

**Genome analysis of *Crassaminicella* sp. SY095, an anaerobic mesophilic marine bacterium isolated from a deep-sea hydrothermal vent on the Southwest Indian Ridge**

Xue-Gong Li, Wei-Jia Zhang, Xiao-Qing Qi, Long Fei Wu

► **To cite this version:**

Xue-Gong Li, Wei-Jia Zhang, Xiao-Qing Qi, Long Fei Wu. Genome analysis of *Crassaminicella* sp. SY095, an anaerobic mesophilic marine bacterium isolated from a deep-sea hydrothermal vent on the Southwest Indian Ridge. *Marine Genomics*, Elsevier, 2019, pp.100733. 10.1016/j.margen.2019.100733 . hal-02473056

**HAL Id: hal-02473056**

**<https://hal-amu.archives-ouvertes.fr/hal-02473056>**

Submitted on 10 Feb 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Genome analysis of *Crassaminicella* sp. SY095, an anaerobic**  
2 **mesophilic marine bacterium isolated from a deep-sea**  
3 **hydrothermal vent on the Southwest Indian Ridge**

4 Xue-Gong Li<sup>1;2;3\*</sup>, Wei-Jia Zhang<sup>1;2;3</sup>, Xiao-Qing Qi<sup>1;2;3</sup>, and Long-Fei Wu<sup>2;4</sup>

5 1 Laboratory of Deep-Sea microbial cell biology, Institute of Deep-sea Science and  
6 Engineering, Chinese Academy of Sciences, China

7 2 France-China Joint Laboratory for Evolution and Development of Magnetotactic  
8 Multicellular Organisms (LIA-MagMC), Marseille, France / IDSSE-CAS, Sanya,  
9 China

10 3 Laboratory for Experimental Study under Deep-sea Extreme Conditions, Institute of  
11 Deep-sea Science and Engineering, Chinese Academy of Sciences, China

12 4 Aix-Marseille Université, CNRS, LCB UMR 7257, IMM, Marseille, France

13 Running title: The first complete genome sequence of *Crassaminicella*.

14 \*Corresponding author

15 Xue-Gong Li

16 Deep Sea microbial cell biology, IDSSE, CAS,

17 Sanya 572000, China.

18 E-mail: xuegongli@idsse.ac.cn

19 Phone: 0086-898-8821 2769

20

21

22 **Abbreviations:** CDS, coding sequence; COG, clusters of orthologous groups; GO,  
23 gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; ORF, open  
24 reading frame

25 **Abstract**

26 *Crassaminicella* sp. strain SY095 is an anaerobic mesophilic marine bacterium that  
27 was recently isolated from a deep-sea hydrothermal vent on the Southwest Indian  
28 Ridge. Here, we present the complete genome sequence of strain SY095. The genome  
29 consists of a chromosome of 3,046,753 bp (G+C content of 30.81%) and a plasmid of  
30 36,627 bp (G+C content of 31.29%), encodes 2966 protein, 135 tRNA genes, and 34  
31 rRNA genes. Numerous genes are related to peptide transport, amino acid metabolism,  
32 motility, and sporulation. This agrees with the observation that strain SY095 is a  
33 spore-forming, motile, and chemoheterotrophic bacterium. Further, the genome  
34 harbors multiple prophages that carry all the genes necessary for viral particle  
35 synthesis. Some prophages carry additional genes that may be involved in the  
36 regulation of sporulation. This is the first reported genome of a bacterium from the  
37 genus *Crassaminicella*, providing insights into the microbial adaptation strategies to  
38 the deep-sea hydrothermal vent environment.

