Clostridium pacaense: a new species within the genus Clostridium

M. Hosny, R. Abou Abdallah, J. Bou Khalil, A. Fontanini, E. Baptiste, N. Armstrong, B. La Scola

Clostridium pacaense: a new species within the genus *Clostridium*

Michel Hosny[1†], Rita Abou Abdallah[2†], Jacques Bou Khalil[1], Anthony Fontanini[1], Emeline Baptiste[1], Nicholas Armstrong[1] and Bernard La Scola[1]*.

[1]Aix-Marseille Université UM63, Institut de Recherche pour le Développement IRD 198, Assistance Publique – Hôpitaux de Marseille (AP-HM), Microbes, Evolution, Phylogeny and Infection (MEΦI), Institut Hospitalo-Universitaire (IHU) - Méditerranée Infection, France.

[2]Aix-Marseille Université UM63, Institut de Recherche pour le Développement IRD 198, Assistance Publique – Hôpitaux de Marseille (AP-HM), Vecteurs – Infections TROpicales et MÉditerrannéennes (VITROME), Service de Santé des Armées, Institut Hospitalo-Universitaire (IHU) - Méditerranée-Infection, France.

† contributed equally to this article

* **Correspondence:** Bernard La Scola, MD, PhD, Pôle des Maladies Infectieuses, Aix-Marseille Université, IRD, Assistance Publique – Hôpitaux de Marseille (AP-HM), Microbes, Evolution, Phylogeny and Infection (MEΦI), Institut Hospitalo-Universitaire (IHU) - Méditerranée Infection, France, 19-21 Boulevard Jean Moulin, 13385 Marseille Cedex 05, France. Telephone: +33 4 91 32 43 75, Fax: +33 4 91 38 77 72, E-mail: bernard.la-scola@univ-amu.fr

Running title: *Clostridium pacaense*

Keywords: Culturomics, taxono-genomics, *Clostridium pacaense*

Clostridium pacaense*: a new species within the genus *Clostridium

Running title: *Clostridium pacaense*

Keywords: Culturomics, taxono-genomics, *Clostridium pacaense*

Abstract

Using the taxonogenomic strategy, we described *Clostridium pacaense* sp. nov. strain Marseille-P3100^{T}, a Gram-variable, non-motile, spore-forming anaerobic bacillus. This strain was isolated from a 3.3-months-old Senegalese girl, with clinical aspects of marasmus. The closest species based on 16S rRNA was *Clostridium aldenense* with a similarity of 98.4%. The genome length was 2,672,129 bp, with a 50% of GC content and 2,360 proteins were predicted. Finally, predominant fatty acids were Hexadecanoic acid, Tetradecanoic acid, 9-Hexadecenoic acid.

1. Introduction

Human intestinal flora is incorporated mainly in the terminal part of small intestine and colon. It consists of about 100,000 billion bacteria, grouped into 500 species including 90% of anaerobic bacteria [1,2]. Oxygen-tolerant species such as lactobacilli, thus aerobic organisms such as *E. coli* and enterococci represent minorities within intestinal microbiota [3]. It appears that each adult own a signature of microbial community, and increasingly influenced on human health [4–6]. Clostridiaceae is a family of Clostridia and has traditionally been described by anaerobic growth and spore formation [4,7]. Clostridia included major composition of mammalian gastrointestinal tract microbiomes [8]. For that purpose, culturomics combined with taxonogenomics revealed main tools for the isolation and characterization of new bacterial species. These techniques allowed the accessibility to study their phenotypes like antibiotics resistance and biochemical features, thus genomics characteristics and the potential impact on human health [9,10]. In the present work, we propose *Clostridium pacaense* sp. nov. strain Marseille-P3100^{T} (CSUR P3100) as a new species within the *Clostridium* genus. This strain was isolated from a 3.3-months-old Senegalese girl, with clinical aspects of marasmus [11].

2. Materials and Methods

2.1. Phenotypic, biochemical and antibiotics susceptibility

Gram staining, motility, catalase and oxidase were determined as described by Lagier et al. [12]. Sporulation was tested using a thermal shock on bacterial colonies (diluted in phosphate buffered saline) for 20 minutes at 80°C. For electronic microscopy, a colony was collected from agar and immersed into a 2.5 % glutaraldehyde fixative solution. The slide was gently washed in water; air dried and examined with approximately 60 centimeters in height and 33 cm in width to evaluate bacteria structure on TM4000 microscope. Mass spectrum was obtained from *C. pacaense* colonies using MALDI-TOF (Figure 1). Biochemical characteristics were tested using API 50CH, API ZYM and API 20A strips (bioMérieux, Marcy l'Etoile, France). Antibiotic susceptibility referred to European Committee on Antimicrobial Susceptibility Testing 2018 recommendations.

