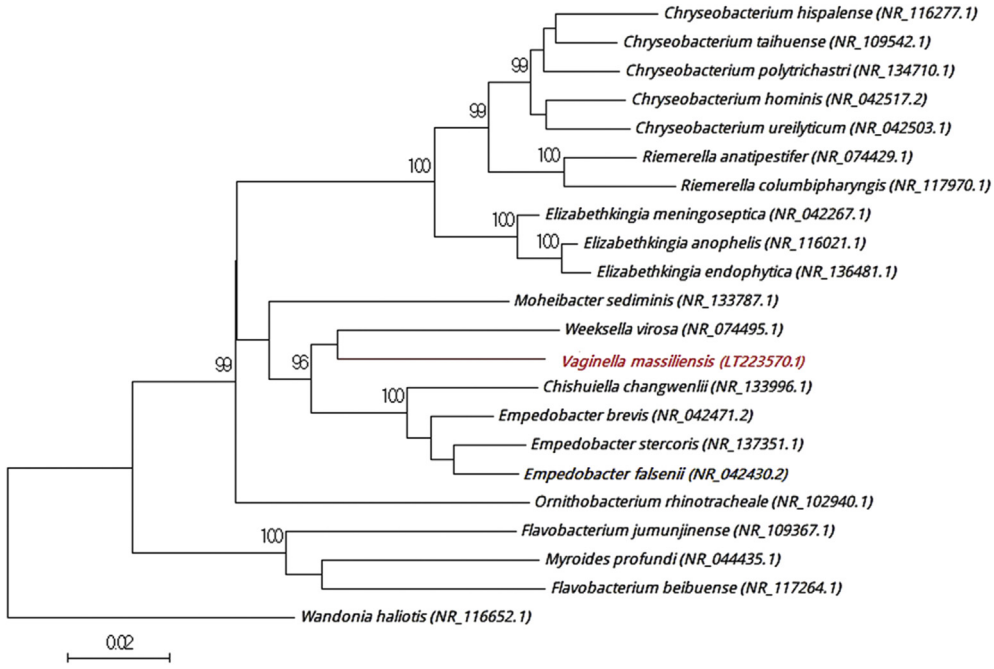

FIG. 1. Phylogenetic tree highlighting position of *Vaginella massiliensis* strain Marseille P2517^T relative to other close strains. GenBank accession numbers of each 16S rRNA are noted immediately after name. Sequences were aligned using Muscle 3.8.31 with default parameters, and phylogenetic inferences were obtained by neighbour-joining method with 500 bootstrap replicates within MEGA6 software. Only bootstraps >95% are shown. Scale bar represents 0.02% nucleotide sequence divergence.

indices 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.

The antibiotic susceptibility of strain Marseille P2517^T was tested using the disk diffusion method (Sirscan discs, Perols, France).

Genomic DNA preparation

Strain Marseille P2517^T was grown in aerobic conditions at 37°C using Columbia agar enriched with 5% sheep's blood (bioMérieux) after 48 hours on four petri dishes. Bacteria were resuspended in 500 µL of Tris-EDTA (TE) buffer; 150 µL of this suspension was diluted in 350 µL 10× TE buffer, 25 µL proteinase K and 50 µL sodium dodecyl sulfate for lysis treatment. This preparation was incubated overnight at 56°C. DNA was purified using phenol/chloroform/isoamylalcohol successively for extraction, followed by ethanol precipitation at −20°C for at least 2 hours each. After centrifugation, the DNA was suspended in 65 µL TE buffer.

TABLE 1. Classification and general features of *Vaginella massiliensis* strain Marseille P2517^T

Characteristic	Term
Current classification	Domain: *Bacteria* Phylum: *Bacteroidetes* Class: *Flavobacteriia* Order: *Flavobacteriales* Family: *Flavobacteriaceae* Genus: *Vaginella* Species: *Vaginella massiliensis* Type strain: Marseille P2517
Gram stain	Negative
Cell shape	Rod
Motility	Nonmotile
Sporulation	Non−spore forming
Temperature range	Aerobic
Optimum temperature	37°C

Genome sequencing and assembly

Genomic DNA (gDNA) of strain Marseille P2517^T was sequenced on the MiSeq Technology (Illumina, San Diego, CA, USA) using the mate-pair strategy. The gDNA was barcoded in order to be mixed with 11 other projects with the Nextera Mate Pair sample prep kit (Illumina). gDNA was quantified using a Qubit assay with the High Sensitivity kit (Life Technologies, Carlsbad, CA, USA) to 153 ng/µL. 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 Bio-Analyzer (Agilent Technologies, Santa Clara, CA, USA) with a DNA 7500 labchip. The DNA fragments ranged in size from 1.5 to 11 kb with an optimal size of 7.455 kb. No size selection was