39 **Keywords:** chemoheterotroph, chemoorganotroph, extreme environment, genome  
40 sequencing, prophage, sporulation

41

## 42 **1. Introduction**

43 Deep-sea hydrothermal vent environment is characterized by steep temperature and  
44 geochemical gradients, which provide a large range of habitats for chemotrophic  
45 microorganisms [1]. Chemosynthetic bacteria and archaea utilize chemical energy to  
46 fix inorganic carbon into organic carbon for microbial growth, and form the  
47 foundation of vent ecosystems [2, 3]. Chemotrophs are an important component of the  
48 microbial community in the deep-sea hydrothermal vent, and play critical roles in  
49 deep-sea carbon, nitrogen and sulfur cycling [4, 5].

50 The genus *Crassaminicella* is affiliated with Clostridiales and was first reported in  
51 2015 [6]. It represents a novel type of mesophilic chemoorganotrophic bacteria from  
52 the deep-sea hydrothermal environment. Cells from this genus are gram stain-positive,  
53 motile, straight or curved rods, and form terminal endospores. *Crassaminicella* is an  
54 obligate anaerobic and heterotrophic bacterium that ferments carbohydrates and  
55 proteinaceous substrates [6]. To date, this genus contains only one species,  
56 *Crassaminicella profunda*, the sole strain Ra1766HT, as the type strain of this species,  
57 was isolated from sediments of the Guaymas Basin at a depth of 2002 m [6]. During  
58 the recent cruise, we have isolated an anaerobic mesophilic marine bacterium, strain  
59 SY095, from a deep-sea hydrothermal vent on the Southwest Indian Ridge.  
60 Phylogenetic analysis based on 16S rRNA gene sequences indicated that strain SY095  
61 was most closely related to the type strain of *C. profunda* Ra1766HT (96.05%  
62 similarity). Notably, the reported bacteria from the deep-sea hydrothermal vent  
63 environment mainly represent gram-negative bacteria. Only a few represent

64 gram-positive bacteria, e.g., *Clostridium tepidiprofundum* SG 508<sup>T</sup> [7], *Vulcanibacillus*  
65 *modesticaldus* BR<sup>T</sup> [8], and *Sulfobacillus acidophilus* NAL<sup>T</sup> [9]. Here, we report the  
66 first genome of a bacterium from the genus *Crassaminicella*. The availability of  
67 genome sequence will promote the understanding not only of bacteria from this genus  
68 but also the adaptation strategies of gram-positive bacteria to the deep-sea  
69 hydrothermal vent.

## 70 **2. Data description**

71 General features of this strain and the MGS mandatory information are shown in  
72 Table 1. High-quality total genomic DNA was extracted using a MagAttract DNA kit  
73 (Qiagen, USA) according to the manufacturer's instructions. Whole genome sequence  
74 was obtained by sequencing using GridION platform (Nextomics, China). Over 3 Gb  
75 of processed reads were generated for an approximately 950-fold depth of coverage.  
76 Clean reads were assembled using Canu v1.7 to generate the complete genome  
77 sequence [10], and the results were corrected using Pilon v1.22  
78 (<https://github.com/broadinstitute/pilon>) based on the MGI-SEQ 2000 sequence. Gene  
79 predictions were made using Prodigal v2.6.3 (<https://github.com/hyattpd/Prodigal>).  
80 Functional information for each predicted gene was obtained based on  
81 sequence-similarity search against the non-redundant protein database available from  
82 the National Center for Biotechnology Information, Clusters of Orthologous Groups  
83 (COG) database [11], Gene Ontology (GO) database [12], and Kyoto Encyclopedia of  
84 Genes and Genomes (KEGG) [13]. tRNA genes were predicted by using tRNA  
85 scan-SE v2.0 (<http://lowelab.ucsc.edu/tRNAscan-SE>). rRNA genes were predicted by

86 using RNAmmer 1.2 (<http://www.cbs.dtu.dk/services/RNAmmer>). Prophages were  
 87 predicted by using PHASTER web server (<http://phaster.ca/>). Presence of genomic  
 88 islands was investigated by using Islander v1.2 program  
 89 (<https://bioinformatics.sandia.gov/islander>) and CRISPR arrays were analyzed by  
 90 using the Minced program v0.3.0 (<https://github.com/ctSkennerton/minced>). Circular  
 91 representations of the SY095 genome were prepared by using Circos v1.7.11  
 92 (<http://circos.ca/>).  
 93

