

Non-contiguous finished genome sequence and description of *Megasphaera massiliensis* sp. nov.

Roshan Padmanabhan¹, Jean-Christophe Lagier¹, Nicole Prisca Makaya Dangui¹, Caroline Michelle¹, Carine Couderc¹, Didier Raoult¹ and Pierre-Edouard Fournier^{1*}

¹Aix-Marseille Université, URMITE, Faculté de médecine, Marseille, France

*Correspondence: Pierre-Edouard Fournier (pierre-edouard.fournier@univ-amu.fr)

Keywords: *Megasphaera massiliensis*, genome, culturomics, taxonogenomics

Megasphaera massiliensis strain NP3^T sp. nov. is the type strain of *Megasphaera massiliensis* sp. nov., a new species within the genus *Megasphaera*. This strain, whose genome is described here, was isolated from the fecal flora of an HIV-infected patient. *M. massiliensis* is a Gram-negative, obligate anaerobic coccobacillus. Here we describe the features of this organism, together with the complete genome sequence and annotation. The 2,661,757 bp long genome (1 chromosome but no plasmid) contains 2,577 protein-coding and 61 RNA genes, including 5 rRNA genes.

Introduction

Megasphaera massiliensis sp. nov. strain NP3^T (= CSUR P245 = DSM 26228) is the type strain of *M. massiliensis* sp. nov. This bacterium is a Gram-negative, non-sporulating, anaerobic and non-motile coccobacillus that was isolated from the stool of an HIV-infected patient as part of a culturomics study designed to cultivate individually all bacterial species within human feces [1,2].

The current classification of prokaryotes is based on a combination of phenotypic and genotypic characteristics [3,4] including 16S rRNA gene phylogeny, G + C content and DNA-DNA hybridization (DDH). Despite being considered as a “gold standard”, these tools exhibit several drawbacks [5,6]. To date, almost 4,000 bacterial genomes have been sequenced [7] and the cost of genomic sequencing is constantly decreasing. Therefore, we recently proposed the addition of genomic information to phenotypic criteria, including the protein profile, for the description of new bacterial species [8-29].

The genus *Megasphaera* (Rogosa 1971), created in 1971 [30], currently contains 5 species including *M. cerevisiae* (Engelmann and Weiss 1986) [31], *M. elsdenii* (Gutierrez *et al.* 1959) [30], *M. micronuciformis* (Marchandin *et al.* 2003) [32], *M. paucivorans* (Juvonen and Suihko 2006) [33] and *M. sueciensis* (Juvonen and Suihko 2006) [33]. The type species, *M. elsdenii* (Gutierrez *et al.* 1959)

[30], originally classified in the *Peptostreptococcus* genus (Gutierrez *et al.* 1959), was later reclassified within a new genus, *Megasphaera* (Rogosa 1971), in the family *Veillonellaceae* (Rogosa 1971) [30]. It is an obligately anaerobic, lactate-fermenting, gastrointestinal microbe of ruminant and non-ruminant mammals, including humans. It was also isolated in a case of human endocarditis [34]. The genome from *M. elsdenii* strain DSM 20460, isolated from the rumen of sheep, was recently sequenced [35]. *M. cerevisiae* [31], *M. micronuciformis* [32], *M. paucivorans* and *M. sueciensis* [33] are brewery-associated species. Here we present a summary classification and a set of features for *M. massiliensis* sp. nov. strain NP3^T (= CSUR P245 = DSM 26228) together with the description of the complete genome sequencing and annotation. These characteristics support the circumscription of the species *M. massiliensis*.

Classification and features

A stool sample was collected from a 32-year-old HIV-infected patient living in Marseille, France. The patient gave written informed consent for the study. The study was approved by the Ethics Committee of the Institut Fédératif de Recherche IFR48, Faculty of Medicine, Marseille, France, under agreement number 09-022.

The fecal specimen was preserved at -80°C after collection. Strain NP3^T (Table 1) was isolated in January 2012 by cultivation on 5% sheep blood agar in anaerobic condition at 37°C, following a 7-day preincubation of the stool specimen in an anaerobic blood culture bottle enriched with sterile 5% sheep rumen fluid and 5% sheep blood. The strain exhibited a nucleotide sequence similarity with other members of the genus *Megasphaera*

ranging from 91.5% with *M. cerevisiae* strain PAT1^T to 95.8% with *M. elsdenii* strain ATCC 25940^T, its closest validated phylogenetic neighbor (Figure 1). These values were lower than the 98.7% 16S rRNA gene sequence threshold recommended by Stackebrandt and Ebers to delineate a new species without carrying out DNA-DNA hybridization [4].

Table 1. Classification and general features of *Megasphaera massiliensis* strain NP3^T according to the MIGS recommendations [36]

MIGS ID	Property	Term	Evidence code ^a
		Domain <i>Bacteria</i>	37]
		Phylum <i>Firmicutes</i>	TAS [38-40]
		Class <i>Negativicutes</i>	TAS [41]
	Current classification	Order <i>Selenomonadales</i>	TAS [41]
		Family <i>Veillonellaceae</i>	TAS [41-43]
		Genus <i>Megasphaera</i>	TAS [30,32,43,44]
		Species <i>Megasphaera massiliensis</i>	IDA
		Type strain NP3 ^T	IDA
	Gram stain	Negative	IDA
	Cell shape	coccobacilli	IDA
	Motility	non motile	IDA
	Sporulation	nonsporulating	IDA
	Temperature range	mesophilic	IDA
	Optimum temperature	37°C	IDA
MIGS-6.3	Salinity	unknown	
MIGS-22	Oxygen requirement	anaerobic	IDA
	Carbon source	unknown	
	Energy source	unknown	
MIGS-6	Habitat	human gut	IDA
MIGS-15	Biotic relationship	free living	IDA
	Pathogenicity	unknown	
	Biosafety level	2	
MIGS-14	Isolation	human feces	
MIGS-4	Geographic location	France	IDA
MIGS-5	Sample collection time	January 2012	IDA
MIGS-4.1	Latitude	43.296482	IDA
MIGS-4.1	Longitude	5.36978	IDA
MIGS-4.3	Depth	Surface	IDA
MIGS-4.4	Altitude	0 m above sea level	IDA

Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [45]. If the evidence is IDA, then the property was directly observed for a live isolate by one of the authors or an expert mentioned in the acknowledgements.

