Taxono-genomics description of '*Lactobacillus raoultii* sp. nov.', strain Marseille-P4006^T, a new *Lactobacillus* species isolated from the female genital tract of a patient with bacterial vaginosis

B. Nicaise[1,2], M. Maaloum[1,2,3], C. I. Lo[1,2], N. Armstrong[2,4], F. Bretelle[2,4,5], P.-É. Fournier[1,2], K. Diop[1,2] and F. Fenollar[1,2]
1) *Aix Marseille University, IRD, AP-HM, SSA, VITROME,* 2) *IHU-Méditerranée Infection, Marseille, France,* 3) *Faculty of Sciences Ben M'Sik, Laboratory of Biology and Health, Hassan II University, Casablanca, Morocco,* 4) *Aix-Marseille University, IRD, AP-HM, MEPHI and* 5) *Department of Gynaecology and Obstetrics, Gynépole, Hôpital Nord, AP-HM, Marseille, France*

Abstract

Strain Marseille-P4006^T, a Gram-stain-positive, rod-shaped, non-sporulating, facultatively anaerobic bacterium, was isolated from the vaginal swab of a 45-year-old woman with recurrent bacterial vaginosis. We studied its phenotypic characteristics and sequenced its whole genome. The major fatty acids were $C_{16:0}$ (48%), $C_{19:1n9}$ (14%) and $C_{18:0}$ (11%). The 3 070 142-bp-long genome contains 2855 protein-coding genes and 68 RNAs. Strain Marseille-P4006^T exhibited 98.1% 16S rRNA similarity with *Lactobacillus farraginis*, the closest species phylogenetically. Thus, strain Marseille-P4006 is distinct enough to represent a new species for which we propose the name *Lactobacillus raoultii* sp. nov. The type strain is Marseille-P4006^T.
© 2019 Published by Elsevier Ltd.

Keywords: Bacterial vaginosis, culturomics, *Lactobacillus raoultii*, taxono-genomics, vaginal flora
Original Submission: 11 February 2019; **Accepted:** 14 March 2019
Article published online: 22 March 2019

Corresponding author: F. Fenollar, VITROME, Institut hospitalo-universitaire Méditerranée-infection, 19-21 Boulevard Jean Moulin, 13005, Marseille, France.
E-mail: florence.fenollar@univ-amu.fr

Introduction

The vaginal microbiota is a dynamic ecosystem containing various types of microorganisms. Doderleïn (1892) and Beijerink (1901) were the first to describe its composition, asserting that the healthy vaginal flora is dominated by species of *Lactobacillus*, mainly *L. cripastus, L. iners, L. jensenii* and *L. gasseri* [1,2]. Other bacteria—such as *Bacteroides, Corynebacterium,* and *Peptostreptococcus* spp.—are also present in the vaginal microbiota [3]. This healthy ecosystem protects the vaginal environment by producing antimicrobial compounds such as lactic acid, hydrogen peroxide and bacteriocins that block or inhibit the growth of pathogens [1,4].

Bacterial vaginosis involves a disturbance of this vaginal ecosystem with a decrease in *Lactobacillus* spp. and an increase in anaerobic bacteria such as *Atopobium vaginae* and species of *Bacteroides*, *Anaerococcus*, *Prevotella*, *Peptoniphilus*, and *Mobiluncus* [5–7]. Bacterial vaginosis is a common cause of vaginal discharge in women of childbearing age and can be responsible for preterm deliveries [8–10]. The vaginal microbiota has been studied by conventional culture methods that are limited because 80% of the bacterial microbiota is considered fastidious or not cultivable [11]. Understanding of the human vaginal microbiota and its components has been enhanced by molecular techniques such as sequencing and phylogenetic analysis of the 16S rRNA gene. Thanks to these techniques, some fastidious and uncultured bacteria—named bacterial vaginosis-associated bacteria types 1, 2 and 3 (BVAB1, BVAB2, and BVAB3)—can be detected [12].

In a study on the diversity of vaginal microbiota in patients with bacterial vaginosis using the concept of culturomics [13], we isolated a previously unknown bacterium, closely related to members of the genus *Lactobacillus* in the family Lactobacillaceae, designated Marseille-P4006. The family Lactobacillaceae, created in 1917 by Winslow, contains 240 species grouped into

© 2019 Published by Elsevier Ltd
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
https://doi.org/10.1016/j.nmni.2019.100534

three genera (www.bacterio.net/lactobacillaceae.html), of which the genus *Lactobacillus* alone includes 224 species (www.bacterio.net/lactobacillus.html).

Herein, we report the taxono-genomic description [14,15] of *Lactobacillus raoultii,* strain Marseille-P4006^T (= CSUR P4006^T = CCUG 71848^T), a new species of the genus *Lactobacillus* isolated in the vaginal flora of a 45-year-old woman with recurrent bacterial vaginosis, including its complete annotated genome.

Material and methods

Sample collection

In February 2017 a vaginal sample from a 45-year-old French woman was collected at the Institut Hospitalo-Universitaire Méditerranée Infection (France). The woman was suffering from recurrent bacterial vaginosis, which was diagnosed as previously reported [16]. Since 2015 she had been treated with secnidazole, metronidazole, dalacin, and azithromycin associated with different probiotics, but at the time of sample collection she had not been treated with any antibiotics for more than a month. She gave her written consent. This study was authorized by the local IFR48 ethics committee (Marseille, France) under number 09-022. The sample was collected and transported using a Sigma Transwab (Medical Wire, Corsham, United Kingdom) and processed within an hour.

Strain identification by matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry

After collection, the vaginal swab was preincubated in a customized medium under anaerobic conditions at 37°C. The customized medium comprised pancreatic digest of casein (10 g/L), yeast extract (5 g/L), peptone (5 g/L), meat extract (3 g/L), L-cysteine HCl (0.1 g/L), dextrose (2.5 g/L), NaCl (5 g/L), $MgSO_4$ (0.1 g/L), $FeSO_4$ (0.02 g/L), K_2HPO_4 (0.83 g/L), tris(hydroxymethyl) aminomethane (3.69 g/L) and 10% horse blood, pH 5. The supernatant was then inoculated on 5% sheep-blood-enriched Columbia agar (bioMérieux, Marcy l'Etoile, France) and the agar plate was incubated for 2 days at 37°C under anaerobic conditions. Isolated colonies were deposited in duplicate on an MTP 96 MALDI-TOF target plate (Bruker Daltonics, Leipzig, Germany) [17]. Briefly, a 1.5-μL sample of matrix solution—containing a solution of α-cyano-4-hydroxycinnaic acid diluted in 500 μL acetonitrile, 250 μL HPLC water and 250 μL 10% trifluoroacetic acid—was deposited on each spot for crystallization and ionization. Spectra obtained were compared with those in the MALDI-TOF database. If the score was ≥2.0, the strain was considered identified. Otherwise, identification failed.