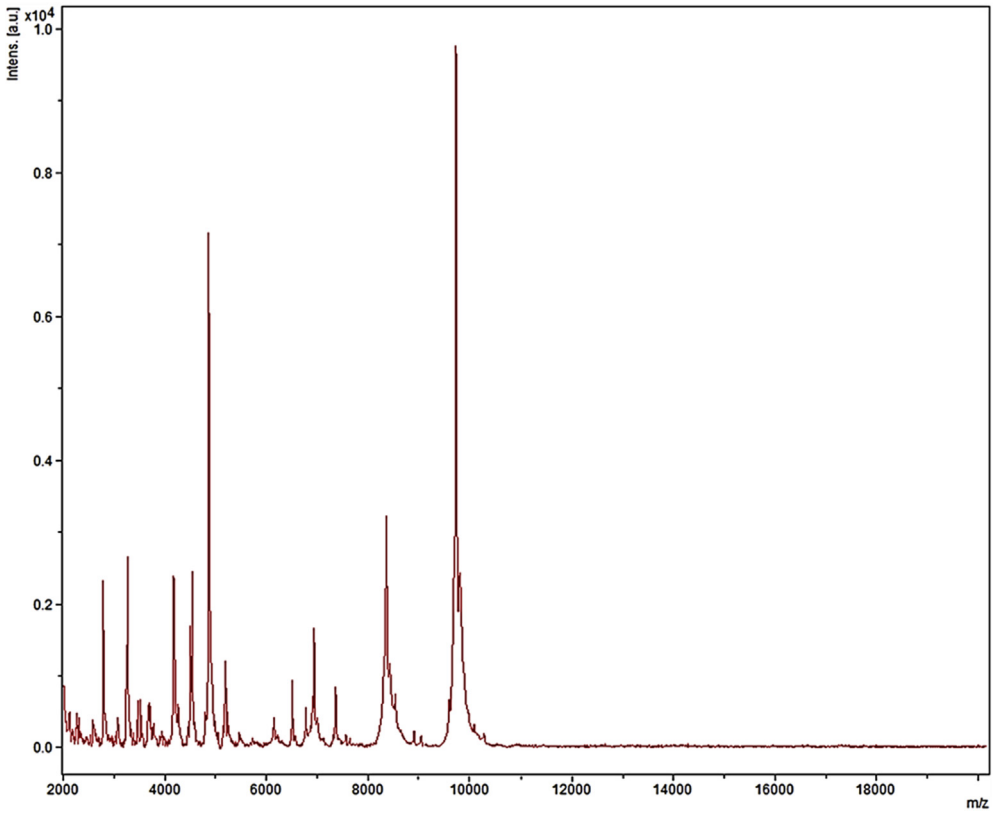

FIG. 2. Reference mass spectrum from *Vaginella massiliensis* strain Marseille P2517^T.

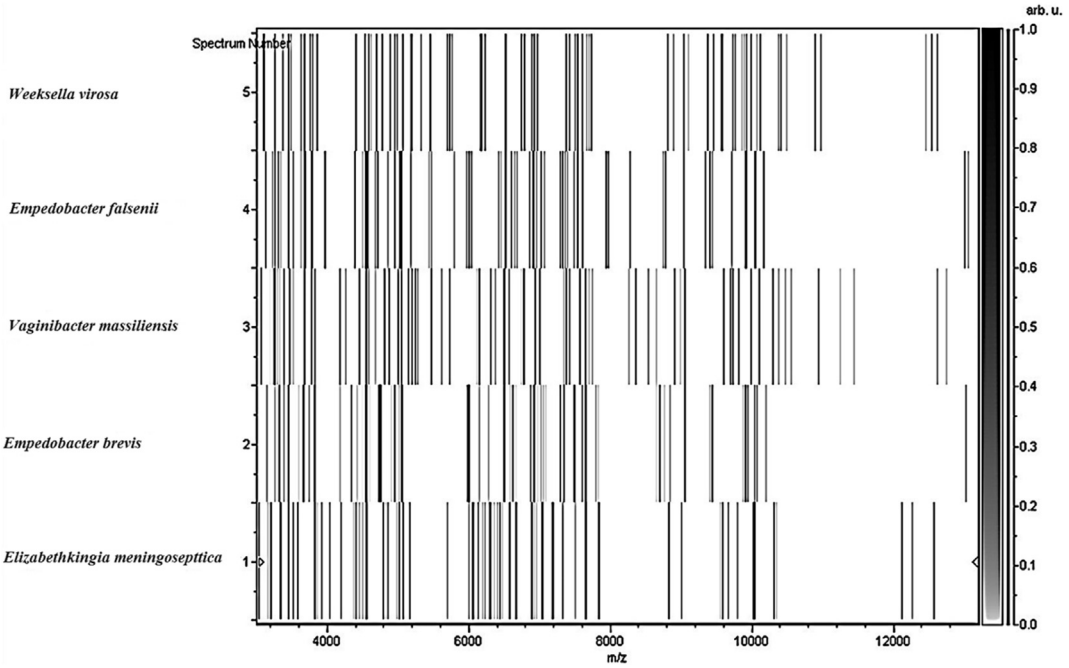

FIG. 3. Gel view comparing *Vaginella massiliensis* strain Marseille P2517^T to other species within genera *Empedobacter*, *Elizabethkingia* and *Weeksella*. Gel view displays raw spectra of loaded spectrum files arranged in pseudo-gel-like look. *x*-axis records *m*/*z* value. Left *y*-axis displays running spectrum number originating from subsequent spectra loading. Peak intensity is expressed by greyscale scheme code. Right *y*-axis indicates relation between color of peak and its intensity in arbitrary units. Displayed species are indicated at left.

