Noncontiguous finished genome sequence and description of Enterococcus massiliensis sp. nov.

S. Le Page, T. Cimmino, A. Togo, M. Million, C. Michelle, S. Khelaifia, J.-C. Lagier, D. Raoult and J.-M. Rolain
Aix-Marseille Université, URMITE UM 63 CNRS 7278 IRD 198 INSERM U1905, IHU Méditerranée Infection, Facultés de Médecine et de Pharmacie, Marseille, France

Abstract

Enterococcus massiliensis strain sp. nov. (= CSUR P1927 = DSM 100308) is a new species within the genus Enterococcus. This strain was first isolated from a fresh stool sample of a man during culturomics study of intestinal microflora. Enterococcus massiliensis is a Gram-positive cocci, facultative anaerobic and motile. E. massiliensis is negative for mannitol and positive for β-galactosidase, contrary to E. gallinarum. The complete genome sequence is 2,712,841 bp in length with a GC content of 39.6% and contains 2,617 protein-coding genes and 70 RNA genes, including nine rRNA genes.

Keywords: Culturomics, Enterococcus massiliensis, genome, new species, taxonogenomics

Organism Information

A stool sample was collected in 2015 from a voluntary patient as a negative control and isolated on Columbia agar supplemented with 5% sheep’s blood (bioMérieux, Marcy-l’Étoile, France) in aerobic and anaerobic condition using GasPak EZ Anaerobe Container System Sachets (Becton Dickinson (BD), San Diego, CA, USA) at 37°C. Enterococcus massiliensis was sequenced as part of a culturomics study aiming to isolate all bacterial species colonizing the human gut [9]. Enterococcus massiliensis strain AM1T (GenBank accession no. LN833866) exhibited a 97% 16S
rRNA nucleotide sequence similarity with Enterococcus gallinarum (JF915769), the phylogenetically closest validly published bacterial species (Fig. 1) after comparison with National Center for Biotechnology Information (NCBI) database. This value is lower than 98.7% 16S rRNA gene sequence similarity set as a threshold recommended by Stackebrandt and Ebers [3] to delineate a new species without carrying out DNA-DNA hybridization.

Growth occurred between 25°C and 37°C, but optimal growth was observed at 37°C, 24 hours after inoculation. Colonies were smooth and whitish, approximately 1 mm in diameter on 5% sheep’s blood–enriched agar (bioMérieux). Growth of the strain was tested under anaerobic and microaerophilic conditions using GasPak EZ Anaerobe pouch (BD) and CampyGen Compact (Oxoid, Basingstoke, UK) systems, respectively, and in aerobic conditions, with or without 5% of CO2. Growth was achieved under aerobic (with and without CO2), microaerophilic and anaerobic conditions. Gram staining showed Gram-positive cocci without sporation (Fig. 2A). A motility test was positive and realized with API M Medium (bioMérieux), a semisolid medium with an inoculation performed by swabbing one colony into the medium. After 24 hours of incubation, the growth of E. massiliensis was away from this stabbed line, characteristic of positive motility. Cells grown on agar exhibited a mean diameter of 0.5 μm and a mean length ranging from 1.1 to 1.3 μm (mean 1.2 μm), determined by negative staining transmission electron microscopy (Fig. 2B).

Differential phenotypic characteristics using API 50CH and API Zym system (bioMérieux) between E. massiliensis sp. nov. AM1T and other Enterococcus species [9] are presented in Table 1. Antibiotic susceptibility testing was performed by the disk diffusion method on Müller-Hinton agar with blood (bioMérieux). E. massiliensis is susceptible to vancomycin, teicoplanin, linezolid, gentamicin, ciprofloxacin, doxycycline, rifampicin and pristinamycin and resistant or intermediate to penicillin G, oxacillin, cefotaxime, cefoxitin, trimethoprim/sulfamethoxazole, fosfomycin, erythromycin and clindamycin.