2.2. Fatty acid methyl ester analysis

Cellular fatty acid methyl ester (FAME) analysis was performed by GC/MS. Two samples were prepared with approximately 35 mg of bacterial biomass per tube harvested from several culture plates. Fatty acid methyl esters were prepared as described previously [13]. GC/MS analyses were carried out as described before [14]. Briefly, fatty acid methyl esters were separated using an Elite 5-MS column and monitored by mass spectrometry (Clarus 500 - SQ 8 S, Perkin Elmer, Courtaboeuf, France). Spectral database search was performed using MS Search 2.0 operated with the Standard Reference Database 1A (NIST, Gaithersburg, USA) and the FAMEs mass spectral database (Wiley, Chichester, UK).

2.3. Genome sequencing, assembly and annotation

Genomic DNA was sequenced on MiSeq sequencer (Illumina, Inc., San Diego, CA) using paired-end strategy as described previously [7]. Spades software was used for genome assembly [15]. Contaminations were eliminated after performing BLASTn. Open reading

frames (ORFs) were predicted and annotated using Prokka software [16]. *C. pacaense* genome was for protein functions against the Clusters of Orthologous Groups (COG) database using BLASTP (E-value of $1e^{-03}$, coverage 0,7 and identity percent 30%). The genome is available on EMBL-EBI scaffolds accession numbers: LS999944-LS999965.

2.4. Comparative genomics

Species to be compared were those with higher similarity based on 16s RNA (Figure 2), provided the genome is available. The following bacterial species were used in this analysis and their genomics features are summarized in supplementary Table S1: *Clostridium bolteae* (GCA_002234575.2), *Clostridium lavalense* (GCA_003024655.1), *Clostridium saccharolyticum* (GCA_000144625.1), *Clostridium aldenense* (GCA_003434055.1), *Lachnoclostridium citroniae* (GCA_000233455.1), *Clostridium amygdalinum* (GCA_900205965.1), *Clostridium celerecrescens* (GCA_000732605.1). Amino-acids and ORF sequences were predicted using Prodigal software to obtain optimized prediction within all genomes [17]. Then, for each couple of genomes, a similarity percentage was computed using the OrthoANI software [18].

3. Results

3.1. Phenotypic and biochemical characterization

C. pacaense is a Gram-variable, spore-forming, non-motile, anaerobic bacillus, with no catalase and oxidase activities. Electron microscopy revealed that its length was 3.5 μm and 0.5 μm of diameter (Figure 3). *C. pacanese* produced α-glucosidase and Napthol-AS-Bl-phosphohydrolase. General feature and biochemical characteristics are summarized in Table 1. Antibiotic susceptibility testing revealed that *C. pacaense* was susceptible to amoxicillin, amoxicillin–clavulanic acid, ceftriaxone, ceftazidime, cefepime, ertapenem, metronidazole and vancomycin.

3.2. Predominant fatty acids

The major fatty acids were Hexadecanoic acid (59 %), Tetradecanoic acid (20 %) and 9-Hexadecenoic acid (9 %). No branched structures were detected (Table 2).

3.3. Genomes properties and comparison

C. pacaense draft genome consisted of 22 scaffolds, genome length was 2,672,129 bp, with a 50% of GC content. 2,360 proteins were predicted. The draft genome sequence of *C. pacaense* owned the smallest genome. Its GC content was same as *C. aldenense*, but smaller than *C. lavalense* and greater than others. Additionally, *C. pacaense* owned the smallest number of predicted genes. Carbohydrate transport and metabolism thus secondary metabolites biosynthesis, transport and catabolism were the predominant COG categories identified within *C. pacaense* (Table 3). Based on 16s RNA similarity the closest species was *C. aldenese* (Table 4). This was in agreement with genome data as *C. aldenese* was also the closest species with OrthoANI value of 89.9744 % (*C. aldenese*) but below the 95% cut-off defining a species (Figure 4).

Description of *Clostridium pacaense* sp. nov.

Clostridium pacaense (pa.ca.en'se L. masc. adj. *pacaense*, of PACA, the abbreviation of Provence Alpes Cote d'Azur, the French area where the strain was isolated). In addition to the characteristics in the genus description, cells are Gram-variable with a length of 3.5 μm and a width of 0.5 μm. Produces α-glucosidase and Napthol-AS-Bl-phosphohydrolase. The major fatty acids are $C_{16}H_{32}O_2$, $C_{14}H_{28}O_2$ and $C_{16}H_{30}O_2$. The type strain Marseille-P3100^T has been deposited in the CSUR and CCUG culture collections under numbers CSUR P3100 and CCUG 71489, respectively. The type strain was isolated from stool sample of a Senegalese girl with marasmus. The draft genome of the type strain is 2,672,129 bp long

with a DNA G + C content of 50% and is available on EMBL-EBI scaffolds accession numbers: LS999944-LS999965.

Acknowledgments

We thank the genomics and biobank platforms of the *Institut Hospitalo-Universitaire (IHU)-Méditerranée Infection* for technical assistance.

Funding

This work was supported by a grant from the French State managed by the National Research Agency under the "Investissements d'avenir (Investments for the Future)" program with the reference ANR-10-IAHU-03 (Méditerranée Infection) and by Région Provence Alpes Côte d'Azur and European funding FEDER PRIMI.

Conflict of interest

The authors have no conflict of interest to declare.

References

[1] Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Diversity of the human intestinal microbial flora. Science 2005;308:1635–8. doi:10.1126/science.1110591.

