

EXTENDED GENOME REPORT

Open Access



# High-quality draft genome sequence and description of *Haemophilus massiliensis* sp. nov.

Cheikh Ibrahima Lo<sup>1,2</sup>, Senthil Alias Sankar<sup>1</sup>, Bécaye Fall<sup>3</sup>, Bissoume Sambe-Ba<sup>3</sup>, Silman Diawara<sup>3</sup>, Mamadou Wague Gueye<sup>3</sup>, Oleg Mediannikov<sup>1,2</sup>, Caroline Blanc-Tailleur<sup>1</sup>, Boubacar Wade<sup>3</sup>, Didier Raoult<sup>1,2,4</sup>, Pierre-Edouard Fournier<sup>1</sup> and Florence Fenollar<sup>1,2\*</sup>

## Abstract

Strain FF7<sup>T</sup> was isolated from the peritoneal fluid of a 44-year-old woman who suffered from pelvic peritonitis. This strain exhibited a 16S rRNA sequence similarity of 94.8 % 16S rRNA sequence identity with *Haemophilus parasuis*, the phylogenetically closest species with a name with standing in nomenclature and a poor MALDI-TOF MS score (1.32 to 1.56) that does not allow any reliable identification. Using a polyphasic study made of phenotypic and genomic analyses, strain FF7<sup>T</sup> was a Gram-negative, facultatively anaerobic rod and member of the family *Pasteurellaceae*. It exhibited a genome of 2,442,548 bp long genome (one chromosome but no plasmid) contains 2,319 protein-coding and 67 RNA genes, including 6 rRNA operons. On the basis of these data, we propose the creation of *Haemophilus massiliensis* sp. nov. with strain FF7<sup>T</sup> (= CSUR P859 = DSM 28247) as the type strain.

**Keywords:** *Haemophilus massiliensis*, Genome, Taxono-genomics, Culturomics

## Introduction

The genus *Haemophilus* (Winslow et al. 1917) was described in 1917 [1] and currently meningitis, bacteremia, sinusitis, and/or pneumonia [2].

The current taxonomic classification of prokaryotes relies on a combination of phenotypic and genotypic characteristics [3, 4]; including 16S rRNA sequence similarity, G + C content and DNA-DNA hybridization. However, these tools suffer from various drawbacks, mainly due to their threshold values that are not applicable to all species or genera [5, 6]. With the development of cost-effective, high-throughput sequencing techniques, dozens of thousands of bacterial genome sequences have been made available in public databases [7]. Recently, we developed a strategy named taxono-genomics in which genomic and phenotypic characteristics, notably the MALDI-TOF-MS spectrum, are

systematically compared to the phylogenetically-closest species with a name with standing in nomenclature [8, 9].

The strain FF7<sup>T</sup> was isolated from the peritoneal fluid of a Senegalese woman suffering from pelvic peritonitis complicating a ruptured ovarian abscess. She was admitted to Hôpital Principal in Dakar, Senegal. *Haemophilus massiliensis* is a Gram-negative, facultatively anaerobic, oxidase and catalase-positive and non-motile rod shaped bacterium. This microorganism was cultivated as part of the MALDI-TOF-MS implementation in Hôpital Principal in Dakar, aiming at improving the routine laboratory identification of bacterial strains in Senegal [10].

Here, we present a summary classification and a set of features for *Haemophilus massiliensis* sp. nov. together with the description of the complete genome sequencing and annotation. These characteristics support the circumscription of the species *Haemophilus massiliensis*.

## Organism information

### Classification and features

In June 2013, a bacterial strain (Table 1) was isolated by cultivation on 5 % sheep blood-enriched Columbia agar (BioMérieux, Marcy l'Etoile, France) of a peritoneal fluid

\* Correspondence: florence.fenollar@univ-amu.fr

<sup>1</sup>Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, Marseille, France