Item	Description
MIGS data	
Project_name	<i>Crassaminicella</i> sp. SY095 genome sequencing
Geographic location	Indian Ocean
Latitude and longitude	37° 46' 58.69" S, 49° 38' 59.96" E
Water depth	2730 m
Collection date	12–2018
Environment (biome)	Ocean biome ENVO:01000048
Environment (feature)	Marine hydrothermal vent ENVO:01000122
Environment (material)	Marine hydrothermal vent chimney ENVO:01000129
Observed biotic relationship	Free living
Trophic_level	Chemoheterotroph
Sequencing method	Oxford Nanopore GridION
Assembly method	Canu version 1.7
Coverage	950.0x
Finishing_strategy	Complete
General feature	
	Domain <i>Bacteria</i>
	Phylum <i>Firmicutes</i>
Classification	Class <i>Clostridia</i>
	Order <i>Clostridiales</i>
	Family <i>Clostridiaceae</i>
	Genus <i>Crassaminicella</i>
Gram stain	Positive
Cell shape	Rod
Motility	Motile

Relationship to oxygen	Obligate anaerobe	
Optimal temperature	50 °C	
Genomic features	Chromosome	Plasmid
Size (bp)	3,046,753	36,627
G+C content (%)	30.81	31.29
Number of predicted CDSs	2,907	59
Genes assigned to COG	1,959	8
Number of rRNA genes	12, 11, 11 (5S, 16S, 23S)	0
Number of tRNAs	135	0
Number of CRISPRs	2	0

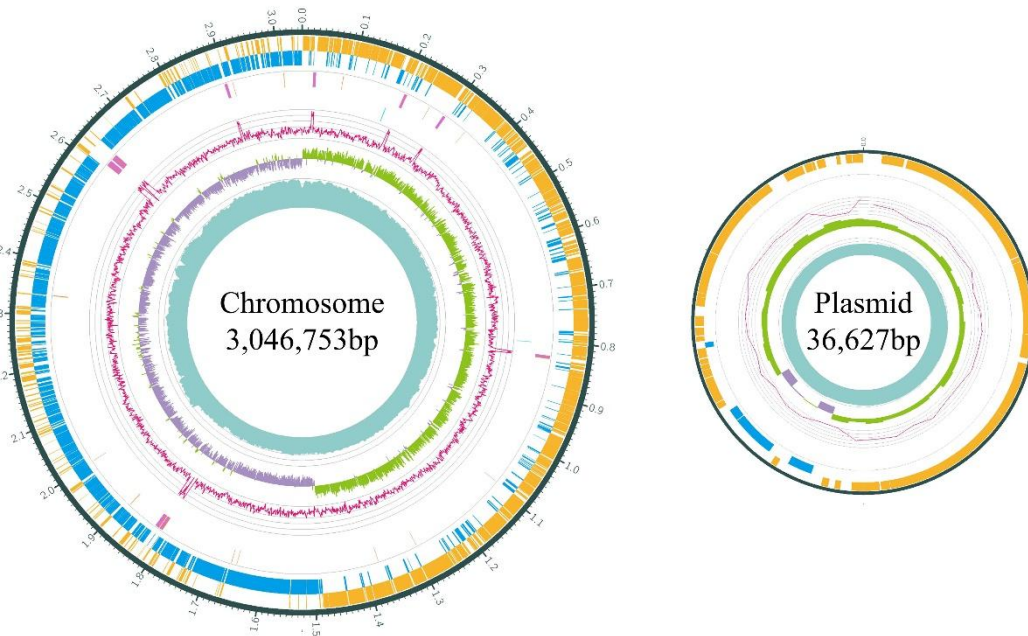
94 Table 1. General features and genome sequencing project information for  
95 *Crassaminicella* sp. SY095 according to MlGS recommendations

