



HAL
open science

Complete genome sequence of *Crassaminicella* sp. 143-21 isolated from a deep-sea hydrothermal vent

Xue-Gong Li, Ze-Xi Jiao, Huan-Huan Zhang, Jun Xu, Wei-Jia Zhang,
Xiao-Qing Qi, Long-Fei Wu

► To cite this version:

Xue-Gong Li, Ze-Xi Jiao, Huan-Huan Zhang, Jun Xu, Wei-Jia Zhang, et al.. Complete genome sequence of *Crassaminicella* sp. 143-21 isolated from a deep-sea hydrothermal vent. *Marine Genomics*, 2021, pp.100899. 10.1016/j.margen.2021.100899 . hal-03335844

HAL Id: hal-03335844

<https://amu.hal.science/hal-03335844v1>

Submitted on 8 Sep 2021

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

1 **Complete genome sequence of *Crassaminicella* sp. 143-21,**
2 **isolated from a deep-sea hydrothermal vent**

3 Xue-Gong Li^{1;2;3;4*}; Ze-Xi Jiao^{1;5}; Huan-Huan Zhang⁶; Jun Xu^{7*}; Wei-Jia Zhang^{1;2;3;4};
4 Xiao-Qing Qi^{1;2;3;4}; and Long-Fei Wu^{2;8}

5 1 Laboratory of Deep-sea Microbial Cell Biology, Institute of Deep-sea Science and
6 Engineering, Chinese Academy of Sciences, Sanya, China

7 2 International Associated Laboratory of Evolution and Development of
8 Magnetotactic Multicellular Organisms, CNRS-Marseille/ CAS- Sanya

9 3 CAS Key Laboratory for Experimental Study under Deep-sea Extreme Conditions,
10 Institute of Deep-sea Science and Engineering, Chinese Academy of Sciences, Sanya,
11 China

12 4 Institution of Deep-sea Life Sciences, IDSSE-BGI, Hainan Deep-sea Technology
13 Laboratory

14 5 University of Chinese Academy of Sciences, Beijing, China

15 6 School of Oceanography, Shanghai Jiao Tong University, Shanghai, China

16 7 State Key Laboratory of Microbial Metabolism, School of Life Sciences and
17 Biotechnology, Shanghai Jiao Tong University, Shanghai, China

18 8 Aix Marseille Univ, CNRS, LCB, IMM, IM2B, Marseille, France

19 Running title: Genome analysis of *Crassaminicella* sp. 143-21

20 *Corresponding author

21 Xue-Gong Li: xuegongli@idsse.ac.cn; Jun Xu: xujunn@sjtu.edu.cn

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

26 **Abstract**

27 *Crassaminicella* sp. 143-21, a putative new species isolated from deep-sea
28 hydrothermal vent chimney on the Central Indian Ridge (CIR), is an anaerobic,
29 thermophilic and rod-shaped bacterium belonging to the family *Clostridiaceae*. In this
30 study, we present the complete genome sequence of strain 143-21, comprising
31 2,756,133 bp with a G+C content of 31.1 %. In total, 2,427 protein coding genes, 121
32 tRNA genes and 33 rRNA genes were obtained. Genomic analysis of strain 143-21
33 revealed that numerous genes related to organic matter transport and catabolism,
34 including peptide transport, amino acid transport, saccharide transport, ethanolamine
35 transport and corresponding metabolic pathways. Further, the genome contains a large
36 proportion of genes involved in translation, ribosomal structure, and signal
37 transduction. These genes might facilitate microbial survival in deep-sea
38 hydrothermal vent environment. The genome of strain 143-21 will be helpful for
39 further understanding its adaptive strategies in the deep-sea hydrothermal vent
40 environment.

41 **Keywords:** *Crassaminicella*, Chemoorganotroph, Hydrothermal vent, Adaptation,
42 Complete genome.

43

44 1. Introduction

45 Hydrothermal vents are a typical deep-sea ecosystem that possesses a steep
46 gradient in terms of temperature, redox potential, pH, and substrate concentration
47 when hot venting fluids mix with the surrounding cold sea water [1]. Microorganisms
48 inhabiting such ecosystems are mainly chemolithotrophic, which exploiting the
49 chemical disequilibrium provided by the mixing of reducing hydrothermal fluids and
50 oxidizing seawater, and harnessing this energy to fix inorganic carbon into biomass [2,
51 3]. These microbial communities are important primary producers and fuel the vent
52 ecosystems. Culture-dependent and culture-independent approaches have exposed a
53 vast diversity of chemotrophic microorganisms in hydrothermal vent ecosystems [3-6],
54 however, only a few of them are culturable [7]. Most of investigations of deep-sea
55 hydrothermal vents focused on thermophilic and hyperthermophilic microorganisms,
56 only a few number of moderate thermophiles were isolated [8]. Therefore, these
57 cultivatable moderate thermophiles are likely to provide us with valuable insights into
58 life in the deep-sea hydrothermal systems.

