Non-contiguous finished genome sequencing and description of *Enterococcus timonensis* sp. nov. isolated from human sputum

M. D. Mbogning Fonkou[1], M. Bilen[1], N. Gouba[2], S. Khelaifia[1], F. Cadoret[1], T. T. Nguyen[1], M. Richez[1], F. Bittar[1], P.-E. Fournier[3], D. Raoult[1,4] and G. Dubourg[1]

1) *MEPHI, UMR, IRD, Aix-Marseille Univ, Marseille, France, Facultés de Médecine et de Pharmacie, Marseille, France,* 2) *Institut Supérieur des Sciences de La Santé (INSSA), Université Polytechnique de Bobo Dioulasso, Burkina Faso,* 3) *Aix Marseille Univ, IRD, VITROME, IHU - Méditerranée Infection, 19-21 Boulevard Jean Moulin, 13005, Marseille, France and* 4) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Enterococcus timonensis sp. nov., strain Marseille-P2817^T, is a facultatively anaerobic, motile and non-spore-forming Gram-positive coccus which was isolated from the sputum of a healthy adult man in Marseilles. We present herein its phenotypic description together with MALDI-TOF (matrix-assisted laser-desorption/ionization time-of-flight) mass spectrometry analysis and genome sequencing and comparison. The genome of *Enterococcus timonensis* is 2 123 933 bp long with 38.46 mol% of G+C content, and it contains 1983 protein-coding genes and 65 RNA genes (including nine rRNA genes).
© 2019 The Authors. Published by Elsevier Ltd.

Keywords: Culturomics, *Enterococcus timonensis*, genome, sputum, taxono-genomics
Original Submission: 8 November 2018; **Accepted:** 14 March 2019
Article published online: 21 March 2019

Corresponding author: G. Dubourg, IHU Méditerranée Infection, 19-21 Bd Jean Moulin, 13385 Marseille cedex 5. France.
E-mail: greg.dubourg@gmail.com

Introduction

As a part of the rebirth of culture, the culturomics approach has enabled the discovery of hundreds of new species isolated from human gut [1], thus contributing to a dramatic increase in the repertoire of bacteria associated with humans. Taxonogenomics was recently introduced [1] to describe these new taxa, combining phenotypic characteristics such as mass spectrum by matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) analysis, biochemical properties, and genomic characteristics including 16S rRNA gene phylogeny, DNA–DNA hybridization (DDH), and G+C content [2,3]. Among strategies adopted to enhance the diversification of the specimen is to improve culturomic performances [1,4]. In particular, inclusion of sputum samples allows description of the respiratory microbiota of healthy people, thus expanding the known repertoire of bacteria colonizing the respiratory tract.

We recently isolated *Enterococcus timonensis* strain Marseille-P2817^T. This new species, belonging to the genus *Enterococcus*, was cultured from a sputum sample from a healthy man in Marseille as part of a culturomics study [5]. The history of the enterococci began in 1899 with Thiercelin [6,7] when they were classified in the genus *Streptococcus* (Lancefield group D) until 1984 [6,7]. Based on genome analysis, *Streptococcus faecalis* and *Streptococcus faecium* have been transferred to a new genus [5]. Since then, urinary tract infections, diverticulitis, bacterial endocarditis, bacteraemia, and meningitis are important clinical infections caused by *Enterococcus* spp. [6,7]. However, enterococcus-associated lower respiratory tract infections are very rare [8]. At the time of writing, according to the List of Prokaryotic Names with Standing in Nomenclature (LPSN; http://www.bacterio.net), the genus *Enterococcus* consists of 58 species and two subspecies.

Herein we present a summary classification and a set of characteristics for *Enterococcus timonensis* strain Marseille-P2817^T (= DSM 103162, = CSUR P2817). In addition, we

© 2019 The Authors. 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.2019.100532

propose the description of the complete genome sequence and annotation.

Materials and methods

Ethics and sample collection

In February 2016, a sputum sample was obtained from a healthy 37-year-old French adult man living in Marseille, France. Informed and signed consent was obtained from the patient and the study was approved by the Institut Fédératif de Recherche 48, Faculty of Medicine, Marseilles, France, under agreement number 2016-011.

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

Sputum samples were cultured using 18 different culture conditions of culturomics [4]. Strain Marseille-P2817^T was isolated on 5% sheep-blood-enriched Columbia agar (bioMérieux, Marcy l'Étoile, France), 10 days after being cultured in a culture bottle containing a blood-enriched Columbia agar liquid medium (BACTEC™ Plus Aerobic/F Culture Vials (Becton, Dickinson and Company)) with 4 mL rumen fluid at 30°C. This bacterium was tested for identification with a Microflex spectrometer (Brüker Daltonics, Leipzig, Germany) as previously described [9,10] and the 12 spectra generated were compared to the 7,567 different bacteria found in our database by standard pattern matching (for which default parameter settings were applied), using MALDI BioTyper database software (version 2.0, Brüker). A resulting score of <1.7 didn't enable identification, and a 16S rRNA gene sequencing was performed as previously described [11]. Codon Code Aligner software (http://www.codoncode.com) was used to assemble and correct sequences, and BLASTn searches were performed in the National Center for Biotechnology Information (NCBI; http://blast.ncbi.nlm.nih.gov.gate1.inist.fr/Blast.cgi). Concerning similarity levels, for thresholds of 98.65% and 95%, a new species or a new genus was suggested respectively as proposed by Meier-Kolthoff et al., 2013 [12] and Kim et al., 2014 [13].

Phylogenetic analyses

We used a custom python script to automatically retrieve all species belonging to the family of the new species and then downloaded 16S sequences from the NCBI by parsing NCBI eUtils results and NCBI taxonomy page. Only sequences from type strains were kept. In case of multiple sequences for one type strain, the sequence with the best identity rate from the BLASTn alignment with our sequence was selected. The 16S sequences are then separated into two groups: one containing the sequences of strains from the same genus (group a) and one containing the others (group b). The 15 closest strains from group a and the closest one from group b are finally kept. If the script is unable to get 15 sequences from group a, it selects more sequences from group b to get at least nine strains from both groups.

Growth conditions

To assess its range of growth temperatures, strain Marseille-P2817^T was cultured at different temperatures (25, 30, 37, 42 and 57°C) on 5% sheep-blood-enriched Columbia agar (bioMérieux) under aerobic, anaerobic and microaerophilic conditions using GENbag Anaer and GENbag miroaer systems (bioMerieux). Aerobic growth was achieved with and without 5% CO_2. Also, a salinity test was performed at different concentrations (5, 10, 50, 75 and 100 g/L), and four different pHs (6, 6.5, 7 and 8.5) were tested.

Biochemical, sporulation and motility assays

API Gallery systems—API ZYM, API 20 NE and API 50CH—were used to investigate biochemical analyses according to the manufacturer's instructions (bioMérieux). Catalase (bioMérieux) and oxidase (Becton Dickinson, Franklin Lakes, NJ, USA) tests were also performed separately. A thermal shock at 80°C for 30 min was done on bacterial colonies (diluted in phosphate-buffered saline) to check for sporulation ability in this bacterium. A DM1000 photonic microscope (Leica Microsystems, Nanterre, France) was used to assess the motility of the bacteria by observing the fresh colony between blades and slats with a 40 × objective lens.