<sup>2</sup>Campus international UCAD-IRD, Dakar, Senegal

Full list of author information is available at the end of the article

**Table 1** Classification and general features of *Haemophilus massiliensis* strain FF7<sup>T</sup> [13]

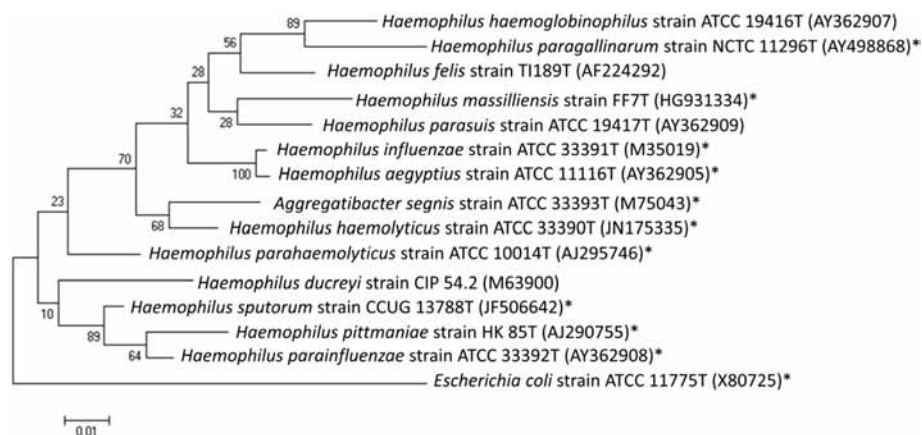
MIGS ID	Property	Term	Evidence code <sup>a</sup>
	Current classification	Domain: <i>Bacteria</i>	TAS [26]
		Phylum: <i>Proteobacteria</i>	TAS [27]
		Class: <i>Gammaproteobacteria</i>	TAS [28, 29]
		Order: <i>Pasteurellales</i>	TAS [29, 30]
		Family: <i>Pasteurellaceae</i>	TAS [31, 32]
		Genus: <i>Haemophilus</i>	TAS [1, 33]
		Species: <i>Haemophilus massiliensis</i>	IDA
	Type strain: FF7 <sup>T</sup>	IDA	
	Gram stain	Negative	IDA
	Cell shape	Rod	IDA
	Motility	Not motile	IDA
	Sporulation	Non-spore forming	IDA
	Temperature range	Mesophile	IDA
	Optimum temperature	37 °C	IDA
	pH range; Optimum	7.2-7.4; 7.3	
	Carbon source	Unknown	NAS
	Energy source	Unknown	NAS
MIGS-6	Habitat	Human peritoneal fluid	IDA
MIGS-6.3	Salinity	Unknown	
MIGS-22	Oxygen requirement	Facultatively anaerobic	IDA
MIGS-15	Biotic relationship	Free living	IDA
MIGS-14	Pathogenicity	Unknown	
	Biosafety level	2	
	Isolation	Human	IDA
MIGS-4	Geographic location	Senegal	IDA
MIGS-5	Sample collection time	June 2013	IDA
MIGS-4.1	Latitude	14° 40' N	IDA
MIGS-4.1	Longitude	17° 26' W	IDA
MIGS-4.3	Depth	Surface	IDA
MIGS-4.4	Altitude	12 m above sea level	IDA

<sup>a</sup>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 [34]. 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

specimen obtained from a 44-year-old Senegalese woman who suffered from pelvic peritonitis that had complicated a ruptured ovarian abscess [10] and

hospitalized in Hôpital Principal de Dakar, Senegal. The strain could not be identified using MALDI-TOF-MS. Strain FF7<sup>T</sup> exhibited a 94.8 % 16S rRNA sequence identity with *Haemophilus parasuis* strain ATCC 19417<sup>T</sup> (GenBank accession number AY362909), the phylogenetically-closest bacterial species with a validly published name (Fig. 1). These values were lower than the 98.7 % 16S rRNA gene sequence threshold recommended by Meier-Kolthoff et al., 2013 to delineate a new species within phylum *Proteobacteria* without carrying out wet lab or digital DNA-DNA hybridization [11].

Different growth temperatures (25 °C, 30 °C, 37 °C, 45 °C, and 56 °C) were tested. Growth was obtained between 25 and 45 °C, with the optimal growth temperature being 37 °C. Colonies were 0.5 mm in diameter and non-hemolytic on 5 % sheep blood-enriched Columbia agar (BioMérieux). Gram staining showed rod-shaped Gram-negative bacilli that were not motile and unable to form spores (Fig. 2). In electron microscopy, cells had a mean length of 2.6 µm (range 2.0-3.2 µm) and width of 0.35 µm (range 0.2-0.5 µm) (Fig. 2). Growth of the strain was tested under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems, respectively (BioMérieux), and under aerobic conditions, with or without 5 % CO<sub>2</sub>. Optimal growth was observed at 37 °C under aerobic and microaerophilic conditions. Strain FF7<sup>T</sup> exhibited oxidase and catalase activities. Using an API ZYM strip (BioMérieux), positive reactions were observed for acid phosphatase, leucine arylamidase, esterase, alkaline phosphatase and Naphthol-AS-BI-phosphohydrolase. Negative reactions were noted for α-chymotrypsin, cystine arylamidase, valine arylamidase, trypsin, α-glucosidase, β-glucosidase, esterase-lipase, leucine arylamidase, α-galactosidase, β-galactosidase, β-glucuronidase, α-mannosidase, α-fucosidase, and N-acetyl-β-glucosaminidase. Using API 20NE (BioMérieux), positive reactions were obtained for L-arginine, esculin, ferric citrate and urea but negative reactions were observed for D-glucose, L-arabinose, D-maltose, D-mannose, D-mannitol, potassium gluconate and N-acetylglucosamine. *Haemophilus massiliensis* strain FF7<sup>T</sup> is susceptible to penicillin, amoxicillin, amoxicillin/clavulanic acid, imipenem, gentamicin, ceftriaxone and doxycycline but resistant to vancomycin, nitrofurantoin, and trimethoprim/sulfamethoxazole. The minimum inhibitory concentrations for some antibiotics tested with *Haemophilus massiliensis* strain FF7<sup>T</sup> sp. nov. are listed in Additional file 1: Table S1. Five species validly published names in the *Haemophilus* genus were selected to make a phenotypic comparison with our new species named *Haemophilus massiliensis* detailed in Additional file 2: Table S2.