[2] Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. PLoS Biol 2007;5:1556–73. doi:10.1371/journal.pbio.0050177.

[3] Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. PLoS Biol 2007;5:1556–73. doi:10.1371/journal.pbio.0050177.

[4] Hosny M, Cassir N, La Scola B. Updating on gut microbiota and its relationship with the occurrence of necrotizing enterocolitis. Hum Microbiome J 2017;4:14–9. doi:10.1016/j.humic.2017.04.001.

[5] Cassir N, Simeoni U, Scola B La. Gut microbiota and the pathogenesis of necrotizing enterocolitis in preterm neonates. Future Microbiol 2016;11:273–92. doi:10.2217/fmb.15.136.

[6] Lagier J-C, Million M, Hugon P, Armougom F, Raoult D. Human gut microbiota: repertoire and variations. Front Cell Infect Microbiol 2012;2:136. doi:10.3389/fcimb.2012.00136.

[7] Hosny M, Benamar S, Durand G, Armstrong N, Michelle C, Cadoret F, et al. Description of *Clostridium phoceensis* sp. nov., a new species within the genus Clostridium. New Microbes New Infect 2016;14:85–92. doi:10.1016/j.nmni.2016.09.008.

[8] 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;28:237–64. doi:10.1128/CMR.00014-14.

[9] Abdallah RA, Beye M, Diop A, Bakour S, Raoult D, Fournier PE. The impact of culturomics on taxonomy in clinical microbiology. Antonie van Leeuwenhoek, Int J Gen Mol Microbiol 2017;110:1327–37. doi:10.1007/s10482-017-0871-1.

[10] Fournier P-E, Lagier J-C, Dubourg G, Raoult D. Anaerobes in the microbiome From culturomics to taxonomogenomics: A need to change the taxonomy of prokaryotes in clinical microbiology. Anaerobe 2015;36:73–8. doi:10.1016/j.anaerobe.2015.10.011.

[11] Pham TPT, Cadoret F, Alou MT, Brah S, Diallo BA, Diallo A, et al. '*Urmitella timonensis*' gen. nov., sp. nov., '*Blautia marasmi*' sp. nov., '*Lachnoclostridium pacaense*' sp. nov., '*Bacillus marasmi*' sp. nov. and '*Anaerotruncus rubiinfantis*' sp. nov., isolated from stool samples of undernourished African children. New Microbes New Infect 2017;17:84–8. doi:10.1016/j.nmni.2017.02.004.

[12] Lagier JC, Edouard S, Pagnier I, Mediannikov O, Drancourt M, Raoult D. Current and past strategies for bacterial culture in clinical microbiology. Clin Microbiol Rev 2015;28:208–36. doi:10.1128/CMR.00110-14.

[13] Sasser M. Bacterial Identification by Gas Chromatographic Analysis of Fatty Acids Methyl Esters (GC-FAME). 2006.

[14] Dione N, Sankar SA, Lagier JC, Khelaifia S, Michele C, Armstrong N, et al. Genome sequence and description of *Anaerosalibacter massiliensis* sp. nov. New Microbes New Infect 2016;10:66–76. doi:10.1016/j.nmni.2016.01.002.

[15] 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.

[16] Seemann T. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 2014;30:2068–9. doi:10.1093/bioinformatics/btu153.

[17] Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:119. doi:10.1186/1471-2105-11-119.

[18] Lee I, Kim YO, 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.

[19] Hall T. BioEdit : An important software for molecular biology. GERF Bull Biosciences 2011;2:60–1. doi:10.1002/prot.24632.

[20] Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 2016;33:1870–4. doi:10.1093/molbev/msw054.

176 **Table 1: General feature and biochemical tests of *Lachnoclostridium pacaense***

Current classification	Domain: Bacteria Phylum: Firmicutes Class: Clostridia Order: Clostridiales Family: Clostridiaceae Genus: *Clostridium* Species: *Clostridium pacaense* Type strain: Marseille-P3100^{T}
Gram staining:	Variable
Cell shape:	Bacillus
Diameter:	0.5 µm
Cell length:	3.5 µm
Motility:	No
Sporulation:	Yes
Indole:	No

Production of:	
Alkaline phosphatase:	No
Catalase:	No
Oxidase:	No
Nitrate reductase:	No
Urease:	No
β-Galactosidase:	No
α-glucosidase	Yes
N-acetyl-glucosamine:	No
Esterase:	No
Acid from:	
L-Arabinose:	No
Ribose:	No
Mannose:	No
Mannitol:	No
Sucrose:	No

D-Glucose:	No
D-Fructose:	No
D-Maltose:	No
D-Lactose:	No

177

178 **Table 2: Cellular fatty acids of *Clostridium pacaense***

Fatty acids	**Name**	**Mean relative %[a]**
16:0	Hexadecanoic acid	58.5 ± 0.5
14:0	Tetradecanoic acid	19.7 ± 0.3
16:1n7	9-Hexadecenoic acid	8.9 ± 0.2
18:1n9	9-Octadecenoic acid	5.5 ± 0.2
18:1n7	11-Octadecenoic acid	4.4 ± 0.3
18:0	Octadecanoic acid	1.0 ± 0.1
15:0	Pentadecanoic acid	TR
16:1n9	7-Hexadecenoic acid	TR