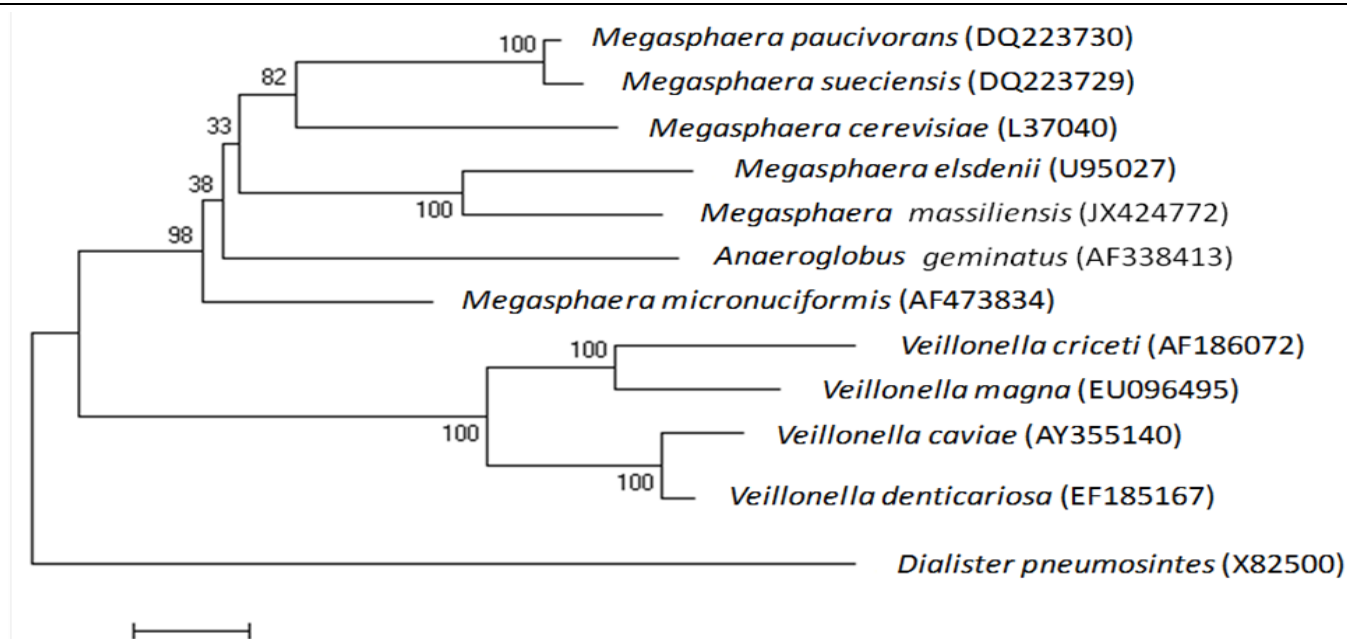


Figure 1. Phylogenetic tree highlighting the position of *Megasphaera massiliensis* strain NP3^T relative to other type strains within the genus *Megasphaera* and other members of the family *Veillonellaceae*. GenBank accession numbers are indicated in parentheses. Sequences were aligned using CLUSTALW, and phylogenetic inferences obtained using the maximum-likelihood method within the MEGA software. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 500 times to generate a majority consensus tree. *Dialister pneumosintes* was used as outgroup. The scale bar indicates a 1% nucleotide sequence divergence.



Figure 2. Gram staining of *Megasphaera massiliensis* strain NP3^T

Different growth temperatures (25, 30, 37, 45°C) were tested. Growth occurred between 30 and 45°C, and optimal growth was observed at 37°C. Colonies were transparent and smooth with a diameter of 0.5 to 1 mm on blood-enriched Columbia agar (BioMérieux). Growth of the strain was tested in 5% sheep blood agar (BioMérieux) under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems, respectively (BioMérieux), and in the presence of air, with or without 5% CO₂. Growth only occurred in anaerobic atmosphere. No growth was observed under aerobic conditions and microaerophilic conditions. A motility test was negative. Cells grown on agar are Gram-negative coccobacilli (Figure 2), with a mean diameter of 0.87 µm and the presence of phages in electron microscopy (Figure 3).

Strain NP3^T exhibited oxidase, but no catalase activity. Using RAPID 32A identification strips (BioMérieux), positive reactions were observed for α-glucosidase and β-glucosidase. Negative reactions were observed for urease, arginine dihydrolase, α and β-galactosidase, β-galactosidase-6-phosphate, α-arabinosidase, β-glucuronidase, N-acetyl-β-glucosanimidase, mannose and raffinose fermentation, α-fucosidase, alkanine phosphatase, arginine arylamidase,

proline arylamidase, leucyl glycine arylamidase, phenylalanine arylamidase, leucine arylamidase, pyroglutamic acid arylamidase, tyrosine arylamidase, alanine arylamidase, glycine arylamidase, histidine arylamidase, glutamyl glutamic acid arylamidase and serine arylamidase. Carbohydrate metabolism was examined using an API 50CH strip (BioMérieux). Positive reactions were observed for potassium gluconate, potassium 5-cetogluconate, aesculin, salicine, N-acetylglucosamine, and arbutine production, and L-arabinose, D-ribose, D-xylose, D-galactose, D-glucose, D-fructose, D-mannose, L-rhamnose, D-mannitol, D-sorbitol, D-celiobiose, D-maltose, D-lactose, D-trehalose, gentiobiose, L-fucose and D-arabitol fermentation. Weak reactions were observed for amygdaline and potassium 2-cetogluconate production, and glycerol and D-arabinose fermentation. Table 2 summarizes the differential phenotypic characteristics of *M. massiliensis*, *M. elsdenii* and *M. micronuciformis*. *M. massiliensis* strain NP3^T was susceptible to amoxicillin, amoxicillin-clavulanic acid, ceftriaxone, imipenem and doxycycline but resistant to vancomycin, erythromycin, rifampicin, trimethoprim-sulfamethoxazole, metronidazole and ciprofloxacin.

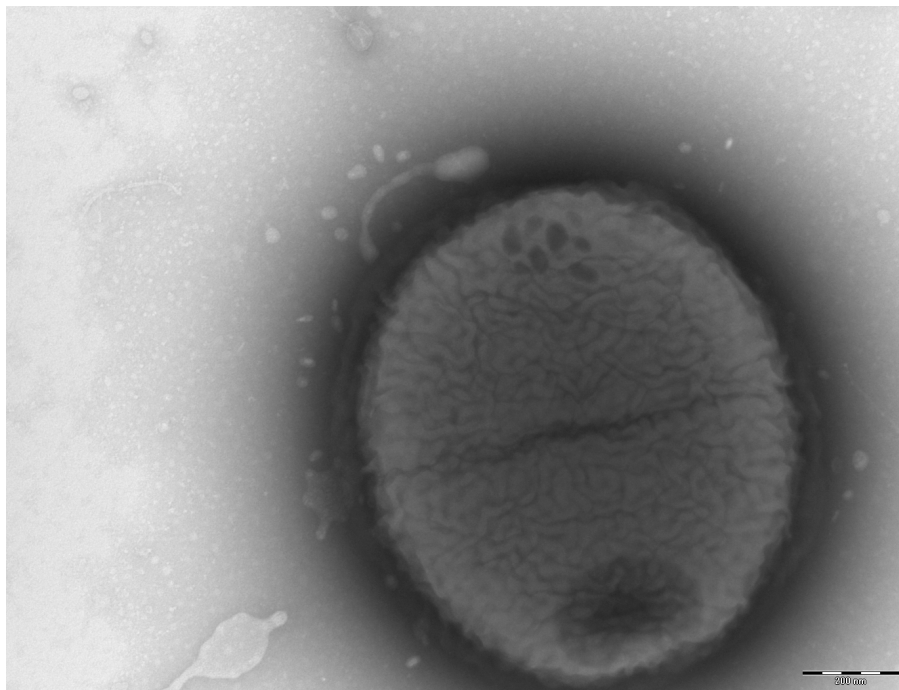


Figure 3. Transmission electron microscopy of *M. massiliensis* strain NP3^T, using a Morgani 268D (Philips) at an operating voltage of 60kV. The scale bar represents 200 nm.