Antibiotic susceptibility

The antibiotic susceptibility of strain Marseille-P2817^T was tested using the E-test strips method according to EUCAST 2015 recommendations (http://www.eucast.org/). Eighteen different antibiotics were used, including teicoplanin (TP) 0.016–256 μg/mL, daptomycin (DPC) 0.016–256 μg/mL, metronidazole (MZ) 0.016–256 μg/mL, rifampicin (RI) 0.002–32 μg/mL, imipenem (IP) 0.002–32 μg/mL, minocycline (MC) 0.016–256 μg/mL, benzylpenicillin (PG) 0.002–32 μg/mL, benzylpenicillin (PG) 0.016–256 μg/mL, vancomycin (VA) 0.016–256 μg/mL, cefotaxime (CT) 0.002–32 μg/mL, amikacin (AK) 0.016–256 μg/mL, erythromycin (EM) 0.016–256 μg/mL, ceftriaxone (TX) 0.016–256 μg/mL, amoxicillin(AC) 0.016–256 μg/mL, tobramycin (TM) 0.016–256 μg/mL, fosfomycin (FM) 0.064–1024 μg/mL, doxycycline (DC) 0.016–256 μg/mL and ertapenem (ETP) 0.002–32 μg/mL. Breakpoint tables for the interpretation of MICs and inhibition zone diameters (version 7.1, 2017) were used to interpret the results; these are available at *http://www.eucast.org*.

© 2019 The Authors. Published by Elsevier Ltd, *NMNI*, **29**, 100532
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Microscopy

The cells were first fixed with 2.5% glutaraldehyde in 0.1 M cacodylate buffer for at least 1 h at 4°C. A drop of cell suspension was deposited for approximately 5 min on glow-discharged formvar carbon film on 400 mesh nickel grids (FCF400-Ni, EMS). The latter were then dried on blotting paper and cells were negatively stained for 10 s with 1% ammonium molybdate solution in filtered water at room temperature. We then acquire electron micrographs using a Tecnai G20 Cryo (FEI) transmission electron microscope operated at 200 keV.

Fatty acid methyl ester (FAME) analysis by gas chromatography/mass spectrometry (GC/MS)

Two samples were prepared with approximately 20 mg of bacterial biomass per tube harvested from several petri dish cultures. The method described by Sasser et al. (2006) was used to prepare FAMEs [14]. We conducted GC/MS as previously described [15]. Briefly, we separated FAMEs using an Elite 5-MS column monitored by mass spectrometry (Clarus 500 - SQ 8 S, Perkin Elmer, Courtaboeuf, France). We performed a spectral database search using MS Search 2.0 operated with the Standard Reference Database 1A (NIST, Gaithersburg, USA) and the FAMEs mass spectral database (Wiley, Chichester, UK).

DNA extraction and genome sequencing

DNA of strain Marseille-P2817^T was extracted on the EZ1 biorobot (Qiagen) with EZ1 DNA tissues kit following a pre-treatment by a lysozyme incubation at 37°C for 2 h. The elution volume was 50 μL. Genomic DNA (gDNA) was quantified by a Qubit assay with the high-sensitivity kit (Life technologies, Carlsbad, CA, USA) to 46.6 ng/μL. Then gDNA was sequenced by 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 using the Nextera Mate Pair sample prep kit (Illumina). We prepared the mate pair library with 1.5 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 size of the DNA fragments ranged from 1.5 kb to 11 kb with an optimal size of 7.710 kb. No size selection was performed, and 600 ng of tagmented fragments were circularized. The circularized DNA obtained was then mechanically sheared into small fragments with optima on a bimodal curve at 843 and 1565 bp on the Covaris device S2 in T6 tubes (Covaris, Woburn, MA, USA). We visualized the library profile on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies Inc, Santa Clara, CA, USA). The final concentration library was measured at 59.91 nmol/L. The libraries were normalized at 2 nM and then 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 runs were performed in a single 39-h run in a 2 × 151-bp. We obtained a total information of 8.3 Gb from a 910 K/mm^2 cluster density with a cluster passing quality control filters of 92.8% (16 316 000 passing filter paired reads). We determined the index representation for strain Marseille-P2817^T within this run, which was of 8.06%. Finally, the 1 315 710 paired reads were trimmed and then assembled.

Genome assembly, annotation and comparison

The assembly of the genome was carried out using a pipeline combining different softwares (Velvet [16], Spades [17] and Soap Denovo [18]) on trimmed (MiSeq and Trimmomatic [19] softwares) or untrimmed data (MiSeq software only). For each of the six assemblies performed, GapCloser [20] was used to reduce gaps. We identified contamination with Phage Phix (BLASTn against Phage Phix174 DNA sequence) which was then eliminated. The scaffolds (<800 bp) were then removed, and scaffolds with a depth value < 25% of the mean depth were removed as they were identified as possible contaminants. The best assembly was selected by using different criteria (number of scaffolds, N50, number of N). For the strain Marseille-P2817^T, Spades gave the best assembly, with depth coverage of 239x.

We predicted open reading frames (ORFs) using Prodigal [20] with default parameters, but the predicted ORFs were excluded if they spanned a sequencing gap region (contains N). We searched for the predicted bacterial protein sequences against the clusters of orthologous groups (COGs) database using BLASTP (*E*-value 1e^{-03}, coverage of 0.7 and identity percentage of 30%). If no hit was found, we searched against the NR database using BLASTP with *E*-value of 1e^{-03}, coverage of 0.7 and identity percentage of 30%. If sequence lengths were <80 amino acids, an *E* value of 1e^{-05} was used. The tRNA genes and rRNAs were found using tRNAScanSE tool [21] and RNAmmer [22], respectively, while we predicted lipoprotein signal peptides and the number of transmembrane helices using Phobius [22]. ORFans were identified if all the performed BLASTP did not give positive results (*E*-value <1e^{-03} for ORFs with sequence size >80 aa or *E*-value <1e^{-05} for ORFs with sequence length <80 amino acids). The HMMscan of the HMMER3 suite [23] were used to search PFAM conserved domains (PFAM-A and PFAM-B domains) on each protein. We searched PKS and NRPS against the ClusterMine360 [24] and analysed the resistome by using ARG-ANNOT database [25]. Toxin and antitoxin were found using a database composed of TADB database [26] and homemade toxin and antitoxin of

© 2019 The Authors. Published by Elsevier Ltd, *NMNI*, **29**, 100532
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

TABLE 1. Classification and general features of *Enterococcus timonensis* strain Marseille-P2817^T

Properties	Term
Current classification	Domain: *Bacteria* Phylum: *Firmicutes* Class: *Bacilli* Order: *Lactobacillales* Family: *Enterococcaceae* Genus: *Enterococcus* Species: *Enterococcus timonensis* Type strain: Marseille-P2817^T
Gram stain	Positive
Cell shape	Coccus
Motility	Motile
Sporulation	Non-sporulating
Temperature range	Mesophile
Optimum temperature	37°C
Oxygen requirement	Facultative anaerobe
Habitat	Human lung
Isolation	Human sputum

Rickettsia and *Wolbachia* (which were provided by Cristina Socolovschi). The mobilome was analysed using Phantome [27], ICEs (ICEberg) [28], ACLAME [29], GYSPSYDB [30] and CRISPR [31]. For the latter we used spacer and repeat database to reconstruct CRISPR module. A homemade database [32] was used to find bacteriocin. Virulence factors were analysed by using both VFDB [33] and MvirDB [34]. Species which had to be compared were automatically retrieved from the 16s RNA tree using Phylopattern [35]. For each selected species, complete genome sequence, proteome and ORFeome sequences were retrieved from the FTP of NCBI. If no complete and available genome was found for one specific strain, a complete genome of the same species was used. If ORFeome and proteome were not predicted, Prodigal was used with default parameters to predict them. Protein Ortho was used to analyse all proteomes [36]. Then for each couple of genomes, a similarity score was computed. This score is the mean value of nucleotide similarity between all couples of orthologous genes between the two genomes studied (AGIOS) [3]. An annotation of the entire proteome was performed to define the distribution of functional classes of predicted genes according to the COG of proteins (using the same method as for the genome annotation). Two parameters were determined to evaluate the genomic similarity among the compared strains, dDDH that exhibits a high correlation with DDH [12,37] and AGIOS [3] which was designed to be independent from DDH. We used the GGDC web server to perform Genome-to-Genome Distance Calculator (GGDC) [12]. Annotation and comparison processes were performed in the Multi-Agent software system DAGOBAH [38] that included Figenix [39] libraries which provided pipeline analysis.