Extended Features Descriptions

MALDI-TOF MS protein analysis was carried out as previously described [2] using a Microflex spectrometer (Bruker Daltonics, Leipzig, Germany). Twelve distinct deposits were done for strain AM1T from 12 isolated colonies. Twelve distinct deposits were done for strain AM1T from 12 isolated colonies. Spectra were imported into the MALDI BioTyper software, version 2.0 (Bruker), and analysed by standard pattern matching against 7765 bacterial spectra, including 92 spectra from 31 Enterococcus species, in the BioTyper database. Interpretation of scores was as follows: a score of ≥2 enabled the identification at the species level, a score of ≥1.7 but <2 enabled the identification at the genus level and a score of <1.7 did not enable any identification (scores established by the manufacturer, Bruker). For strain AM1T, no significant MALDI-TOF MS score was obtained against the Bruker database, thus suggesting that our isolate was a new species. We incremented our database with the spectrum from strain AM1T (Fig. 3).

**FIG. 1.** Consensus phylogenetic tree highlighting position of Enterococcus massiliensis relative to other type strains within genus Enterococcus by 16S. GenBank accession numbers appear in brackets. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using maximum-likelihood method in MEGA6 software package. Numbers at nodes are percentages of bootstrap values from 1000 replicates that support nodes. Streptococcus pneumoniae and Staphylococcus aureus were used as outgroups. Scale bar = 1% nucleotide sequence divergence.
**Genome Sequencing Information**

*Enterococcus massiliensis* sp. nov. (GenBank accession no. CVRN00000000) is the 54 species described within *Enterococcus* genus.

After DNA extraction by the phenol–chloroform method, genomic DNA of *E. massiliensis* was sequenced on a MiSeq instrument (Illumina, San Diego, CA, USA) using paired end and mate pair strategies.

For genome annotation, open reading frames (ORFs) were predicted using Prodigal (http://prodigal.oml.gov/) with default parameters, but the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched for against the GenBank database (http://www.ncbi.nlm.nih.gov/genbank) and the Clusters of Orthologous Groups (COGs) databases using BLASTP. The tRNAscanSE tool [10] was used to find tRNA genes, whereas ribosomal RNAs were detected using RNAmmer [11] and BLASTn against the GenBank database.

The ARG-ANNOT database for acquired antibiotic resistance genes (ARGs) was used for a BLAST search using the Bio-Edit interface [12]. The assembled sequences were searched against the ARG database under moderately stringent conditions (e-value of $10^{-5}$) for the *in silico* ARG prediction.

*E. massiliensis* presents the *Lsa* gene, encoding a putative ABC protein *Lsa* with an identity to 72% with *Lsa* family ABC-F of *E. faecalis* in NCBI, which phenotypically confirms its resistance to clindamycin.

Analysis of presence of polyketide synthase (PKS) and non-ribosomal polyketide synthesis (NRPS) was performed by

---

**TABLE I. Differential characteristics of Enterococcus massiliensis sp. AM1, E. faecalis, E. casseliflavus, E. gallinarum, E. haemoperoxidus, E. cecorum, E. sulfureus and E. caccae**