96

97 The strain SY095 genome contains a single circular chromosome (3,046,753 bp,  
98 30.81% G+C) and one circular plasmid (36,627 bp, 31.29% G+C). Maps of the  
99 chromosome and plasmid are shown in Fig. 1. CheckM analysis affiliated it with  
100 Clostridiales and showed a high quality of the genome with a completeness of 97.87%  
101 and a contamination of 1.03%. The genome codes eleven 16S rRNA genes with  
102 sequence identities higher than 98.9% between each other. Overall, 2966  
103 protein-coding sequences (CDS) were predicted, which cover approximately 87.03%  
104 of the entire genome. The genome also contains 135 tRNA genes, 34 rRNA genes,  
105 and two CRISPR loci. No genomic island was identified. Approximately half CDSs  
106 were annotated using the COG (66.32%), GO (58.23%), and KEGG (50.23%)  
107 databases. Upon COG classification, 1967 genes were assigned to 23 functional  
108 categories. The major categories were amino acid transport and metabolism (10.46%);  
109 translation, ribosomal structure and biogenesis (9.95%); signal transduction  
110 mechanisms (7.76%); general function prediction only (7.04%); transcription (6.99%);  
111 cell wall/membrane/envelope biogenesis (6.33%); coenzyme transport and



112 metabolism (5.97%); carbohydrate transport and metabolism (5.87%); and energy  
113 production and conversion (5.72%).



114  
115 Fig. 1. Schematic representation of the *Crassaminicella* sp. SY095 genome. The  
116 genome comprises one chromosome and one plasmid. Labeling from the outside to  
117 the center is as follows: circle 1, genes on the forward strand; circle 2, genes on  
118 reverse strand; circle 3, RNA genes (tRNAs orange, rRNAs purple); circle 4,  
119 CRISPRs (blue) and predicted genomic islands (green); circle 5, GC content; circle 6,  
120 GC skew; and circle 7, sequencing depth.

121 Viruses, especially phages, are the most abundant life forms in the deep-sea  
122 hydrothermal vent ecosystems, and have been proposed to play a pivotal role in the  
123 regulation of microbial abundance and metabolism, thus driving the biogeochemical  
124 cycles [14, 15]. In addition to acting as microbial predators, phages might confer an  
125 enhanced level of fitness to the microbial host, enhancing host survival in the extreme

126 environment [16]. In the strain SY095 genome, four putative prophages were  
 127 identified, named CTV1 to CTV4. CTV1, CTV2, and CTV3 are integrated into the  
 128 chromosome, while CTV4 exists in the form of an extrachromosomal plasmid (Table  
 129 2). The four prophage genomes contain all the genes necessary for viral particle  
 130 synthesis and have similar gene organization. The prophage attachment sites (*attL* and  
 131 *attR*) were detected at the ends of proviral sequences. In addition, large numbers of  
 132 genes coding for functionally unknown proteins were identified in the prophage  
 133 genomes, whose functions in phage-host interaction or adaptation of the host cells to  
 134 extreme environments require further investigation.

Prophage	Size(kb)	GC content	ORFs	Location
CTV1	60.5	31.66%	96	chromosome
CTV2	56.6	31.53%	93	chromosome
CTV3	48.0	31.09%	91	chromosome
CTV4	36.6	31.28%	59	plasmid

135 Table 2. Predicted prophage clusters in *Crassaminicella* sp. SY095.

136

137 Strain SY095 forms terminal endospores at late stage of growth. *Spo0E* gene encodes  
 138 an aspartyl-phosphate phosphatase that controls the precise timing and progression of  
 139 sporulation [17]. Seven *spo0E* genes harboring a classical conserved motif were  
 140 identified in the SY095 genome. Three of them were carried by chromosomal  
 141 prophages CTV1, CTV2, and CTV3, while none was identified in the CTV4 prophage.  
 142 It is therefore likely that auxiliary *spo0E* genes of prophages are involved in the  
 143 sporulation of the host cell. However, the interaction among the prophages and their  
 144 relationship with the host cell remains largely unknown. Further efforts will be

145 required to elucidate their biological function and ecological roles, especially in the  
146 deep-sea hydrothermal vent environment.