12:0	Dodecanoic acid	TR

[a] Mean peak area percentage; TR = trace amounts < 1 %

181 **Table 3: *Clostridium pacaense* number of genes associated to COGs categories**

COGs description	Total
C: Chromatin structure and dynamics	119
D: Cell cycle control, mitosis and meiosis	17
E: Amino acid transport and metabolism	110
F: Nucleotide transport and metabolism	48
G: Carbohydrate transport and metabolism	280
H: Coenzyme transport and metabolism	44
I: Lipid transport and metabolism	31
J: Translation	41
K: Transcription	169
L: Replication, recombination and repair	73
M: Cell wall/membrane biogenesis	73
N: Cell motility	18
O: Posttranslational modification, protein turnover, chaperones	28

P: Inorganic ion transport and metabolism	76
Q: Secondary metabolites biosynthesis, transport and catabolism	7
R: General function prediction only	222
S: Function unknown	98
T: Signal transduction mechanisms	93
U: Intracellular trafficking and secretion	4
V: Defense mechanisms	55

182

183

184 **Table 4: *Clostridium pacaense* matrix of similarity based on 16S rRNA gene**

	C. pacaense	***C. lavalense***	***C. citroniae***	***C. celerecrescens***	***C. bolteae***	***C. amygdalinum***	***C. aldenense***	***C. saccharolyticum***
C. pacaense	ID							
C. lavalense	96.3	ID						
C. citroniae	96.7	96.1	ID					
C. celerecrescens	93.7	92.9	93.5	ID				
C. bolteae	95.7	97	96.8	94.1	ID			
C. amygdalinum	94.2	93.2	93.7	97.9	94.3	ID		
C. aldenense	98.4	95.9	96.7	93.9	95.8	94.1	ID	
C. saccharolyticum	94.2	93.2	93.6	98.5	94.1	98.8	94	ID

185 **16S rRNA sequences were aligned and similarity matrix were calculated using Bioedit Software** [19]**.**

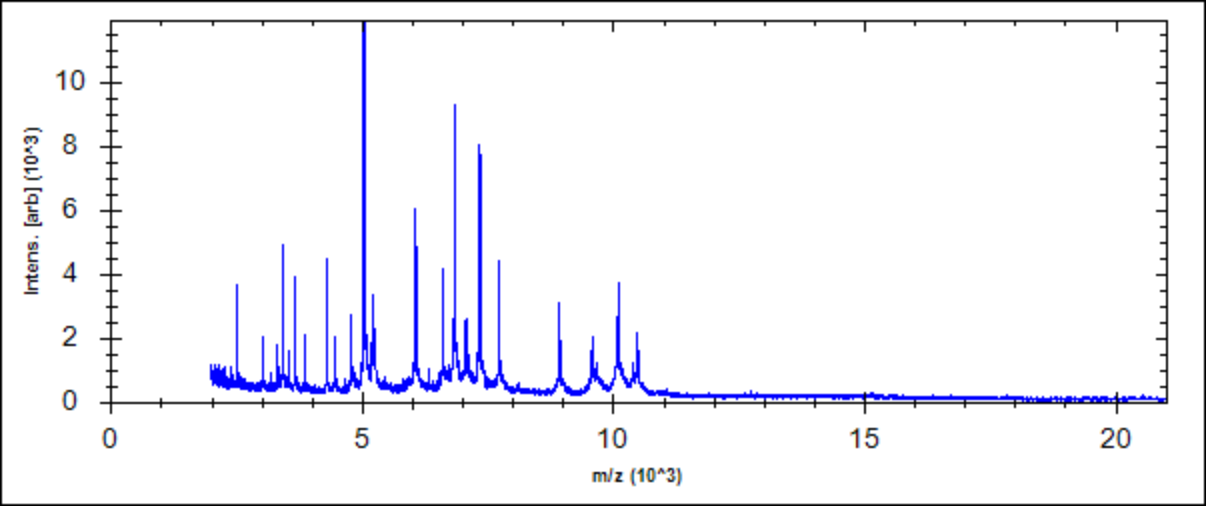
Figure 1: Reference mass spectrum (MALDI-TOF) from *Clostridium pacaense* strain Marseille-P3100