Table 2. Differential characteristics of *M. massiliensis* strain NP3^T, *M. elsdenii* strain DSM 20460 and *M. micronuciformis* strain AIP 412-00^{T,†}

Properties	<i>M. massiliensis</i>	<i>M. elsdenii</i>	<i>M. micronuciformis</i>
Cell diameter (µm)	0.87	1.5-3.0	0.4-0.6
Oxygen requirement	anaerobic	anaerobic	Anaerobic
Pigment production	+	+	–
Gram stain	–	–	–
Motility	–	–	–
Endospore formation	–	–	–
Indole production	–	na	–
Production of			
Catalase	–	–	–
Oxidase	+	+	na
Nitrate reductase	na	–	–
Urease	–	–	na
β-galactosidase	–	–	na
N-acetyl-glucosamine	na	–	na
Acid production from			
Arabinose	w	–	–
Ribose	+	–	na
Mannose	–	–	–
Mannitol	+	+	–
Raffinose	–	–	–
Sucrose	–	–	–
Glycerol	w	–	–
Sorbitol	+	–	na
Arabitol	+	–	na
Galactose	+	+	–
D-glucose	+	+	–
D-fructose	+	+	–
D-maltose	+	+	–
D-lactose	+	+	–
Hydrolysis of gelatin			
	+	+	–
Habitat	Human gut	Sheep rumen	Liver abscess, whitlow

na = data not available; w = weak

Matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) MS protein analysis was carried out as previously described [46] using a Microflex spectrometer (Brüker Daltonics, Germany). Briefly, a pipette tip was used to pick one isolated bacterial colony from a culture agar plate and spread it as a thin film on a MTP 384 MALDI-TOF target plate (Bruker Daltonics). Twelve distinct deposits were done for strain NP3^T from 12 isolated colonies. Each smear was overlaid with 2 µL of matrix solution (a saturated solution of alpha-cyano-4-hydroxycinnamic acid) in 50% acetonitrile, 2.5% tri-fluoroacetic acid and allowed to dry for five minutes. Spectra were recorded in the positive linear mode for the mass range from 2,000 to 20,000 Da (parameter settings: ion source 1 (ISI), 20kV; IS2, 18.5 kV; lens, 7 kV). A spectrum was obtained after 675 shots with variable laser power. The time of acquisition was between 30 seconds and 1 minute per spot. The 12 NP3^T spectra were imported into the MALDI Bio Typer software (version 2.0, Bruker) and analyzed by standard pattern matching (with default parameter settings) against the

main spectra of 3,769 bacteria, including the spectra from *M. micronuciformis*, *Veillonella atypica*, *V. caviae*, *V. criceti*, *V. denticariosi*, *V. dispar*, *V. montpellierensis*, *V. parvula*, *V. ratti* and *V. rogosae*, that were used as reference data (Figures 4 and 5). The method of identification included the m/z from 3,000 to 15,000 Da. For every spectrum, 100 peaks at most were taken into account and compared with the spectra in the database. The MALDI-TOF score enabled the predictive identification and discrimination of the tested species from those in a database: a score > 2 with a validated species enabled identification at the species level, and a score < 1.7 did not enable any identification. No significant score was obtained for strain NP3^T against the Brüker database, suggesting that our isolate was not a member of a known species. We added the spectrum from strain NP3^T to our database for future reference (Figure 4). Figure 5 shows the MALDI-TOF MS spectrum differences between *M. massiliensis* and other *Megasphaera* and *Veillonella* species (Figure 5).

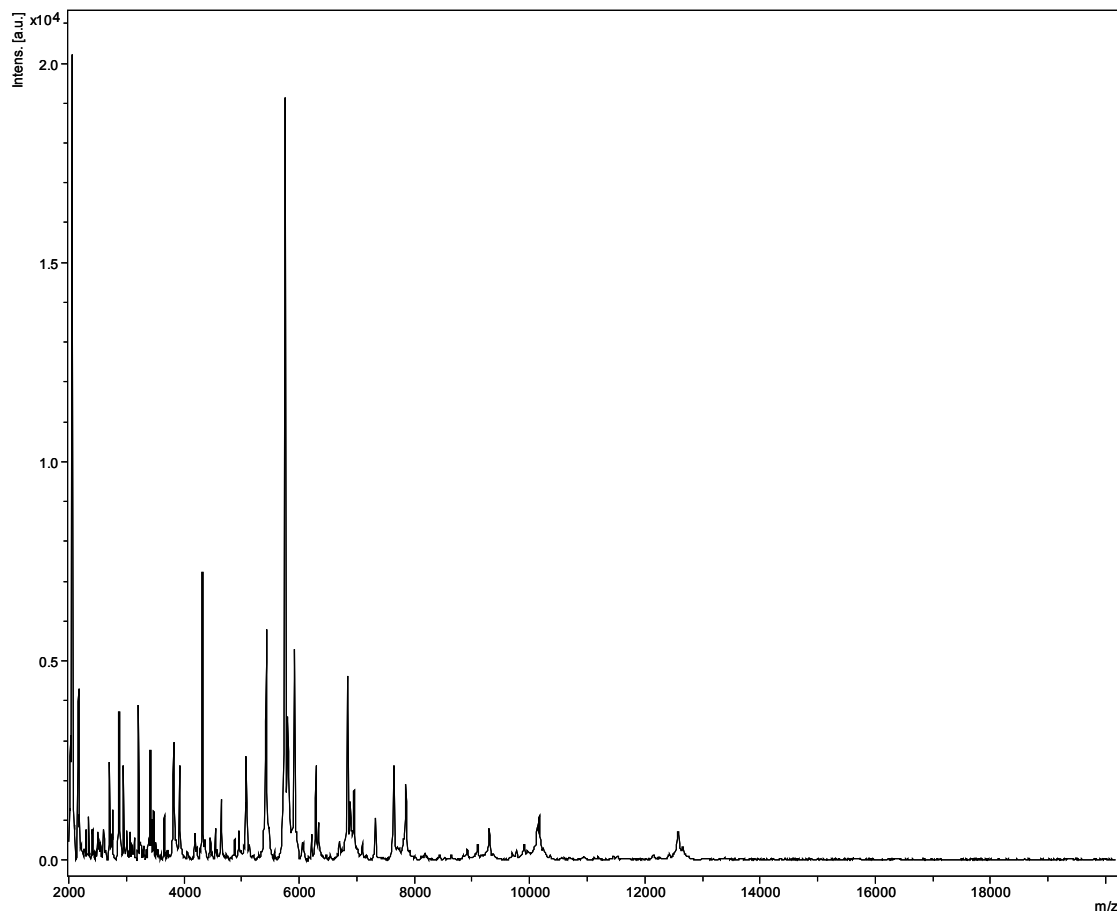


Figure 4. Reference mass spectrum from *M. massiliensis* strain NP3. Spectra from 12 individual colonies were compared and a reference spectrum was generated.

