Description of Clostridium cagae sp. nov., Clostridium rectalis sp. nov. and Hathewaya massiliensis sp. nov., new anaerobic bacteria isolated from human stool samples

M. L. Tall, C. I. Lo, E. Kuete Yimagou, S. Ndongo, T. P. T. Pham, D. Raoult, P. -E. Fournier, F. Fenollar, A. Levasseur

Description of *Clostridium cagae* sp. nov., *Clostridium rectalis* sp. nov., and *Hathewaya massiliensis* sp. nov., new anaerobic bacteria isolated from human stool samples

Mamadou Lamine TALL[1,2], Cheikh Ibrahima LO[1,3] , Edmond KUETE YIMAGOU[1,2], Sokhna NDONGO[1,2], Thi Phuong Thao Pham[1,2], Didier RAOULT[1,2], Pierre-Edouard FOURNIER[1,2], Florence FENOLLAR[1,3], Anthony LEVASSEUR[1,2,4]*

1 Aix-Marseille Université, UMR MEPHI (Microbes, Evolution, Phylogeny and Infections), IRD, APHM, Faculté de Médecine, Marseille 13005, France.

2 IHU-Méditerranée Infection, Marseille, France.

3 Aix Marseille Univ, IRD, AP-HM, SSA, VITROME, Marseille, France

4 Institut Universitaire de France (IUF), Paris, France

*** Corresponding author :** Pr. Anthony Levasseur, AMU, MEPHI (Microbes, Evolution, Phylogeny and Infection), UM63, IRD, APHM, IHU - Méditerranée Infection, 19-21 Boulevard Jean Moulin, 13385, Marseille cedex 05, France.

E-mail: anthony.levasseur@univ-amu.fr

Running title: *Clostridium cagae* sp. nov., *Clostridium rectalis* sp. nov. and *Hathewaya massiliensis* sp. nov.

Keywords: *Clostridium cagae* sp. nov., *Clostridium rectalis* sp. nov., *Hathewaya massiliensis* sp. nov.; culturomics; taxonogenomics; malnutrition; HIV.

Abstract:

Using culturomics method, three strains were isolated, identified and characterized following the taxono-genomics concept. *Clostridium cagae* strain Marseille-P4344^T (=CSURP4344), *Clostridium rectalis* strain Marseille-P4200^T (=CSURP4200), and *Hathewaya massiliensis* strain Marseille-P3545^T (=CSURP3545) were isolated from human stool samples. The phylogenic reconstruction, phenotypic criteria and genomic analyses were carried out and demonstrated that these three bacteria are different from previously known bacterial species with standing in nomenclature and were classified as new members of *Clostridiaceae* family.

Introduction

The taxonomy of the genus *Clostridium* has undergone many changes and revisiting over the course of its history [1-3]. It was first described by Prazmowski in 1880 and classified in the phylum *Firmicutes* [4]. It is mainly composed of Gram positive bacteria, strictly anaerobic and able to produce spores. Members of this genus are frequently encountered in various environments such as soils [5, 6], commensal digestive flora of mammals [7], algae [8]. Moreover, some species can fix nitrogen and play an important role in agriculture, such as *C. butyricum* and *C. pasteurianum* [9, 10]. Several *Clostridium* species are also involved in human pathologies. For example, *Clostridium difficile* causes nosocomial infections [11, 12]. Furthermore, others as *C. botulinum*, *C. tetani* and *C. perfringens* are responsible for neuroparalytic [13], tetanus [14] and gastrointestinal [15] diseases in human or animal. Withal, the wide diversity of the genus *Clostridium* led to the creation of the genus *Hathewaya* which regroups *Hathewaya histolytica*, *Hathewaya limosa* and *Hathewaya proteolytica* named in the past as *C. histolytica*, *C. limosa* and *C. proteolytica*, respectively [16].

In order to identify most bacteria living in the human gut, even the most fastidious ones, our laboratory has developed a culturomic strategy based on diversified culture conditions (temperature, medium, atmosphere and pH) and followed by rapid screening by Matrix Assisted Desorption Ionization - Time of flight Mass Spectrometry (MALDI-TOF MS) [17-19]. We describe here three new bacterial species using the taxonogenomic method that combines phenotypic characteristics and whole genome sequencing analysis as previously described [20, 21]. In the present study, we aimed at comparing strains Marseille-P4344, Marseille-P4200 and Marseille-P3545 to their closely related phylogenetic neighbors, and at proposing respectively the creation of new species *Clostridium cagae* sp. nov., *Clostridium rectalis* sp. nov., and *Hathewaya massiliensis* sp. nov.

Material and methods

Strains isolation

In 2016, as part of a culturomic study investigating the human microbiome, we isolated two bacterial strains from stool samples of patients in Niger with marasmus. These were strains Marseille-P3545 and Marseille-P4200. Strain Marseille-P4334 was isolated from a stool sample from an HIV-positive patient at La Timone Hospital in Marseille, France. All the patients imparted endorsed an informed accord, and the study was approved by the ethics committee of the Institut Federatif de Recherche IFR48 under number 09-022.

Initial growth was obtained for strain Marseille-P3545 after ten days in blood culture bottle at 37°C in an anaerobic atmosphere. Further, under anaerobic condition at 37°C, strains Marseille-P4200 and Marseille-P4334 had formed colonies after 3 days of preincubation in an anaerobic blood culture bottle (Thermo Scientific, Villebon sur Yvette, France) supplemented with 5 mL of 0.2 µm-filtered rumen. For each strain, the enriched liquid medium in which it was initially incubated was then inoculated on 5% Columbia agar enriched in sheep blood (BioMérieux, Marcy l'Etoile, France) followed by a new incubation at 37 ° C in an anaerobic atmosphere (Thermo Scientific, Dardilly, France) for 24 hours.

Strains identification

Strains identification was realized using MALDI-TOF MS according to the process previously reported [22-23]. A Microflex LT spectrometer (Bruker, Daltonics, Bremen, Germany) was used for this purpose. Spectra (Figure 1) were output and analyzed using the Biotyper 3.0 software against the Bruker database that was regularly incremented with the local URMS database (https://www.mediterranee-infection.com/urms-data-base). If any identification was not provided, the 16S rRNA gene was amplified using the primer pair fD1 and rP2 (Eurogentec, Angers, France) and sequenced using the Big Dye® Terminator v1.1 Cycle Sequencing Kit and 3500xLGenetic Analyzer capillary sequencer (Thermofisher, Saint-

Aubin, France), as previously described [24]. All 16S rRNA nucleotide sequences were assembled and corrected using CodonCode Aligner software (http://www.codoncode.com).