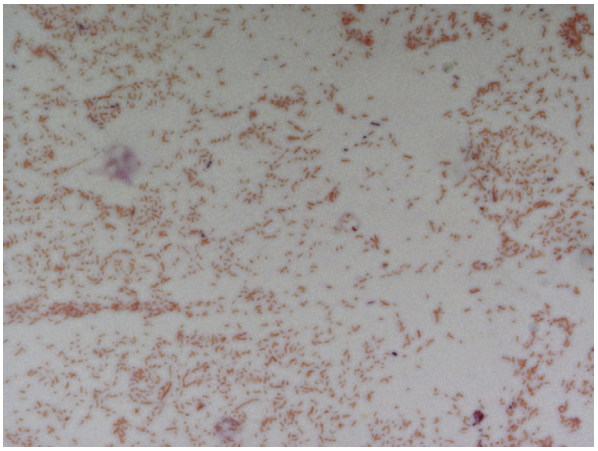

FIG. 4. Gram staining of *Vaginella massiliensis* strain Marseille P2517^T.

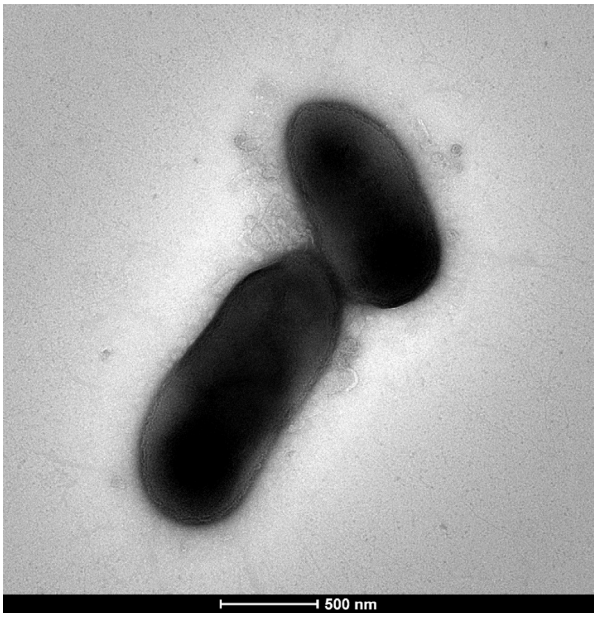

FIG. 5. Electron micrographs of *Vaginella massiliensis* strain Marseille P2517^T using Tecnai G^{20} Cryo (FEI Company) transmission electron microscope operated at 200 keV. Scale bar = 500 nm.

performed, and 410.7 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal size of 1115 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), and the final concentration library was measured at 12.49 nmol/L.

The libraries were normalized at 2 nM and pooled. After a denaturation step and dilution at 15 pM, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and sequencing run were performed in a single 39-hour run at a 2 × 251 bp read length.

A total of 9.2 Gb of information was obtained from a 1042K/mm^2 cluster density with a cluster passing quality control filters of 91.6% (18 078 000 passing filter paired reads). Within this run, the index representation for strain P2517^T was determined to 6.87%. The 1 241 784 paired reads were trimmed, then assembled in nine scaffolds.

TABLE 2. **Cellular fatty acid composition (%)**

Fatty acid	Name	Mean relative %[a]
15:0 iso	13-Methyl-tetradecanoic acid	60.2 ± 0.6
17:1n5 anteiso	14-Methyl-11-hexadecenoic acid	8.5 ± 0.3
17:0 3-OH	3-Hydroxy-heptadecanoic acid	6.2 ± 0.4
15:0 2-OH	2-Hydroxy-pentadecanoic acid	5.1 ± 0.3
17:1n5 iso	15-Methyl-11-hexadecenoic acid	4.9 ± 0.1
17:0 iso	15-Methyl-hexadecanoic acid	3.7 ± 0.1
5:0 iso	3-Methyl-butanoic acid	3.2 ± 0.7
16:0	Hexadecanoic acid	2.2 ± 0.1
15:0 3-OH	3-Hydroxy-pentadecanoic acid	1.8 ± 0.1
15:1n5 iso	13-Methyltetradec-9-enoic acid	1.1 ± 0.1
18:2n6	9,12-Octadecadienoic acid	1.0 ± 0.1
15:0 anteiso	12-Methyl-tetradecanoic acid	TR
18:1n9	9-Octadecenoic acid	TR
18:0	Octadecanoic acid	TR
16:0 iso	14-Methyl-pentadecanoic acid	TR
16:1n6 iso	14-Methylpentadec-9-enoic acid	TR
15:0	Pentadecanoic acid	TR
17:1n7 anteiso	14-Methylhexadec-9-enoic acid	TR
14:0	Tetradecanoic acid	TR
16:0 3-OH	3-Hydroxy-hexadecanoic acid	TR

TR, trace amounts <1%
[a]Mean peak area percentage ± SD.

Genome annotation and analysis

Open reading frames (ORFs) were predicted using Prodigal software [15] with default parameters. Predicted ORFs spanning a sequencing gap region (containing N) were excluded. We predicted the bacterial proteome sequences using BLASTP (*E* value 1e-03, coverage 0.7 and identity percentage 30) against the Clusters of Orthologous Groups (COGs) database. A search against the NR database [16] was performed if no hit was found using BLASTP with an *E* value of 1e-03 coverage of 0.7 and an identity percentage of 30. An *E* value of 1e-05 was used with sequence lengths smaller than 80 aa. The hmmscan analysis tools were used for searching PFam conserved domains (PFAM-A and PFAM-B domains) on each protein. We used RNAmmer [17] and the tRNAScanSE tool [18] to find ribosomal RNAs genes and tRNA genes respectively. For visualization and for data management of genomic features, we used Artemis [19] and DNA Plotter [20] respectively. For the mean level of nucleotide sequence similarity analysis at the genome level, we used MAGI homemade software. It calculated the average genomic identity of orthologous gene sequences (AGIOS) among compared genomes [21]. The Proteinortho [22] software was incorporated into the MAGI homemade software for detecting orthologous proteins in pairwise genomic comparisons. The corresponding