**Fig. 1** Phylogenetic tree showing the position of *Haemophilus massiliensis* strain FF7<sup>T</sup> relative to the most closely related type strains other type strains (type = <sup>T</sup>) within the genus *Haemophilus*. The GenBank accession numbers for 16S rRNA genes are indicated in parentheses. An asterisk marks strains that have a genome sequence in the NCBI database. Sequences were aligned using MUSCLE [35], and a phylogenetic tree inferred using the Maximum Likelihood method with Kimura 2-parameter model using the MEGA software. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 1,000 times to generate a majority consensus tree. Only bootstrap values equal to or greater than 70 % are displayed. The scale bar represents a rate of substitution per site of 1 %. *Escherichia coli* strain ATCC 11775<sup>T</sup> was used as outgroup

MALDI-TOF protein analysis was carried out as previously described [12] using a Microflex LT (Bruker Daltonics, Leipzig, Germany). For strain FF7<sup>T</sup>, scores ranging from 1.32 to 1.56 were obtained with spectra available in the Brüker database. Therefore the isolate could not be classified within any known species. The reference mass spectrum from strain FF7<sup>T</sup> was incremented in our database (Additional file 3: Figure S1). Finally, the gel view showed that all members of the genus *Haemophilus* for which spectra were available in the database could be discriminated (Additional file 4: Figure S2).

## Genome sequencing information

### Genome project history

The strain was selected for sequencing on the basis of its 16S rRNA similarity, phylogenetic position, and phenotypic differences with other members of the genus *Haemophilus*, and is part of a study aiming at using MALDI-TOF-MS for the routine identification of bacterial isolates in Hôpital Principal in Dakar [10]. It is the eleventh genome of a *Haemophilus* species and the first genome of *Haemophilus massiliensis* sp. nov. The Genbank accession number is CCFL00000000 and consists of 46 contigs. Table 2 shows the project information and its association with MIGS version 2.0 compliance [13].

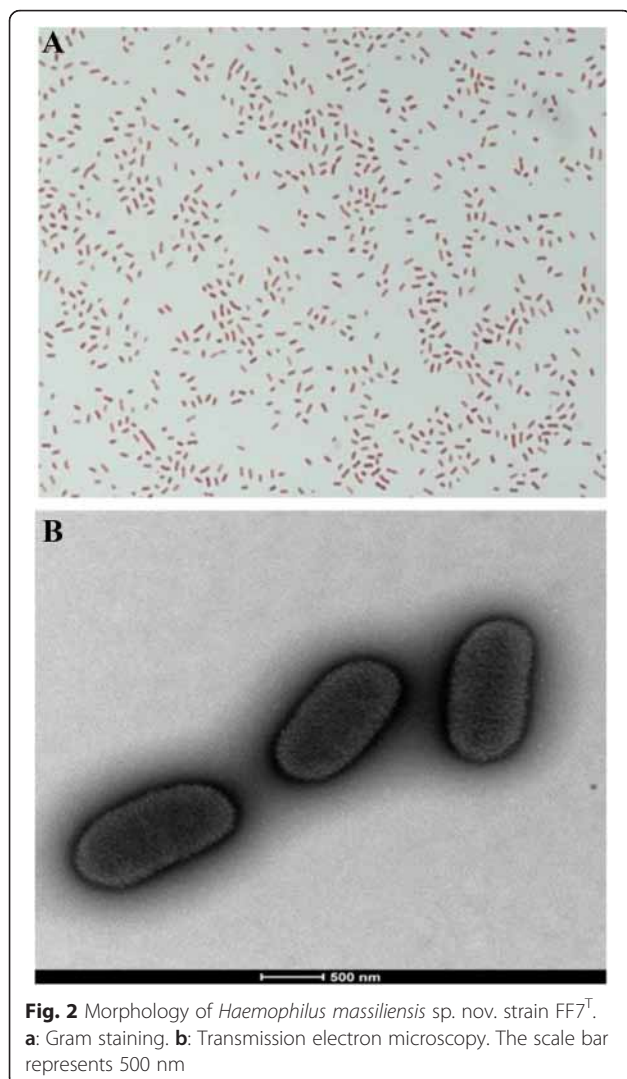
### Growth conditions and genomic DNA preparation

*Haemophilus massiliensis* sp. nov., strain FF7<sup>T</sup> (= CSUR P859= DSM 28247) was grown aerobically on 5 % sheep blood-enriched Columbia agar (BioMérieux) at 37 °C. Bacteria grown on four Petri dishes were resuspended in 5x100 µL of TE buffer; 150 µL of this suspension was

diluted in 350 µL TE buffer 10X, 25 µL proteinase K and 50 µL sodium dodecyl sulfate for lysis treatment. This preparation was incubated overnight at 56 °C. Extracted DNA was purified using 3 successive phenol-chloroform extractions and ethanol precipitations. Following centrifugation, the DNA was suspended in 65 µL EB buffer. The genomic DNA (gDNA) concentration was measured at 14.7 ng/µl using the Qubit assay with the high sensitivity kit (Life Technologies, Carlsbad, CA, USA).

### Genome sequencing and assembly

Genomic DNA of *Haemophilus massiliensis* FF7<sup>T</sup> was sequenced on the MiSeq sequencer (Illumina, San Diego, CA, USA) with the Mate-Pair strategy. The gDNA was barcoded in order to be mixed with 11 other projects with the Nextera Mate-Pair sample prep kit (Illumina). The Mate-Pair library was prepared with 1 µg of genomic DNA 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, Santa Clara, CA, USA) with a DNA 7500 labchip. The DNA fragments ranged in size from 1 kb up to 10 kb with an optimal size at 4.08 kb. No size selection was performed and only 464 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal at 569 bp on the Covaris S2 device in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies) and the final library concentration was measured at 24.42 nmol/L. The libraries were normalized at 2nM