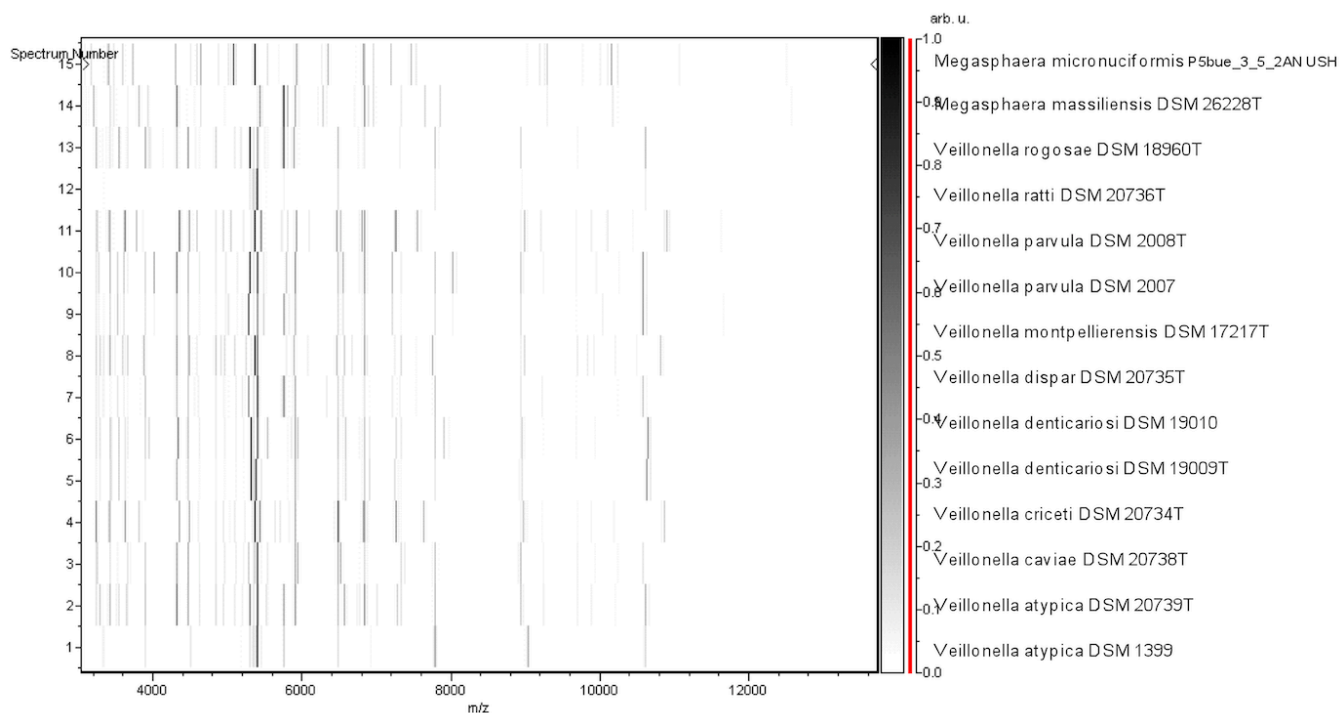


Figure 5. Gel view comparing the *M. massiliensis* NP3^T spectrum with those of *M. micronuciformis* and *Veillonella* species. The Gel View displays the raw spectra of all loaded spectrum files arranged in a pseudo-gel like look. The x-axis records the m/z value. The left y-axis displays the running spectrum number originating from subsequent spectra loading. The peak intensity is expressed by a Gray scale scheme code. The color bar and the right y-axis indicate the relation between the color a peak is displayed and the peak intensity in arbitrary units.

Genome sequencing information

Genome project history

The organism was selected for sequencing on the basis of its phenotypic differences, phylogenetic position and 16S rRNA similarity to other members of the genus *Megasphaera*, and is part of a study of the human digestive flora aiming at isolating all bacterial species within human feces [1,2]. It was the third genome of a *Megasphaera* species and the first sequenced genome of *M. massiliensis* sp. nov. The GenBank ID is CAVO00000000 and consists of 106 large contigs. Table 3 shows the project information and its association with MIGS version 2.0 compliance [47].

Growth conditions and DNA isolation

Megasphaera massiliensis strain NP3^T sp. nov. (= CSUR P245 = DSM 26228) was grown anaerobically on 5% sheep blood-enriched agar (BioMérieux) at 37°C. Ten petri dishes were spread and resuspended in 3 × 100 µl of G2 buffer (EZ1 DNA Tissue kit, Qiagen). A first mechanical lysis was performed using glass powder on the Fastprep-24

<http://standardsingenomics.org>

device (MP Biomedicals, Ilkirch, France) during 2 × 20 seconds. DNA was then treated with 2.5 µg/µL lysozyme treatment (30 minutes at 37°C) and extracted using a BioRobot EZ 1 Advanced XL (Qiagen). The DNA was then concentrated and purified using a Qiasm kit (Qiagen). The yield and the concentration were measured using the Quant-it Picogreen kit (Invitrogen) on the Genios_Tecan fluorometer at 82.2 ng/µl.

Genome sequencing and assembly

A paired-end sequencing strategy was used (Roche). The library was pyrosequenced on a GS FLX Titanium sequencer (Roche). This project was loaded on a 1/4 region on PTP Picotiterplate (Roche). Five µg of DNA were mechanically fragmented on the Covaris device (KBioScience-LGC Genomics, Teddington, UK) using miniTUBE-Red 5Kb. The DNA fragmentation was visualized through the Agilent 2100 BioAnalyzer on a DNA labchip 7500 with an optimal size of 3.3 kb.