TABLE 3. Differential characteristics of *Vaginella massiliensis, Weeksella virosa, Empedobacter brevis, Empedobacter falsenii, Chishuiella changwenlii* and *Moheibacter sediminis* [26–29]

Property	*Vaginella massiliensis*	*Weeksella virosa*	*Empedobacter brevis*	*Empedobacter falsenii*	*Chishuiella changwenlii*	*Moheibacter sediminis*
Cell diameter (μm)	0.54–0.68	0.6	NA	NA	0.5–0.6	0.2–0.3
Oxygen requirement	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic
Gram stain	–	–	–	–	–	–
Indole	–	+	NA	+	+	NA
Major fatty acid	iso-C15:0 (60.2%) anteiso-C17:1n5 (8.5%)	iso-C15:0 (46.8%) iso-C17:0 3-OH (13.6%)	iso-C15:0 (24.5%) iso-C17:0 3-OH (17.9%)	iso-C15:0 (24.5%) iso-C17:0 3-OH (17.6%)	iso-C15:0 (19.6%) iso-C17:0 3-OH (17.8%)	iso-C15:0 (43.2%) iso-C17:0 3-OH (24.0%)
%G+C (%mol/L)	38.16	35.9	32.8	32.1	30.0	38.2
Production of:						
Alkaline phosphatase	+	+	NA	+	+	+
Catalase	–	+	NA	+	+	+
Oxidase	+	+	NA	+	+	+
Nitrate reductase	–	–	NA	–	–	–
Urease	–	–	–	+	–	–
β-Galactosidase	–	–	NA	–	–	–
N-acetyl-glucosamine	–	+	–	–	–	–
Acid from:						
L-Arabinose	–	+	–	–	–	–
Mannose	–	–	–	–	+	–
Mannitol	–	–	–	–	+	–
D-Glucose	+	–	–	–	+	–
D-Fructose	–	–	–	–	+	–
D-Maltose	–	+	+	–	+	+
Habitat	Human vagina	Human urinogenital tract	Clinical material	Surgical wound	Freshwater	Sediment

+, positive result; –, negative result; NA, data not available.

TABLE 4. Nucleotide content and gene count levels of genome

Attribute	Value	% of total[a]
Size (bp)	2 434 475	100%
G+C content (bp)	928 861	38.16%
Coding region (bp)	2 208 924	90.73%
Total genes	2395	100%
RNA genes	71	2.96%
Protein-coding genes	2324	100%
Genes with function prediction	1618	69.62%
Genes assigned to COGs	1320	56.8%
Genes with peptide signals	524	22.55%
Genes with transmembrane helices	447	19.23%

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.

genes were then retrieved, and the mean percentage of nucleotide sequence identity among orthologous ORFs was determined using the Needleman-Wunsch global alignment algorithm. Genomes from the genus *Weeksella, Empedobacter* and *Elizabethkingia* were used for the calculation of AGIOS values. The genome of strain Marseille P2517^T (European Molecular Biology Laboratory (EMBL)/European Bioinformatics Institute (EBI) accession number FLMR00000000) was compared to those of *Weeksella virosa* DSM 16922 (NC_015144.1), *Empedobacter brevis* ATCC 43319 (NZ_ARNT00000000.1), *Empedobacter falsenii* strain 282 (NZ_JSYQ00000000.1), *Elizabethkingia anophelis* strain B2D (JNCG00000000.1), and *Elizabethkingia meningoseptica* ATCC 13253 (BARD00000000.1); all these genomes were reannotated with Prodigal. The multiagent software system DAGOBAH [23] was used to perform the annotation and comparison process, including Figenix [24] libraries, which provide pipeline analysis. Genome-to-Genome Distance Calculator (GGDC) analysis was performed using the GGDC Web server, as previously reported [25].

Results

Strain characterization

Strain identification. Strain Marseille P2517^T was first cultivated in January 2016 after 7 days of preincubation in a blood culture bottle supplemented with sheep's blood and rumen under aerobic conditions and then inoculated on PVX agar incubated for 2 days at 37°C in aerobic conditions. Scores of 1.35 and 1.5 were obtained with the MALDI-TOF MS identification, suggesting that this isolate was not in the database and consequently was not a known species. The 16S rRNA sequence (accession no. LT223570) of our strain showed 93.03% nucleotide sequence similarity with *Weeksella virosa,* the phylogenetically closest species with a validly published name (Fig. 1). Because this 16S rRNA nucleotide sequence similarity was lower than 95%, the threshold recommended by Stackebrandt and Ebers [14] for delineating a new genus, strain Marseille P2517^T was classified as a new genus, *Vaginella,* with *Vaginella*