147 Collectively, the complete genomic data of strain SY095 provides additional genetic  
148 information for bacteria from the genus *Crassaminicella* and also contributes to the  
149 expansion of knowledge on microbial adaptations to the deep-sea hydrothermal vent  
150 environment.

### 151 **3. Genome sequence accession numbers**

152 The complete genome sequence of *Crassaminicella* sp. SY095 is available in the  
153 Genbank database (<https://www.ncbi.nlm.nih.gov/genbank/>) under the accession  
154 number CP042243 (Chromosome) and CP042244 (Plasmid). BioSample data is  
155 available in the NCBI BioSample database (<http://www.ncbi.nlm.nih.gov/biosample/>)  
156 under accession number SAMN12368985. The data have been deposited with links to  
157 BioProject accession number PRJNA556796 in the NCBI BioProject database  
158 (<https://www.ncbi.nlm.nih.gov/bioproject/>). The strain is available at Marine Culture  
159 Collection of China, MCCC (<http://www.mccc.org.cn/>) with accession number  
160 MCCC 1K04191.

### 161 **Acknowledgements**

162 This work was supported by grants from the National Key R&D Program of China  
163 (2018YFC0309904, 2016YFC0304905, and 2016YFC0302502), the National Natural  
164 Science Foundation of China (NSFC 41806174, 91751202, and 91751108), and a  
165 grant for LIA0858-MagMC from the Centre National de la Recherche Scientifique.

166 We are very grateful to all the participants of the TS10 cruise of R/V *Tan Suo Yi Hao*

167 and the pilots of the *Shen Hai Yong Shi* for their professional service during this cruise.  
168 The LABGeM (CEA/Genoscope & CNRS UMR8030), the France Génomique and  
169 French Bioinformatics Institute national infrastructures (funded as part of  
170 Investissement d'Avenir program managed by Agence Nationale pour la Recherche,  
171 contracts ANR-10-INBS-09 and ANR-11-INBS-0013) are acknowledged for support  
172 within the MicroScope annotation platform.

173 **Declarations of interest:** none

174

## 175 **References**

176 [1] G.J. Dick, The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped  
177 locally, *Nat Rev Microbiol*, 17 (2019) 271-283.

178 [2] J. McNichol, H. Stryhanyuk, S.P. Sylva, F. Thomas, N. Musat, J.S. Seewald, S.M. Sievert,  
179 Primary productivity below the seafloor at deep-sea hot springs, *P Natl Acad Sci USA*, 115 (2018)  
180 6756-6761.

181 [3] L.B. Nadine, M. Yucel, A. Das, S.M. Sievert, P.R. Girguis, Hydrothermal energy transfer and  
182 organic carbon production at the deep seafloor, *Frontiers in Marine Science*, 5 (2018) 531.

183 [4] J. Ding, Y. Zhang, H. Wang, H. Jian, H. Leng, X. Xiao, Microbial Community Structure of  
184 Deep-sea Hydrothermal Vents on the Ultraslow Spreading Southwest Indian Ridge, *Front*  
185 *Microbiol*, 8 (2017) 1012.

186 [5] H.L. Cao, Y. Wang, O.O. Lee, X. Zeng, Z.Z. Shao, P.Y. Qian, Microbial Sulfur Cycle in Two  
187 Hydrothermal Chimneys on the Southwest Indian Ridge, *Mbio*, 5 (2014).