Strain identification by 16S rRNA sequencing

In the case of bacteria unidentified by MALDI-TOF mass spectrometry (MS), we achieved identification using 16S rRNA sequencing. When the 16S rRNA gene sequence similarity value was <98.7%, the strain was defined as a new species [18].

Morphological observation and growth conditions

Optimal growth of the strain was tested at different temperatures (25, 28, 37, 45, and 56°C) in aerobic, anaerobic and microaerobic atmospheres using GENbag Anaer and GENbag microaer systems (bioMérieux), respectively. To view cell morphology, electron microscopy was performed as previously described [19]. Standard procedures were used to determine Gram-stain reaction, motility, sporulation, and oxidase and catalase production [20].

Biochemical analysis and antibiotic susceptibility tests

Cellular fatty acid methyl ester (FAME) analysis was performed by gas chromatography/mass spectrometry (GC/MS). Two samples were prepared with approximately 2 mg of bacterial biomass per tube harvested from several culture plates. FAMEs were prepared as described by Sasser [21]. GC/MS analyses were carried out as described previously [22]. Briefly, FAMEs were separated using an Elite 5-MS column and monitored by mass spectrometry (Clarus 500 SQ 8S, Perkin Elmer, Courtaboeuf, France). Spectral database search was performed using MS Search 2.0 operated with the Standard Reference Database 1A (NIST, Gaithersburg, USA) and the FAMEs mass spectral database (Wiley, Chichester, UK).

In order to test biochemical characteristics, we used API ZYM, API 20A and API 50CH strips (bioMérieux) according to the manufacturer's instructions. Strips were incubated under anaerobic conditions for 4, 24 and 72 h. We performed the E-test gradient strips method (bioMérieux) to test antibiotic susceptibility and determined the minimal inhibitory concentration (MIC) of each antibiotic tested (benzylpenicillin, amoxicillin, ceftriaxone, imipenem, rifampicin, and vancomycin). A bacterial inoculum of turbidity 0.5 McFarland was prepared by suspending the culture in sterile saline (0.85% NaCl) using Marseille-P4006^T previously grown on Columbia agar (bioMérieux). Antibiotic susceptibility was performed conforming to EUCAST recommendations [23,24]. Elliptical zones of inhibition were formed around the strip, and the intersection with the strip indicated the MIC [23]. We interpreted MICs according to the EUCAST recommendations [25].

© 2019 Published by Elsevier Ltd, *NMNI*, **29**, 100534
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

DNA extraction and genome sequencing

Genomic DNA (gDNA) of strain Marseille-P4006^T was extracted in two steps. A mechanical treatment was first performed with acid-washed glass beads (G4649-500g Sigma) using a FastPrep BIO 101 instrument (Qbiogene, Strasbourg, France) at maximum speed for 90 s. Then, after 2 h of lysozyme incubation at 37°C, DNA was extracted on the EZ1 biorobot (Qiagen, Hilden, Germany) with an EZ1 DNA tissues kit. The elution volume was 50 μL. Genomic DNA was quantified by a Qubit assay with the high-sensitivity kit (Life Technologies, Carlsbad, CA, USA) to 25.6 ng/μL, and sequenced using 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 Inc.). The mate-pair library was prepared with 1.2 μg gDNA using the Nextera mate-pair Illumina guide. The gDNA sample was simultaneously fragmented and tagged with a mate-pair junction adapter. The pattern of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies Inc., Santa Clara, CA, USA) with a DNA 7500 LabChip. The DNA fragments ranged in size from 1.5 kb to 11 kb with an optimal size of 5.96 kb. No size selection was performed, and 600 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared into small fragments with an optimal at 703 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.) and the final concentration library was measured at 37.45 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-h run in a 2 × 151 bp. Total information of 3.07 Gb was obtained from a 479 K/mm^2 cluster density with a cluster passing quality control filters of 97.2% (9 282 000 passing filters paired reads). Within this run, the index representation for *Lactobacillus raoultii* was determined at 10.26%. The 952 663 paired reads were trimmed and then assembled.