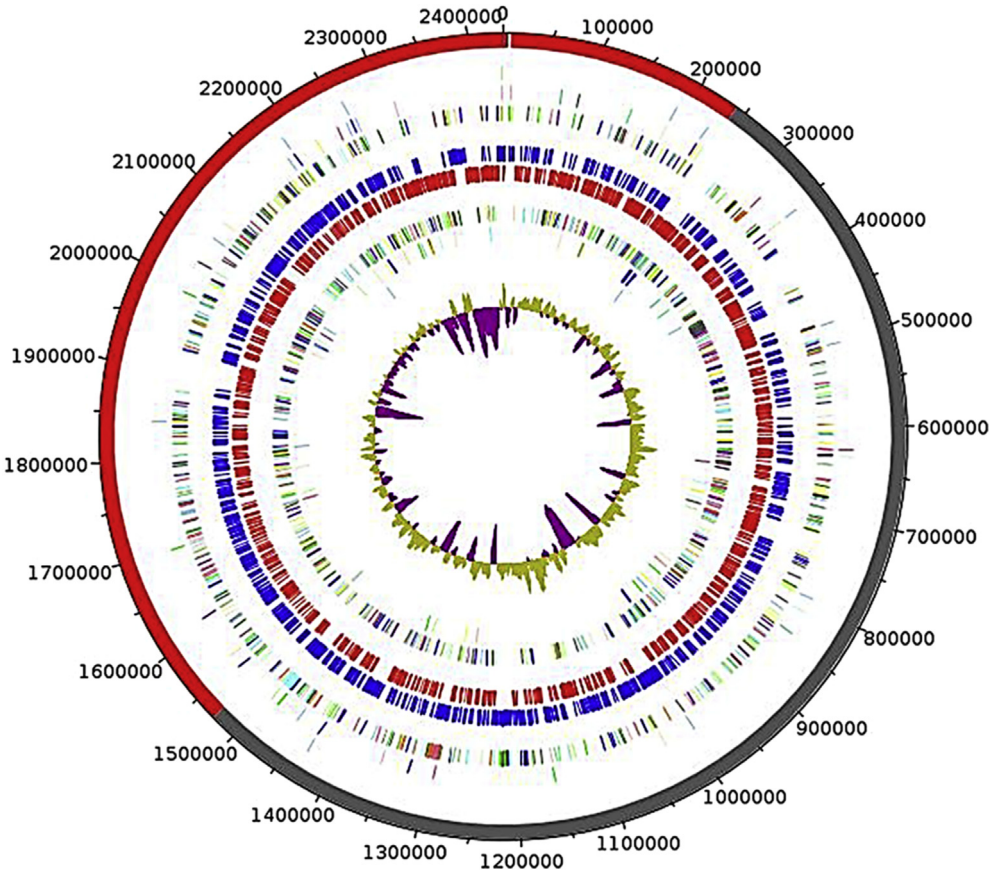

FIG. 6. Graphical circular map of chromosome. From outside to center: contigs (red/grey), COGs category of genes on forward strand (three circles), genes on forward strand (blue circle), genes on reverse strand (red circle), COGs category on reverse strand (three circles), G+C content. COGs, Clusters of Orthologous Groups database.

massiliensis as the type species (Table 1). The reference spectrum of strain Marseille P2517^T (Fig. 2) was then incremented in our database and compared to other known species of the family *Flavobacteriaceae*. Their differences are shown in a gel view in Fig. 3.

Phenotypic characteristics. Cultivated on blood agar for 2 days at 37°C under aerobic conditions, colonies of strain Marseille P2517^T are yellow, opaque, circular and smooth with a diameter of 1.7 to 2 mm. The strain grows only under aerobic conditions at 25, 28 and 37°C, but optimal growth was observed at 37°C after 48 hours of incubation. Its growth also requires a pH ranging from 6.5 to 8.5 and a NaCl concentration lower than 5 g/L. Nonmotile and non−spore forming, strain Marseille P2517^T exhibits positive oxidase activity; however, the catalase activity was negative. Under the microscope, bacterial cells are Gram negative and rod shaped (Fig. 4), and individual cells have a diameter ranging 0.54 to 0.68 µm and a length ranging 1.2 to 1.5 µm (Fig. 5, Table 1).

Using an API 20NE strip, we observed that nitrate and nitrite were not reduced, and urease and indole activities were absent. β-Glucosidase and esculin were not hydrolyzed, unlike gelatin. There was also no assimilation from D-glucose, L-arabinose, D-mannose, D-mannitol, D-maltose and *N*-acetyl-glucosamine. API 50CH showed that strain Marseille P2517^T metabolized only D-glucose; and acid was not produced from other carbohydrates: arabinose, D-ribose, D-xylose, D-adonitol, D-galactose, D-fructose, D-mannose, L-rhamnose, D-mannitol, D-sorbitol, inositol, methyl-αD-mannopyranoside, methyl- αD-glucopyranoside, amygdalin, arbutin, esculin ferric citrate, salicin, D-cellobiose, D-maltose, D-lactose, D-melibiose, D-trehalose, inulin, D-melezitose, D-raffinose, starch, glycogen, xylitol, gentiobiose, D-turanose, D-lyxose, D-tagatose, fucose, arabitol, potassium gluconate, potassium 2-ketogluconate and potassium 5-ketogluconate. API ZYM revealed positive reactions for alkaline phosphatase, esterase, esterase lipase, lipase, leucine arylamidase, cystine arylamidase, valine arylamidase, trypsin, α-chymotrypsin, acid phosphatase and