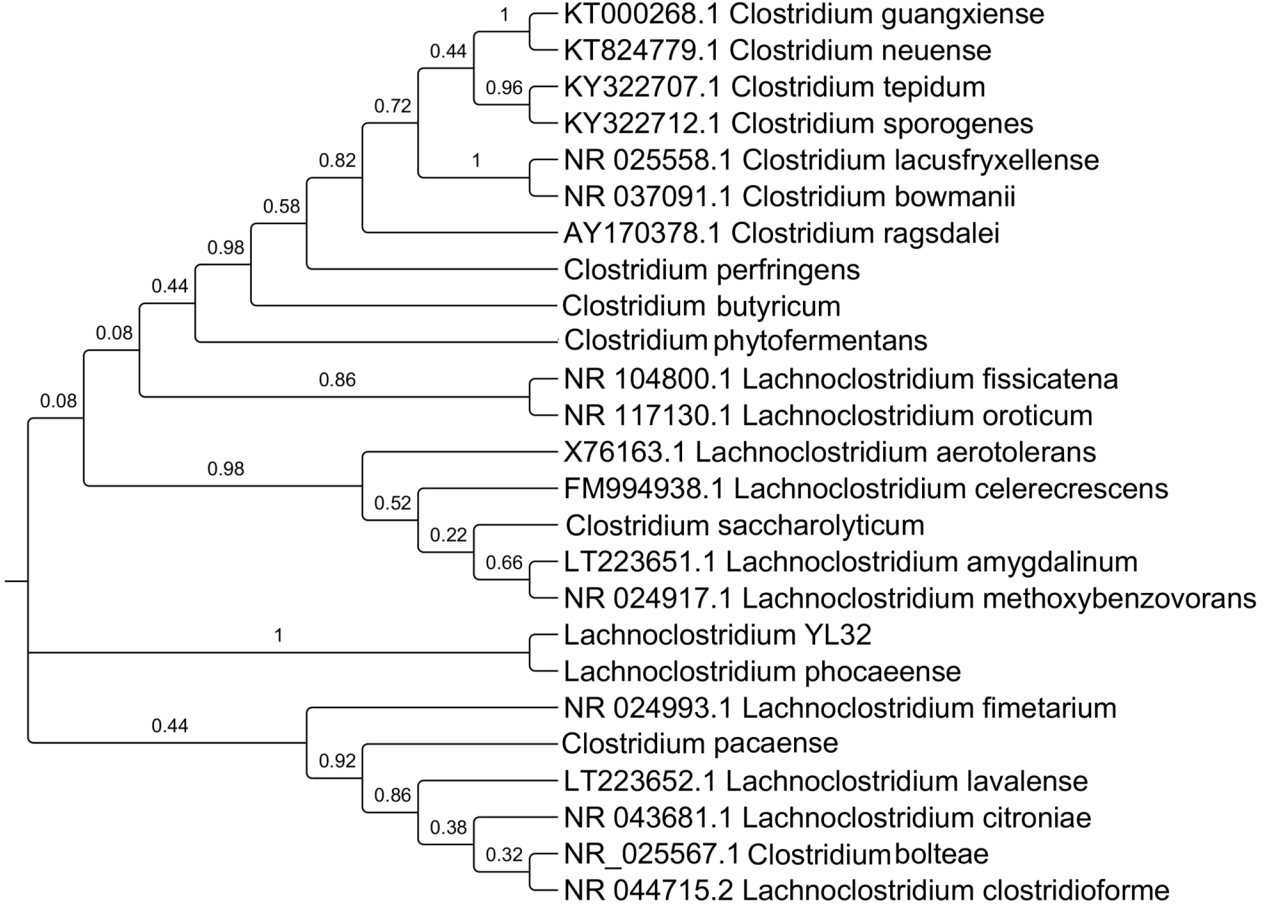
Figure 2: Phylogenetic tree analysis based on 16S rRNA gene sequences

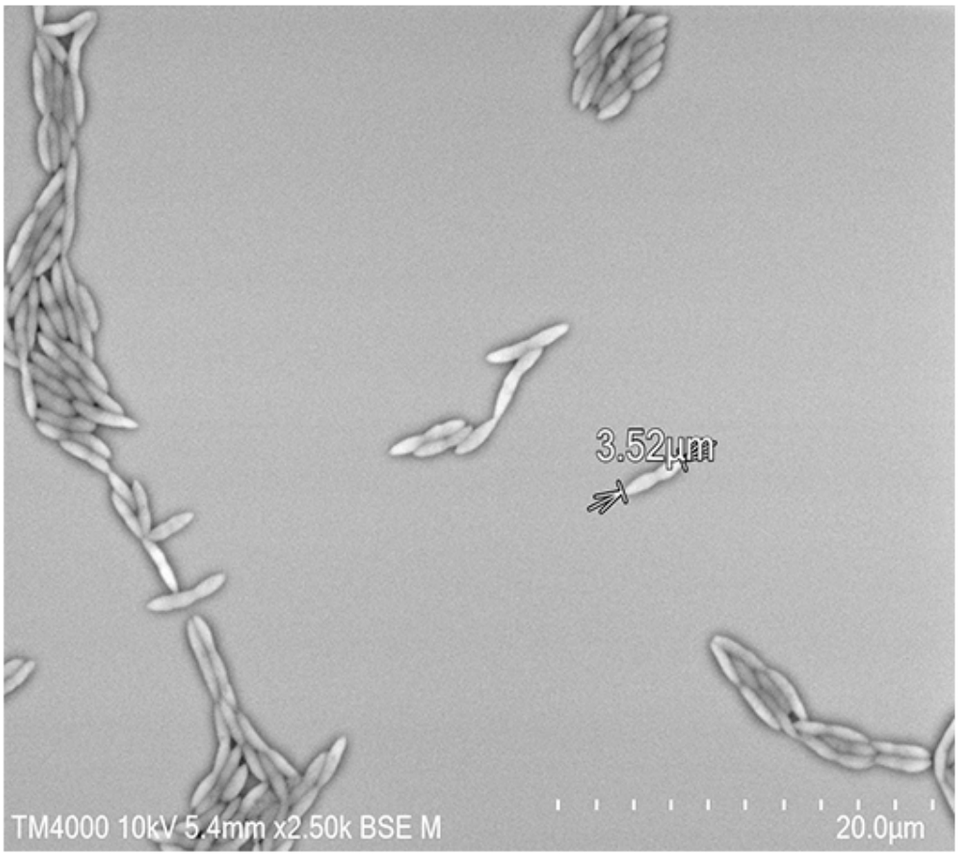
16S rRNA genes were aligned using CLUSTALW and phylogenetic tree was generated using MEGA 7 software [20]