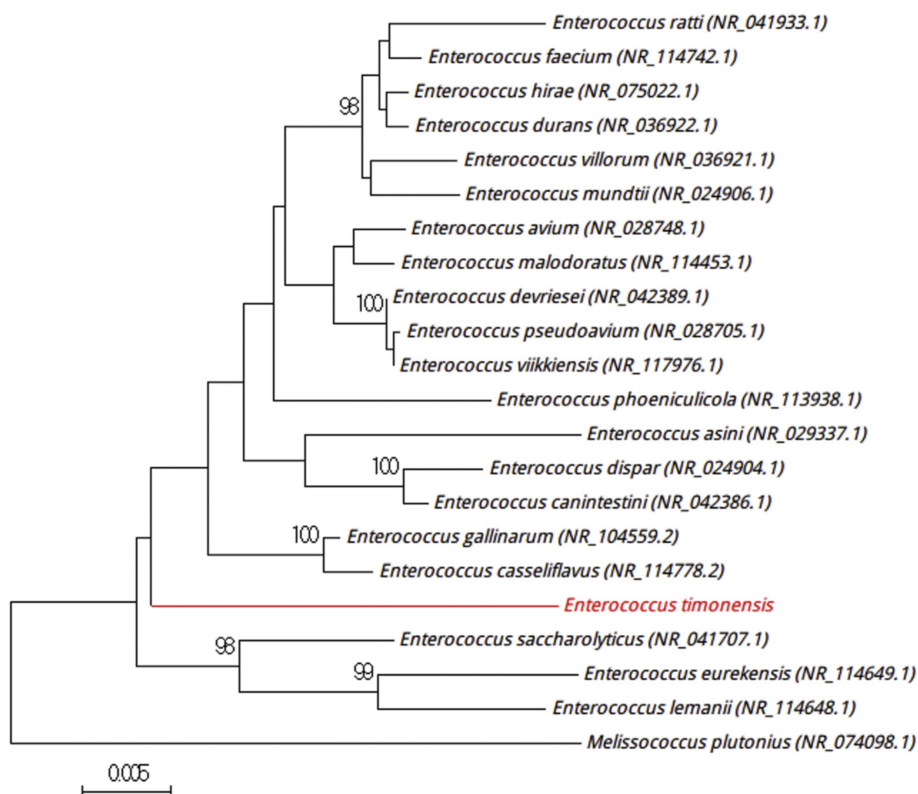

FIG. 1. Phylogenetic tree showing the position of *Enterococcus timonensis* sp. nov. strain Marseille-P2817^T (= CSUR P2817, = DSM 103162) with respect to other close species. Sequences were aligned using Muscle v3.8.31 with default parameters, and phylogenetic inferences were obtained using the neighbour-joining method with 500 bootstrap replicates, within MEGA6 software. Only bootstraps >95% were kept. Nevertheless, the scale bar represents a 0.5% nucleotide sequence divergence.

© 2019 The Authors. Published by Elsevier Ltd, *NMNI*, **29**, 100532
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

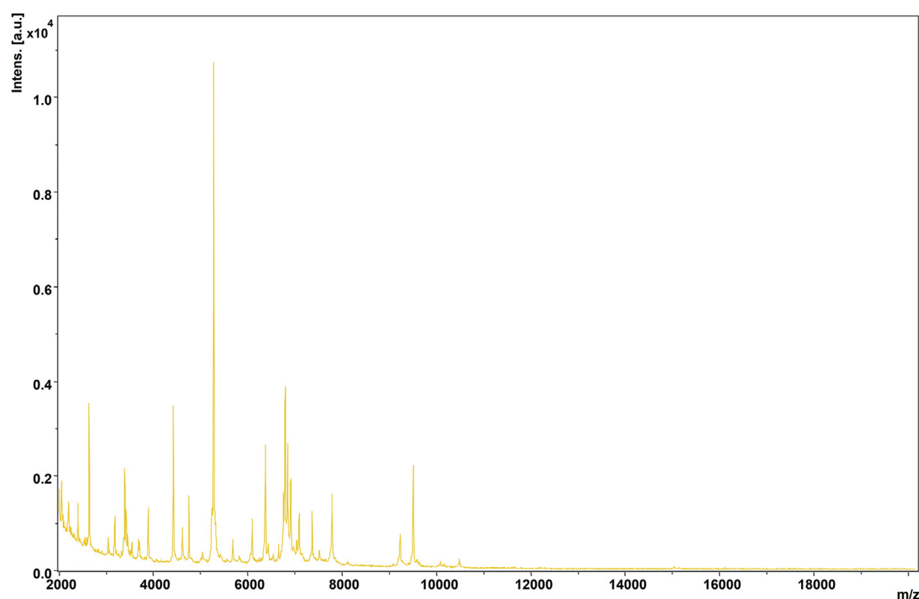

FIG. 2. Reference mass spectrum from *Enterococcus timonensis* strain Marseille-P2817^T. Spectra from 17 individual colonies were compared and the reference spectrum was generated.

Results

Strain identification and phylogenetic analysis

Strain Marseille-P2817^T (Table 1) was first isolated in March 2016 by a 10-day preincubation in a blood culture bottle supplemented with 4 mL rumen fluid. This bacterium was isolated and cultivated on 5% sheep-blood-enriched Columbia agar in an aerobic atmosphere at 30°C.

Strain Marseille-P2817^T was not identified by MALDI-TOF MS because its spectrum did not match any of the spectra present in our database. The 16S rRNA nucleotide sequence (accession number LT576388) exhibited 95.99% sequence identity with *Enterococcus hirae* strain ATCC9790 (Genbank accession number NR_075022), the closest validated species. Thus, we can classify strain Marseille-P2817^T as a new species in the *Enterococcus* genus (Fig. 1). A representative reference spectrum was therefore added to our IHU Méditerranée

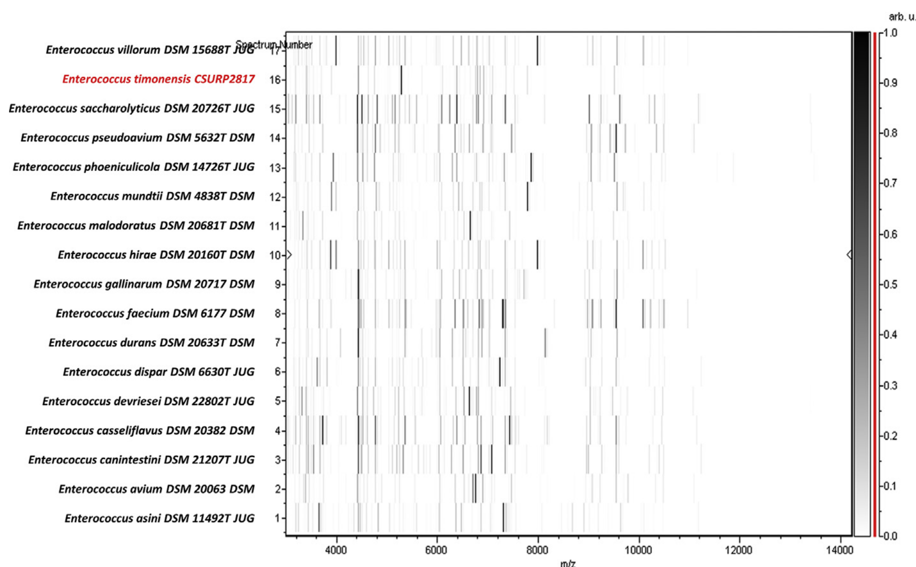

FIG. 3. Gel view comparing *Enterococcus timonensis* strain Marseille-P2817^T to other species within the *Enterococcus* genus. The gel view displays the raw spectra of loaded spectrum files arranged in a pseudo-gel. The x-axis records the m/z value. The left y-axis displays the running spectrum number originating from subsequent spectra loading. The peak intensity is expressed by a grey scale scheme code. The colour bar and the right y-axis indicate the relation between the peak colour and the intensity in arbitrary units. Displayed species are indicated on the left.