**Table 2** Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality draft
MIGS-28	Libraries used	Mate-Pair 3.1 kb library
MIGS-29	Sequencing platforms	Illumina Miseq
MIGS-31.2	Fold coverage	42.54
MIGS-30	Assemblers	CLC GENOMICSWB4
MIGS-32	Gene calling method	Prodigal
	Locus Tag	Not indicated
	Genbank ID	CCFL00000000
	Genbank Date of Release	August 22, 2014
	GOLD ID	Ga0059233
	BIOPROJECT	PRJEB5521
MIGS-13	Source material identifier	CSUR P859, DSM 28247
	Project relevance	MALDI-TOF-MS implementation in Dakar

and pooled. After a denaturation step and dilution at 15 pM, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and sequencing run were performed in a single 39-h-run in a 2x251-bp. Total information of 10.1Gb was obtained from a 1,189 K/mm<sup>2</sup> cluster density with a cluster passing quality control filters of 99.1 % (22,579,000 clusters). Within this run, the index representation for *Haemophilus massiliensis* was 9.72 %. The 1,976,771 paired reads were filtered according to the read qualities. These reads were trimmed, then assembled using the CLC genomicsWB4 software. Finally, the draft genome of *Haemophilus massiliensis* consists of 9 scaffolds with 46 contigs and generated a genome size of 2.4 Mb with a 46.0 % G + C content.

#### Genome annotation

Open Reading Frames were predicted using Prodigal [14] with default parameters but the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched against the GenBank database [15] and the Clusters of Orthologous Groups databases using BLASTP. The tRNAScanSE tool [16] was used to find tRNA genes, whereas ribosomal RNAs were found using RNAmmer [17] and BLASTn against the GenBank database. Lipoprotein signal peptides and the number of transmembrane helices were predicted using SignalP [18] and TMHMM [19] 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 [20] was used for data management and DNA Plotter [21] for visualization of genomic features. The Mauve alignment tool (version 2.3.1) was used for multiple genomic sequence alignment [22]. To estimate the mean level of nucleotide sequence similarity at the genome level, we used the AGIOS home-made software [9]. Briefly, this software combines the Proteinortho software [23] for detecting orthologous proteins in pairwise genomic comparisons, then retrieves the corresponding genes and determines the mean percentage of nucleotide sequence identity among orthologous ORFs using the Needleman-Wunsch global alignment algorithm. The script created to calculate AGIOS values was named MAGi and is written in perl and bioperl modules. GGDC analysis was also performed using the GGDC web server as previously reported [24, 25].

**Table 3** Genome statistics

Attribute	Value	% of total <sup>a</sup>
Genome size (bp)	2,442,548	100
DNA coding (bp)	2,181,795	89.35
DNA G + C (bp)	1,123,572	46.0
DNA scaffolds	46	-
Total genes	2,386	100
Protein coding genes	2,319	97.19
RNA genes	67	2.80
Pseudo genes	N/D <sup>b</sup>	-
Genes in internal clusters	N/D <sup>b</sup>	-
Genes with function prediction	1,885	79.00
Genes assigned to COGs	2,093	87.72
Genes with Pfam domains	1,419	59.47
Genes with signal peptides	188	7.87
Genes with transmembrane helices	445	18.65
ORFan genes	36	1.50
CRISPR repeats	2	0.08

<sup>a</sup>The 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

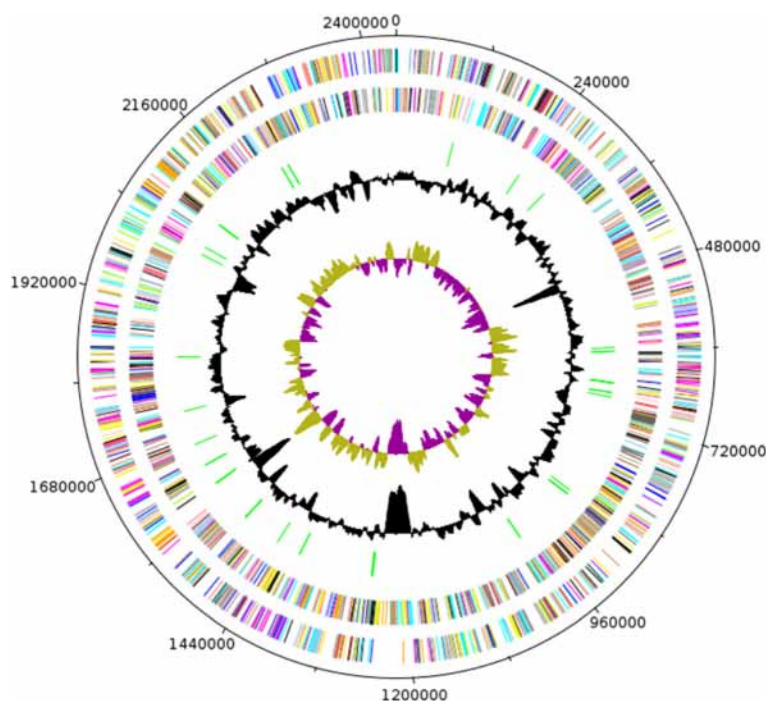
<sup>b</sup>N/D = not determined

### Genome properties

The genome of *Haemophilus massiliensis* strain FF7<sup>T</sup> is 2,442,548 bp-long with a 46.0 % G + C content. Of the 2,386 predicted genes, 2,319 were protein-coding genes and 67 were RNA genes, including six complete rRNA operons. A total of 1,885 genes (79.5 %) were assigned a putative function. A total of 36 genes were identified as ORFans (1.5 %). The remaining genes were annotated as hypothetical proteins. The properties and statistics of the genome are summarized in Table 3 and Fig. 3. The distribution of genes into COGs functional categories is presented in Table 4 and Fig. 4. The distribution of genes into COGs categories was similar for most of the compared species (Fig. 4). However, *H. influenzae* and *H. aegyptius* were over-represented for category N (cell motility), and *H. ducreyi* was under-represented for category W (extracellular structures) (Fig. 4).