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

Code	Value	% of total	Description
J	177	7.6	Translation
A	0	0	RNA processing and modification
K	53	2.2	Transcription
L	78	3.3	Replication, recombination and repair
B	0	0	Chromatin structure and dynamics
D	24	1.0	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	47	2.0	Defense mechanisms
T	36	1.5	Signal transduction mechanisms
M	116	4.9	Cell wall/membrane biogenesis
N	16	0.6	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	19	0.8	Intracellular trafficking and secretion
O	81	3.4	Post-translational modification, protein turnover, chaperones
X	10	0.4	Mobilome: prophages, transposons
C	81	3.4	Energy production and conversion
G	56	2.4	Carbohydrate transport and metabolism
E	111	4.7	Amino acid transport and metabolism
F	51	2.1	Nucleotide transport and metabolism
H	94	4.0	Coenzyme transport and metabolism
I	87	3.7	Lipid transport and metabolism
P	106	4.5	Inorganic ion transport and metabolism
Q	32	1.3769363	Secondary metabolites biosynthesis, transport and catabolism
R	109	4.6901894	General function prediction only
S	55	2.3666093	Function unknown
—	1004	43.201378	Not in COGs

COGs, Clusters of Orthologous Groups database.

naphtol-AS-BI-phosphohydrolase. Reactions for other enzymes such as galactosidase (α and β) and oxidases were negative. FAME analysis demonstrated that the most abundant compound was 13-methyl-tetradecanoic acid (60%). This strain shows many saturated and unsaturated branched fatty acids (iso and anteiso). Several hydroxy fatty acids were also present (Table 2).

Resistant to gentamicin (CN 500 µg) and metronidazole (Met 4), cells were sensitive to ceftriaxone (CRO 30 µg), ciprofloxacin (CIP 5 µg), clindamycin (DA 15 µg), colistin (CT 50 µg), doxycycline (DO 30 µg), erythromycin (E 15 µg), fosfomycin (POS 50 µg), oxacillin (OX 5 µg), penicillin (P 10 µg), rifampicin (RA 30 µg), teicoplasmin (TEC 30 µg), trimethoprim/sulfamethoxazole (SXT 25 µg), vancomycin (VA 30 µg) and imipenem (MP 10).

All the phenotypic characteristics of strain Marseille P2517^T were compared to those of the closely related *Flavobacteriaceae* species [26–29] (Table 3).

Genome properties

The draft genome of strain Marseille P2517^T is 2 434 475 bp long with 38.16 mol% G+C content (Table 4, Fig. 6). It is composed of nine scaffolds (composed of ten contigs). On the 2395 predicted genes, 2324 were protein-coding genes and 71 were RNAs (five 5S rRNA, four 16S rRNA, two 23S rRNA and 60 tRNA genes). A total of 1618 genes (69.62%) were assigned a putative function (by COGs or by NR blast), and 123 genes were identified as ORFans (5.29%). The remaining genes were annotated as hypothetical proteins (528 genes, 22.72%). Table 5 provides the distribution of the genes into COGs functional categories.

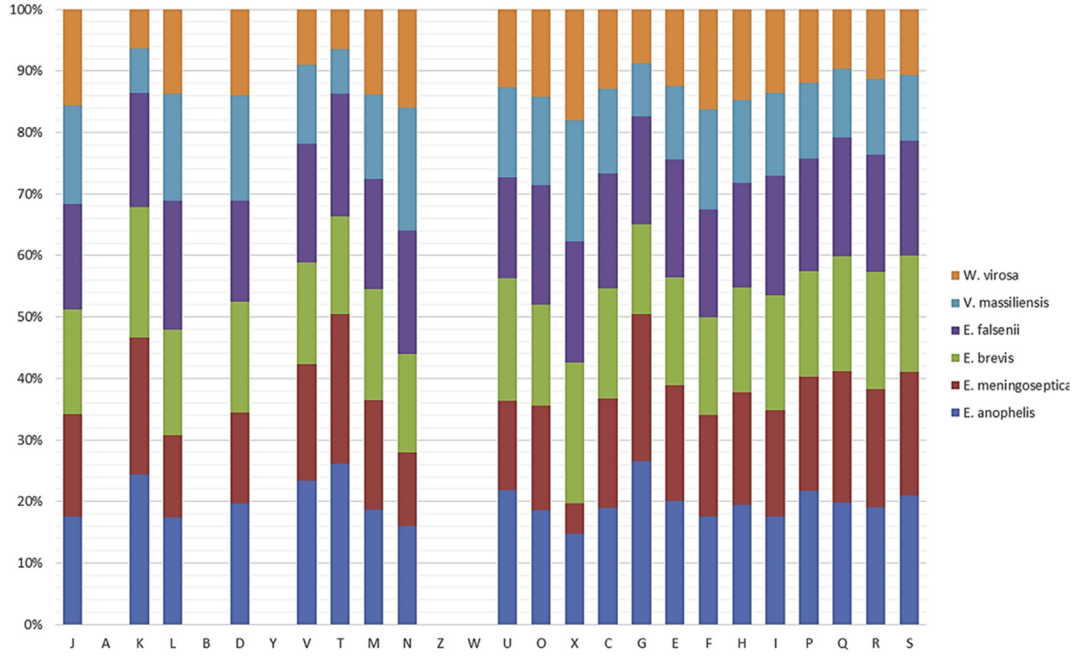

FIG. 7. Distribution of functional classes of predicted genes according to clusters of orthologous groups of proteins of *Vaginella massiliensis* strain Marseille P2517^T among other species.