After PCR amplification through 17 cycles followed by double size selection, the single stranded paired-end library was then loaded on a DNA labchip RNA pico 6000 on the BioAnalyzer. The pattern showed an optimum at 613 bp and the concentration was quantified on a Genios Tecan fluorometer at 3.48 pg/ μ L. The library concentration equivalence was calculated at 5.21E+09 molecules/ μ L. The library was stored at -20°C until further use, and the library was clonally amplified with 0.5 cpb in 3 emPCR reactions with the GS Titanium SV emPCR Kit (Lib-L) v2 (Roche). The yield of the emPCR was 9.99%, in the range of 5 to 20% from the Roche procedure. Approximately 790,000 beads were loaded on the GS Titanium PicoTiterPlates PTP Kit 70x75 and sequenced with the GS FLX Titanium Sequencing Kit XLR70 (Roche). The run was performed overnight and then analyzed on the cluster through the gsRunBrowser and Newbler Assembler (Roche). A total of 186,153 passed filter wells generated 61.97 Mb with a length average of 332 bp. The filter-passed sequences were assembled using Newbler with 90% identity and 40 bp as overlap. The final assembly identified 114 large contigs (>1,500 bp) arranged into 28 scaffolds and generated a genome size of 2.66 Mb, which corresponds to a coverage of 23.3 \times genome equivalent.

Genome annotation

Prodigal [48] with default parameters was used to predict the Open Reading Frames (ORFs). The predicted ORFs were excluded if they spanned a sequencing gap region. Protein functional assessment was obtained by comparison with sequences in the GenBank [49] and Clusters of Orthologs Groups (COG) databases using BLASTP. The rRNA and tRNA were identified using RNAmmer [50] and tRNAscan-SE 1.21 [51] respectively. SignalP [52]

and TMHMM [53] were used to predict signal peptides and transmembrane helices, respectively. ORFans were identified if their BLASTP E-value was lower than 1e-03 for alignment length greater than 80 amino acids. If alignment lengths were smaller than 80 amino acids, we used an E-value of 1e-05. Such parameter thresholds have already been used in previous works to define ORFans. Artemis [54] was used for data management and DNA Plotter [55] was used for visualization of genomic features. PHAST was used to identify, annotate and graphically display prophage sequences within bacterial genomes or plasmids [56]. To estimate the mean level of nucleotide sequence similarity at the genome level between *M. massiliensis* and another 5 members of the family *Veillonellaceae*, orthologous proteins were detected using the Proteinortho software with the following parameters: e-value 1e-5, 30% percentage of identity, 50% coverage and algebraic connectivity of 50% [57], and genomes compared two by two. For each pair of genomes, we determined the mean percentage of nucleotide sequence identity among orthologous ORFs using BLASTn.

Genome properties

The genome of *M. massiliensis* strain NP3^T is 2,661,757 bp long (in 28 scaffolds, 1 chromosome, and no plasmid) with a 50.2% GC content (Table 3 and Figure 6). Of the 2,577 predicted genes, 2,516 were protein-coding genes and there were 61 RNA genes. A total of 1,697 genes (65.8%) were assigned a putative function. A total of 248 genes (9.6%) were annotated as hypothetical proteins. The properties and the statistics of the genome are summarized in Tables 4 and 5. The distribution of genes into COGs functional categories is presented in Table 5.

Table 3. Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality draft
MIGS-28	Libraries used	Paired-end 3kb library
MIGS-29	Sequencing platforms	454 GS FLX Titanium
MIGS-31.2	Fold coverage	19 \times
MIGS-30	Assemblers	Newbler version 2.5.3
MIGS-32	Gene calling method	Prodigal
	INSDC ID	PRJEB645
	Genbank ID	CAVO00000000
	Genbank Date of Release	June 4, 2013
	Project relevance	Study of the human gut microbiome

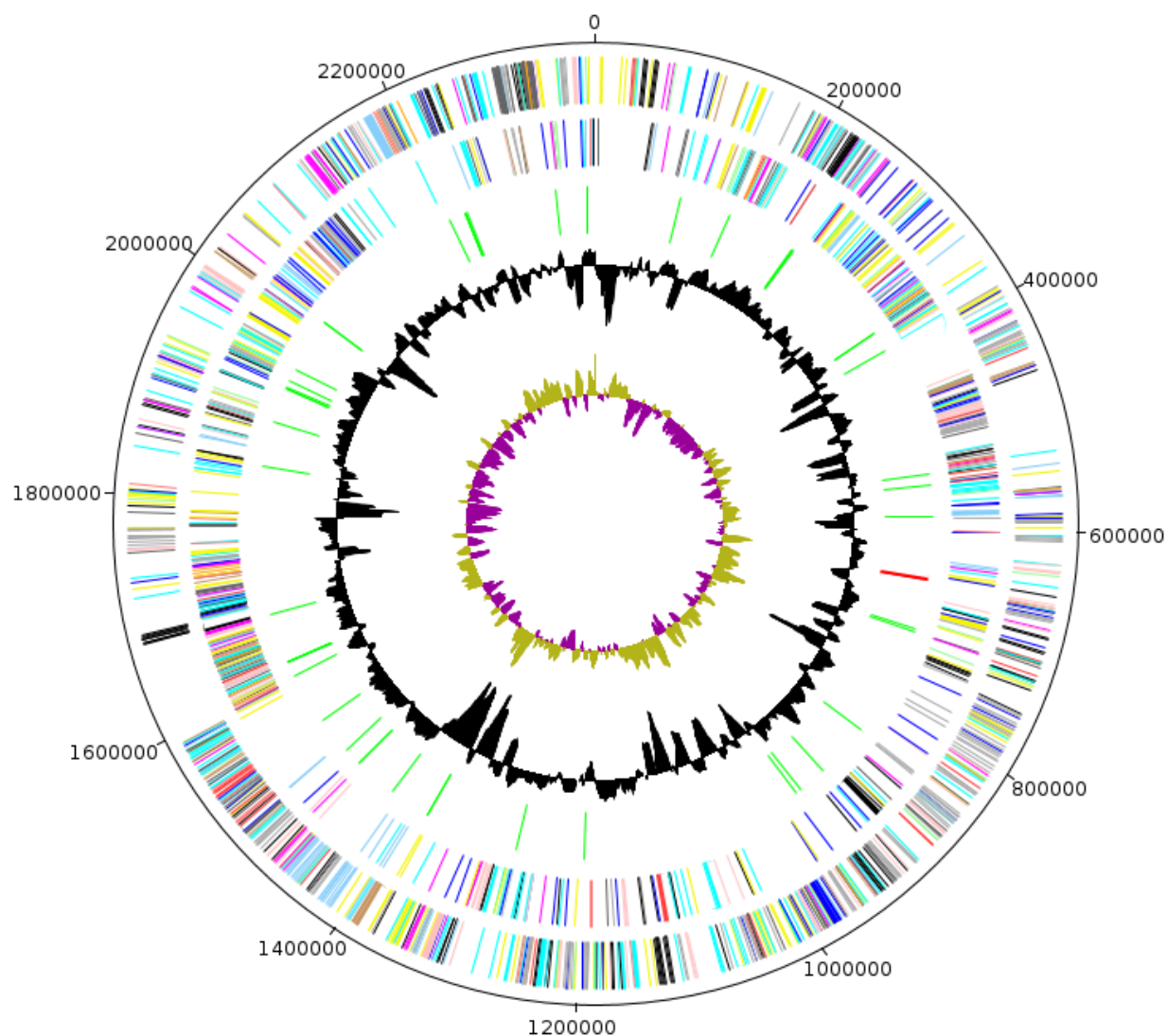


Figure 6. Graphical circular map of the *M. massiliensis* strain NP3 chromosome. From the outside in: the outer two circles show open reading frames oriented in the forward and reverse (colored by COG categories) directions, respectively. The third circle displays the rRNA gene operon (red) and tRNA genes (green). The fourth circle shows the G+C% content plot. The inner-most circle shows the GC skew, purple and olive indicating negative and positive values, respectively.