| Property                  | *E. massiliensis* | *E. faecalis* | *E. casseliflavus* | *E. gallinarum* | *E. haemoperoxidus* | *E. cecorum* | *E. sulfureus* | *E. caccae* |
|---------------------------|------------------|--------------|--------------------|----------------|---------------------|--------------|----------------|-------------|
| Oxygen requirement        | Faculative       | Faculative   | Faculative         | Faculative     | Faculative          | Faculative   | Faculative     | Faculative   |
| Gram stain                | Positive         | Positive     | Positive           | Positive       | Positive            | Positive     | Positive       | Positive     |
| Motility                  | −                | −            | +                  | −              | −                   | +            | −              | −           |
| Pigment                   | −                | −            | +                  | −              | −                   | +            | −              | −           |
| Production of:            |                  |              |                    |                |                     |              |                |             |
| Alkaline phosphatase      | −                | −            | −                  | −              | −                   | −            | −              | −           |
| Catalase                  | +                | +            | +                  | +              | +                   | +            | +              | +           |
| Oxidase                   | +                | −            | −                  | −              | −                   | −            | −              | −           |
| β-Glucuronidase           | −                | +            | −                  | −              | −                   | −            | −              | −           |
| ω-Galactosidase           | −                | −            | −                  | −              | −                   | −            | −              | −           |
| β-Galactosidase           | +                | −            | −                  | −              | −                   | −            | −              | −           |
| N-acetyl-glucosamine      | −                | +            | +                  | +              | +                   | +            | +              | +           |
| Acid form:                |                  |              |                    |                |                     |              |                |             |
| Mannitol                  | −                | +            | +                  | +              | +                   | +            | +              | +           |
| Sorbose                   | −                | −            | −                  | −              | −                   | −            | −              | −           |
| L-Arabinose               | +                | −            | −                  | −              | −                   | −            | −              | −           |
| Sorbitol                  | +                | +            | +                  | +              | +                   | +            | +              | +           |
| α-Raffinose               | +                | +            | +                  | +              | +                   | +            | +              | +           |
| Xylose                    | +                | −            | −                  | −              | −                   | −            | −              | −           |
| G+C content (%)           | 39.6             | 37.3         | 42.7               | 40.7           | 35.8                | 36.3         | 37.8           | 35.8        |
| Habitat                   | Human stool      | Intestine of mammals | Intestine of mammals | Intestine of mammals | Water | Commensal chicken | Plants | Human stool |

*+, positive result; −, negative result; v, variable result; NA, data not available.*
discriminating the gene with a large size using a database realized in our laboratory; predicted proteins were compared against the nonredundant (nr) GenBank database using BLASTP and finally examined using antiSMASH [13]. Analysis of the genome revealed the absence of NRPKs and PKS. Lipoprotein signal peptides and the number of transmembrane helices were predicted using SignalP [14] and TMHMM [15], respectively. ORFans were identified if their BLASTP E value was lower than $10^{-3}$ for alignment length >80 amino acids.

We used the Genome-to-Genome Distance calculator (GGDC) web server (http://ggdc.dsmz.de) to estimate the overall similarity among the compared genomes and to replace the wet-lab DDH by a digital DDH [16,17]. GGDC 2.0 BLAST+ was chosen as alignment method, and the recommended formula 2 was taken into account to interpret the results.

We compared the genome of *E. massiliensis* with nine other genomes of *Enterococcus* strains. The genome is 2 712 841 bp long (one chromosome, no plasmid) with a GC content of 39.6% (Table 2). The properties and statistics of the genome are summarized in Table 2. The draft genome of *E. massiliensis* is smaller than those of *E. moraviensis*, *E. haemoperoxidus*, *E. cacaoe*, *E. casseliflavus*, *E. gallinarum* and *E. faecalis* (3.60, 3.58, 3.56, 3.43, 3.42).

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

| Attribute                        | Value     | % of total |
|----------------------------------|-----------|------------|
| Genome size (bp)                 | 2 712 841 | 100        |
| DNA G+C content (bp)             | 1 075 567 | 39.6       |
| DNA coding region (bp)           | 2 408 151 | 88.77      |
| Total genes                      | 2 571     | 100        |
| RNA genes                        | 70        | 2.60       |
| Protein-coding genes             | 2 564     | 97.39      |
| Genes with function prediction   | 1 089     | 42.12      |
| Genes assigned to COGs           | 1 083     | 42.12      |
| Genes with peptide signals       | 250       | 9.55       |
| Genes with transmembrane helices | 630       | 24.07      |

COGs, Clusters of Orthologous Groups database.