TABLE 6. Numbers of orthologous proteins shared between genomes (upper right) and AGIOS values obtained (lower left)

	Vm	Eb	Ef	Wv	Ea	Em
Vm	**2324**	1214	1200	1142	702	690
Eb	72.18%	**3567**	1426	1208	926	902
Ef	72.54%	86.26%	**3500**	1199	890	876
Wv	75.10%	72.37%	72.80%	**2118**	685	686
Ea	67.43%	68.71%	68.73%	67.63%	**4042**	941
Em	67.26%	68.37%	68.29%	67.38%	83.65%	**3459**

Numbers of proteins per genome are indicated in bold.
AGIOS, average genomic identity of orthologous gene sequences; Ea, *Elizabethkingia anophelis* B2D; Eb, *Empedobacter brevis* ATCC 43319; Ef, *Empedobacter falsenii* 282; Em, *Elizabethkingia meningoseptica* ATCC13253; Vm, *Vaginella massiliensis* Marseille P2517^T; Wv, *Weeksella virosa* DSM16922.

Genomic comparison

Comparison of the genome of our strain Marseille P2517^T with those of the closest species revealed that the draft genome sequence of strain Marseille P2517^T (2.34 Mb) was smaller than those of *Empedobacter falsenii* (3.71 Mb), *Empedobacter brevis* (3.79 Mb), *Elizabethkingia meningoseptica* (3.84 Mb) and *Elizabethkingia anophelis* (4.02 Mb) but larger than those of *Weeksella virosa* (2.27 Mb). The G+C content of strain Marseille P2517^T was larger than those of all the compared genomes: *E. meningoseptica* (36.4%), *W. virosa* (35.9%), *E. anophelis* (35.6%), *E. brevis* (32.8%) and *E. falsenii* (32.1%).

The gene content of strain Marseille P2517^T (2395) was smaller than those of *E. falsenii, E. brevis, E. anopheles* and *E. meningoseptica* (3610, 3633, 4108 and 3500 respectively) but larger than those of *W. virosa* (2192). Nevertheless, the distribution of genes into COGs categories was similar among all compared genomes (Fig. 7). Otherwise, the AGIOS analysis showed that strain Marseille P2517^T shares between 1214 and 690 orthologous genes with its closely related species: 1214, 1200, 1142, 702 and 690 with *E. brevis, E. falsenii, W. virosa, E. anopheles* and *E. meningoseptica* respectively (Table 6). Analysis of the average percentage of nucleotide sequence identity between strain Marseille P2517^T and other species ranged from 67.26% with *E. meningoseptica* and 75.10% with *W. virosa* (Table 6). Similar results were also observed for analysis of digital DNA-DNA hybridization (Table 7).

Conclusion

Phenotypic, phylogenetic analyses and genomic results enable us to propose that strain Marseille P2517^T may be the representative of a novel genus, *Vaginella,* with *Vaginella massiliensis* as the type strain. It was isolated from the normal vaginal flora of a 22-year-old Frenchwoman.

Taxonomic and nomenclatural proposals

Description of Vaginella *gen. nov. Vaginella* (va.gi.nel′la, L. fem. n. *vagina,* 'vagina,' part of the female genital tract; L. dim. suff. *-ella;* N.L. dim. fem. n. *Vaginella,* 'small vagina,' referring to the source of the isolation of the type strain).

The organism is an obligate aerobic, Gram-negative and rod-shaped bacilli. It is nonmotile and non–spore forming. It has negative catalase activity, and nitrate not reduced, with no urease production and positive oxidase activity. Habitat is human vagina flora. The type species is *Vaginella massiliensis* strain P2517^T.

Description of Vaginella massiliensis *strain Marseille P2517^T gen. nov., sp. nov. Vaginella massiliensis* (mas.si.li.en′sis, L. gen. adj. *massiliensis,* from Massilia, the Latin name of Marseille, France, where the organism was first grown, identified, and characterized).

The organism is obligate aerobic, nonmotile, nonsporulating and mesophilic, with optimal growth at 37°C. *Vaginella* cells are Gram negative and rod shaped, with a mean diameter of 0.61 µm and a length of 1.35 µm; it is oxidase positive and catalase negative; the major fatty acid is 13-methyl-tetradecanoic acid (60%). On Columbia agar, colonies are yellow, opaque, circular, smooth and approximately 1.85 mm in diameter. Nitrate reduction, urease and indole formation are negative. They are asaccharolytic; acid is produced only from glucose. Gelatin is hydrolyzed. Cells are susceptible to ceftriaxone, ciprofloxacin, clindamycin, colistin, doxycycline, erythromycin, fosfomycin, oxacillin, penicillin, rifampicin, teicoplasmin, trimethoprim/sulfamethoxazole, vancomycin and imipenem but are resistant to gentamicin and metronidazole.

TABLE 7. Digital DNA-DNA hybridization values obtained by comparison of all studied genomes

	Vm	Eb	Ef	Wv	Ea	Em
Vm	**100%**	21.10 ± 2.3%	19.10 ± 2.3%	21.40 ± 2.35%	26.10 ± 2.4%	25.50 ± 2.4%
Eb		**100%**	28.40 ± 2.4%	18.90 ± 2.3%	21.30 ± 2.35%	16.80 ± 2.2%
Ef			**100%**	19.90 ± 2.3%	24.90 ± 2.4%	19.30 ± 2.3%
Wv				**100%**	24.80 ± 2.4%	22.60 ± 2.35%
Ea					**100%**	23.70 ± 2.35%
Em						**100%**

Ea, *Elizabethkingia anophelis* B2D; Eb, *Empedobacter brevis* ATCC 43319; Ef, *Empedobacter falsenii* 282; Em, *Elizabethkingia meningoseptica* ATCC13253; Vm, *Vaginella massiliensis* Marseille P2517^T; Wv, *Weeksella virosa* DSM16922.