© 2019 The Authors. Published by Elsevier Ltd, *NMNI*, **29**, 100532
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Infection database (http://www.mediterranee-infection.com/article.php?laref=256&titre=urms-database) (Fig. 2), and a gel view shows the comparison between the spectrum of strain Marseille-P2817^T and that of the phylogenetically closest species (Fig. 3).

Phenotypic description

Growth was observed at 25, 30 and 37°C under aerobic, microaerophilic and anaerobic conditions on blood-enriched Columbia agar, with optimal growth being obtained aerobically at 37°C after 24 h of incubation. Strain Marseille-P2817^T grew in saline conditions of 5 g/L, and weak growth was observed also at 10 g/L. The strain tolerated pH values of 6, 6.5, 7 and 8.5. The cells were motile and non-sporulating, and they formed smooth, convex, grey colonies with a mean diameter of 1 mm on blood-enriched Columbia agar. Under electron microscopy, the bacteria had a mean diameter of 0.65 µm and a length of 1.1 µm (Fig. 4).

The major fatty acids were 9-octadecenoic acid (34%) and hexadecanoic acid (33%). Several fatty acids composed of 18 carbon atoms were also listed: 18:2n6 (16%); 18:0 (11%); 18:1n6 (1%) and 18:1n7 (TR). No branched structures were found (Table 2).

Strain Marseille-P2817^T was both catalase-negative and oxidase-negative. Using API® ZYM, positive reactions were recorded for alkaline phosphatase, esterase (C4), esterase lipase (C8), lipase (C14), valine arylamidase, leucine arylamidase, cystine arylamidase, α-chymotrypsin, acid phosphatase, naphthol-AS-BI-phosphohydrolase, α-galactosidase, β-glucuronidase, α-glucosidase, α-fucosidase and α-mannosidase, and negative reactions were observed for trypsin, β-galactosidase, β-glucosidase and N-acetyl-β-glucosaminidase. Using API® 20NE reactions recorded were positive for urease and β-galactosidase and negative for tryptophan and arginine dihydrolase, nitrate was reduced, glucose fermented, aesculin and

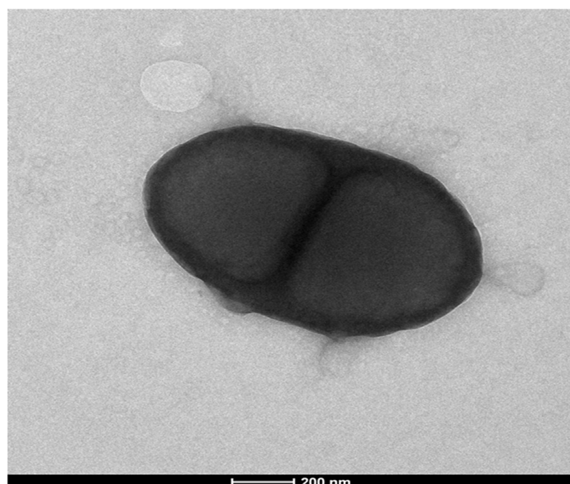

FIG. 4. Electron micrographs of *Enterococcus timonensis* strain Marseille-P2817^T, using a Tecnai G^{20} Cryo (FEI) transmission electron microscope operated at 200 keV. The scale bar represents 200 nm.

TABLE 2. Cellular fatty acid composition (%)

Fatty acids	Name	Mean relative %[a]
18:1n9	9-Octadecenoic acid	34.3 ± 0.5
16:0	Hexadecanoic acid	32.5 ± 0.5
18:2n6	9,12-Octadecadienoic acid	15.6 ± 0.6
18:0	Octadecanoic acid	11.1 ± 0.3
14:0	Tetradecanoic acid	3.1 ± 0.1
18:1n6	12-Octadecenoic acid	1.3 ± 0.1
12:0	Dodecanoic acid	TR
18:1n7	11-Octadecenoic acid	TR
10:0	Decanoic acid	TR
16:1n7	9-Hexadecenoic acid	TR
15:0	Pentadecanoic acid	TR

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

TABLE 3. *Enterococcus timonensis* strain Marseille-P2817^T reactions results with API® ZYM, API® 20NE and API® 50CH

Oxidase	—	API 50CH	
Catalase	—	Glycerol	+
API ZYM		Erythritol	—
Alkaline phosphatase	+	D-Arabinose	—
Esterase (C4)	+	L-Arabinose	—
Esterase lipase (C8)	+	D-Ribose	—
Lipase (C14)	+	D-Xylose	—
Leucine arylamidase	+	L-Xylose	—
Valine arylamidase	+	D-Adonitol	—
Cystine arylamidase	+	Methyl-βD-xylopyranoside	—
Trypsin	—	D-Galactose	+
α-Chymotrypsin	+	D-Glucose	+
Acid phosphatase	+	D-Fructose	+
Naphthol-AS-BI-phosphohydrolase	+	D-Mannose	+
α-Galactosidase	+	L-Sorbose	—
β-Galactosidase	—	L-Rhamnose	—
β-Glucuronidase	+	Dulcitol	—
α-Glucosidase	+	Inositol	—
β-Glucosidase	—	D-Mannitol	+
N-Acetyl-β-glucosaminidase	—	D-sorbitol	—
α-Mannosidase	+	Methyl-αD-mannopyranoside	—
α-Fucosidase	+	Methyl-α D-glucopyranoside	—
		N-Acetylglucosamine	+
API 20NE		Amygdalin	—
Reduction of nitrates	+	Arbutin	—
Indole production (tryptophan)	—	Aesculin ferric citrate	+
Fermentation (Glucose)	+	Salicin	+
Arginine dihydrolase	—	D-Cellobiose	+
Urease	—	D-Maltose	+
Hydrolysis (β-glucosidase) (aesculin)	+	D-Lactose	+
Hydrolysis (protease) (gelatin)	+	D-Melibiose	—
β-Galactosidase	+	D-Saccharose	+
Assimilation (glucose)	—	D-Trehalose	+
Assimilation (arabinose)	—	Inulin	—
Assimilation (mannose)	+	D-Melezitose	—
Assimilation (mannitol)	+	D-Raffinose	—
Assimilation (N-acetylglucosamine)	—	Amidon	—
Assimilation (maltose)	—	Glycogen	—
Assimilation (potassium gluconate)	+	Xylitol	—
Assimilation (capric acid)	—	Gentiobiose	—
Assimilation (adipic acid)	—	D-Turanose	—
Assimilation (malate)	—	D-Lyxose	—
Assimilation (trisodium citrate)	—	D-Tagatose	—
Assimilation (phenylacetic acid)	—	D-Fucose	—
		L-Fucose	—
		D-Arabitol	—
		L-Arabitol	—
		Potassium gluconate	—
		Potassium 2-ketogluconate	—
		Potassium 5-ketogluconate	—