Comparison with the genomes from *M. elsdenii*, *Megasphaera* species, *Veillonella dispar*, *V. parvula* and *Anaeroglobus geminatus*

The draft genome of *M. massiliensis* strain NP3^T (2.66 Mb) has a larger size than that of *M. elsdenii* (2.47 Mb), *V. parvula* (2.13 Mb), *V. dispar* (2.12 Mb), *A. geminatus* (1.79 Mb) and *M. micronuciformis* (1.77 Mb) respectively. *M. massiliensis* has a lower G + C content (50.2%) than *M. elsdenii* (52.8%) but higher than *V. parvula*, *V. dispar*, *M. micronuciformis* and *A. geminatus* (38.6, 38.8, 46.8 and 48.7%, respectively). *M. massiliensis* (2,516) has more predicted protein-coding genes than *M. elsdenii*, *A. geminatus*, *V. dispar*, *V. parvula*, and *M.*

micronuciformis (2,219, 2,148, 1,954, 1,844 and 1,774, respectively) (Table 6). In addition, *M. massiliensis* shared a mean genomic sequence similarity of 81.84, 69.44, 63.68, 62.92 and 70.27% with *M. elsdenii*, *M. micronuciformis*, *V. dispar*, *V. parvula* and *A. geminatus* respectively (Table 6).

M. massiliensis harbors two intact bacteriophages. Based on PHAST results, phage 1 of *M. massiliensis* was most closely related to *Clostridium* phage phi CD119 whereas phage 2 was most similar to *Bacillus* phage BCJA1c.

Table 4. Nucleotide content and gene count levels of the genome

Attribute	Value	% of total ^a
Genome size (bp)	2,661,757	
DNA coding region (bp)	1,479,861	93.98
DNA G+C content (bp)	1,337,412	50.2
Coding region (bp)	1,479,861	93.98
Number of replicons	1	
Extrachromosomal elements	0	
Total genes	2,577	100
RNA genes	61	2.39
rRNA operons	2	
Protein-coding genes	2,516	97.63
Genes with function prediction	1,697	65.8
Genes assigned to COGs	1,892	73.41
Genes with peptide signals	60	2.38
Genes with transmembrane helices	530	21.0
CRISPR repeats	7	

^aThe 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 5. Number of genes associated with the 25 general COG functional categories

Code	Value	%age ^a	Description
J	138	5.48	Translation
A	0	0	RNA processing and modification
K	120	4.77	Transcription
L	118	4.69	Replication, recombination and repair
B	0	0	Chromatin structure and dynamics
D	23	0.91	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	28	1.11	Defense mechanisms
T	27	1.07	Signal transduction mechanisms
M	103	4.09	Cell wall/membrane biogenesis
N	0	0	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	24	0.95	Intracellular trafficking and secretion
O	52	2.07	Post-translational modification, protein turnover, chaperones
C	147	5.84	Energy production and conversion
G	118	4.69	Carbohydrate transport and metabolism
E	163	6.48	Amino acid transport and metabolism
F	52	2.07	Nucleotide transport and metabolism
H	87	3.46	Coenzyme transport and metabolism
I	46	1.83	Lipid transport and metabolism
P	79	3.14	Inorganic ion transport and metabolism
Q	14	0.56	Secondary metabolites biosynthesis, transport and catabolism
R	217	8.62	General function prediction only
S	141	5.60	Function unknown
-	195	7.75	Not in COGs

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

Table 6. Orthologous gene comparison and average nucleotide identity of *M. massiliensis* with other compared genomes[†]

Species (GenBank accession number)	<i>M. massiliensis</i>	<i>M. elsdenii</i>	<i>M. micronuciformis</i>	<i>V. dispar</i>	<i>V. parvula</i>	<i>A. geminatus</i>
<i>M. massiliensis</i> (CAVO000000000)	2,516	1,289	1,189	987	999	1,159
<i>M. elsdenii</i> (HE576794)	81.84	2,219	1,175	980	989	1,145
<i>M. micronuciformis</i> (AECS000000000)	69.44	69.01	1,774	933	939	1,167
<i>V. dispar</i> (ACIK000000000)	63.68	63.08	64.92	1,954	1,081	893
<i>V. parvula</i> (ADFU000000000)	62.92	62.01	64.43	67.62	1,844	899
<i>A. geminatus</i> (AGCJ000000000)	70.27	70.50	74.22	63.87	62.99	2,148

[†]Upper right, numbers of orthologous genes; lower left, mean nucleotide identities of orthologous genes. Bold numbers indicate the numbers of genes or each genome.

Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Megasphaera massiliensis* sp. nov. that contains the strain NP3^T. This bacterial strain has been found in Marseille, France.

Description of *Megasphaera massiliensis* sp. nov.

Megasphaera massiliensis (mas.il.ien'sis. L. gen. fem. n. *massiliensis*, of Massilia, the Latin name of Marseille where was cultivated strain NP3^T). It has been isolated from the feces of a 32-year-old HIV-infected French patient.

Colonies were smooth and transparent with 0.5 to 1 mm in diameter on blood-enriched Columbia agar. Optimal growth is only achieved anaerobically and grows between 30 and 45°C, with optimal growth observed at 37°C. The strain is a Gram-negative, non-endospore forming, non motile coccobacillus. Positive for α -glucosidase, β -glucosidase, potassium gluconate, potassium 5-cetogluconate, aesculin, salicine, N-acetylglucosamine, and arbutine production. Positive for L-arabinose, D-ribose, D-xylose, D-galactose, D-glucose, D-fructose, D-mannose, L-rhamnose, D-mannitol, D-sorbitol, D-celiobiose, D-maltose, D-lactose, D-trehalose, gentiobiose, L-

fucose and D-arabitol fermentation. Negative for urease, arginine dihydrolase, α and β -galactosidase, β -galactosidase-6-phosphate, α -arabinosidase, β -glucuronidase, N-acetyl- β -glucosanimidase, mannose and raffinose fermentation, α -fucosidase, alkanine phosphatase, arginine arylamidase, proline arylamidase, leucyl glycine arylamidase, phenylalanine arylamidase, leucine arylamidase, pyroglutamic acid arylamidase, tyrosine arylamidase, alanine arylamidase, glycine arylamidase, histidine arylamidase, glutamyl glutamic acid arylamidase and serine arylamidase. Weak reactions observed for amygdaline and potassium 2-cetogluconate production, and glycerol and D-arabinose fermentation. Cells are susceptible to amoxicillin, amoxicillin-clavulanic acid, ceftriaxone, imipenem and doxycycline, but resistant to vancomycin, erythromycin, rifampicin, trimethoprim/sulfamethoxazole, metronidazole, and ciprofloxacin. The G+C content of the genome is 50.2%. The 16S rRNA and genome sequences are deposited in Genbank under accession numbers JX424772 and CAVO000000000, respectively. The type strain NP3^T (= CSUR P245 = DSM 26228) was isolated from the fecal flora of an HIV-infected patient in Marseille, France.