Phenotypic characterization

Growth of strains was tested under aerobic, microaerophilic and anaerobic atmospheres (Thermo Scientific, Dardilly, France). Several growth temperatures (28, 37, 45 and 55 °C) were also studied in order to know the optimal temperature of growth on 5% sheep blood-enriched Columbia agar medium (BioMérieux, Marcy l'Etoile, France). In addition, API ZYM and API 50CH strips (bioMérieux) were used to evaluate the biochemical characteristics of each strain test following the manufacturer's recommendations. Using API 50CH and API ZYM strips, the three strains were incubated for 48 and 4 hours respectively under anaerobic conditions as indicated by the manufacturer. Gram staining as well as catalase and oxidase tests were performed. Also, the spore formation tests were carried out on each strain as previously reported [25]. The morphology of these three new species was performed with an electron microscope (Hitachi High-Technologies, Tokyo, Japan) (Figure 2). A bacterial colony from each strain was collected from agar and immersed into a 2.5% glutaraldehyde fixative solution. Then, a drop of the suspension was directly deposited on a poly-L-lysine coated microscope slide for 5 minutes and treated with 1% phosphotungstic acid (PTA) aqueous solution (pH 2.0) for 2 minutes to increase scanning electron microscope (SEM) image contrast. The slide was gently washed in water, air dried and examined in a tabletop TM4000 SEM.

Genome characteristics

EZ1 biorobot with the EZ1 DNA tissue kit (Qiagen, Hilden, Germany) was used for genomic DNA extraction. Then, sequencing was performed with a MiSeq sequencer (Illumina Inc, San Diego, CA, USA) using the Nextera Mate Pair sample prep kit and Nextera XT Paired End (Illumina), as previously reported [24]. The assembly was carried out using a set of software

such as Velvet [26], Spades [27] and Soap Denovo [28]. Sequences are trimmed with MiSeq and Trimmomatic [29] software, whereas untrimmed data were processed only using MiSeq software. To decrease assembly gaps, we used GapCloser software [30]. Scaffolds that have a nucleotide number <800 (bp) and scaffolds that have a depth value lower than 25% of the mean depths were removed. The best assembly was selected using different criteria (number of scaffolds, N50, number of N).

Genomes of these three species were annotated as previously described [31]. Furthermore, the Genome-to-Genome Distance Calculator (GGDC) web server available at (http://ggdc.dsmz.de) was used to evaluate the overall similarity among the compared genomes and to replace the wet-lab DNA–DNA hybridization (DDH) by a digital DDH (dDDH) [32]. Average nucleotide identity analysis was also estimated using the OAT software [33]. Finally, the Clustage software [34] is employed to group the genomes of these three strains in order to study distribution of accessory elements within each strain.

Results

Strain identification and phylogenetic analysis

Mass spectrometry identification of single colony from each strain first cultivated on blood agar in anaerobic atmosphere at 37°C was failed. This suggested that these isolates were not referenced in the database and may be unknown species. Their MALDI-TOF MS spectra were added to our database to expand its content. 16S rDNA-based similarity analysis of strain Marseille-P4344^T, strain Marseille-P4200^T and strain Marseille-P3545^T against GenBank yielded highest nucleotide sequence similarities of 97.21% sequence identity with *Clostridium uliginosum* strain CK55 (GenBank accession no. NR_028920.1), 97.02% sequence identity with *Clostridium tetanomorphum* strain DSM 4474 (GenBank accession no. NR_043671.1) and 97.63% sequence identity with *Hathewaya histolytica* strain JCM 1403 (GenBank accession no. NR_113187.1). As these similarity values were under the threshold of 98.65%,

established to delineate new bacterial species [35-36], strain Marseille-P4344^T, strain Marseille-P4200^T and strain Marseille-P3545^T were considered as potentially new species within the family *Clostridiaceae*. The phylogenetic tree highlighting the position of these three strains relative to other closely related species with a validly published name is shown in Figure 3.

Biochemical properties of the strains

For all three strains, growth occurred only in anaerobic atmosphere at temperatures ranging from 28 to 55°C, with an optimal growth observed at 37°C. Cells from strain Marseille-P4344^T, strain Marseille-P4200^T and strain Marseille-P3545^T were strictly anaerobic, stained Gram positive and rod shaped bacilli with a mean diameter of 0.95, 0.63 and 0.60 µm, respectively. The colonies of the strains Marseille-P4344 and Marseille-P4200 have a similar appearance on blood agar after 24 hours of anaerobic growth. They are fine, translucent and non-hemolytic with a mean diameter of 1 to 2 mm. Colonies of strain Marseille-P3545 were circular, white and smooth with regular boundaries and a diameter ranging from 2 to 5 mm. They exhibited all catalase positive and oxidase negative. A biochemical comparison was performed between these three new strains and other closest species (Table 1).

Using an API ZYM and API 50 CH (bioMérieux), the three strains tested were positive for acid phosphatase, naphthol-AS-BI-phosphohydrolase, arbutin, D-melezitose, D-turanose, D-lyxose, D-tagalose, D-fucose, L-fucose, D-arabitol, L-arabitol and potassium 5-ketogluconate. In addition, both strains Marseille-P4200^T and Marseille-P3545^T were also positive for methyl α-D-mannopyranoside, methyl α-D-glucopyranoside, N-acetyl-glucosamine, amygdalin, salicin, D-cellobiose, D-melibiose, sucrose, D-trehalose, starch and glycogen while D-glucose, D-mannitol and inulin were fermented by both strains Marseille-

P4344 and Marseille-P3545^{T}. The remaining negative tests were reported in supplementary tables S1 and S2.

Genomic analysis

The genomes of the strains Marseille-P4344 and Marseille-P4200 were 3,738,409 and 3,279,426 bp long with 36.6 and 36.6 mol% G+C content respectively. The assembly was carried out into 10 scaffolds for Marseille-P4344 and into 30 scaffolds for Marseille-P4200. They had 3,480 and 3,233 predicted genes respectively. Their genome also contained 3,299 and 3,083 protein-coding genes and 107 and 90 as RNAs genes (26 and 23 rRNAs, 77 and 63 tRNAs, 4 and 4 ncRNAs) and 74 and 60 pseudo genes. Genes with putative function (by COGs) were 2618 for strain Marseille-P4344 (72%) and 2,592 for strain Marseille-P4200 (74%). Finally, 993 and 916 genes (28% and 26%) were annotated as hypothetical proteins for strain Marseille-P4344 and strain Marseille-P4200, respectively. In contrast, the genome size of *Hathewaya massiliensis* strain Marseille-P3545 was 3,058,214 bp with 29.4 mol% G+C content. The genome assembly of this strain was achieved on 10 contigs. Of the 2,911 predicted genes, 2,738 were protein-coding genes and 112 were RNAs (9 5S rRNA, 7 16S rRNA, 7 23S rRNA, 85 tRNA and 4 ncRNA genes) and 61 pseudo genes. More statistical data from the genomes are listed in Table 2. The properties and distribution of accessory genes into the pan-genome of these strains showed that both strain Marseille-P4344 and strain Marseille-P4200 shared several genes, while *Hathewaya massiliensis* displays a less wide distribution than that of *Clostridium* species (Figure 4).