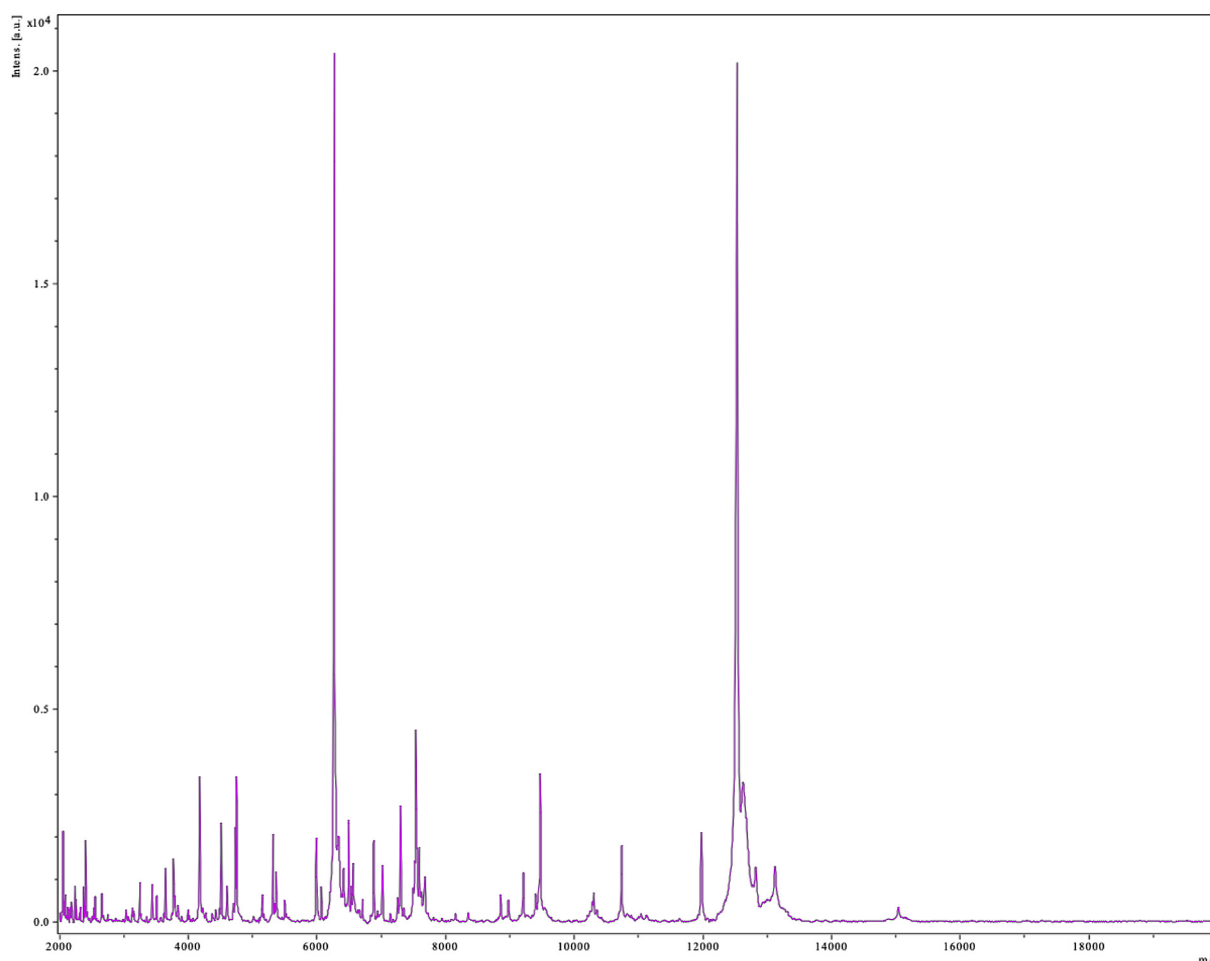

FIG. 1. Reference mass spectrum from *Lactobacillus raoultii* strain Marseille-P4006^T.

© 2019 Published by Elsevier Ltd, *NMNI*, **29**, 100534
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Genome annotation

Open reading frames (ORFs) were predicted using Prodigal [26] with default parameters. However, the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched against the GenBank [27] and clusters of orthologous groups (COGs) databases using BLASTP. DNA G+C content was identified by the RAST server [28], and the tRNAs and rRNAs were predicted using the tRNAScan-SE [29] and RNAmmer tools [30], respectively. Signal peptides and numbers of transmembrane helices were predicted using SignalP [31] and TMHMM [32], respectively; PHAge Search Tool (PHAST) was used to find prophage sequences within bacterial genomes [33]. ORFans were identified if their BLASTP E-value was $<1e^{-03}$ for alignment length >80 amino acids. If alignment lengths were <80 amino acids, we used an E-value of $1e^{-05}$. Such parameter thresholds have already been used in previous work to define ORFans. Artemis [34] and DNA Plotter [35] were used for data management and visualization of genomic features, respectively. To estimate the mean level of nucleotide sequence similarity at the genome level between strain Marseille-P4006 and its closest related species, we used the Average Genomic Identity Of Gene Sequences (AGIOS) inhouse software [36]. Briefly, this software uses the Proteinortho software [37] for the pairwise detection of orthologous proteins between genomes, then retrieves the corresponding genes and determines the mean percentage of nucleotide sequence identity among orthologous ORFs using the Needleman–Wunsch global alignment algorithm. In addition, we calculated between all genomes compared the digital DNA–DNA hybridization (dDDH) using Genome-to-Genome Distance Calculator (GGDC) software, as previously described [38].

Results

Strain characterization

Strain identification by MALDI-TOF. Strain Marseille-P4006^T was first isolated in February 2017 after 3 days of preincubation in a blood-culture bottle containing home-made culture medium (see Materials and methods) enriched with 10% horse blood under anaerobic conditions, and subcultured on a Columbia

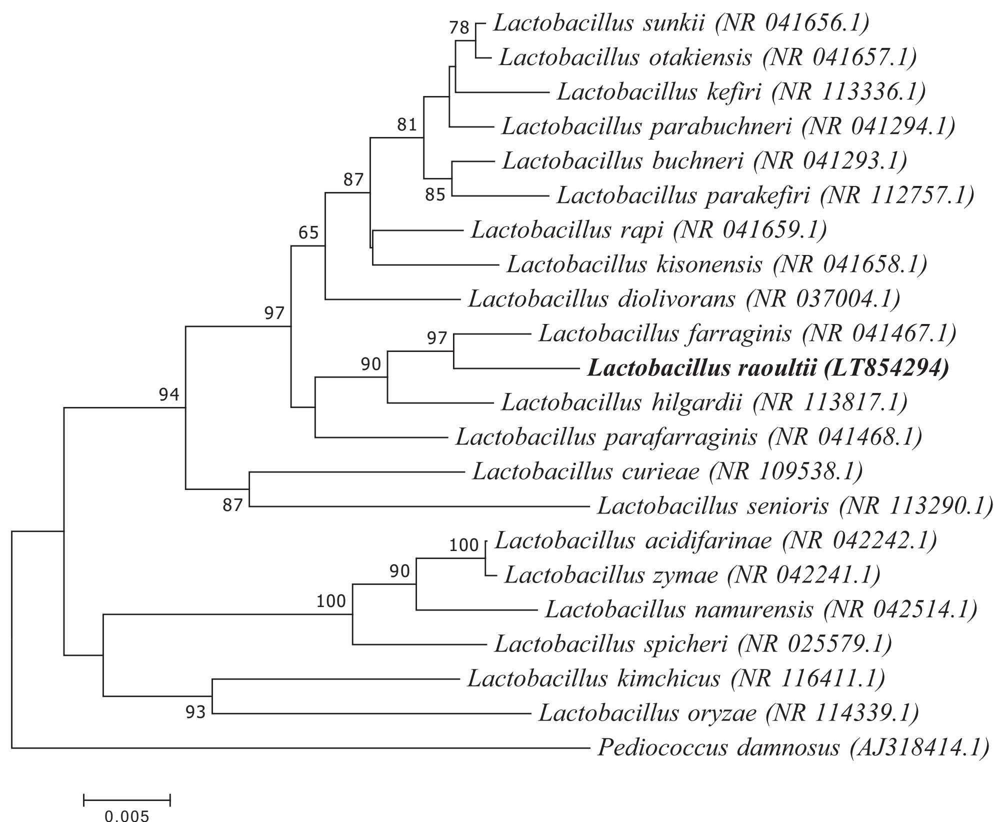

FIG. 2. This phylogenetic tree shows the positioning of '*Lactobacillus raoultii*' strain Marseille-P4006^T relative to its phylogenetically closest bacterial species with standing in nomenclature. 16rRNA sequences were aligned using Muscle v.3.8.31 with default parameters and phylogenetic interferences were obtained using the neighbour-joining method with 500 bootstrap replicates, by the MEGA7 software. The scale bar indicates a 0.5% nucleotide sequence divergence.