188 [6] R. Lakhali, N. Pradel, A. Postec, B. Ollivier, J.L. Cayol, A. Godfroy, M.L. Fardeau, G. Gales,  
189 *Crassaminicella profunda* gen. nov., sp. nov., an anaerobic marine bacterium isolated from  
190 deep-sea sediments, *Int J Syst Evol Microbiol*, 65 (2015) 3097-3102.

191 [7] G.B. Slobodkina, T.V. Kolganova, T.P. Tourova, N.A. Kostrikina, C. Jeanthon, E.A.  
192 Bonch-Osmolovskaya, A.I. Slobodkin, *Clostridium tepidiprofundum* sp. nov., a moderately

193 thermophilic bacterium from a deep-sea hydrothermal vent, *Int J Syst Evol Microbiol*, 58 (2008)  
194 852-855.

195 [8] S. L'Haridon, M.L. Miroshnichenko, N.A. Kostrikina, B.J. Tindall, S. Spring, P. Schumann, E.  
196 Stackebrandt, E.A. Bonch-Osmolovskaya, C. Jeanthon, *Vulcanibacillus modesticaldus* gen. nov.,  
197 sp. nov., a strictly anaerobic, nitrate-reducing bacterium from deep-sea hydrothermal vents, *Int J*  
198 *Syst Evol Microbiol*, 56 (2006) 1047-1053.

199 [9] P.R. Norris, D.A. Clark, J.P. Owen, S. Waterhouse, Characteristics of *Sulfobacillus acidophilus*  
200 sp. nov. and other moderately thermophilic mineral-sulphide-oxidizing bacteria, *Microbiology*,  
201 142 ( Pt 4) (1996) 775-783.

202 [10] S. Koren, B.P. Walenz, K. Berlin, J.R. Miller, N.H. Bergman, A.M. Phillippy, Canu: scalable  
203 and accurate long-read assembly via adaptive k-mer weighting and repeat separation, *Genome Res*,  
204 27 (2017) 722-736.

205 [11] M. Galperin, K. Makarova, Y. Wolf, E. Koonin, Expanded microbial genome coverage and  
206 improved protein family annotation in the COG database. *Nucleic Acids Res*43: D261–D269,  
207 2015.

208 [12] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K.  
209 Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis,  
210 J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, G.O. Consortium, Gene  
211 Ontology: tool for the unification of biology, *Nat Genet*, 25 (2000) 25-29.

212 [13] M. Kanehisa, S. Goto, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, Data, information,  
213 knowledge and principle: back to metabolism in KEGG, *Nucleic Acids Res*, 42 (2014) D199-205.

214 [14] E. Rastelli, C. Corinaldesi, A. Dell'Anno, M. Tangherlini, E. Martorelli, M. Ingrassia, F.L.  
215 Chiocci, M. Lo Martire, R. Danovaro, High potential for temperate viruses to drive carbon cycling  
216 in chemoautotrophy-dominated shallow-water hydrothermal vents, *Environmental microbiology*,  
217 19 (2017) 4432-4446.

218 [15] H.G. Castelan-Sanchez, I. Lopez-Rosas, W.A. Garcia-Suastegui, R. Peralta, A.D.W. Dobson,  
219 R.A. Batista-Garcia, S. Davila-Ramos, Extremophile deep-sea viral communities from  
220 hydrothermal vents: Structural and functional analysis, *Mar Genomics*, 46 (2019) 16-28.

221 [16] T. He, H. Li, X. Zhang, Deep-sea hydrothermal vent viruses compensate for microbial

222 metabolism in virus-host interactions, *Mbio*, 8 (2017) e00893-00817.  
223 [17] G.P. Dubey, A. Narayan, A.R. Mattoo, G.P. Singh, R.K. Kurupati, M.S. Zaman, A. Aggarwal,  
224 R.B. Baweja, S. Basu-Modak, Y. Singh, Comparative genomic study of *spo0E* family genes and  
225 elucidation of the role of Spo0E in *Bacillus anthracis*, *Archives of Microbiology*, 191 (2009)  
226 241-253.