### Insights from the genome sequence

Here, we compared the genome sequences of *Haemophilus massiliensis* strain FF7<sup>T</sup> (GenBank accession number CCFL000000000) with those of *Haemophilus parasuis* strain SH0165 (CP001321), *Haemophilus influenzae* strain Rd KW20 (L42023), *Aggregatibacter segnis* strain ATCC 33393<sup>T</sup> (AEPS000000000), *Haemophilus sputorum* strain CCUG 13788<sup>T</sup> (AFNK000000000), *Haemophilus*



**Fig. 3** Graphical circular map of the *Haemophilus massiliensis* strain FF7<sup>T</sup> chromosome. From the outside in, the outer two circles show open reading frames oriented in the forward (colored by COG categories) and reverse (colored by COG categories) directions, respectively. The third circle marks the tRNA genes (green). The fourth circle shows the G + C% content plot. The inner-most circle shows GC skew, purple indicating negative values whereas olive for positive values

**Table 4** Number of genes associated with general COG functional categories

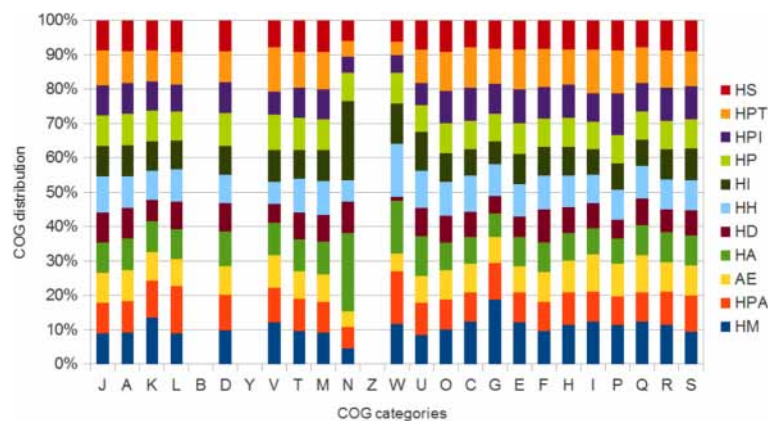
Code	Value	% of total <sup>a</sup>	Description
J	160	6.90	Translation
A	1	0.04	RNA processing and modification
K	150	6.47	Transcription
L	131	5.65	Replication, recombination and repair
B	0	0	Chromatin structure and dynamics
D	28	1.21	Cell cycle control, mitosis and meiosis
V	37	1.60	Defense mechanisms
T	46	1.98	Signal transduction mechanisms
M	135	5.82	Cell wall/membrane biogenesis
N	6	0.26	Cell motility
W	9	0.39	Extracellular structures
U	58	2.50	Intracellular trafficking and secretion
O	109	4.70	Posttranslational modification, protein turnover, chaperones
C	163	7.03	Energy production and conversion
G	228	9.83	Carbohydrate transport and metabolism
E	234	10.09	Amino acid transport and metabolism
F	63	2.72	Nucleotide transport and metabolism
H	117	5.05	Coenzyme transport and metabolism
I	62	2.67	Lipid transport and metabolism
P	152	6.55	Inorganic ion transport and metabolism
Q	34	1.47	Secondary metabolites biosynthesis, transport and catabolism
R	258	11.13	General function prediction only
S	174	7.50	Function unknown
-	226	9.53	Not in COGs

<sup>a</sup>The total is based on the total number of protein coding genes in the annotated genome

*pittmaniae* strain HK 85 (AFUV00000000), *Haemophilus aegyptius* strain ATCC 1111<sup>T</sup> (AFBC00000000), *Haemophilus parainfluenzae* strain ATCC 33392<sup>T</sup> (AEWU00000000), *Haemophilus haemolyticus* strain M21621 (AFQQ00000000), *Haemophilus ducreyi* strain 35000HP (AE017143), and *Haemophilus parahaemolyticus* strain HK385 (AJSW00000000).

The draft genome of *Haemophilus massiliensis* has a larger size than that of *H. parasuis*, *H. influenzae*, *A. segnis*, *H. sputorum*, *H. pittmaniae*, *H. aegyptius*, *H. parainfluenzae*, *H. haemolyticus*, *H. ducreyi*, and *H. parahaemolyticus* (2.44, 2.27, 1.83, 1.99, 2.14, 2.18, 1.92, 2.11, 2.09, 1.7, and 2.03 Mb, respectively). The G + C content of *Haemophilus massiliensis* is higher than those of *H. parasuis*, *H. influenzae*, *A. segnis*, *H. sputorum*, *H. pittmaniae*, *H. aegyptius*, *H. parainfluenzae*, *H. haemolyticus*, *H. ducreyi*, and *H. parahaemolyticus* (46.0, 40.0, 38.2, 42.5, 39.7, 42.5, 38.1, 39.1, 38.4, 38.2, and 40.1 %, respectively). As it has been suggested in the literature that the G + C content deviation is at most 1 % within species, these data are an additional argument for the creation of a new taxon [25].