© 2019 Published by Elsevier Ltd, *NMNI*, **29**, 100534
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

agar plate (bioMérieux) at 37°C also under anaerobic conditions. MALDI-TOF MS analysis gave a low score (<1.6), suggesting that our isolate was not in the database and could be a previously unknown species. The reference mass spectrum represented in Fig. 1 was added in the URMS database (http://www.mediterranee-infection.com/article.php?laref=256&titre=urms-database).

Strain identification by 16S rRNA gene sequencing. The 16S rRNA gene was sequenced, and the sequence obtained (accession number LT854294) exhibited a 98.1% sequence similarity with that of Lactobacillus farraginis strain JCM8627 (GenBank accession AB690214.1), the phylogenetically closest bacterial species with standing in nomenclature (Fig. 2). As this value was <98.7%, the threshold defined by Stackebrandt and Ebers for defining a new species, we classified strain Marseille-P4006^T as a new species of the genus *Lactobacillus*, named *Lactobacillus raoultii* sp. nov. (Table 1).

Phenotypic characteristics. Strain Marseille-P4006^T grew under anaerobic and microaerobic conditions. Growth was observed at temperatures ranging from 25 to 45°C, with optimal growth at 37°C under anaerobic conditions after 48 h of incubation. The bacterium needed an NaCl concentration <0.5%, and the pH for growth ranged from 5 to 7.5. On Columbia agar, colonies were opaque white with a diameter of 1–1.2 mm. Gram staining showed a rod-shaped Gram-positive bacterium. Using electron microscopy, individual cells appeared with a mean diameter of 0.7 μm and a mean length of 1.6 μm (Fig. 3). Strain Marseille-P4006^T is non-motile and non-spore-forming.

The major fatty acids found for Marseille-P4006^T were $C_{16:0}$ (48%), $C_{19:1n9}$ (14%), and $C_{18:0}$ (11%). Minor amounts of unsaturated branched and other saturated fatty acids were also detected. This strain presented an unusual pair of unsaturated C_{19} structures (19:1n7 and 19:1n9) (Table 2).

Strain Marseille-P4006^T exhibited neither catalase nor oxidase activities. Using API ZYM strip, positive reactions were detected for leucine arylamidase, α-galactosidase, β-galactosidase, β-glucuronidase, α-glucosidase and β-glucosidase, and weakly positive reactions were detected for esterase, esterase lipase, valine arylamidase, cystine arylamidase, and naphthol-AS-BI-phosphohydrolase. Negative reactions were observed for alkaline phosphatase, lipase, trypsin, α-chymotrypsin, acid phosphatase, N-acetyl-β-glucosamidase, α-mannosidase, and α-fucosidase. An API 50 CHL strip revealed that strain Marseille-P4006^T metabolized D-galactose, D-fructose, melibiose, 5-ketogluconate, D-xylose, L-arabinose, D-ribose, D-glucose, maltose and potassium gluconate. The same strip exhibited negative reactions for glycerol, erythritol, D-arabinose, L-xylose, D-adonitol, methyl-βD-xylopyranoside, D-mannose, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl-αD-mannopyranoside, methyl-αD-glucopyranoside, N-acetylglucosamine, amygdaline, arbutin, salicilin, D-cellobiose, D-lactose, D-saccharose, D-trehalose, inulin, D-melezitose, D-raffinose, glycogen, xylitol, gentiobiose, D-turanose, D-lyxose, D-tagatose, starch, D-fucose, arabitol (D and L) and potassium 2-ketogluconate. On API 20A and API 20NE strips, nitrate was not reduced and indole formation was negative. API 20A also revealed that aesculin was hydrolysed and gelatine was not.

TABLE 1. Classification and general features of *Lactobacillus raoultii* strain Marseille-P4006^T

Properties	Terms
Taxonomy	***Kingdom:*** *Bacteria* ***Phylum:*** *Firmicutes* ***Class:*** *Bacilli* ***Order:*** *Lactobacillales* ***Family:*** *Lactobacillaceae* ***Genus:*** *Lactobacillus* ***Species:*** *Lactobacillus raoultii*
Type strain	Marseille-P4006
Isolation site	Human vagina
Isolation country	France
Gram stain	Positive
Cell shape	Rod
Motility	Non-motile
Oxygen requirements	Facultatively anaerobic
Optimal temperature	37°C
Temperature range	Mesophilic
Habitat	Host-associated
Biotic relationship	Free-living
Host name	*Homo sapiens*
Sporulation	Non-sporulating

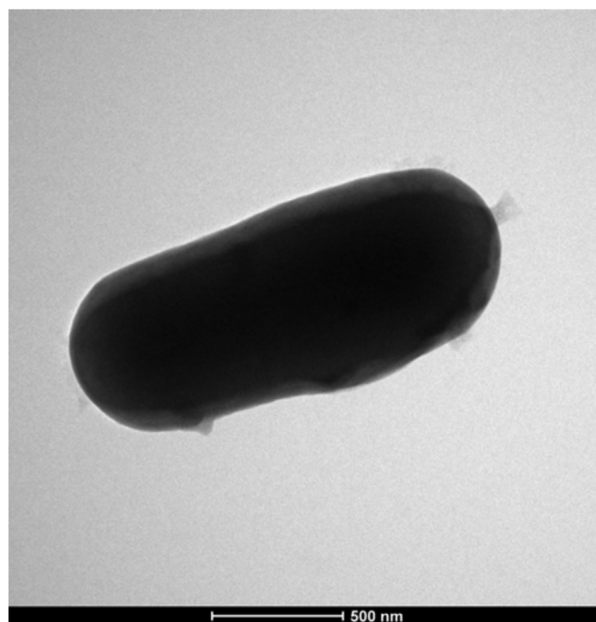

FIG. 3. Transmission electron microscopy of *Lactobacillus raoultii* strain Marseille-P4006^T using a Tecnai G20 transmission electron microscope (FEI Company). The scale bar represents 500 nm.

© 2019 Published by Elsevier Ltd, *NMNI*, **29**, 100534
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

TABLE 2. Cellular fatty acid composition (% ± standard deviation) of *Lactobacillus raoultii* strain Marseille-P4006ᵀ

Fatty acids	Name	Mean relative %[a]
16:0	Hexadecanoic acid	47.7 ± 1.2
19:1n9	10-Nonadecenoic acid	13.5 ± 0.5
18:00	Octadecanoic acid	11.2 ± 3.3
19:1n7	12-Nonadecenoic acid	9.1 ± 0.8
18:1n9	9-Octadecenoic acid	8.3 ± 0.5
18:1n7	11-Octadecenoic acid	3.5 ± 0.4
18:2n6	9,12-Octadecadienoic acid	2.2 ± 0.2
15:0	Pentadecanoic acid	1.6 ± 0.1
14:0	Tetradecanoic acid	1.0 ± 0.2
17:0	Heptadecanoic acid	TR
16:1n9	7-Hexadecenoic acid	TR
16:1n7	9-Hexadecenoic acid	TR
17:0 anteiso	14-methyl-Hexadecanoic acid	TR

TR, trace amounts (<1%).
[a]Mean peak area percentage.