Analysis of the Clusters of Orthologous Groups (COGs) categories shows that the mobile elements of the *H. massiliensis* genome appear to be more numerous than those of the genomes of other Clostridium species (32, 35 and 88 in category [X], respectively). In contrast, Marseille-P4344 and Marseille-P4200 strains exhibit higher ability to transport and

metabolize carbohydrates than Marseille-P3545, with 223, 147 and 97 genes associated to category [G] (Table 3, Figure 5).

Using dDDH analysis, values ranged from 18% between *C. tepidiprofundi* and *C. lundense,* to 27.6% between *C. lundense* and *C. liquoris* (Table 4). These values are lower than the 70% threshold used for delineating prokaryotic species, thus confirming that these three strains represent new species. Finally, OrthoANI analysis among closely related species (Figure 6) show that the higher percentage value was 83.56% shared between *C. liquoris* and *C. lundense,* while the lowest was 67.73% between *H. histolytica* and *C. chromiireducens*. In addition, when *C. rectalis* is compared with other species, the values ranged from 69.09% with *C. cagae* to 75.56% with *C. tetanomorphum*. For *C. cagae*, values ranged from 68.46% with *H. histolytica* to 79.45% with *C. uliginosum* and for *H. massiliensis* they ranged from 67.86% with *C. chromiireducens* to 78.27% with *H. histolytica*.

Conclusion

Based on all the phenotypic, biochemical and genomic tests performed on these bacterial strains, we consequently considered that strains Marseille-P4200, Marseille-P3545 and Maseille-P4344 are new species. Indeed, the taxono-genomic evidences brought during this study, such as the sequence similarity of the 16S rRNA gene below the threshold value of 98.65%, OrthoANI values also <95%, as well as the several phenotypic divergences obtained, have allowed to formally propose *Clostrium cagae* sp. nov., *Clostridium rectalis* sp. nov., and *Hathewaya massiliensis* sp. nov., as new species within the phylum *Firmicutes*.

Description of *Clostridium cagae* sp. nov.

Strain Marseille-P4344^T is the type strain of *'Clostridium cagae* sp. nov. (ca.ga.e, Gr.n. pronounced as follow kakke, refered to 'faeces' the clinical sample from which this bacterium was isolated). *Clostridium cagae* is a strictly anaerobic, spore forming and Gram-positive rod bacterium with a mean diameter of 0.95 µm. It exhibited catalase positive and oxidase

negative activities. Strain Marseille-P4344^T grows under anaerobic conditions at temperatures ranging between 28 and 55°C, with an optimal temperature of 37°C. Colonies of strain Marseille-P4344 are fine, translucent and non-hemolytic with a mean diameter of 1 to 2 mm. The genome of strain Marseille-P4344^T was 3,738,409 bp with 36.6 mol% of G+C content. Strain Marseille-P4344 is able to ferment D-arabitol, D-melezitose, D-lyxose, L-arabitol, D-turanose, D-tagalose, L-fucose, D-fucose, and potassium 5-ketogluconate. The type strain of *Clostridium cagae* sp. nov., strain Marseille-P4344 was isolated from the stool sample of patient with HIV.

Description of *Clostridium rectalis* sp. nov.

Strain Marseille-P4200^T is the type strain of '*Clostridium rectaclis* sp. nov. (rec.ta.lis, N.L. masc. adj. rectalis to rectal refered to rectum the straight bowel from which this bacterium was isolated). *Clostridium rectalis* is an anaerobic bacterium. It can form spores and is Gram-positive bacilli with a mean diameter of 0.63 µm. It exhibited catalase positive and oxidase negative activities. Strain Marseille-P4200^T grows under anaerobic conditions at 37°C. Strain Marseille-P4200 presents translucent, fine and non-hemolytic colonies with a mean diameter of 2 mm. The genome of strain Marseille-P4200^T was 3,279,426 bp with 36.6 mol% of G+C content. Strain Marseille-P4200 is positive for acid phosphatase, naphthol-AS-BI-phosphohydrolase, methyl αD-mannopyranoside, methyl αD-glucopyranoside, N-acetyl-glucosamine, amygdalin, salicin, D-cellobiose, D-melibiose, sucrose, D-trehalose, and glycogen. The type strain of *Clostridium rectalis* sp. nov., strain Marseille-P4200 was isolated from the stool sample of a child with marasmus.

Description of *Hathewaya massiliensis* sp. nov.

'*Hathewaya massiliensis*' sp. nov. (mas.si.li.en'sis N.L. fem. adj. *massiliensis*, to Massilia, the Latin name of Marseille where the type strain was first isolated and characterized) is classified as a member of the family *Clostridiaceae* in the phylum *Firmicutes*. Strain

Marseille-P3545^T is the type strain of the new species *Hathewaya massiliensis* sp. nov. It is an anaerobic Gram-positive rod shaped bacterium, spore-forming and motile. Colonies of strain Marseille-P3545^T observed on blood agar medium are circular, white and smooth with regular edges and a diameter of 3 mm. This bacterial strain possesses catalase positive and oxidase negative. The genome size of *Hathewaya massiliensis* strain Marseille-P3545^T is 3,058,214 bp with 29.4 mol% G+C content. The Genbank Accession number for the 16S rRNA gene sequence of strain Marseille-P3545^T is LT797537 and for the whole genome shotgun project is CABFVD000000000. It was isolated from a stool sample of young patient suffering from malnutrition.

Acknowledgements

The authors thank Aurelia Caputo for submitting the genomic sequence to GenBank and Amael Fadlane for subculturing the strains.

Conflicts of interest:

None to declare

Funding sources:

This study was supported by the Institut Hospitalo-Universitaire (IHU) Méditerranée Infection, the National Research Agency under the program « Investissements d'avenir », reference ANR-10-IAHU-03, the Région Provence-Alpes-Côte d'Azur and European funding FEDER PRIMI. This research was supported by a grant from the Institut Universitaire de France (IUF, Paris, France) to Professor Anthony Levasseur.

Ethics and consent

The study was approved and authorized by the ethics committee of the Institut Hospitalo-Universitaire Méditerranée Infection (IHU-MI) under reference 2016-010.