The gene content of *Haemophilus massiliensis* is larger than those of *H. parasuis*, *H. influenzae*, *A. segnis*, *H. sputorum*, *H. aegyptius*, *H. parainfluenzae*, *H. haemolyticus*, *H. ducreyi*, and *H. parahaemolyticus* (2,319, 2,299, 1,765, 1,956, 2,072, 2,020, 2,068, 2,056, 1,717, and 1,980, respectively) but smaller than that of *H. pittmaniae* (2,390). However the distribution of genes into COG categories was similar in all compared genomes as shown in Fig. 4. In addition, in this last figure, *Haemophilus massiliensis* shared 2,021, 1,956, 2,020, 1,717, 1,977, 1,610, 1,980, 2,010, 2,390, and 2,123 orthologous genes with *H. parasuis*, *A. segnis*, *H. aegyptius*, *H. ducreyi*, *H. haemolyticus*, *H. influenzae*, *H.*



**Fig. 4** Distribution of functional classes of predicted genes in the genomes from *Haemophilus massiliensis* (HM) strain FF7<sup>T</sup>, *H. parasuis* (HPA) strain ATCC 19417<sup>T</sup>, *Aggregatibacter segnis* (AE) strain ATCC 33393<sup>T</sup>, *H. aegyptius* (HA) strain ATCC 11116<sup>T</sup>, *H. ducreyi* (HD) strain CIP 54.2, *H. haemolyticus* (HH) strain ATCC 33390<sup>T</sup>, *H. influenzae* (HI) strain ATCC 33391<sup>T</sup>, *H. parahaemolyticus* (HP), *H. parainfluenzae* (HPI) strain ATCC 10014<sup>T</sup>, *H. pittmaniae* (HPT) strain HK 85<sup>T</sup>, and *H. sputorum* (HS) strain CUG 13788<sup>T</sup> chromosomes according to the clusters of orthologous groups of proteins

**Table 5** dDDH values (upper right) and AGIOS values obtained (lower left)

	HM <sup>a</sup>	HPA <sup>b</sup>	HI <sup>c</sup>	HA <sup>d</sup>	AE <sup>e</sup>	HH <sup>f</sup>	HP <sup>g</sup>	HD <sup>h</sup>	HS <sup>i</sup>	HPT <sup>j</sup>	HPH <sup>k</sup>
HM	<b>2,319</b>	0.248	0.222	0.203	0.204	0.202	0.243	0.285	0.232	0.235	0.201
HPA	70.64	<b>2,021</b>	0.296	0.292	0.244	0.278	0.236	0.237	0.262	0.283	0.251
HI	72.76	72.75	<b>1,610</b>	0.777	0.234	0.433	0.259	0.252	0.279	0.252	0.239
HA	72.80	72.67	97.31	<b>2,020</b>	0.231	0.434	0.237	0.244	0.271	0.242	0.235
AE	74.19	71.93	75.72	75.69	<b>1,956</b>	0.235	0.243	0.267	0.255	0.246	0.232
HH	72.72	72.64	91.85	91.80	75.72	<b>1,977</b>	0.240	0.250	0.284	0.247	0.246
HP	70.16	75.40	74.07	73.98	71.97	73.79	<b>1,980</b>	0.239	0.218	0.293	0.227
HD	70.00	74.81	72.47	72.36	71.34	72.33	75.34	<b>1,717</b>	0.228	0.270	0.251
HS	70.23	75.06	73.46	73.39	72.33	73.54	78.00	75.68	<b>2,123</b>	0.319	0.280
HPT	72.55	71.51	76.38	76.48	74.95	76.64	71.88	71.19	72.88	<b>2,390</b>	0.269
HPH	72.67	72.71	79.69	79.70	76.20	79.96	73.36	72.30	74.14	78.94	<b>2,010</b>

The values printed in bold are gene numbers. Digital DDH similarities between the genomes were calculated using GGDC web server version 2.0. under recommend setting [36, 37]; formula 2 is recommended, particularly for draft genomes. <sup>a</sup>*Haemophilus massiliensis*, <sup>b</sup>*Haemophilus parasuis*, <sup>c</sup>*Haemophilus influenzae*, <sup>d</sup>*Haemophilus aegyptius*, <sup>e</sup>*Aggregatibacter segnis*, <sup>f</sup>*Haemophilus haemolyticus*, <sup>g</sup>*Haemophilus parainfluenzae*, <sup>h</sup>*Haemophilus ducreyi*, <sup>i</sup>*Haemophilus sputorum*, <sup>j</sup>*Haemophilus pittmaniae* and <sup>k</sup>*Haemophilus parahaemolyticus*

*parahaemolyticus*, *H. parainfluenzae*, *H. pittmaniae*, and *H. sputorum*, respectively (Table 5). Among species with standing in nomenclature, AGIOS values ranged from 71.19 between *H. pittmaniae* and *H. ducreyi* to 97.31 % between *H. influenzae* and *H. aegyptius*. When compared to other species, *Haemophilus massiliensis* exhibited AGIOS values ranging from 70.00 with *H. ducreyi* to 74.19 with *A. segnis*. We obtained similar results using the GGDC software, as dDDH values ranged from 0.201 to 0.777 between studied species, and were 0.248 between *Haemophilus massiliensis* and *Haemophilus parasuis*. These values confirm the status of *Haemophilus massiliensis* as a new species.