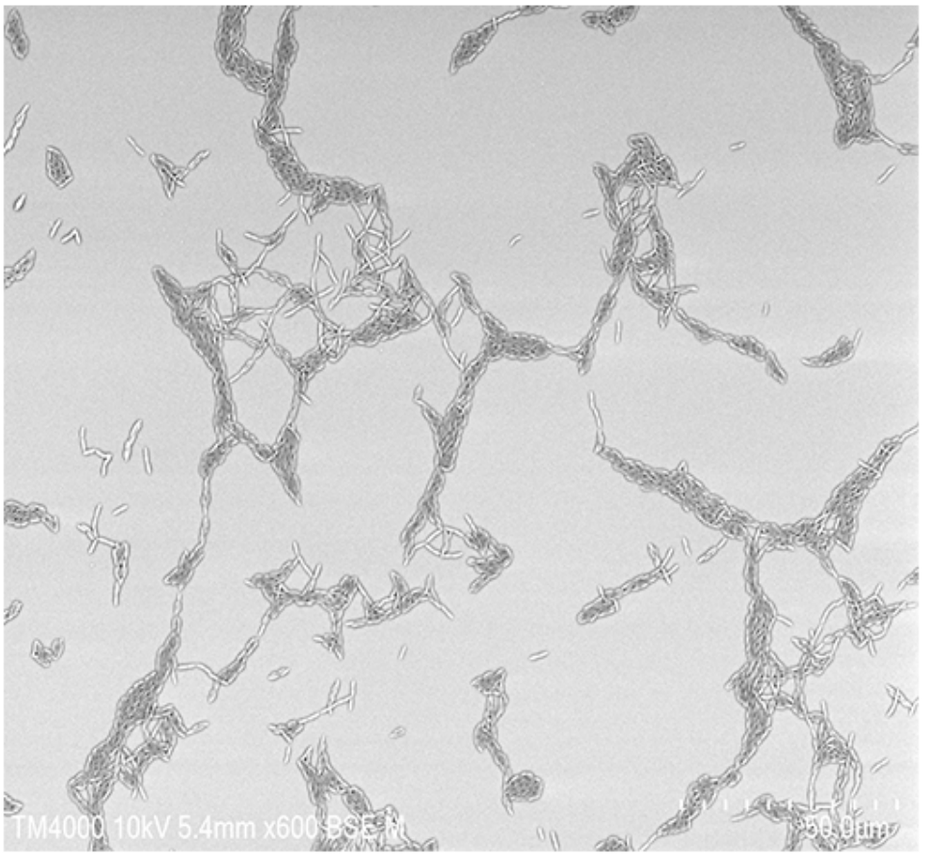
Figure 3: Electron microscopy of *Clostridium pacaense*

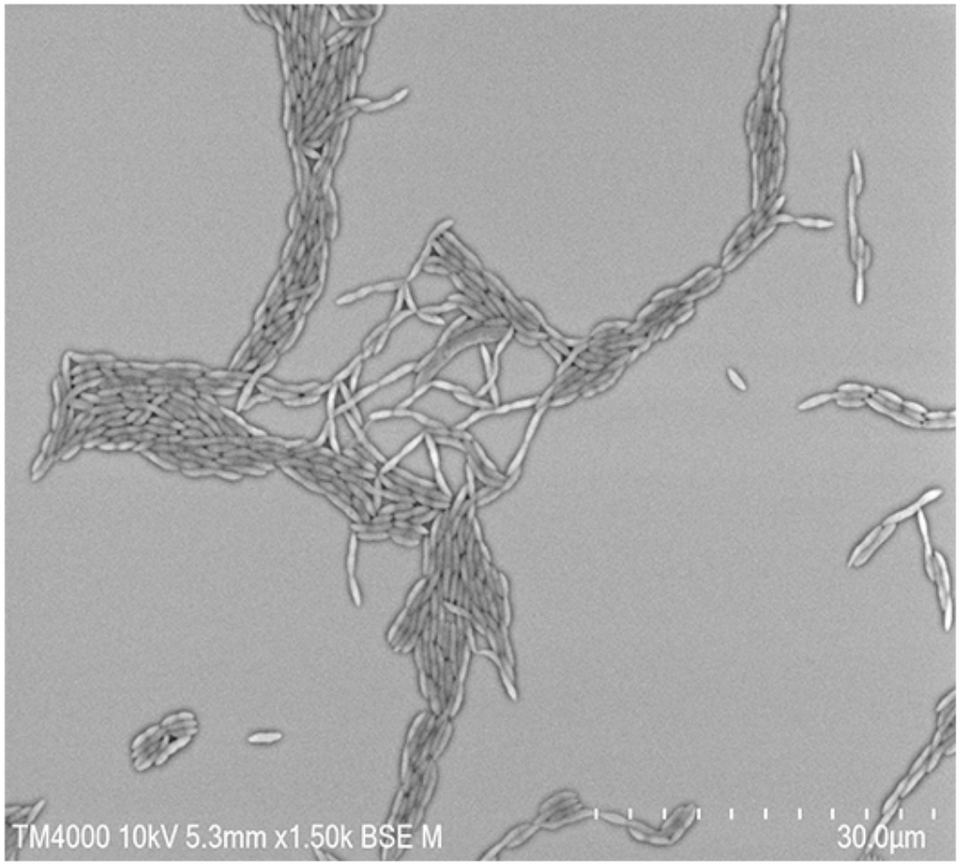
Figure 4: OrthoAni Heatmap of implicated genomes