© 2019 The Authors. Published by Elsevier Ltd, *NMNI*, **29**, 100532
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

gelatine were hydrolysed, assimilations were reported for mannose, mannitol and potassium gluconate, but not for glucose, arabinose, N-acetylglucosamine, maltose, capric acid, adipic acid, malate, trisodium citrate and phenylacetic acid. Using API API® 50CH positive reactions were recorded for glycerol, D-galactose, D-glucose, D-fructose, D-mannose, D-mannitol, N-acetylglucosamine, aesculin ferric citrate, salicin, D-cellobiose, D-maltose, D-lactose, D-saccharose and D-trehalose, and negative reactions were observed for erythritol, D-arabinose, L-arabinose, D-ribose, D-xylose, L-xylose, D-adonitol, methyl-βD-xylopyranoside, L-sorbose, L-rhamnose, dulcitol, inositol, D-sorbitol, methyl-αD-mannopyranoside, methyl-αD-glucopyranoside, amygdalin, arbutin, D-melibiose, inulin, D-melezitose, D-raffinose, amidon, glycogen, xylitol, gentiobiose, D-turanose, D-lyxose, D-tagatose, D-fucose, L-fucose, D-arabitol, L-arabitol, potassium gluconate, potassium 2-ketogluconate and potassium 5-ketogluconate (Table 3).

Susceptibility to antimicrobial agents was interpreted according to the EUCAST recommendations leading to the following MIC results. Cells were susceptible to teicoplanin 0.75 μg/mL, vancomycin 1 μg/mL and amoxicillin 0.038 μg/mL, and intermediate for benzylpenicillin (0.002–32 μg/mL) 0.75 μg/mL, benzylpenicillin (0.016–256 μg/mL) 1 μg/mL and imipenem 0.75 μg/mL. Cells were resistant to metronidazole 256 μg/mL, cefotaxime (0.002–32 μg/mL) 6.0 μg/mL, ceftriaxone (0.016–256 μg/mL) 16 μg/mL and ertapenem 2 μg/mL. MICs toward agents for which breakpoints do not exist are distributed as follows: daptomycin 2.0 μg/mL, rifampicin 1.0 μg/mL, minocycline 0.094 μg/mL, amikacin 48.0 μg/mL, erythromycin 0.75 μg/mL, tobramycin 3.0 μg/mL, fosfomycin 12.0 μg/mL, and doxycycline 0.38 μg/mL.

The biochemical and phenotypic features of strain Marseille-P2817^T were compared to the corresponding features of other close representatives of the *Enterococcus* genus (Table 4). We observed that all the species were facultatively anaerobic, Gram-positive, and positive for N-acetylglucosamine, D-glucose, D-fructose, D-maltose, D-lactose.

Genome properties

The genome of strain Marseille-P2817^T (genome accession no. FNVY00000000) was 2 123 933 bp long with 38.46 mol% of G+C content (Fig. 5). It is composed of four scaffolds (composed of four contigs). Along the 2048 predicted genes, 1983 were protein-coding genes and 65 were RNAs (three genes were 5S rRNA, three genes were 23S rRNA, three genes were 16S rRNA, and 56 genes were tRNA genes). Putative functions were attributed to a total of 1507 genes (76%) (by COGs or by NR blast); 112 genes (5.65%) were identified as ORFans, other genes were annotated as hypothetical proteins (261 genes ≥13.16%) (Table 5). The properties and statistics of

TABLE 4. Differential characteristics of *Enterococcus timonensis* strain Marseille-P2817^T, *E. hirae* strain ATCC 9790 [43], *E. gallinarum* strain NBRC 100675 [44], *E. saccharolyticus* strain ATCC 43076 [45,46], *E. casseliflavus* strain NBRC 100478 [44], *E. rotai* strain LMG 26678 [47], *E. silesiacus* strain LMG 23085 [48] and *E. asini* strain ATCC 700915 [49]

Properties	*E. timonensis*	*E. hirae*	*E. gallinarum*	*E. saccharolyticus*	*E. casseliflavus*	*E. rotai*	*E. silesiacus*	*E. asini*
Cell diameter (μm)	0.65–1.1	na	na	na	na	na	na	na
Oxygen requirement	Facultative anaerobe	Facultative anaerobe	Facultative anaerobe	Facultative anaerobe	Facultative anaerobe	Facultative anaerobe	na	Facultative
Gram stain	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive
Salt requirement	+	+	+	+	+	+	+	—
Motility	Motile	Non-motile	Non-motile	na	Motile	Non-motile	Non-motile	Non-motile
Endospore formation	—	na	—	na	na	—	na	—
Production of:								
Alkaline phosphatase	+	—	na	+	na	—	—	—
Catalase	—	—	—	—	—	—	v	—
Oxidase	—	na	na	na	na	na	na	na
Nitrate reductase	+	na	na	na	na	—	na	na
Urease	—	na	na	—	na	+	na	na
α-Glucosidase	+	na	na	—	na	+	na	na
β-Galactosidase	+	+	+	—	+	+	+	—
Acid from:								
N-Acetylglucosamine	+	+	+	+	+	+	+	+
L-Arabinose	—	—	+	—	+	+	+	—
D-Ribose	—	+	+	—	+	+	+	—
D-Mannose	+	+	+	na	+	+	+	—
D-Mannitol	+	—	+	+	+	+	—	—
Sucrose	—	+	+	+	+	+	—	—
D-Glucose	+	+	+	+	+	+	+	+
D-Fructose	+	+	+	+	+	+	+	+
D-Maltose	+	+	+	+	+	v	+	+
D-Lactose	+	+	+	+	+	+	+	+
G+C content (mol%)	38.46	36.9	39.80	36.90	42.40	36.7	36.40	44.70
Habitat	Human lung	Chicken and pig intestines	Intestines of domestic fowls	Fresh broccoli	Plant material	Drinking water	Surface waters	Caecum of donkeys

na, data not available; v, variable.

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

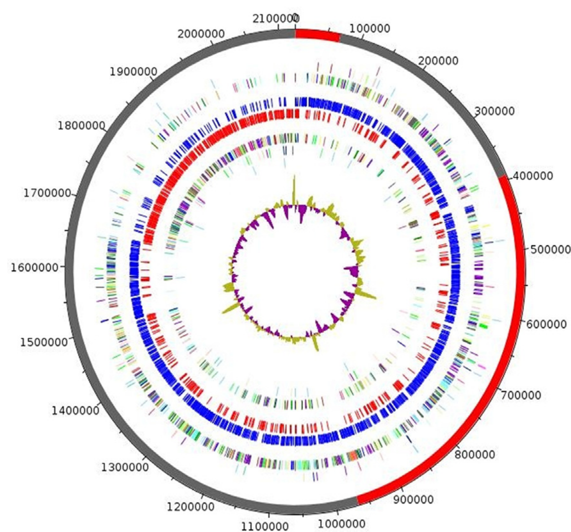

FIG. 5. Graphical circular map of the genome. From outside to the centre: contigs (red/grey), clusters of orthologous groups (COG) category of genes on the forward strand (three circles), genes on the forward strand (blue circle), genes on the reverse strand (red circle), COG category on the reverse strand (three circles), G+C content.

the genome are summarized in Tables 5 and 6. The distribution of genes into COGs functional categories is presented in Table 6.