Acknowledgements

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

study was funded by the Mediterranee-Infection Foundation.

References

1. Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, Bittar F, Fournous G, Gimenez G, Maraninchi M, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect* 2012; **18**:1185-1193. [PubMed](#)
2. Dubourg G, Lagier JC, Armougom F, Robert C, Hamad I, Brouqui P. The gut microbiota of a patient with resistant tuberculosis is more comprehensively studied by culturomics than by metagenomics. *Eur J Clin Microbiol Infect Dis* 2013; **32**:637-645. [PubMed](#)
<http://dx.doi.org/10.1007/s10096-012-1787-3>
3. Tindall BJ, Rossello-Mora R, Busse HJ, Ludwig W, Kampfer P. Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol* 2010; **60**:249-266. [PubMed](#)
<http://dx.doi.org/10.1099/ijs.0.016949-0>
4. Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* 2006; **33**:152-155.
5. Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, Krichevsky MI, Moore LH, Moore WEC, Murray RGE, Stackebrandt E, et al. Report of the ad hoc committee on reconciliation of approaches to bacterial systematic. *Int J Syst Bacteriol* 1987; **37**:463-464.
<http://dx.doi.org/10.1099/00207713-37-4-463>
6. Rossello-Mora R. DNA-DNA Reassociation Methods Applied to Microbial Taxonomy and Their Critical Evaluation. In: Stackebrandt E (ed), *Molecular Identification, Systematics, and population Structure of Prokaryotes*. Springer, Berlin, 2006; p. 23-50.
7. Genome Online Database.
<http://genomesonline.org/cgi-bin/GOLD/index.cgi>.
8. Kokcha S, Mishra AK, Lagier JC, Million M, Leroy Q, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Bacillus timonensis* sp. nov. *Stand Genomic Sci* 2012; **6**:346-355. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.2776064>
9. Lagier JC, El Karkouri K, Nguyen TT, Armougom F, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Anaerococcus senegalensis* sp. nov. *Stand Genomic Sci* 2012; **6**:116-125. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.2415480>
10. Mishra AK, Gimenez G, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Alistipes senegalensis* sp. nov. *Stand Genomic Sci* 2012; **6**:304-314. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.2625821>
11. Lagier JC, Armougom F, Mishra AK, Ngyuen TT, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Alistipes timonensis* sp. nov. *Stand Genomic Sci* 2012; **6**:315-324. [PubMed](#)
12. Mishra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Clostridium senegalense* sp. nov. *Stand Genomic Sci* 2012; **6**:386-395. [PubMed](#)
13. Mishra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Peptoniphilus timonensis* sp. nov. *Stand Genomic Sci* 2012; **7**:1-11. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.2956294>
14. Mishra AK, Lagier JC, Rivet R, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Paenibacillus senegalensis* sp. nov. *Stand Genomic Sci* 2012; **7**:70-81. [PubMed](#)
15. Lagier JC, Gimenez G, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Herbaspirillum massiliense* sp. nov. *Stand Genomic Sci* 2012; **7**:200-209. [PubMed](#)
16. Roux V, El Karkouri K, Lagier JC, Robert C, Raoult D. Non-contiguous finished genome sequence and description of *Kurthia massiliensis* sp. nov. *Stand Genomic Sci* 2012; **7**:221-232. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.3206554>
17. Kokcha S, Ramasamy D, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Brevibacterium senegalense* sp. nov. *Stand Genomic Sci* 2012; **7**:233-245. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.3256677>
18. Ramasamy D, Kokcha S, Lagier JC, N'Guyen TT, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Aeromicrobium massilense* sp. nov. *Stand Genomic Sci* 2012; **7**:246-257. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.3306717>
19. Lagier JC, Ramasamy D, Rivet R, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Cellulomonas massiliensis* sp. nov. *Stand Genomic Sci* 2012; **7**:258-270. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.3316719>
20. Lagier JC, El Karkouri K, Rivet R, Couderc C, Raoult D, Fournier PE. Non-contiguous finished genome