Phenotypic characteristics of strain Marseille-P4006^T compared with those of closely related species are summarized in Table 3.

Strain Marseille-P4006^T was susceptible to imipenem (MIC 0.012 μg/L), rifampicin (MIC 0.032 μg/L), amoxicillin (MIC 0.38 μg/L), benzylpenicillin (MIC 1.5 μg/L) and ceftriaxone (MIC 16 μg/L) but resistant to metronidazole and vancomycin.

Genome properties

The final assembly of strain Marseille-P4006^T genome identified a genome size of 3 070 142 bp with a 41.4% G+C content. Of the 2924 predicted genes, 2855 were protein-coding genes and 68 were RNAs (five 5S rRNA, one 16S rRNA, one 23S rRNA, one tmRNA, and 60 tRNA genes) (Fig. 4, Table 4). In total 2047 genes (70%) were assigned a putative function: 606 genes were identified as ORFans and one prophage region of 5.5 kb with a total of six proteins and five CRISPRs were found. The properties and statistics of the genome are summarized in Table 4. The distribution of genes into COGs functional categories is presented in Table 5.

Genome comparison

We compared the genome of strain Marseille-P4006^T with the genomes of *L. hilgardii*, *L. diolivorans*, *L. curieae*, *L. senioris*, *L. farraginis* and *L. parafarraginis* (Table 6). The genome of strain Marseille-P4006^T (3.2 mega) is larger than those of *L. farraginis*, *L. hilgardii*, *L. curieae* and *L. senioris*, but smaller than those of *L. diolivrans* and *L. parafarraginis*. The GC content of strain Marseille-P4006^T (41.1%) is larger than in *L. diolivorans*, *L. hilgardii* and *L. senioris*, but smaller than in *L. parafarraginis* and *L. farraginis*. For gene content, strain Marseille-P4006^T coding genes (2855) are greater than those of *L. hilgardii* (2599), *L. senioris* (1960) and *L. curieae* (1539), but smaller than those of *L. parafarraginis* (3053), *L. diolivorans* (2962) and *L. farraginis* (3079). However, the distribution of genes into COG categories was similar in the seven genomes compared (Fig. 5), with the exception of the category X (mobilome), which was

TABLE 3. Differential characteristics of *Lactobacillus raoultii* strain Marseille-P4006ᵀ, *Lactobacillus senioris* strain DSM 24302, *Lactobacillus farraginis* strain DSM 18382, *Lactobacillus parafarraginis* strain DSM 18390, *Lactobacillus diolivorans* strain DSM 14421, *Lactobacillus hilgardii* Bergey's manual, *Lactobacillus curieae* strain JCM 18524 [39–43]

Properties	*L. raoultii*	*L. senioris*	*L. farraginis*	*L. parafarraginis*	*L. diolivorans*	*L. hilgardii*	*L. curieae*
Cell diameter (μm)	0.7 × 1.6	0.7 × 1.0–10.0	0.8 x 3–6	0.8 x 2–4	1 × 2	0.5–0.8 x 2–4	0.6 × 1.2–3
Major fatty acid	$C_{16:0}$ (48%)	$C_{18:1\omega9}$ (57%)	na	na	na	na	$C_{18:1\omega7c}$/$C_{18:1\omega9c}$ (46.36%)
DNA G+C content (%mol/L)	41.4	39.1	42.1	45.2	40	39.6	39.8
Production of							
Catalase	—	—	—	—	—	na	—
Oxidase	—	na	na	na	na	na	na
Acid production from							
D-Galactose	+	—	+	+	+	v	+
D-Fructose	+	+	+	+	+	na	+
Melibiose	+	—	+	+	+	—	+
5-Ketogluconate	+	—	na	na	w	na	—
D-Xylose	+	+	—	+	+	+	—
L-Arabinose	+	+	+	+	+	—	—
D-Ribose	+	+	+	+	+	+	na
D-Glucose	+	+	+	+	+	na	na
Aesculin	+	—	na	na	—	—	w
Maltose	+	—	+	+	+	+	+
Gluconate	+	+	w	w	+	na	na
D-Mannose	—	na	—	—	—	—	w
Methyl αD-glucopyranoside	—	—	na	na	w	na	+
Lactose	—	—	—	—	—	na	w
Sucrose	—	—	+	+	—	v	+
Melezitose	—	—	+	+	—	v	+
Raffinose	—	—	+	+	—	—	+
Turanose	—	—	na	na	—	na	+
Methyl βD-xylopyranoside	—	—	na	na	+	na	—
N-acetylglucosamine	—	W	na	na	—	na	—
Habitat	Human vagina	Human Faeces	Shochu residue	Shochu residue	Maize silage	Wine	Stinky tofu brine

+, positive reaction; –, negative reaction; v, variable; w, weakly positive; na, no available data.
Data are from the literature except for DNA G+C content which was calculated by RAST online software.

© 2019 Published by Elsevier Ltd, *NMNI*, **29**, 100534
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

overrepresented in *L. parafarraginis*, underrepresented in *L. senioris*, and absent in *L. curieae*.

Strain Marseille-P4006^T shared 1118, 1658, 1476, 1794, 1743, and 1550 orthologous genes with *L. senioris*, *L. farraginis*, *L. curieae*, *L. diolivorans*, *L. hilgardii, and L. parafarraginis,* respectively. Among the genomes compared, with the exception of the strain Marseille-P4006^T, AGIOS values ranged from 67.07% between *L. parafarraginis* and *L. senioris* to 78.17% between *L. hilgardii* and *L. farraginis*. When strain Marseille-P4006^T was compared to other species, AGIOS values ranged from 68.79% with *L. senioris* to 77.75% with *L. hilgardii* (Table 7). We obtained similar results for the analysis of the digital DNA–DNA hybridization (dDDH) (Table 8). When compared to the most closely related species, strain Marseille-P4006^T exhibited 19.7, 19.8, 20.4, 20.8, and 21.3% dDDH values with *L. parafarraginis*, *L. curieae*, *L. diolivorans*, *L. farraginis*, *L. hilgardii* and *L. senioris,* respectively.