## Conclusions

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Haemophilus massiliensis* sp. nov. that contains strain FF7<sup>T</sup> (CSUR P859 = DSM 28247) which is the type strain. The strain was isolated from a peritoneal fluid specimen from a 44-year-old Senegalese woman admitted to Hôpital Principal in Dakar, Senegal.

## Description of *Haemophilus massiliensis* sp. nov.

*Haemophilus massiliensis* (mas.il.i.en.'sis. L. gen. masc. n. massiliensis, of *Massilia*, the Latin name of Marseille where strain FF7<sup>T</sup> was characterized).

*Haemophilus massiliensis* is a facultatively anaerobic Gram-negative bacterium, non-endospore forming and non-motile. Colonies are not haemolytic, round, and light with a size of 0.5-1 mm on blood-enriched Colombia agar. Cells are rod-shaped with a mean length of 2.6 µm (range 2.0-3.2 µm) and a mean diameter of 0.35 µm (range

0.2-0.5 µm). Growth occurs between 25 and 45 °C, with optimal growth occurring at 37 °C. Catalase and oxidase reactions are positive. Positive reactions are also observed for acid phosphatase, leucine arylamidase, esterase, alkaline phosphatase, Naphthol-AS-BI-phosphohydrolase, L-arginine, esculin, ferric citrate, and urea. *Haemophilus massiliensis* strain FF7<sup>T</sup> is susceptible to penicillin, amoxicillin, amoxicillin/clavulanic acid, imipenem, gentamicin, ceftriaxone and doxycycline but resistant to vancomycin, nitrofurantoin and trimethoprim/sulfamethoxazole.

The type strain is FF7<sup>T</sup> (= CSUR P859 = DSM 28247) and was isolated from the peritoneal fluid of a 44-year-old Senegalese woman suffering from pelvic peritonitis in Dakar, Senegal.

## Additional files

**Additional file 1: Table S1.** Antimicrobial susceptibility and minimum inhibitory concentrations (MIC) of *Haemophilus massiliensis* strain FF7<sup>T</sup> sp. nov. (DOC 30 kb)

**Additional file 2: Table S2.** Differential characteristics of *Haemophilus massiliensis* strain FF7<sup>T</sup> (data from this study), *H. influenzae* strain ATCC 33391T [1, 33], *H. sputorum* strain CCUG 13788T [38, 39], *H. pittmaniae* strain HK 85T [40, 41], *Haemophilus felis* strain T1189T [42, 43] and *H. parasuis* strain ATCC 19417T [33, 41]. na = data not available. (DOC 57 kb)

**Additional file 3: Figure S1.** Reference mass spectrum from *Haemophilus massiliensis* strain FF7<sup>T</sup>. Spectra from 12 individual colonies were compared and a reference spectrum was generated. (DOCX 25 kb)

**Additional file 4: Figure S2.** Gel view comparing *Haemophilus massiliensis* strain FF7<sup>T</sup> to the members of the *Haemophilus* genus. 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 with and the peak intensity in arbitrary units. Displayed species are indicated on the left. (DOCX 192 kb)

### Abbreviations

CSUR: Strains collection of *Rickettsia* Unit; CNERS: National Ethics Committee of Senegal; MALDI-TOF MS: Matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry; MAGi: Marseille Average genomic identity; dDDH: digital DNA-DNA hybridization; URMITE: Unit of Research on Emergent Infectious and Tropical Diseases.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

CIL performed the phenotypic characterization of the bacterium and drafted the manuscript. SAS performed the genomic analyses and drafted the manuscript. BF participated in the design of the study and helped to draft the manuscript. BSB, SD and MWG performed the phenotypic characterization of the bacterium and helped to draft the manuscript. OM participated in its design and helped to draft the manuscript. CBT performed the genomic sequencing and helped to draft the manuscript. BW conceived the study and helped to draft the manuscript. DR conceived the study and helped to draft the manuscript. PEF and FF conceived the study, participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

The authors thank the Xegen Company for automating the genomic annotation process. This study was funded by the Méditerranée-Infection Foundation.

### Author details

<sup>1</sup>Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, Marseille, France. <sup>2</sup>Campus international UCAD-IRD, Dakar, Senegal. <sup>3</sup>Hôpital Principal, Dakar, Senegal. <sup>4</sup>Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia.