References

1. Cruz-Morales P, Orellana CA, Moutafis G, et al. Revisiting the Evolution and Taxonomy of *Clostridia*, a Phylogenomic Update. Genome Biol Evol. 2019; 11(7):2035–2044. doi:10.1093/gbe/evz096
2. Gerritsen J, Fuentes S, Grievink W, et al. Characterization of *Romboutsia ilealis* gen. nov., sp. nov., isolated from the gastro-intestinal tract of a rat, and proposal for the reclassification of five closely related members of the genus *Clostridium* into the genera *Romboutsia* gen. nov., *Intestinibacter* gen. nov., *Terrisporobacter* gen. nov. and *Asaccharospora* gen. nov. Int J Syst Evol Microbiol. 2014; 64 (Pt 5):1600–1616. doi:10.1099/ijs.0.059543-0.
3. Kaur S, Yawar M, Kumar PA, Suresh K. *Hungatella effluvii* gen. nov., sp. nov., an obligately anaerobic bacterium isolated from an effluent treatment plant, and reclassification of *Clostridium hathewayi* as *Hungatella hathewayi* gen. nov., comb. nov. Int J Syst Evol Microbiol. 2014; 64(Pt 3):710–718. doi:10.1099/ijs.0.056986-0
4. Prazmowski A. Untersuchung über die Entwickelungsgeschichte und Fermentwirkung einiger Bacterien-Arten. Inaugural Dissertation. Hugo Voigt, Leipzig, 1880.
5. Horino H, Ito M, Tonouchi A. *Clostridium oryzae* sp. nov., from soil of a Japanese rice field. Int J Syst Evol Microbiol. 2015; 65(Pt 3):943–951. doi:10.1099/ijs.0.000042
6. Shin Y, Kang SS, Paek J, et al. *Clostridium kogasensis* sp. nov., a novel member of the genus *Clostridium*, isolated from soil under a corroded gas pipeline. Anaerobe. 2016; 39:14–18. doi:10.1016/j.anaerobe.2016.02.006
7. Sankar SA, Rathored J, Metidji S, et al. *Clostridium polynesiense* sp. nov., a new member of the human gut microbiota in French Polynesia. Anaerobe. 2015; 36:79–87. doi:10.1016/j.anaerobe.2015.10.004

8. Wu YF, Zheng H, Wu QL, Yang H, Liu SJ. *Clostridium algifaecis* sp. nov., an anaerobic bacterial species from decomposing algal scum. Int J Syst Evol Microbiol. 2014; 64(Pt 11):3844–3848. doi:10.1099/ijs.0.064345-0
9. Keis S, Shaheen R, Jones DT. Emended descriptions of *Clostridium acetobutylicum* and *Clostridium beijerinckii*, and descriptions of *Clostridium saccharoperbutylacetonicum* sp. nov. and *Clostridium saccharobutylicum* sp. nov. Int J Syst Evol Microbiol. 2001; 51(Pt 6):2095–2103. doi:10.1099/00207713-51-6-2095
10. Srivastava N, Srivastava M, Kushwaha D, et al. Efficient dark fermentative hydrogen production from enzyme hydrolyzed rice straw by *Clostridium pasteurianum* (MTCC116). Bioresour Technol. 2017; 238:552–558. doi:10.1016/j.biortech.2017.04.077
11. Cassir N, Fahsi N, Durand G, Lagier JC, Raoult D, Fournier PE. Emergence of *Clostridium difficile* tcdC variant 078 in Marseille, France. Eur J Clin Microbiol Infect Dis. 2017;36(10):1971–1974. doi:10.1007/s10096-017-3022-8
12. Czepiel J, Dróżdż M, Pituch H, et al. *Clostridium difficile* infection: review. Eur J Clin Microbiol Infect Dis. 2019; 38(7):1211–1221. doi:10.1007/s10096-019-03539-6
13. Dembek ZF, Smith LA, Rusnak JM. Botulism: cause, effects, diagnosis, clinical and laboratory identification, and treatment modalities. Disaster Med Public Health Prep. 2007; 1(2):122–134. doi:10.1097/DMP.0b013e318158c5fd
14. Rhinesmith E, Fu L. Tetanus Disease, Treatment, Management. Pediatr Rev. 2018; 39(8):430–432. doi:10.1542/pir.2017-0238
15. Shrestha A, Uzal FA, McClane BA. Enterotoxic Clostridia: *Clostridium perfringens* Enteric Diseases. Microbiol Spectr. 2018; 6(5):10.1128/microbiolspec.GPP3-0003-2017. doi:10.1128/microbiolspec.GPP3-0003-2017

16. Lawson PA, Rainey FA. Proposal to restrict the genus *Clostridium* Prazmowski to *Clostridium butyricum* and related species. Int J Syst Evol Microbiol. 2016;66 (2):1009–1016. doi:10.1099/ijsem.0.000824

17. Lagier JC, Armougom F, Million M, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect. 2012;18 (12):1185–1193. doi:10.1111/1469-0691.12023

18. Lagier JC, Khelaifia S, Alou MT, et al. Culture of previously uncultured members of the human gut microbiota by culturomics. Nat Microbiol. 2016;1:16203. doi:10.1038/nmicrobiol.2016.203

19. Lagier JC, Hugon P, Khelaifia S, Fournier 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. 2015; 28 (1):237–264. doi:10.1128/CMR.00014-14

20. Fournier PE, Drancourt M. New Microbes New Infections promotes modern prokaryotic taxonomy: a new section "TaxonoGenomics: new genomes of microorganisms in humans". New Microbes New Infect. 2015;7:48–49. doi:10.1016/j.nmni.2015.06.001

21. Ramasamy D, Mishra AK, Lagier JC, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol. 2014;64(Pt 2):384–391. doi:10.1099/ijs.0.057091-0

22. Lo CI, Fall B, Sambe-Ba B, Diawara S, Gueye MW, Mediannikov O, et al. MALDI-TOF Mass Spectrometry: A Powerful Tool for Clinical Microbiology at Hôpital Principal de Dakar, Senegal (West Africa). PLoS One. 2015; 30:10-12. doi: 10.1371/journal.pone.0145889.

23. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier P-E, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser

desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 2009;49:543–51. doi:10.1086/600885.

24. Morel A-S, Dubourg G, Prudent E, Edouard S, Gouriet F, Casalta JP, et al. Complementarity between targeted real-time specific PCR and conventional broad-range 16S rDNA PCR in the syndrome-driven diagnosis of infectious diseases. Eur J Clin Microbiol Infect Dis 2015;34:561–70. doi:10.1007/s10096-014-2263-z.
25. Gary P. Wormser, Charles Stratton, Manual of Clinical Microbiology, 9th Edition Edited by Patrick R. Murray, Ellen Jo Baron, James H. Jorgensen, Marie Louise Landry, and Michael A. Pfaller Washington, DC: ASM Press, 2007 2488 pp., illustrated. Clinical Infectious Diseases, Volume 46, Issue 1, 1 January 2008, Page 153, https://doi.org/10.1086/524076 18.
26. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 2008; 18:821–9. doi:10.1101/gr.074492.107.
27. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 2012; 19:455–77. doi:10.1089/cmb.2012.0021.
28. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012; 1:18. doi:10.1186/2047-217X-1-18.
29. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014; 30:2114–20. doi:10.1093/bioinformatics/btu170.
30. Xu GC, Xu TJ, Zhu R, et al. LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. Gigascience. 2019; 8(1):giy157. doi:10.1093/gigascience/giy157

31. Lo CI, Sankar SA, Fall B, et al. High-quality draft genome sequence and description of *Haemophilus massiliensis* sp. nov. Stand Genomic Sci. 2016; 11:31. doi:10.1186/s40793-016-0150-1

32. 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. doi:10.1186/1471-2105-14-60

33. Lee I, Ouk Kim Y, Park SC, Chun J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol 2016; 66:1100–3. doi:10.1099/ijsem.0.000760.