*Total is based on either size of genome in base pairs or total number of protein-coding genes in annotated genome.*
3.16 and 2.96 Mb, respectively), but larger than those of *E. saccharolyticus, E. columbae, E. cecorum* and *E. sulfureus* (2.60, 2.58, 2.34 and 2.31, respectively). The G+C content of *E. massiliensis* is lower than those of *E. casseli flavus* and *E. gallinarum* (42.8 and 40.7) but greater than those of *E. moraviensis, E. haemoperoxidus, E. caccae, E. saccharolyticus, E. columbae, E. cecorum, E. sulfureus and E. faecalis* (39.6, 36.1, 35.7, 35.8, 36.9, 36.6, 36.4, 38.0 and 37.5, respectively).

| Color of COGs class | COGs class | Value | Percentage* | Description |
|---------------------|------------|-------|-------------|-------------|
| A                   | RNA processing and modification | 0     | 0           | RNA processing and modification |
| B                   | Chromatin structure and dynamics | 0     | 0           | Chromatin structure and dynamics |
| C                   | Energy production and conversion | 78    | 2.98        | Energy production and conversion |
| D                   | Cell cycle control, cell division, chromosome partitioning | 22    | 0.84        | Cell cycle control, cell division, chromosome partitioning |
| E                   | Amino acid transport and metabolism | 167   | 6.38        | Amino acid transport and metabolism |
| F                   | Nucleotide transport and metabolism | 67    | 2.56        | Nucleotide transport and metabolism |
| G                   | Carbohydrate transport and metabolism | 248   | 9.48        | Carbohydrate transport and metabolism |
| H                   | Coenzyme transport and metabolism | 46    | 1.76        | Coenzyme transport and metabolism |
| I                   | Lipid transport and metabolism | 53    | 2.03        | Lipid transport and metabolism |
| J                   | Translation, ribosomal structure and biogenesis | 154   | 5.88        | Translation, ribosomal structure and biogenesis |
| K                   | Transcription | 187   | 7.15        | Transcription |
| L                   | Replication, recombination and repair | 155   | 5.98        | Replication, recombination and repair |
| M                   | Cell wall/membrane/envelope biogenesis | 89    | 3.40        | Cell wall/membrane/envelope biogenesis |
| N                   | Cell motility | 5     | 0.19        | Cell motility |
| O                   | Posttranslational modification, protein turnover, chaperones | 54    | 2.06        | Posttranslational modification, protein turnover, chaperones |
| P                   | Inorganic ion transport and metabolism | 104   | 3.97        | Inorganic ion transport and metabolism |
| Q                   | Secondary metabolites biosynthesis, transport and catabolism | 20    | 0.76        | Secondary metabolites biosynthesis, transport and catabolism |
| R                   | General function prediction only | 260   | 9.94        | General function prediction only |
| S                   | Function unknown | 190   | 7.26        | Function unknown |
| T                   | Signal transduction mechanisms | 59    | 2.25        | Signal transduction mechanisms |
| U                   | Intracellular trafficking, secretion, and vesicular transport | 24    | 0.91        | Intracellular trafficking, secretion, and vesicular transport |
| V                   | Defense mechanisms | 60    | 2.29        | Defense mechanisms |
| W                   | Extracellular structures | 0     | 0           | Extracellular structures |
| Y                   | Nuclear structure | 0     | 0           | Nuclear structure |
| Z                   | Cytoskeleton | 0     | 0           | Cytoskeleton |
| —                   | Not in COGs | 754   | 28.81       | Not in COGs |

COGs, Clusters of Orthologous Groups database.

*Total is based on total number of protein-coding genes in annotated genome.

3.16 and 2.96 Mb, respectively), but larger than those of *E. saccharolyticus, E. columbae, E. cecorum* and *E. sulfureus* (2.60, 2.58, 2.34 and 2.31, respectively). The G+C content of *E. massiliensis* is lower than those of *E. casseli flavus* and *E. gallinarum* (42.8 and 40.7) but greater than those of *E. moraviensis, E. haemoperoxidus, E. caccae, E. saccharolyticus, E. columbae, E. cecorum, E. sulfureus and E. faecalis* (39.6, 36.1, 35.7, 35.8, 36.9, 36.6, 36.4, 38.0 and 37.5, respectively). Of the 2687 predicted chromosomal genes, 2617 were protein-
coding genes and 70 were RNAs including 61 tRNAs and nine rRNAs (5S = 4, 23S = 2, 16S = 3). A total of 1889 genes (72.2%) were assigned to a putative function (Fig. 3, Table 3). Seventy-one genes were identified as ORFans (2.71%), and the remaining genes were annotated as hypothetical proteins. The distribution of genes into COGs functional categories is presented in Table 3.