Received: 2 April 2015 Accepted: 4 April 2016

Published online: 14 April 2016

### References

- Winslow CEA, Broadhurst J, Buchanan RE, Krumwiede C, Rogers LA, Smith GH. The families and genera of the bacteria: preliminary report of the committee of the society of american bacteriologists on characterization and classification of bacterial types. *J Bacteriol.* 1917;2:505–66.
- Musher DM. *Haemophilus* species. In: Samuel B, editor. Book: Medical Microbiology, vol. Chapter 30. 4th ed. Galveston: University of Texas Medical Branch; 1996.
- Tindall BJ, Rosselló-Móra R, Busse HJ, Ludwig W, Kämpfer P. Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol.* 2010;60:249–66.
- Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today.* 2006;33:152–5.
- Wayne LG, Brenner DJ, Colwell PR, Grimont PAD, Kandler O, Krichevsky MI, Moore LH, Moore WEC, Murray RGE, Stackebrandt E and others. Report of the ad hoc committee on reconciliation of approaches to bacterial systematic. *Int J Syst Bacteriol.* 1987;37:463–4.
- Rosselló-Móra 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; 2006. p. 23–50.
- Reddy TBK, Thomas A, Stamatis D, Bertsch J, Isbani M, Jansson J, Mallajosyula J, Pagani I, Lobos E and Kyripides N. The Genomes OnLine Database (GOLD) v. 5: a metadata management system based on a four level (meta) genome project classification. *Nucl Acids Res.* 2014;43:1099–106.
- Lo CI, Padhmanabhan R, Mediannikov O, Nguyen TT, Raoult D, Fournier PE, et al. Genome sequence and description of *Pantoea septica* strain FF5. *Stand Genomic Sci.* 2015;10:103.
- Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi-Tamisier M, Sentausa E, Raoult D, Fournier PE. A polyphasic strategy incorporating genomic data for the taxonomic description of new bacterial species. *Int J Syst Evol Microbiol.* 2014;64:384–91.
- Fall B, Lo CI, Samb-Ba B, Perrot N, Diawara S, Gueye MW, et al. The ongoing revolution of maldi-tof mass spectrometry for microbiology reaches tropical Africa. *Am J Trop Med Hyg.* 2015;92:641–7.
- Meier-Kolthoff JP, Göker M, Spröer C, Klenk HP. When should a DDH experiment be mandatory in microbial taxonomy? *Arch Microbiol.* 2013;195:413–8. <http://dx.doi:10.1007/s00203-013-0888-4>.
- Seng P, Abat C, Rolain JM, Colson P, Lagier JC, 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;51:2182–94.
- 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;26:541–7.
- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* 2010;11:119.
- Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res.* 2012;40:D48–53.
- 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.
- Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucl Acids Res.* 2007;35:3100–8.
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol.* 2004;340:783–95.
- 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.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. Artemis: sequence visualization and annotation. *Bioinformatics.* 2000;16:944–5.
- Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics.* 2009;25:119–20.
- Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 2004;14:1394–403.
- 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.
- Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics.* 2013;14:60.
- 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.
- Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci U S A.* 1990;87:4576–9.
- Garrity GM, Bell JA, Lilburn T. Phylum XIV. *Proteobacteria* phyl. nov. In: *Bergey's Manual of Systematic Bacteriology.* 2005. p. 2. Part B: 1.
- Garrity GM, Bell JA, Lilburn T. Class III. *Gammaproteobacteria* class. nov. In: *Bergey's Manual of Systematic Bacteriology.* 2005. p. 2. Part B: 1.
- Euzéby J. Validation of publication of new names and new combinations previously effectively published outside the IJSEM. List no. 106. *Int J of Syst Evol Microbiol.* 2005;55:2235–8.
- Garrity GM, Bell JA, Lilburn T. Order XIV. *Pasteurellales* ord. nov. In: *Bergey's Manual of Systematic Bacteriology.* 2005. p. 2. Part B: 850.
- Pohl S. Reklassifizierung der Gattung *Actinobacillus* Brumpt 1910, *Haemophilus* Winslow, et al. 1917 und *Pasteurella* Trevisan 1887 anhand phänotypischer und molekularer Daten, insbesondere der DNS-Verwandtschaften bei DNS-DNS-hybridisierung in vitro und vorschlag einer neuen Familie, *Pasteurellaceae*. 1979.
- Pohl S. Validation of the publication of new names and new combinations previously effectively published outside the IJSEM. List no. 7. *Int J Syst Bacteriol.* 1981;31:382–3.
- Skerman VBD, McGowan V, Sneath PHA. Approved lists of bacterial names. *Int J Syst Bacteriol.* 1980;30:225–420.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000;25:25–9.



35. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–7.
36. Auch AF, Von Jan M, Klenk HP, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci.* 2010;2:117–34. <http://dx.doi.org/10.4056/sigs.531120>.
37. Auch AF, Klenk HP, Göker M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand Genomic Sci.* 2010;2:142–8. <http://dx.doi.org/10.4056/sigs.541628>.
38. Nørskov-Lauritsen N, Bruun B, Andersen C, Kilian M. Identification of haemolytic *Haemophilus* species isolated from human clinical specimens and description of *Haemophilus sputorum* sp. nov. *Int J Med Microbiol.* 2012;302:78–83.
39. Nørskov-Lauritsen N, Bruun B, Andersen C, Kilian M. List of new names and new combinations previously effectively, but not validly, published. *Int J Syst Evol Microbiol.* 2012;62:1443–5.
40. Nørskov-Lauritsen N, Bruun B, Andersen C, Kilian M. Multilocus sequence phylogenetic study of the genus *Haemophilus* with description of *Haemophilus pittmaniae* sp. nov. *Int J Syst Evol Microbiol.* 2005;55:449–56.
41. Biberstein EL, White DC. A proposal for the establishment of two new *Haemophilus* species. *J Med Microbiol.* 1969;2:75–8.
42. Inzana TJ, Johnson JL, Shell L, Møller K, Kilian M. Isolation and characterization of a newly identified *Haemophilus* species from cats: "*Haemophilus felis*". *J Clin Microbiol.* 1992;30:2108–12.
43. Inzana TJ, Johnson L, Shell L, Møller K, Kilian M. Validation of publication of new names and new combinations previously effectively published outside the IJSB. List No. 69. *Int J Syst Bacteriol.* 1999;49:341–2.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