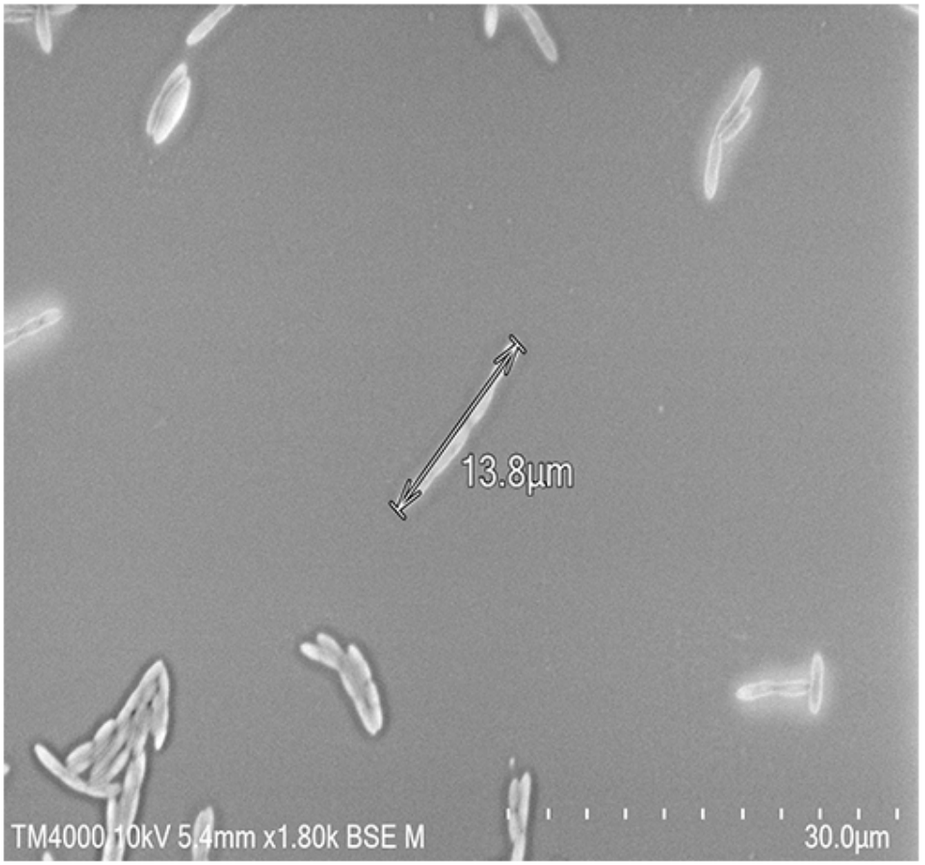

Intens. [arb] (10^3)
10
8
6
4
2
0
0
5
10
15
20
m/z (10^3)