Genome comparison

The draft genome sequence of strain Marseille-P2817^T was compared to that of the closest species in the *Enterococcus* genus: *Enterococcus casseliflavus* strain NBRC 100478

TABLE 5. Nucleotide content and gene count levels of genome

Attribute	Genome (total)[a] Number	% of total[a]
Size (bp)	2 123 933	100
G+C content (mol%)	816 803	38.46
Coding region (bp)	1 887 347	88.86
Total genes	2048	100
RNA genes	65	3.17
Protein-coding genes	1983	100
Genes with function prediction	1507	76.00
Genes assigned to COGs	1398	70.50
Genes with peptide signals	196	9.88
Gene associated to bacteriocine	28	1.41
Genes associated to mobilome	905	45.64
Genes associated to virulence	432	21.79
Genes associated to toxine/antitoxine	86	4.34
Genes with Pfam-A domains	1867	91
ORFan genes	112	5.64
Genes with transmembrane helices	452	22.79
Genes associated with PKS or NRPS	1	0.05
Number of antibiotic resistance genes	0	0

COG, clusters of orthologous groups; PKS, polyketide synthase; NRPS, non-ribosomal peptide-synthetase.
[a]Total is based on either the size of the genome (in base pairs) or the total number of protein coding genes in the annotated genome.

TABLE 6. Number of genes associated with 25 general clusters of orthologous groups (COGs) functional categories

Code	Value	% of total[a]	Description
[J]	185	9.33	Translation
[A]	0	0	RNA processing and modification
[K]	107	5.40	Transcription
[L]	89	4.49	Replication, recombination and repair
[B]	0	0	Chromatin structure and dynamics
[D]	22	1.11	Cell cycle control, mitosis and meiosis
[Y]	0	0	Nuclear structure
[V]	72	3.63	Defence mechanisms
[T]	57	2.87	Signal transduction mechanisms
[M]	73	3.68	Cell wall/membrane biogenesis
[N]	9	0.45	Cell motility
[Z]	0	0	Cytoskeleton
[W]	2	0.10	Extracellular structures
[U]	18	0.91	Intracellular trafficking and secretion
[O]	51	2.571	Post-translational modification, protein turnover, chaperones
[X]	45	2.27	Mobilome: prophages, transposons
[C]	53	2.67	Energy production and conversion
[G]	176	8.88	Carbohydrate transport and metabolism
[E]	92	4.64	Amino acid transport and metabolism
[F]	55	2.77	Nucleotide transport and metabolism
[H]	56	2.82	Coenzyme transport and metabolism
[I]	43	2.17	Lipid transport and metabolism
[P]	73	3.68	Inorganic ion transport and metabolism
[Q]	20	1.01	Secondary metabolites biosynthesis, transport and catabolism
[R]	127	6.40	General function prediction only
[S]	103	5.19	Function unknown
—	585	29.50	Not in COGs

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

(BCPT00000000), *Enterococcus gallinarum* strain FDAARGOS_163 (CP014067), *Enterococcus asini* strain ATCC_700915 (ASVU00000000), *Enterococcus dispar* strain ATCC_51266 (AHYR00000000), *Enterococcus saccharolyticus* subsp. *saccharolyticus* strain ATCC_43076 (ASWN00000000), *Enterococcus rotai* strain LMG_26678 (CP013655), *Enterococcus hirae* strain ATCC 9790 (CP003504), and *Enterococcus silesiacus* strain LMG_23085 (CP013614). It is smaller than the genomes of *E. silesiacus, E. rotai, E. hirae, E. asini, E. saccharolyticus* subsp. *saccharolyticus, E. dispar, E. gallinarum* and *E. casseliflavus* (2124, 3928, 3746, 2883, 2573, 2604, 2813, 3821 and 3498 MB respectively).

The G+C content of strain Marseille-P2817^T is lower than that of *E. asini, E. gallinarum* and *E. casseliflavus* (38.46, 44.72, 42.32 and 42.35% respectively) but larger than that of *E. silesiacus, E. rotai, E. hirae, E. saccharolyticus* subsp. *saccharolyticus* and *E. dispar* (36.41, 36.13, 36.75, 36.89 and 37.17% respectively). The protein-coding gene content of strain Marseille-P2817^T is smaller than that of *E. silesiacus, E. rotai, E. hirae, E. asini, E. saccharolyticus* subsp. *saccharolyticus, E. dispar, E. gallinarum* and *E. casseliflavus* (1983, 3559, 3253, 2669, 2430, 2582, 2637, 3333 and 3353 respectively).

To evaluate the genomic similarity among studied strains of the Enterococcaceae, we determined two parameters: AGIOS (average of genomic identity of orthologous gene sequences) [4], which was designed to be independent from DDH, and

© 2019 The Authors. Published by Elsevier Ltd, *NMNI*, **29**, 100532
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

TABLE 7. Numbers of orthologous proteins shared between genomes (upper right), average percentage similarity of nucleotides corresponding to orthologous proteins shared between genomes (lower left), and numbers of proteins per genome (bold)

	E. saccharolyticus	*E. gallinarum*	*E. timonensis*	*E. casseliflavus*	*E. rotai*	*E. silesiacus*	*E. asini*	*E. dispar*	*E. hirae*
E. saccharolyticus	**2582**	1479	1049	1528	1371	1423	1295	1336	1235
E. gallinarum	60.09	**3333**	1176	1980	1557	1636	1456	1519	1436
E. timonensis.	60.18	58.36	**1983**	1215	1064	1117	1159	1136	1065
E. casseliflavus	61.63	72.83	66.31	**3353**	1571	1652	1,514	1576	1464
E. rotai	61.29	58.61	58.18	58.58	**3253**	1748	1308	1358	1307
E. silesiacus	61.19	58.43	58.69	59.08	84.10	**3559**	1393	1446	1350
E. asini	59.03	61.88	60.06	62.05	56.98	57.49	**2430**	1423	1252
E. dispar	58.27	58.69	59.22	59.47	60.24	60.14	61.59	**2637**	1352
E. hirae	61	56.89	58.57	59.51	62.44	62.21	57.31	62.49	**2669**

TABLE 8. Pairwise comparison of *Enterococcus timonensis* strain Marseille-P2817^T with other species using the Genome-to-Genome Distance Calculator (GGDC), formula 2 (DNA–DNA hybridization (DDH) estimates based on identities/high-scoring segment pairs (HSP) length)

	E. timonensis.	*E. gallinarum*	*E. saccharolyticus*	*E. casseliflavus*	*E. rotai*	*E. silesiacus*	*E. asini*	*E. dispar*	*E. hirae*
E. timonensis.	100% ± 00	21.90% [19.6–24.3%]	20.00% [17.8–22.4%]	24.20% [21.9–26.7%]	23.40% [21.1–25.9%]	23.40% [21.1–25.8%]	25.50% [23.1–28%]	25.80% [23.5–28.3%]	25.00% [22.7–27.5%]
E. gallinarum		100% ± 00	21.20% [19–23.7%]	21.30% [19.1–23.8%]	21.10% [18.9–23.6%	22.70% [20.4–25.1%]	24.20% [21.9–26.6%]	25.30% [23–27.8%]	25.00% [22.6–27.4%]
E. saccharolyticus			100% ± 00	20.50% [18.3–22.9%]	20.70% [18.4–23.1%]	21.00% [18.8–23.5%]	25.00% [22.7–27.5%]	22.00% [19.8–24.5%]	22.40% [20.1–24.8%]
E. casseliflavus				100% ± 00	22.80% [20.5– 5.2%]	23.80% [21.5–26.3%]	23.50% [21.2–26%]	25.70% [23.4–28.2%]	23.50% [21.2–25.9%]
E. rotai					100% ± 00	26.90% [24.6–29.4%]	25.20% [22.9–27.7%]	22.40% [20.1–24.8%]	23.40% [21.1–25.9%]
E. silesiacus						100% ± 00	23.60% [21.3–26%]	22.40% [20.1–24.9%]	24.20% [21.9–26.7%]
E. asini							100% ± 00	25.20% [22.9–27.7%]	28.30% [25.9–30.8%]
E. dispar								100% [00–00%]	28.50% [26.1–31%]
E. hirae									100% [00–00%]