Discussion

Strain Marseille-P4006^T was isolated as part of a 'culturomics' study of the vaginal flora with the aim of isolating all bacterial species within the vagina in a physiological situation and in bacterial vaginosis. A polyphasic taxono-genomics strategy [14,15], based on the combination of phenotypic and genomic analyses, was used to characterize strain Marseille-P4006^T. Strain Marseille-P4006^T was considered a new bacterial species of the genus *Lactobacillus* based on its unique MALDI-TOF MS spectrum, genome comparison, and its low 16S rRNA similarity level. The latter value was 98.1% with *Lactobacillus farraginis*, which was lower than the 98.7% threshold for defining a new species [18]. Strain Marseille-P4006^T is a member of the family Lactobacillaceae belonging to the phylum Firmicutes. This family comprises 240 species grouped in three validated genera (www.bacterio.net/lactobacillus.html). Most members of the Lactobacillaceae are Gram-positive and non-spore-forming rods. Species of the *Lactobacillus* genus have been detected in diverse habitats such as milk products, vegetation, food products

TABLE 4. Nucleotide content and gene count levels of the genome of *Lactobacillus raoultii* strain Marseille-P4006^T

Attribute	Value	% of total
Size (bp)	3 070 142	100%
G+C content (bp)	1 271 038	41.4%
Coding region (bp)	2 627 322	85.57%
Total genes	2924	100%
RNA genes	68	2.35%
Protein-coding genes	2855	97.64%
Genes with function prediction	2047	70.00%
Genes assigned to COGs	2195	75.06%
Genes with peptide signals	210	7.18%
Genes with transmembrane helices	828	28.31%
Protein associated with ORFans	606	20.72%

COGs, clusters of orthologous groups.

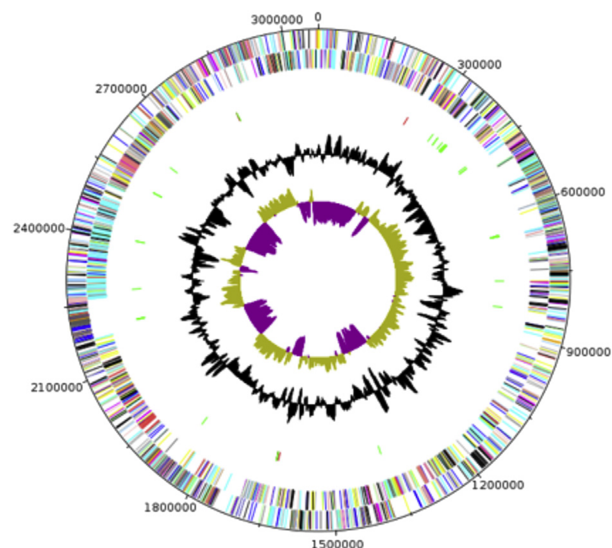

FIG. 4. Graphic circular map of the chromosome of *Lactobacillus raoultii* strain Marseille-P4006T. From the outside in: open reading frames oriented in the forward direction (coloured by COG categories), open reading frames oriented in the reverse direction (coloured by COG categories), RNA operon (red), and tRNAs (green), GC content plot, and GC skew (purple: negative values, olive: positive values).

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

Code	Value	% value[a]	Description
J	141	4.82	Translation
A	0	0	RNA processing and modification
K	178	6.09	Transcription
L	119	4.07	Replication, recombination and repair
B	0	0	Chromatin structure and dynamics
D	22	0.75	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	60	2.05	Defence mechanisms
T	61	2.09	Signal transduction mechanisms
M	112	3.83	Cell wall/membrane biogenesis
N	8	0.27	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	15	0.51	Intracellular trafficking and secretion
O	74	2.53	Posttranslational modification, protein turnover, chaperones
X	4	0.14	Mobilome: prophages, transposons
C	109	3.73	Energy production and conversion
G	200	6.84	Carbohydrate transport and metabolism
E	250	8.55	Amino acid transport and metabolism
F	87	2.98	Nucleotide transport and metabolism
H	107	3.66	Coenzyme transport and metabolism
I	65	2.22	Lipid transport and metabolism
P	117	4.00	Inorganic ion transport and metabolism
Q	25	0.85	Secondary metabolites biosynthesis, transport, and catabolism
R	239	8.17	General function prediction only
S	202	6.91	Function unknown
_	729	24.93	Not in COGs

COGs, clusters of orthologous groups.
[a]The total is based on the total number of protein-coding genes in the annotated genome.

© 2019 Published by Elsevier Ltd, *NMNI*, **29**, 100534
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

TABLE 6. Genomic comparison between *Lactobacillus raoultii* and closely related *Lactobacillus* species

Species	Strain	Genome accession number	Genome size (Mb)	GC content (%)	Gene content
Lactobacillus raoultii	**Marseille-P4006**	**OVSN00000000**	**3.07**	**41.4**	**2855**
Lactobacillus parafarraginis	DSM 18390	NZ_BBAR00000000	3.10	45.2	3053
Lactobacillus farraginis	DSM 18382	NZ_BAKI00000000	2.84	42.1	3079
Lactobacillus diolivorans	DSM 14421	NZ_AZEY00000000	3.26	40	2962
Lactobacillus hilgardii	DSM 20176	NZ_ACGP00000000	2.61	39.6	2599
Lactobacillus curieae	JCM 18524	NZ_CP018906	2.09	39.8	1960
Lactobacillus senioris	DSM 24302	NZ_AYZR00000000	1.56	39.1	1539