34. Ozer EA. ClustAGE: a tool for clustering and distribution analysis of bacterial accessory genomic elements. BMC Bioinformatics. 2018; 19(1):150. doi:10.1186/s12859-018-2154-x

35. 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. doi:10.1186/1471-2105-14-60 9.

36. Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes [published correction appears in Int J Syst Evol Microbiol. 2014 May;64(Pt 5):1825]. Int J Syst Evol Microbiol. 2014; 64(Pt 2):346–351. doi:10.1099/ijs.0.059774-0

Table 1: Different phenotypic characteristics between the studied strains and their closest relative species. **1**, *Clostridium cagae* strain Marseille-P4344; **2**, *Clostridium uliginosum*; **3**, *Clostridium rectalis* strain Marseille-P4200; **4**, *Clostridium tetanomorphum* ; **5**, *Clostridium liquoris*; **6**, *Clostridium polynesiense*; **7**, *Hathewaya massiliensis* strain Marseille P-3545; **8**, *Hathewaya histolytica* and **9**, *Hathewaya limosa* . NF, not found; +, positive test; -, negative test.

Properties	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
Cell diameter (µm)	0.8-1.1	1.0	0.5-0.6	0.7-1.0	0.5-1.0	0.62	0.5-0.6	0.5–0.9	0.6–1.6
Oxygen requirement	Strictly anaerobic	Anaerobic	Strictly anaerobic	Strictly anaerobic	Strictly anaerobic	Obligately anaerobic	Strictly anaerobic	Strictly anaerobic	Strictly anaerobic
Gram stain	+	-	+	+	+	+	+	+	+
Motility	-	+	-	+	-	+	+	+	+
Endospore formation	+	+	+	+	+	+	+	+	+
Alkaline phosphatase	-	NF		NF	NF	+	+	+	NF
Catalase	+	NF	+	-	-	-	+	NF	NF
Oxidase	-	NF	-	NF	NF	-	-	NF	NF
β-Galactosidase	-	NF	-	NF	NF	+	-	NF	NF
N-acetyl-glucosamine	-	NF	-		NF	+	-	NF	NF
Arabinose	-	-	-	-	NF	-	+	NF	NF
Esterase lipase (C8)	-	NF	-	+	NF	NF	-	NF	NF
Mannose	-	+	-	-	weakly +	+	+	NF	NF
Mannitol	+	+	-	NF	+	+	+	NF	NF
Sucrose	-	+	+	-	NF	+	+	NF	NF
D-Glucose	+	+	-	+	+	+	+	NF	+
D-Fructose	-	NF	-	+	NF	+	+	NF	NF
D-Maltose	-	-	-	+	+	+	+	NF	NF
G+C (mol%)	27.3	28.0	28.1	27.5	34.4	34.0	29.5	NF	24.0
Source	Stool	Acidic forest bog	Stool	Septic wounds	Old fermentation pit	Fecal flora	Stool	Soil, Human	Human, Animal

Table 2: Statistical details from genome of *Clostridium cagae* strain Marseille-P4344, *Clostridium rectalis* strain Marseille-P4200 and *Hathewaya strain massiliensis* Marseille-P3545.

Characteristics	*Strain Marseille-P4200*	*Strain Marseille-P3545*	*Strain Marseille-P4344*
GenBank accession numbers	UYZZ00000000	CABFVD000000000	OKRA00000000
Size (bp)	3,279,426	3,058,214	3,738,409
RNAs	90	112	107
Protein-coding genes	3,083	2,738	3,299
Genes	3,233	2,911	3,480
G + C content (%)	36.6	29.4	36.6

Table 3 : Number of genes associated with the 25 general clusters of orthologous group (COG) functional categories of *Clostridium cagae* strain Marseille- P4344, *Clostridium rectalis* strain Marseille-P4200 and *Hathewaya massiliensis* strain Marseille-P3545

Code	***strain Marseille-P4344***	***strain Marseille-P3545***	***strain Marseille-P4200***	***Description***
[J]	224	218	209	Translation, ribosomal structure and biogenesis
[A]	1	0	0	RNA processing and modification
[K]	245	175	220	Transcription
[L]	118	128	130	Replication, recombination and repair
[B]	1	1	1	Chromatin structure and dynamics
[D]	44	40	47	Cell cycle control, cell division, chromosome partitioning
[Y]	0	0	0	Nuclear structure
[V]	72	88	94	Defense mechanisms
[T]	202	146	187	Signal transduction mechanisms
[M]	162	143	162	Cell wall/membrane/envelope biogenesis
[N]	79	60	74	Cell motility
[Z]	0	0	2	Cytoskeleton
[W]	9	4	10	Extracellular structures
[U]	33	25	23	Intracellular trafficking, secretion, and vesicular transport
[O]	106	111	97	Posttranslational modification, protein turnover, chaperones
[X]	32	88	35	Mobilome: prophages, transposons
[C]	164	114	144	Energy production and conversion
[G]	223	94	147	Carbohydrate transport and metabolism
[E]	194	191	195	Amino acid transport and metabolism
[F]	102	76	83	Nucleotide transport and metabolism
[H]	135	138	105	Coenzyme transport and metabolism
[I]	92	65	78	Lipid transport and metabolism
[P]	135	117	131	Inorganic ion transport and metabolism
[Q]	44	25	26	Secondary metabolites biosynthesis, transport and catabolism
[R]	261	218	243	General function prediction only
[S]	167	121	149	Function unknown
-	993	786	916	Not in cog

Table 4: Numerical DNA-DNA hybridization values (%) obtained by comparison between *Clostridium rectalis* strain Marseille-P4200, *Hathewaya massiliensis* strain Marseille-P3545, *Clostridium cagae* strain Marseille-P4344 and other closely related species using GGDC formula 2 software (DDH estimates based on HSP identities / length) [https://ggdc.dsmz.de/ggdc.php#], top right.