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

dDDH, which exhibited a high correlation with DDH (Tables 7 and 8) [25,26]. Strain Marseille-P2817^T shared 1049, 1176, 1215, 1064, 1117, 1159, 1065 and 1136 orthologous genes with *E. saccharolyticus, E. gallinarum, E. casseliflavus, E. rotai, E. silesiacus, E. asini, E. hirae* and *E. dispar* respectively (Table 7). AGIOS values ranged from 56.89 between *E. hirae* and *E. gallinarum* to 84.10% between *E. rotai* and *E. silesiacus* among compared species with standing in nomenclature (except strain Marseille-P2817^T). AGIOS values of strain Marseille-P2817^T ranged from 58.18% with *E. rotai* to 66.31% with *E. casseliflavus* when compared to other species (Table 7). Concerning dDDH values of compared species (except strain Marseille-P2817^T), they ranged from 20.50% (18.3–22.9) when estimated between *E. casseliflavus* and *E. saccharolyticus* to 28.50% (26.1–31%) when estimated between *E. dispar* and *E. hirae*. dDDH values of strain Marseille-P2817^T ranged from 20.00% (17.8–22.4%) with *E. saccharolyticus* to 25.80% (23.5–28.3%) with *E. dispar* when compared to other species (Table 8). Those values are under the 70% threshold, thus confirming the new species status [40–42]. Furthermore, the distribution of genes into COGs categories was similar in all compared genomes (Fig. 6).

Conclusion

In conclusion, based on the phenotypic, phylogenetic and genomic analyses, we suggest the creation of a new species, *Enterococcus timonensis* sp. nov., that contains the strain Marseille-P2817^T. This bacterial strain has been isolated from the sputum of a healthy adult man from Marseilles, France.

Taxonomic and nomenclatural proposals

Description of *Enterococcus timonensis* sp. nov

Enterococcus timonensis (ti.mo.nen'sis, N.L. masc. adj., *timonensis* from the Latin name of the Hôpital de la Timone, hospital in Marseille, where strain Marseille-P2817^T was isolated) exhibited smooth, convex, grey colonies with a diameter of 1 mm on 5% sheep-blood-enriched Columbia agar. Cells showed a mean diameter of 0.65 μm and a length of 1.1 μm. This bacterium is a Gram-positive, non-spore-forming and motile coccus. Optimal growth was obtained aerobically at 37°C after 24 h of incubation. It is both oxidase- and catalase-negative. The reactions were positive for alkaline phosphatase, esterase (C4), esterase lipase (C8), lipase (C14), leucine arylamidase, valine arylamidase, cystine arylamidase, α-chymotrypsin, acid phosphatase, naphthol-AS-BI-phosphohydrolase, α-galactosidase, β-glucuronidase, α-glucosidase, α-mannosidase, α-fucosidase, β-galactosidase, nitrate reduction, aesculin hydrolysis, gelatine hydrolysis, assimilation of mannose, mannitol and potassium gluconate and fermentation of glycerol, D-galactose, D-glucose, D-fructose, D-mannose, D-mannitol, N-acetylglucosamine, aesculin ferric citrate, salicin, D-cellobiose, D-maltose, D-lactose, D-saccharose and D-trehalose. The major fatty acids were 9-octadecenoic acid (34%) and hexadecanoic acid (33%).

This strain exhibited a G+C genome content of 38.46 mol%. The 16S rRNA gene sequence and whole-genome shotgun sequence have been deposited in EMBL-EBI under accession numbers LT576388 and FNVY00000000, respectively. The type strain Marseille-P2817^T (= CSUR P2817 = DSM 103162) was isolated from the sputum of a French adult man living in Marseille.

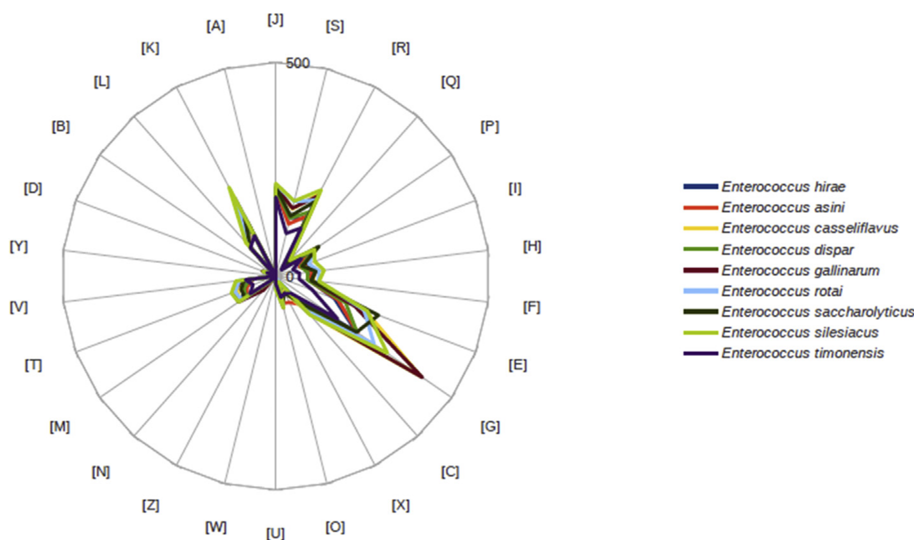

FIG. 6. Distribution of functional classes of predicted genes in genomes from *Enterococcus timonensis, E. hirae, E. gallinarum, E. saccharolyticus, E. casseliflavus, E. rotai, E. silesiacus* and *E. asini* chromosomes according to clusters of orthologous groups of proteins.

© 2019 The Authors. Published by Elsevier Ltd, *NMNI*, **29**, 100532
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Transparency declaration

The authors declare no conflict of interest. This work has benefited from French State support, managed by the 'Agence Nationale pour la Recherche' including the 'Programme d'Investissement d'avenir' under the reference Méditerranée Infection 10-IAHU-03. This work was also supported by the Région Provence-Alpes-Côte d'Azur and Fonds Européen de Développement Regional - Plateformes de Recherche et d'Innovation Mutualisées Méditerranée Infection (FEDER PRIMI).

Acknowledgements

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

We thank Magdalen Lardiere for reviewing the English language.

References

[1] Lagier J-C, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P, et al. Culture of previously uncultured members of the human gut microbiota by culturomics. Nat Microbiol 2016;1:16203.

[2] Mahato NK, Gupta V, Singh P, Kumari R, Verma H, Tripathi C, et al. Microbial taxonomy in the era of OMICS: application of DNA sequences, computational tools and techniques. Antonie Van Leeuwenhoek 2017;110(10):1357–71.

[3] 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(Pt 2):384–91.

[4] Lagier J-C, Hugon P, Khelaifia S, Fournier P-E, La Scola B, Raoult D. The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. Clin Microbiol Rev 2015;28(1):237–64.