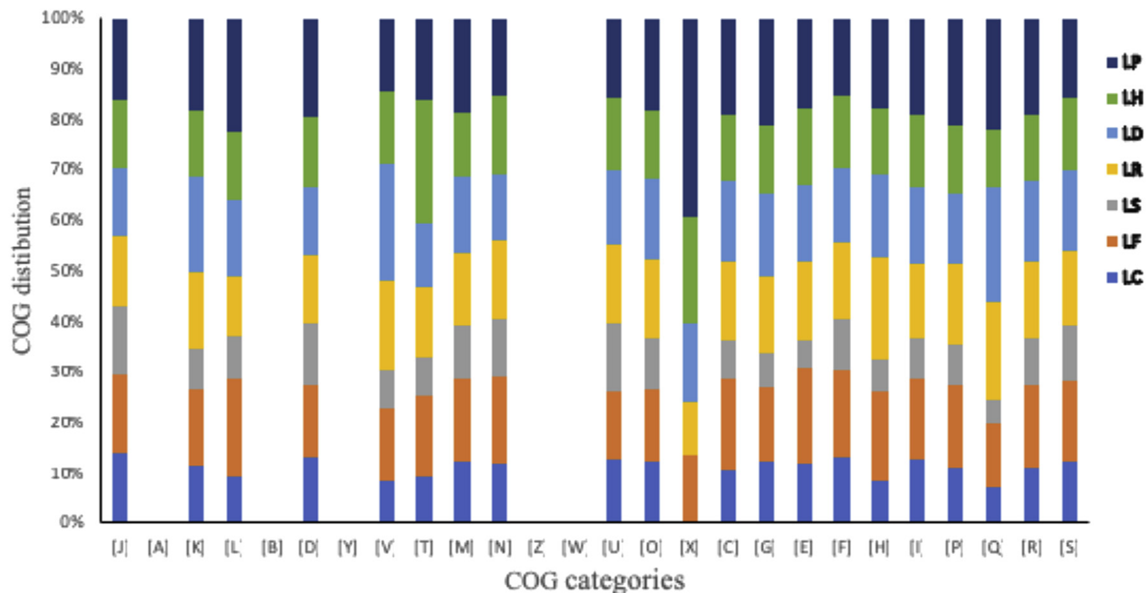

FIG. 5. Distribution of predicted genes into COG categories among other *Lactobacillus* species. LP, *L. parafarraginis*; LH, *L. hilgardii*; LD, *L. diolivorans*; LR, *L. raoultii*; LS, *L. senioris*; LF, *L. farraginis*; LC, *L. curieae*.

(especially fermented ones) and humans (genital tract, intestinal tract etc.) [39–42] (www.bacterio.net/lactobacillus.html). None of the phylogenetically most closely related species of *Lactobacillus* have pathogenic capabilities. Phenotypically, strain Marseille-P4006^T differs from the other closest bacterial species studied in their ability to ferment carbohydrate. Most of the most closely related species of *Lactobacillus* studied (*L. curieae*, *L. diolivorans*, *L. hilgardii*, *L. farraginis*, *L. parafarraginis*, *L. senioris*) ferment fructose, glucose, ribose and gluconate, while most do not metabolize 5-ketogluconate and aesculin ferric citrate.

The 16S rRNA gene nucleotide similarity below the threshold set to define new species (98.1%, >98.7%), the difference in G+C content (between 39.6 and 45.2%), and the values of AGIOS and dDDH values under the 70% threshold established to separate distinct species, confirmed that the strain Marseille-P4006^T is different from other *Lactobacillus* species.

TABLE 7. Genomic comparisons of *L. raoultii* with closely related *Lactobacillus* species[a]

Strains	*L. raoultii*	*L. senioris*	*L. farraginis*	*L. curieae*	*L. diolivorans*	*L. hilgardii*	*L. parafarraginis*
Lactobacillus raoultii	**2855**	1118	1658	1476	1794	1743	155
Lactobacillus senioris	68.79%	**1569**	974	1080	1112	1094	930
Lactobacillus farraginis	76.87%	68.00%	**3170**	1298	1572	1593	1385
Lactobacillus curieae	69.56%	69.20%	68.80%	**1992**	1470	1426	1199
Lactobacillus diolivorans	73.46%	69.16%	73.00%	70.03%	**3053**	1706	1482
Lactobacillus hilgardii	77.75%	68.82%	78.17%	69.85%	74.26%	**2592**	1461
Lactobacillus parafarraginis	73.50%	67.07%	74.05%	68.38%	71.84%	74.40%	**3629**

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

© 2019 Published by Elsevier Ltd, *NMNI*, **29**, 100534
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

TABLE 8. Digital DNA–DNA hybridization (dDDH) values obtained by comparing all genomes studied using the Genome-to-Genome Distance Calculator (GGDC), Formula 2

Strain	*L. raoultii*	*L. senioris*	*L. farraginis*	*L. curieae*	*L. diolivorans*	*L. hilgardii*	*L. parafarraginis*
Lactobacillus raoultii	100% ± 00	21.3% ± 2.30	20.4% ± 2.30	19.8% ± 2.30	19.8% ± 2.30	20.8% ± 2.30	19.7% ± 2.30
Lactobacillus senioris		100% ± 00	20.0% ± 2.30	18.7% ± 2.30	18.8% ± 2.25	19.6% ± 2.30	20.0% ± 2.30
Lactobacillus farraginis			100% ± 00	19.4% ± 2.30	20.5% ± 2.35	19.0% ± 2.30	21.2% ± 2.35
Lactobacillus curieae				100% ± 00	19.9% ± 2.30	21.5% ± 2.35	21.1% ± 2.35
Lactobacillus diolivorans					100% ± 00	20.8% ± 2.35	20.5% ± 2.35
Lactobacillus hilgardii						100% ± 00	20.4% ± 2.35
Lactobacillus parafarraginis							100% ± 00

Conclusion

Based on taxono-genomics results (phenotypic analysis as well as phylogenetic and genomic results), strain Marseille-P4006^T can be considered as a new species within the *Lactobacillus* genus. The name *Lactobacillus raoultii* sp. nov. is proposed for this new bacterium isolated from the vaginal microbiota.

Taxonomic and nomenclature proposal

Description of *Lactobacillus raoultii* sp. nov

The name *Lactobacillus raoultii* (*ra.oul'ti.i.* N. L. masc. gen. n. *raoultii)* was chosen to honour Professor Didier Raoult for his outstanding contribution in the field of medical microbiology.

L. raoultii is a facultatively anaerobic, mesophilic, non-motile and non-sporulating Gram-stain-positive rod. Its optimal growth occurs at 37°C. On Columbia agar, colonies are opaque white with a diameter of approximately 0.5 mm. The cells are rod-shaped with a mean width of 0.7 μm and a mean length of 1.6 μm. The species exhibits neither catalase nor oxidase activities; nitrate is not reduced. Positive reactions are observed for D-galactose, D-fructose, melibiose, 5-ketogluconate, D-xylose, L-arabinose, D-ribose, D-glucose, maltose, potassium gluconate, leucine arylamidase, α-galactosidase, β-galactosidase, β-glucuronidase, α-glucosidase and β-glucosidase and weakly positive reactions were detected for esterase, esterase lipase, valine arylamidase, cystine arylamidase and naphthol-AS-BI-phosphohydrolase. Its habitat is the human vaginal microbiota. The major fatty acids are $C_{16:0}$ (48%), $C_{19:1n9}$ (14%), and $C_{18:0}$ (11%). It is susceptible to imipenem, rifampicin, amoxicillin, benzylpenicillin and ceftriaxone but is resistant to metronidazole and vancomycin.

The *L. raoultii* genome measures 3 070 142 bp long and exhibits 41.4% DNA G+C content. Its 16S rRNA gene and whole genome sequences are both deposited in GenBank under accession numbers LT854294 and OVSN00000000, respectively. The type strain, Marseille-P4006^T (= CSUR P4006^T = CCUG 71848^T), was isolated from the vaginal flora of a patient with recurrent bacterial vaginosis.