- sequence and description of *Senegalemassilia anaerobia* sp. nov. *Stand Genomic Sci* 2013; **7**:343-356. <http://dx.doi.org/10.4056/sigs.3246665>
21. Mishra AK, Hugon P, Lagier JC, Nguyen TT, Robert C, Couderc C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Peptoniphilus obesi* sp. nov. *Stand Genomic Sci* 2013; **7**:357-369. <http://dx.doi.org/10.4056/sigs.32766871>
 22. Mishra AK, Lagier JC, Nguyen TT, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Peptoniphilus senegalensis* sp. nov. *Stand Genomic Sci* 2013; **7**:370-381. <http://dx.doi.org/10.4056/sigs.3366764>
 23. Lagier JC, El Karkouri K, Mishra AK, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Enterobacter massiliensis* sp. nov. *Stand Genomic Sci* 2013; **7**:399-412. <http://dx.doi.org/10.4056/sigs.3396830>
 24. Hugon P, Ramasamy D, Lagier JC, Rivet R, Couderc C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Alistipes obesi* sp. nov. *Stand Genomic Sci* 2013; **7**:427-439. <http://dx.doi.org/10.4056/sigs.3336746>
 25. Mishra AK, Hugon P, Robert C, Couderc C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Peptoniphilus grossensis* sp. nov. *Stand Genomic Sci* 2012; **7**:320-330. [PubMed](http://pubmed.ncbi.nlm.nih.gov/22811111/)
 26. Mishra AK, Hugon P, Lagier JC, Nguyen TT, Couderc C, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Enorma massiliensis* gen. nov., sp. nov., a new member of the Family *Coriobacteriaceae*. *Stand Genomic Sci* 2013; **8**:290-305. <http://dx.doi.org/10.4056/sigs.3426906>
 27. Ramasamy D, Lagier JC, Gorlas A, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Bacillus massiliosenegalensis* sp. nov. *Stand Genomic Sci* 2013; **8**:264-278. <http://dx.doi.org/10.4056/sigs.3496989>
 28. Ramasamy D, Lagier JC, Nguyen TT, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Dielma fastidiosa* gen. nov., sp. nov., a new member of the Family *Erysipelotrichaceae*. *Stand Genomic Sci* 2013; **8**:336-351. <http://dx.doi.org/10.4056/sigs.3567059>
 29. Mishra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Genome sequence and description of *Timonella senegalensis* gen. nov., sp. nov., a new member of the suborder *Micrococccinae*. *Stand Genomic Sci* 2013; **8**:318-335. <http://dx.doi.org/10.4056/sigs.3476977>
 30. Rogosa M. Transfer of *Peptostreptococcus elsdenii* Gutierrez et al. to a New Genus, *Megasphaera* [*M. elsdenii* (Gutierrez et al.) comb. nov.]. *Int J Syst Bacteriol* 1971; **21**:187-189. <http://dx.doi.org/10.1099/00207713-21-2-187>
 31. Engelmann U, Weiss N. *Megasphaera cerevisiae* sp. nov.: a new Gram-negative obligately anaerobic coccus isolated from spoiled beer. *Syst Appl Microbiol* 1985; **6**:287-290. [http://dx.doi.org/10.1016/S0723-2020\(85\)80033-3](http://dx.doi.org/10.1016/S0723-2020(85)80033-3)
 32. Marchandin H, Jumas-Bilak E, Gay B, Teyssier C, Jean-Pierre H, de Buochberg MS, Carriere C, Carlier JP. Phylogenetic analysis of some *Sporomusa* sub-branch members isolated from human clinical specimens: description of *Megasphaera micronuciformis* sp. nov. *Int J Syst Bacteriol* 2003; **53**:547-553. [PubMed](http://pubmed.ncbi.nlm.nih.gov/12345678/)
 33. Juvonen R, Suihko ML. *Megasphaera paucivorans* sp. nov., *Megasphaera sueciensis* sp. nov. and *Pectinatus haikarae* sp. nov., isolated from brewery samples, and emended description of the genus *Pectinatus*. *Int J Syst Bacteriol* 2006; **56**:695-702. [PubMed](http://pubmed.ncbi.nlm.nih.gov/16789012/)
 34. Brancaccio M, Legendre GG. *Megasphaera elsdenii* endocarditis. *J Clin Microbiol* 1979; **10**:72-74. [PubMed](http://pubmed.ncbi.nlm.nih.gov/71234567/)
 35. Marx H, Graf AB, Tatto NE, Thallinger GG, Mattanovich D, Sauer M. Genome sequence of the ruminal bacterium *Megasphaera elsdenii*. *J Bacteriol* 2011; **193**:5578-5579. [PubMed](http://pubmed.ncbi.nlm.nih.gov/21234567/) <http://dx.doi.org/10.1128/JB.05861-11>
 36. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. [PubMed](http://pubmed.ncbi.nlm.nih.gov/18765432/) <http://dx.doi.org/10.1038/nbt1360>
 37. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. [PubMed](http://pubmed.ncbi.nlm.nih.gov/10987654/) <http://dx.doi.org/10.1073/pnas.87.12.4576>
 38. Gibbons NE, Murray RGE. Proposals Concerning the Higher Taxa of Bacteria. *Int J Syst Bacteriol* 1978; **28**:1-6. <http://dx.doi.org/10.1099/00207713-28-1-1>
 39. Garrity GM, Holt JG. The Road Map to the Manual. In: Garrity GM, Boone DR, Castenholz RW (eds), *Bergey's Manual of Systematic Bacteriology*,

- Second Edition, Volume 1, Springer, New York, 2001, p. 119-169.
40. Murray RGE. The Higher Taxa, or, a Place for Everything...? In: Holt JG (ed), Bergey's Manual of Systematic Bacteriology, First Edition, Volume 1, The Williams and Wilkins Co., Baltimore, 1984, p. 31-34.
 41. Marchandin H, Teyssier C, Campos J, Jean-Pierre H, Roger F, Gay B, Carlier JP, Jumas-Bilak E. *Negativicoccus succinicivorans* gen. nov., sp. nov., isolated from human clinical samples, emended description of the family *Veillonellaceae* and description of *Negativicutes* classis nov., *Selenomonadales* ord. nov. and *Acidaminococcaceae* fam. nov. in the bacterial phylum *Firmicutes*. *Int J Syst Evol Microbiol* 2010; **60**:1271-1279. [PubMed](#) <http://dx.doi.org/10.1099/ijs.0.013102-0>
 42. Rogosa M. Transfer of *Veillonella* Prévot and *Acidaminococcus* Rogosa from *Neisseriaceae* to *Veillonellaceae* fam. nov. and the inclusion of *Megasphaera* Rogosa in *Veillonellaceae*. *Int J Syst Bacteriol* 1971; **21**:231-233. <http://dx.doi.org/10.1099/00207713-21-3-231>
 43. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; **30**:225-420. <http://dx.doi.org/10.1099/00207713-30-1-225>
 44. Rogosa M. Genus III. *Megasphaera* Rogosa 1971, 187. In: Buchanan RE, Gibbons NE (eds), Bergey's Manual of Determinative Bacteriology, Eighth Edition, The Williams and Wilkins Co., Baltimore, 1974, p. 448-449.
 45. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29. [PubMed](#) <http://dx.doi.org/10.1038/75556>
 46. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, Raoult D. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin Infect Dis* 2009; **49**:543-551. [PubMed](#) <http://dx.doi.org/10.1086/600885>
 47. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. [PubMed](#) <http://dx.doi.org/10.1038/nbt1360>
 48. Prodigal. <http://prodigal.ornl.gov/>
 49. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res* 2012; **40**:D48-D53. [PubMed](#) <http://dx.doi.org/10.1093/nar/gkr1202>
 50. 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**:3100-3108. [PubMed](#) <http://dx.doi.org/10.1093/nar/gkm160>
 51. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997; **25**:955-964. [PubMed](#)
 52. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004; **340**:783-795. [PubMed](#) <http://dx.doi.org/10.1016/j.jmb.2004.05.028>
 53. 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**:567-580. [PubMed](#) <http://dx.doi.org/10.1006/jmbi.2000.4315>
 54. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. Artemis: sequence visualization and annotation. *Bioinformatics* 2000; **16**:944-945. [PubMed](#) <http://dx.doi.org/10.1093/bioinformatics/16.10.944>
 55. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 2009; **25**:119-120. [PubMed](#) <http://dx.doi.org/10.1093/bioinformatics/btn578>
 56. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. *Nucleic Acids Res* 2011; **39**:W347-W352. [PubMed](#) <http://dx.doi.org/10.1093/nar/gkr485>
 57. Lechner M, Findeib S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho. Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics* 2011; **12**:124. [PubMed](#) <http://dx.doi.org/10.1186/1471-2105-12-124>