[5] Schleifer KH, Kilpper-Bälz R. Transfer of *Streptococcus faecalis* and *Streptococcus faecium* to the genus *Enterococcus* nom. rev. as *Enterococcus faecalis* comb. nov. and *Enterococcus faecium* comb. nov. Int J Syst Evol Microbiol 1984;34(1):31–4.

[6] Ryan KJ, Ray CG. Medical microbiology: an introduction to infectious diseases. Mcgraw-Hill; 2004.

[7] Fisher K, Phillips C. The ecology, epidemiology and virulence of *Enterococcus*. Microbiol Read Engl 2009;155(Pt 6):1749–57.

[8] Grupper M, Kravtsov A, Potasman I. Enterococcal-associated lower respiratory tract infections: a case report and literature review. Infection 2009;37(1):60–4.

[9] Dubourg G, Cimmino T, Senkar SA, Lagier J-C, Robert C, Flaudrops C, et al. Noncontiguous finished genome sequence and description of *Paenibacillus antibioticophila* sp. nov. GD11(T), the type strain of *Paenibacillus antibioticophila*. New Microbe. New Infect 2015;8:137–47.

[10] Seng P, Abat C, Rolain JM, Colson P, Lagier J-C, Gouriet F, et al. Identification of rare pathogenic bacteria in a clinical microbiology laboratory: impact of matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol 2013;51(7):2182–94.

[11] Drancourt M, Bollet C, Carlioz A, Martelin R, Gayral J-P, Raoult D. 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. J Clin Microbiol 2000;38(10):3623–30.

[12] 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;14:60.

[13] Kim M, Oh H-S, Park S-C, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol 2014;64(Pt 2):346–51.

[14] Sasser M. Bacterial identification by gas chromatographic analysis of fatty acids methyl esters (GC-FAME) [Internet]. NY: Microbial ID Inc. Newark; 2006 [cited 2016 Aug 4]. Available from: http://youngin.com/application/AN-0702-0013EN.pdf.

[15] 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.

[16] Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 2008;18(5):821–9.

[17] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 2012;19(5):455–77.

[18] Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience [Internet] 2012 Dec;1(1) [cited 2017 Feb 12] Available from: https://academic.oup.com/gigascience/article-lookup/doi/10.1186/2047-217X-1-18.

[19] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30(15):2114–20.

[20] Prodigal: Microbial Gene Prediction Software [Internet]. [cited 2016 Jul 28]. Available from: http://prodigal.ornl.gov/.

[21] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25(5):955–64.

[22] Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 2007;35(9):3100–8.

[23] Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol 2011;7(10):e1002195.

[24] Conway KR, Boddy CN. ClusterMine360: a database of microbial PKS/NRPS biosynthesis. Nucleic Acids Res 2013;41(D1):D402–7.

[25] Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 2014;58(1):212–20.

[26] Shao Y, Harrison EM, Bi D, Tai C, He X, Ou H-Y, et al. TADB: a web-based resource for Type 2 toxin-antitoxin loci in bacteria and archaea. Nucleic Acids Res 2011;39:D606–11. Database.

[27] Edwards R. PHANTOME: PHage ANnotation TOols and MEthods. Grantome [Internet]. [cited 2017 Feb 12]; Available from: http://grantome.com/grant/NSF/DBI-0850356.

[28] Bi D, Xu Z, Harrison EM, Tai C, Wei Y, He X, et al. ICEberg: a web-based resource for integrative and conjugative elements found in bacteria. Nucleic Acids Res 2012;40(D1):D621–6.

[29] Leplae R. ACLAME: a CLAssification of mobile genetic elements. Nucleic Acids Res 2004;32(90001):45D–9D.

[30] Llorens C, Futami R, Covelli L, Dominguez-Escriba L, Viu JM, Tamarit D, et al. The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res 2011;39:D70–4. Database.

[31] Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinform 2007;8(1):172.

[32] Drissi F, Buffet S, Raoult D, Merhej V. Common occurrence of antibacterial agents in human intestinal microbiota. Front Microbiol

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

[Internet] 2015 May 7;6 [cited 2017 Feb 12] Available from: http://www.frontiersin.org/Evolutionary_and_Genomic_Microbiology/10.3389/fmicb.2015.00441/abstract.

[33] Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. Nucleic Acids Res 2016;44(D1):D694–7.

[34] Zhou CE, Smith J, Lam M, Zemla A, Dyer MD, Slezak T. MvirDB—a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. Nucleic Acids Res 2007;35:D391–4. Database.

[35] Gouret P, Thompson JD, Pontarotti P. PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinform 2009;10:298.

[36] 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.

[37] Auch AF, von Jan M, Klenk H-P, Göker M. Digital DNA–DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci 2010;2(1):117–34.

[38] 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 [Internet]. Springer Berlin Heidelberg; 2011 [cited 2016 Oct 23]. pp. 71–87. Available from: http://link.springer.com/chapter/10.1007/978-3-642-20763-1_5.

[39] 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 Bioinform 2005;6(1):1.

[40] Stackebrandt E, Goebel BM. Taxonomic note: a place for DNA–DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Evol Microbiol 1994;44(4):846–9.

[41] Busse H-J, Tindall BJ, Ludwig W, Rosselló-Móra R, Kämpfer P. Notes on the characterization of prokaryote strains for taxonomic purposes. Int J Syst Evol Microbiol 2010;60(1):249–66.

[42] Moore WEC, Stackebrandt E, Kandler O, Colwell RR, Krichevsky MI, Truper HG, et al. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int J Syst Evol Microbiol 1987;37(4):463–4.

[43] Farrow JA, Collins MD. *Enterococcus hirae*, a new species that includes amino acid assay strain NCDO 1258 and strains causing growth depression in young chickens. Int J Syst Evol Microbiol 1985;35(1):73–5.

[44] Collins MD, Jones D, Farrow JAE, Kilpper-Balz R, Schleifer KH. *Enterococcus avium* nom. rev., comb. nov.; *E. casseliflavus* nom. rev., comb. nov.; E. *durans* nom. rev., comb. nov.; E. *gallinarum* comb. nov.; and E. *malodoratus* sp. nov. Int J Syst Evol Microbiol 1984;34(2):220–3.

[45] Chen Y, Lin Y, Pan S, Ji S, Chang Y, Yu C, et al. *Enterococcus saccharolyticus* subsp. *taiwanensis* subsp. nov., isolated from broccoli. Int J Syst Evol Microbiol 2013;63(12):4691–7.

[46] Farrow J, Kruze J, Phillips B, Bramley A, Collins M. Taxonomic studies on *Streptococcus bovis* and *Streptococcus equinus*: description of *Streptococcus alactolyticus* sp. nov. and *Streptococcus saccharolyticus* sp. nov. Syst Appl Microbiol 1984;5(4):467–82.

[47] Sedláček I, Holochová P, Mašlaňová I, Kosina M, Spröer C, Bryndová H, et al. *Enterococcus ureilyticus* sp. nov. and *Enterococcus rotai* sp. nov., two urease-producing enterococci from the environment. Int J Syst Evol Microbiol 2013;63(2):502–10.

[48] Švec P, Vancanneyt M, Sedláček I, Naser SM, Snauwaert C, Lefebvre K, et al. *Enterococcus silesiacus* sp. nov. and *Enterococcus termitis* sp. nov. Int J Syst Evol Microbiol 2006;56(3):577–81.

[49] de Vaux A, Laguerre G, Diviès C, Prévost H. *Enterococcus asini* sp. nov. isolated from the caecum of donkeys (*Equus asinus*). Int J Syst Evol Microbiol 1998;48(2):383–7.

© 2019 The Authors. Published by Elsevier Ltd, *NMNI*, **29**, 100532
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).