Transparency declaration

The authors declare no conflict of interest. This study was funded by the Institut Hospitalo-Universitaire (IHU) Méditerranée Infection, the National Research Agency under the program 'Investissements d'avenir', reference ANR-10-IAHU-03, the Région Provence-Alpes-Côte d'Azur and the European funding FEDER PRIMI.

Acknowledgements

We thank Frédéric Cadoret for technical help.

References

[1] Lepargneur JP, Rousseau V. Protective role of the Doderleïn flora. J Gynecol Obstet Biol Reprod 2002;31:485–94.
[2] Chaban B, Links MG, Jayaprakash TP, Wagner EC, Bourque DK, Lohn Z, et al. Characterization of the vaginal microbiota of healthy Canadian women through the menstrual cycle. Microbiome 2014;2:23.
[3] Kumar N, Behera B, Sagiri SS, Pal K, Ray SS, Roy S. Bacterial vaginosis: etiology and modalities of treatment—a brief note. J Pharm Bioallied Sci 2011;3:496–503.
[4] Borges S, Silva J, Teixeira P. The role of lactobacilli and probiotics in maintaining vaginal health. Arch Gynecol Obstet 2014;289:479–89.
[5] Pépin J, Deslandes S, Giroux G, Sobéla F, Khonde N, Diakité S, et al. The complex vaginal flora of West African women with bacterial vaginosis. PloS One 2011;6:e25082.
[6] 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.
[7] Onderdonk AB, Delaney ML, Fichorova RN. The human microbiome during bacterial vaginosis. Clin Microbiol Rev 2016;29:223–38.
[8] Leitich H, Kiss H. Asymptomatic bacterial vaginosis and intermediate flora as risk factors for adverse pregnancy outcome. Best Pract Res Clin Obstet Gynaecol 2007;21:375–90.

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

[9] Leitich H, Bodner-Adler B, Brunbauer M, Kaider A, Egarter C, Husslein P. Bacterial vaginosis as a risk factor for preterm delivery: a meta-analysis. Am J Obstet Gynecol 2003;189:139–47.
[10] Svare JA, Schmidt H, Hansen BB, 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.
[11] Lamont RF, Sobel JD, Akins RA, Hassan SS, Chaiworapongsa T, Kusanovic JP, et al. The vaginal microbiome: new information about genital tract flora using molecular based techniques. BJOG 2011;118: 533–49.
[12] Fredricks DN, Fiedler TL, Marrazzo JM. Molecular identification of bacteria associated with bacterial vaginosis. N Engl J Med 2005;353: 1899–911.
[13] 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.
[14] 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.
[15] Fournier PE, Drancourt M. New Microbes New Infections promotes modern prokaryotic taxonomy: a new section 'TaxonoGenomics: new genomes of microorganisms in humans. New Microbe New Infect 2015;7:48–9.
[16] 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;47:33–43.
[17] 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.
[18] Kim M, Oh HS, 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.
[19] Diop K, Diop A, Bretelle F, Cadoret F, Michelle C, Richez M, et al. *Olegusella massiliensis* gen. nov., sp. nov., strain KHD7^T, a new bacterial genus isolated from the female genital tract of a patient with bacterial vaginosis. Anaerobe 2017;44:87–95.
[20] Standards for microbiology investigations (SMI) https://www.gov.uk/government/collections/standards-for-microbiology-investigations-smi#test-procedures.
[21] Sasser M. Bacterial identification by gas chromatographic analysis of fatty acids methyl esters (GC-FAME). Newark, NY: MIDI; 2006. Technical Note #101, http://youngin.com/application/AN-0702-0013EN.pdf.
[22] Dione N, Sankar SA, Lagier J-C, Khelaifia S, Michele C, Armstrong N, et al. Genome sequence and description of *Anaerosalibacter massiliensis* sp. nov. New Microbe. New Infect 2016;10:66–76.
[23] Citron DM, Ostovari MI, Karlsson A, Goldstein EJ. Evaluation of the E test for susceptibility testing of anaerobic bacteria. J Clin Microbiol 1991;29:2197–203.
[24] Matuschek E, Brown DFJ, Kahlmeter G. Development of the EUCAST disk diffusion antimicrobial susceptibility testing method and its implementation in routine microbiology laboratories. Clin Microbiol Infect 2014;20:O255–66.
[25] EUCAST: MIC distributions and ECOFFs http://www.eucast.org/mic_distributions_and_ecoffs/.
[26] Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform 2010;11:119.
[27] Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank Nucleic Acids Res 2012;40:D48–53.
[28] Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Serv: rapid annotations using subsystems technology. BMC Genomics 2008;9:75.
[29] Avni D, Biberman Y, Meyuhas O. The 5′ terminal oligopyrimidine tract confers translational control on TOP mRNAs in a cell type- and sequence context-dependent manner. Nucleic Acids Res 1997;25: 995–1001.
[30] Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 2007;35:3100–8.
[31] Dyrløv Bendtsen J, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340:783–95.
[32] 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.
[33] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res 2011;39:W347–52.
[34] Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. Bioinformatics 2000;16:944–5.
[35] Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 2009;25:119–20.
[36] 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.
[37] Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (Co-)orthologs in large-scale analysis. BMC Bioinform 2011;12:124.
[38] Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinform 2013;14:1.
[39] Oki K, Kudo Y, Watanabe K. *Lactobacillus saniviri* sp. nov. and *Lactobacillus senioris* sp. nov., isolated from human faeces. Int J Syst Evol Microbiol 2012;62:601–7.
[40] Endo A, Okada S. *Lactobacillus farraginis* sp. nov. and *Lactobacillus parafarraginis* sp. nov., heterofermentative lactobacilli isolated from a compost of distilled shochu residue. Int J Syst Evol Microbiol 2007;57: 708–12.
[41] Krooneman J, Faber F, Alderkamp AC, Elferink SJ, Driehuis F, Cleenwerck I, et al. *Lactobacillus diolivorans* sp. nov., a 1,2-propanediol-degrading bacterium isolated from aerobically stable maize silage. Int J Syst Evol Microbiol 2002;52:639–46.
[42] Lei X, Sun G, Xie J, Wei D. *Lactobacillus curieae* sp. nov., isolated from stinky tofu brine. Int J Syst Evol Microbiol 2013;63:2501–5.
[43] Hammes WP, Hertel C. Genus *Lactobacillus* Beijerinck 1901, 212. In: Bergey's Manual of systematic bacteriology, vol. 3. Springer; 2009.

© 2019 Published by Elsevier Ltd, *NMNI*, **29**, 100534
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).