	(Hh)	*(Cu)*	*(Ct)*	*(Cte)*	***(Cre)***	*(Clu)*	*(Clq)*	***(Hm)***	*(Cch)*	***(Cca)***
H. histolytica (Hh)	100	23.1	21.0	22.5	21.3	20.1	22.7	20.7	23.3	23.0
C. uliginosum (Cu)		100	23.5	20.1	19.8	21.5	21.4	24.3	23.6	24.9
C. tetanomorphum (Ct)			100	19.8	19.7	27.5	24.2	19.4	21.9	19.2
C. tepidiprofundi (Cte)				100	19.0	18.0	21.7	21.9	22.2	21.4
C. rectalis (Cre)					100	19.9	19.3	21.1	21.2	19.5
C. lundense (Clu)						100	27.6	21.8	21.3	19.3
C. liquoris (Clq)							100	23.2	22.2	19.2
H. massiliensis (Hm)								100	23.3	20.3
C. chromiireducens (Cch)									100	22.0
C. cagae (Cca)										100

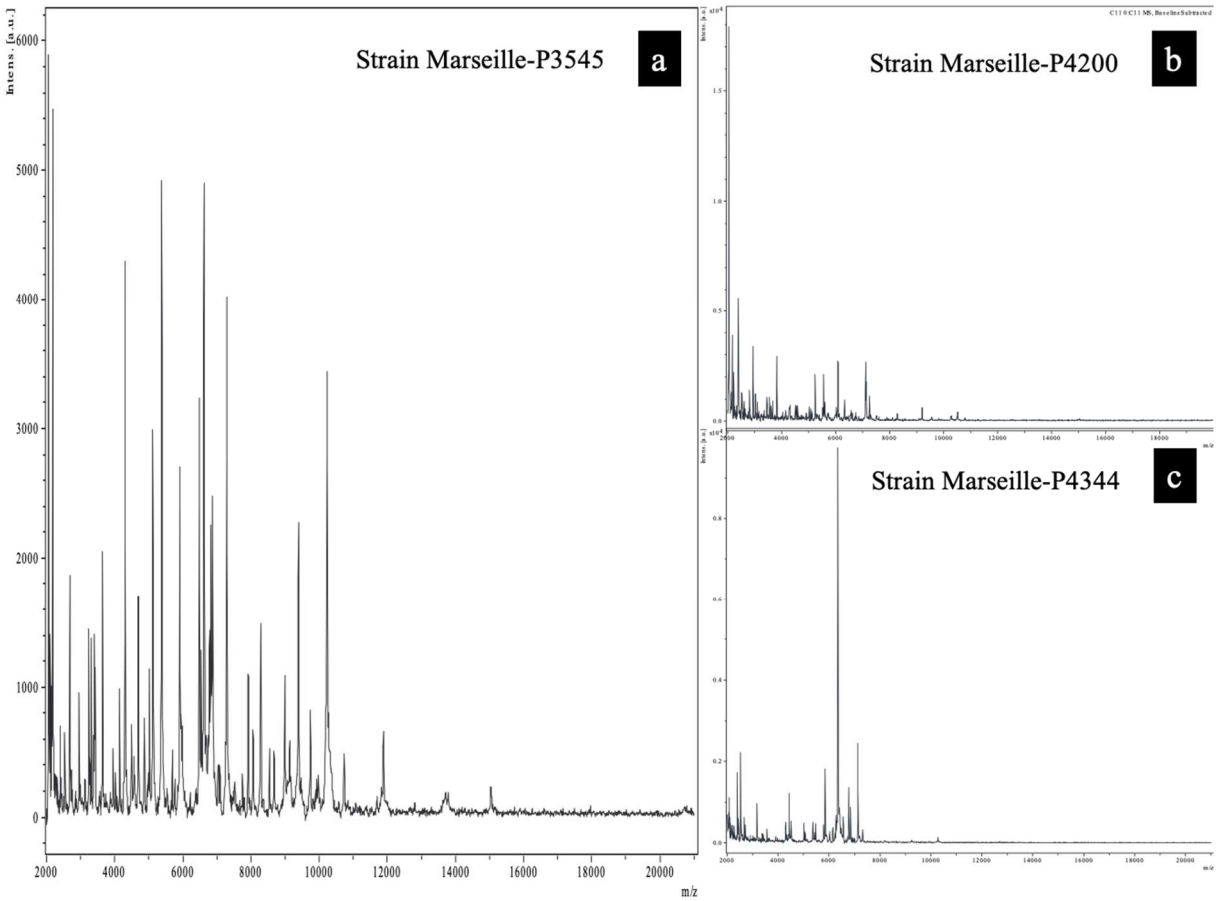

Figure 1: MALDI-TOF MS reference spectrum of a)*Hathewaya massiliensis* sp. nov., b) *Clostridium rectalis* sp. nov., and c) *Clostridium cagae* sp. nov.,. The reference spectrum was generated by comparison of spectra from 12 individual colonies.

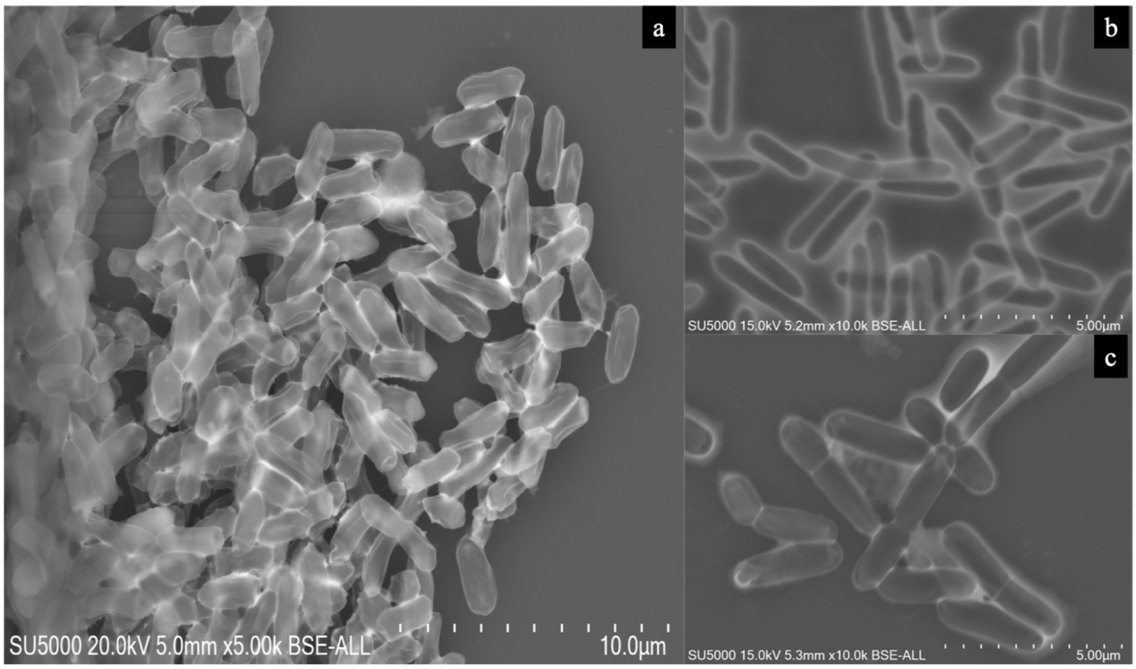

Figure 2: Scanning electron microscopy of stained a) *Clostridium cagae* sp. nov., b) *Hathewaya massiliensis* sp. nov., and c) *Clostridium rectalis* sp. nov. A colony was collected from agar and immersed into a 2.5 % glutaraldehyde fixative solution. Then, a drop of the suspension was directly deposited on a poly-L-lysine coated microscope slide for 5 minutes and treated with 1 % phosphotungstic acid (PTA) aqueous solution (pH 2.0) for 2 minutes to increase SEM image contrast. The slide was gently washed in water; air dried and examined in a tabletop SEM (Hitachi SU5000) with approximately 60 centimeters in height and 33 cm in width to evaluate bacteria structure. Scales and acquisition settings are shown in figures.