The genome of *Vaginella massiliensis* is 2 434 475 bp long and exhibits 38.16% G+C content. Its 16S rRNA gene sequence and that of draft genome are both deposited in EMBL/EBI under accession numbers LT223570 and FLMR00000000, respectively. The type strain Marseille P2517^T (= DSM 102346^T = CSUR P2517) was isolated from the vaginal swab of a healthy Frenchwoman.

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 also thank K. Griffiths for English-language review.

Conflict of Interest

None declared.

References

[1] Onderdonk AB, Delaney ML, Fichorova RN. The human microbiome during bacterial vaginosis. Clin Microbiol Rev 2016;29:223–38.
[2] Shipitsyna E, Roos A, Datcu R, Hallén A, Fredlund H, Jensen JS, et al. Composition of the vaginal microbiota in women of reproductive age—sensitive and specific molecular diagnosis of bacterial vaginosis is possible? PLoS One 2013;8:e60670.
[3] Svare J, Schmidt H, Hansen B, Lose G. Bacterial vaginosis in a cohort of Danish pregnant women: prevalence and relationship with preterm delivery, low birthweight and perinatal infections. BJOG 2006;113:1419–25.
[4] Bradshaw CS, Tabrizi SN, Fairley CK, Morton AN, Rudland E, Garland SM. The association of *Atopobium vaginae* and *Gardnerella vaginalis* with bacterial vaginosis and recurrence after oral metronidazole therapy. J Infect Dis 2006;194:828–36.
[5] Bretelle F, Rozenberg P, Pascal A, Favre R, Bohec C, Loundou A, et al. High *Atopobium vaginae* and *Gardnerella vaginalis* vaginal loads are associated with preterm birth. Clin Infect Dis 2015;60:860–7.
[6] 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.
[7] Bernardet JF, Nakagawa Y, Holmes B. Proposed minimal standards for describing new taxa of the family *Flavobacteriaceae* and emended description of the family. Int J Syst Evol Microbiol 2002;52:1049–70.
[8] Bernardet JF, Segers P, Vancanneyt M, Berthe F, Kersters K, Vandamme P. Cutting a Gordian knot: emended classification and description of the genus *Flavobacterium*, emended description of the family *Flavobacteriaceae*, and proposal of *Flavobacterium hydatis* nom. nov. (basonym, *Cytophaga aquatilis* Strohl and Tait 1978). Int J Syst Evol Microbiol 1996;46:128–48.
[9] Fournier PE, Lagier JC, Dubourg G, Raoult D. From culturomics to taxonomogenomics: a need to change the taxonomy of prokaryotes in clinical microbiology. Anaerobe 2015;36:73–8.
[10] Fournier PE, Drancourt M. *New Microbes New Infections* promotes modern prokaryotic taxonomy: a new section 'Taxonogenomics: new genomes of microorganisms in humans'. New Microb New Infect 2015;7:48–9.
[11] Menard JP, Fenollar F, Henry M, Bretelle F, Raoult D. Molecular Quantification of *Gardnerella vaginalis* and *Atopobium vaginae* loads to predict bacterial vaginosis. Clin Infect Dis 2008;20:33–43.
[12] 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.
[13] 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.
[14] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152.
[15] 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.
[16] Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2016;44:D67–72.
[17] 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.
[18] 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.
[19] Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 2012;28:464–9.
[20] Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 2009;25:119–20.
[21] 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.
[22] 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.
[23] Gouret P, Paganini J, Dainat J, Louati D, Darbo E, Pontarotti P, et al. Integration of evolutionary biology concepts for functional annotation and automation of complex research in evolution: the multi-agent software system DAGOBAH. In: Pontarotti P, editor. Evolutionary biology—concepts, biodiversity, macroevolution and genome evolution. Berlin: Springer; 2011. p. 71–87.
[24] 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.
[25] 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.
[26] Zhang RG, Tan X, Liang Y, Meng TY, Liang HZ, Lv J. Description of *Chishuiella changwenlii* gen. nov., sp. nov., isolated from freshwater, and transfer of *Wautersiella falsenii* to the genus *Empedobacter* as *Empedobacter falsenii* comb. nov. Int J Syst Evol Microbiol 2014;64:2723–8.
[27] Kampfer P, Avesani V, Janssenes M, Charlier J, De Baere T, Vaneechouette M. Description of *Wautersiella falsenii* gen. nov., sp. nov., to accommodate clinical isolates phenotypically resembling members of the genera *Chryseobacterium* and *Empedobacter*. Int J Syst Evol Microbiol 2006;56:2323–9.
[28] Zhang RG, Tan X, Zhao X-M, Deng J, Lv J. *Moheibacter sediminis* gen. nov., sp. nov., a member of the family *Flavobacteriaceae* isolated from sediment, and emended descriptions of *Empedobacter brevis*, *Wautersiella falsenii* and *Weeksella virosa*. Int J Syst Evol Microbiol 2014;64:1481–7.
[29] Holmes B, Snell JJS, Lapage SP. Revised description, from clinical strains, of *Flavobacterium breve* (Lustig) Bergey et al. 1923 and proposal of the neotype strain. Int J Syst Evol Microbiol 1978;28:201–8.