Conclusion and Perspectives

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of Enterococcus massiliensis sp. nov. AM1T. This strain was isolated in Marseille, France.

Taxonomic and Nomenclatural Proposals

Description of Enterococcus massiliensis sp. nov.

Enterococcus massiliensis (massiliensis because this strain was isolated in Massilia, the Latin name of Marseille, where the strain was sequenced).

Colonies were whitish and approximately 1 mm diameter on 5% sheep’s blood–enriched agar. Cells are Gram-positive, non-haemolytic, facultative anaerobic with a mean length of 1.2 μm and a mean diameter of 0.6 μm. Growth occurred between 25°C to 37°C, but optimal growth was observed at 37°C. Alkaline phosphatase, esterase (C4), esterase lipase (C8), leucine arylamidase, acid phosphatase, β-galactosidase and N-acetyl-β-glucosaminidase activities were present. Esulin activity was also positive, but catalase, oxidase, β-galactosidase and N-acetyl-β-glucosaminidase were negative. Positive reaction were obtained for d-ribose, d-glucose, d-fructose, d-mannose and N-acetylglucosamine. E. massiliensis was susceptible to vancomycin, teicoplanin, linezolid, gentamicin, ciprofloxacin, doxycycline, rifampicin and pristinamycin, but resistant to trimethoprim/sulfamethoxazole, fosfomycin, erythromycin and clindamycin.

The G+C content of the genome is 39.6%. The 16S rRNA and genome sequences are deposited in GenBank under accession numbers LN833866 and CVRN00000000, respectively. The type strain AM1T (= CSUR P1927 = DSM 100308) was isolated from a fresh stool sample of a patient in Marseille, France.

Acknowledgements

The authors thank the Xegen Company (http://www.xegen.fr) for automating the genomic annotation process.

Conflict of Interest

None declared.

References

[1] Lagier JC, Hugon P, Khelafia S, Fourrier PE, La Scola B, Raoult D. The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. Clin Microbiol Rev 2013;28:237–64.
[2] Seng P, Drancourt M, Gouriet F, La Scola B, Fourrier PE, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 2009;49:543–51.
[3] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152–5.
[4] Rosello-Mora R. DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation. In: Stackebrandt E, editor. Molecular identification, systematics, and population structure of prokaryotes. Berlin: Springer; 2015. p. 23–50.
[5] Thiercelin E, Jouhaud L. Reproduction of the enteroococque; taches centrales; granulations périphériques et microblastes. Compt Rend Sci Biol 1903;55:686–8.
[6] Schleifer KH, Kilpper-Balz R. Transfer of Streptococcus faecalis and Streptococcus fecium to the genus Enterococcus nom. rev. as Enterococcus faecalis comb. nov. and Enterococcus fecium comb. nov. Int J Syst Bacteriol 1984;34:31–4.
[7] Bereket W, Hemalatha K, Getenet B, Wondwossen T, Solomon A, Zeynudin A, et al. Update on bacterial nosocomial infections. Eur Rev Med Pharmacol Sci 2012;16:1039–44.
[8] Miller WR, Munita JM, Arias CA. Mechanisms of antibiotic resistance in enterococci. Expert Rev Anti Infect Ther 2014;12:1221–36.
[9] 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.
[10] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955–64.
[11] Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 2007;35:3100.
[12] Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955–64.
[13] Weber T, Bliin D, Duddela S, Krug D, Kim HU, Bruccoleri R, et al. antiSMASH 3.0—an comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res 2015;43:W237–43.
[14] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340:783–95.
[15] 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.
[16] Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence–based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 2016;21:14–60.
[17] Meier-Kolthoff JP, Klenk HP, Göker M. Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. Int J Syst Evol Microbiol 2014;64:352–6.