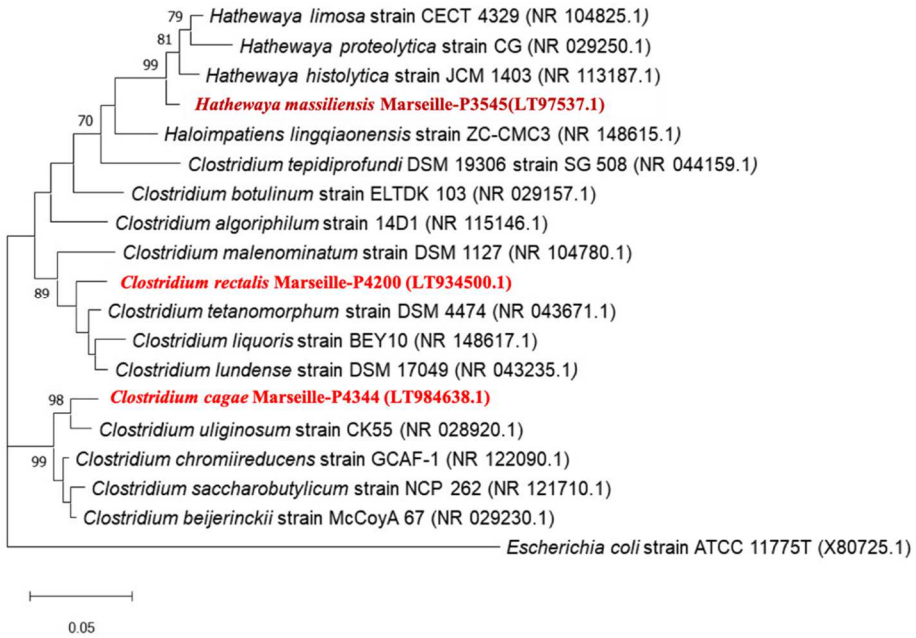

Figure 3: Phylogenetic tree highlighting the position of *Clostridium rectalis* sp. nov., *Clostridium cagae* sp. nov., and *Hathewaya massiliensis* sp. nov., with regard to other closely related species. Genbank accession numbers of 16S rRNA are indicated in parentheses. Bootstrap values obtained by repeating the analysis 1,000 times to generate a majority consensus tree are indicated at the nodes. The scale bar indicates a 5 % nucleotide sequence divergence.

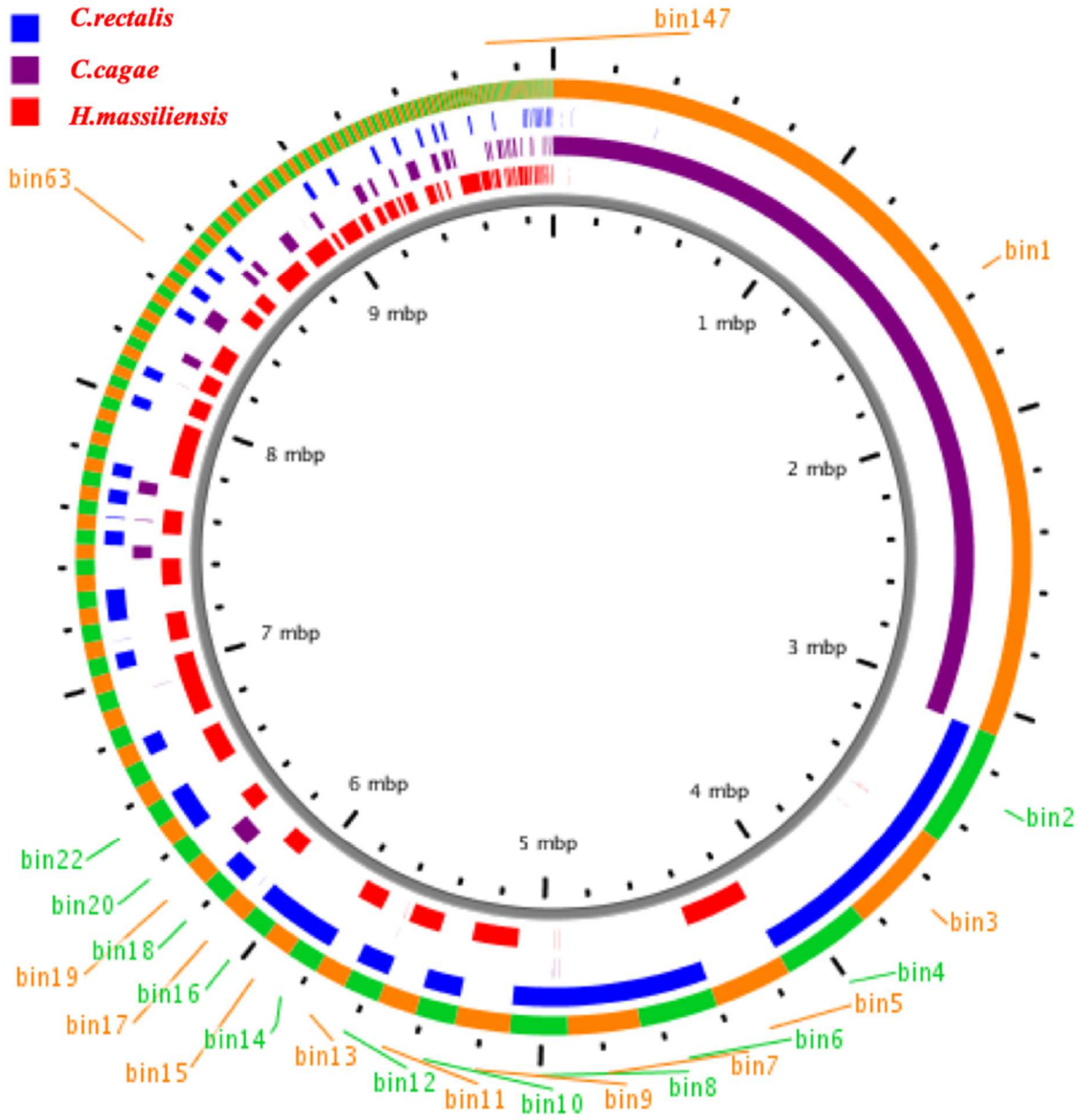

Figure 4: Pan accessory genome distribution among *Clostridium rectalis* sp. nov., *Clostridium cagae* sp. nov., and *Hathewaya massiliensis* sp. nov. The outer ring shows the ClustAGE trays, classified from the largest to the smallest, with alternating orange and green colours indicating the boundaries of the trays. The concentric inner rings show the distribution of accessory elements within each strain. The ruler in the centre of the figure shows the cumulative size of the representatives of the accessory trays of the genome.