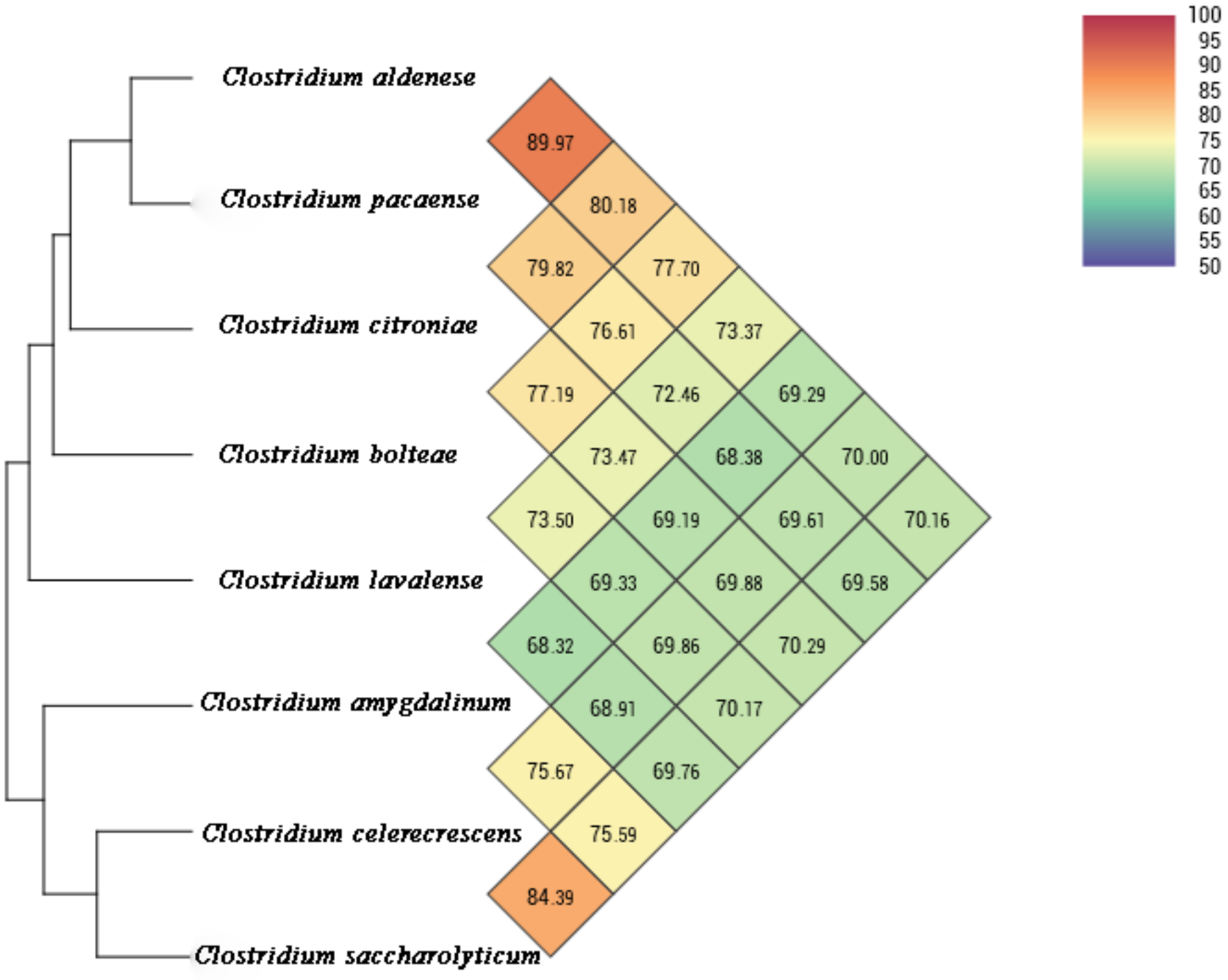

KT000268.1 Clostridium guangxiense
KT824779.1 Clostridium neuense
KY322707.1 Clostridium tepidum
KY322712.1 Clostridium sporogenes
NR 025558.1 Clostridium lacusfryxellense
NR 037091.1 Clostridium bowmanii
AY170378.1 Clostridium ragsdalei
Clostridium perfringens
Clostridium butyricum
Clostridium phytofermentans
NR 104800.1 Lachnoclostridium fissicatena
NR 117130.1 Lachnoclostridium oroticum
X76163.1 Lachnoclostridium aerotolerans
FM994938.1 Lachnoclostridium celerecrescens
Clostridium saccharolyticum
LT223651.1 Lachnoclostridium amygdalinum
NR 024917.1 Lachnoclostridium methoxybenzovorans
Lachnoclostridium YL32
Lachnoclostridium phocaeense
NR 024993.1 Lachnoclostridium fimetarium
Clostridium pacaense
LT223652.1 Lachnoclostridium lavalense
NR 043681.1 Lachnoclostridium citroniae
NR_025567.1 Clostridium bolteae
NR 044715.2 Lachnoclostridium clostridioforme
1
0.44
0.96
0.72
0.82
1
0.58
0.98
0.44
0.08
0.86
0.08
0.98
0.52
0.22
0.66
1
0.44
0.92
0.86
0.38
0.32


3.52µm
TM4000 10kV 5.4mm x2.50k BSE M
20.0µm


TM4000 10kV 5.4mm x600 BSE M
50.0µm


TM4000 10kV 5.3mm x1.50k BSE M
30.0µm


13.8µm
TM4000 10kV 5.4mm x1.80k BSE M
30.0µm


Clostridium aldenese
Clostridium pacaense
Clostridium citroniae
Clostridium bolteae
Clostridium lavalense
Clostridium amygdalinum
Clostridium celerecrescens
Clostridium saccharolyticum
89.97
80.18
79.82
77.70
76.61
73.37
77.19
72.46
69.29
73.47
68.38
70.00
73.50
69.19
69.61
70.16
69.33
69.88
69.58
68.32
69.86
70.29
68.91
70.17
75.67
69.76
75.59
84.39
100
95
90
85
80
75
70
65
60
55
50