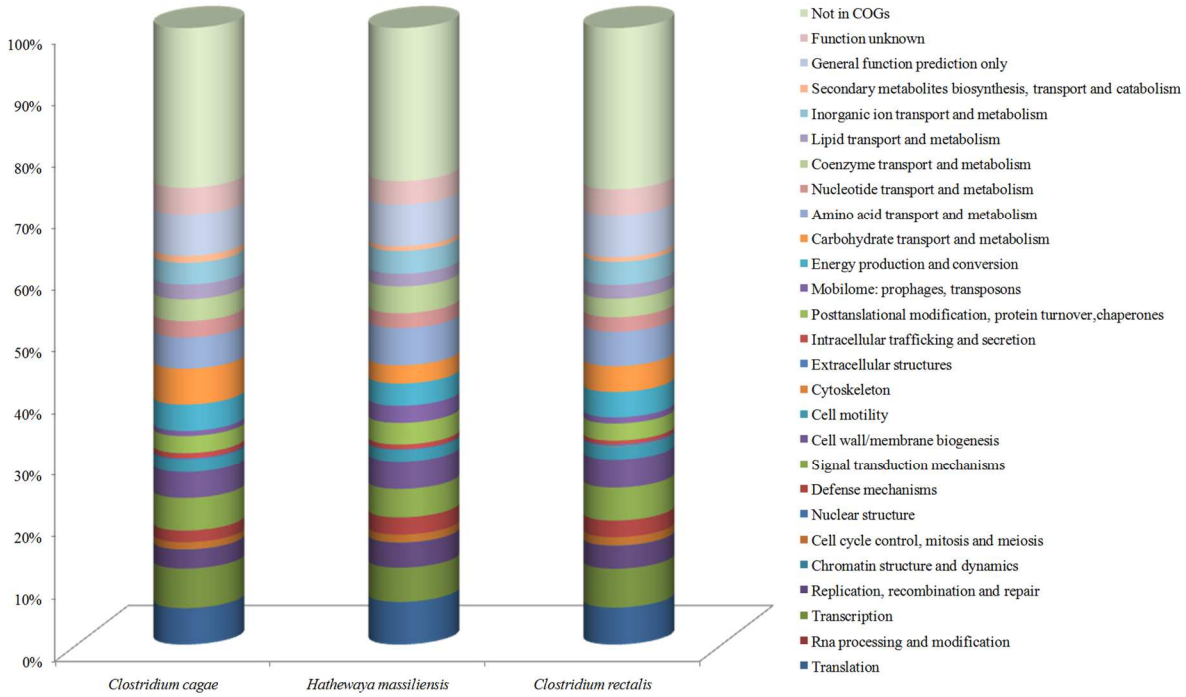

Figure 5: Distribution of functional classes of the predicted genes in *Clostridium cagae*, *Clostridium rectalis* and *Hathewaya massiliensis* chromosomes according to the clusters of orthologous groups of proteins.

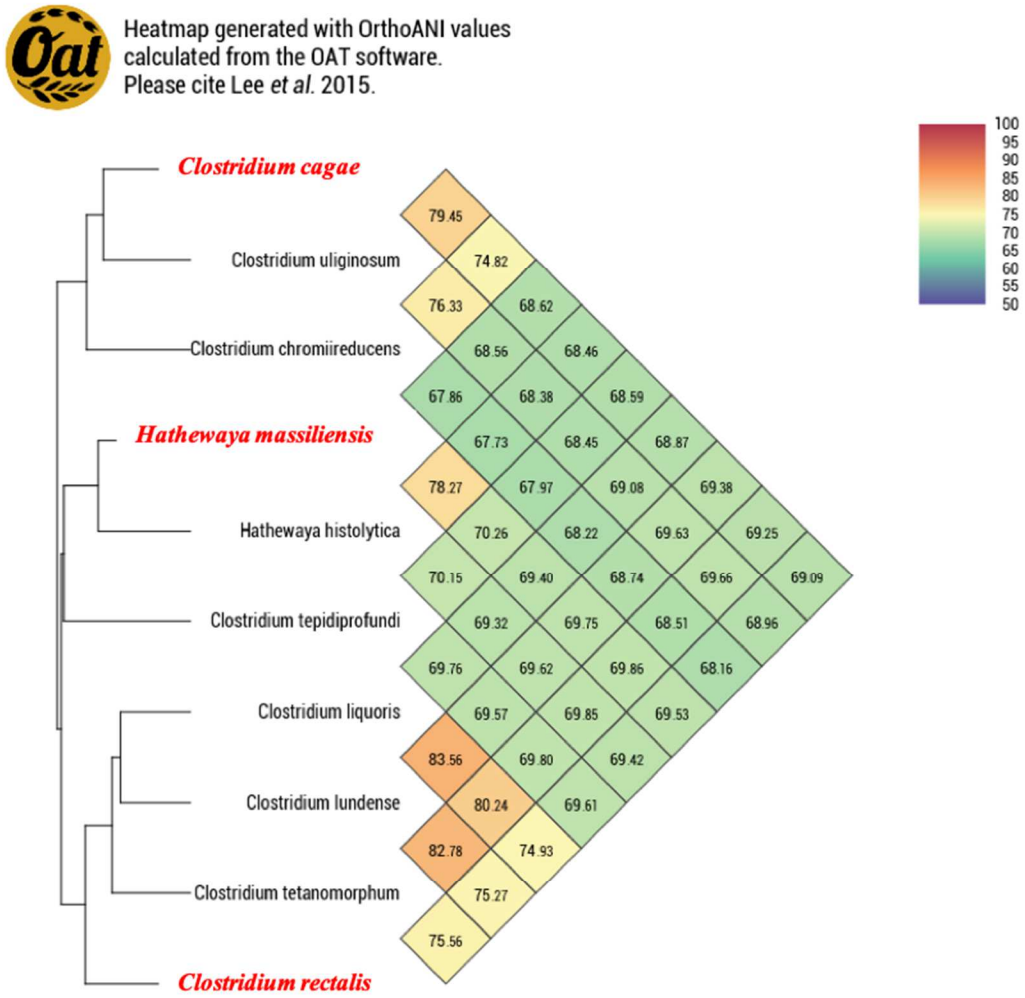

Figure 6: Heatmap generated with OrthoANI values calculated using the OAT software between *Clostridium cagae* sp. nov., *Clostridium rectalis* sp. nov., and *Hathewaya massiliensis* with other closely related species validly described.