59 The phylum *Firmicutes* includes a very heterogeneous assemblage of bacteria,
60 some of them were identified as dominant groups in the deep-sea hydrothermal vents
61 [9, 10]. The genus *Crassaminicella* belongs to the *Clostridiaceae* family within
62 phylum *Firmicutes* and was first reported in 2015 [11]. Members of this genus are
63 often obligate anaerobes and heterotrophs that use organic matter such as tryptone and
64 glucose [11, 12]. At the time of writing, this genus comprises only one species of
65 hydrothermal origin, *Crassaminicella profunda* Ra1766H^T, which was isolated from

66 geothermal sediments of the Guaymas Basin at a depth of 2,002 m [11]. Recently, we
67 have isolated a potential new species of genus *Crassaminicella*, strain SY095, from a
68 deep-sea hydrothermal vent on the Southwest Indian Ridge. Genomic analysis
69 showed that strain SY095 genome harbors multiple prophages, some of them carry
70 additional genes that may be involved in the regulation of sporulation [12]. The
71 genome information expanded our understanding of the adaptation strategies of genus
72 *Crassaminicella* to the deep-sea hydrothermal vent high pressure and high
73 temperature environments. In this study, strain 143-21 was isolated from a chimney
74 sample that was collected at the depth of 2,440 m from the site Kali (25.32° S, 70.04°
75 E) located in the *Kairei* hydrothermal field of the intermediate spreading Central
76 Indian Ridge (CIR) using the manned submersible “*Shen Hai Yong Shi*” during the
77 TS10-3 cruise of R/V “*Tan Suo Yi Hao*” on February 2019. Physiological analysis
78 showed that strain 143-21 is obligate anaerobic and heterotrophic moderate
79 thermophiles with an optimum growth temperature of 50 °C. 16S rRNA gene sequence
80 identities showed that it was most closely related to the *C. profunda* Ra1766H^T
81 (95.91 % identity), which indicated that strain 143-21 represent a novel species within
82 the genus *Crassaminicella*. Here we report the complete genome sequence of strain
83 143-21 for facilitating an understanding of microbial adaptation to deep-sea
84 hydrothermal vent habitats.

85 **2. Data description**

86 Strain 143-21 grown on JCM 909 medium (<https://jcm.brc.riken.jp>) at 50 °C was
87 harvested in late exponential-phase by centrifugation (4,500 g, 20 min). Genomic

88 DNA was extracted using a MagAttract DNA Kit (Qiagen, USA) according to the
 89 manufacturer's instructions and sequenced on the PromethION and MGI-SEQ 2000
 90 systems at Wuhan Nextomics Biosciences Co., Ltd. (Wuhan, China). About 2.5 Gb
 91 clean data were generated to reach approximately 837-fold depth of coverage. Clean
 92 reads were assembled using the Unicycler version 0.4.8 [13] and polished using Pilon
 93 v1.22 [14]. Gene predictions were made using Prodigal version 2.6.3 [15]. Gene
 94 annotation was performed by a sequence similarity search against the non-redundant
 95 protein database available from the National Center for Biotechnology Information,
 96 Clusters of Orthologous Groups (COG) database [16] and Gene Ontology (GO)
 97 database [17]. The carbohydrate-active enzymes (CAZymes) were predicted by the
 98 dbCAN meta server [18] using the CAZy database [19]. Metabolic pathways were
 99 predicted by using KEGG Automatic Annotation Server (KAAS) [20]. The rRNA and
 100 tRNA genes were identified using RNAmmer version 1.2 [21] and tRNAscan-SE
 101 version 2.0 [22], respectively. Clustered regularly interspaced short palindromic
 102 repeat (CRISPR) arrays were analyzed using the Minced program version 0.4.2 [23],
 103 and genomic island was identified using Islander program version 1.2 [24]. The
 104 graphic circular map of the genome was generated using Circos version 0.69 [25].

105

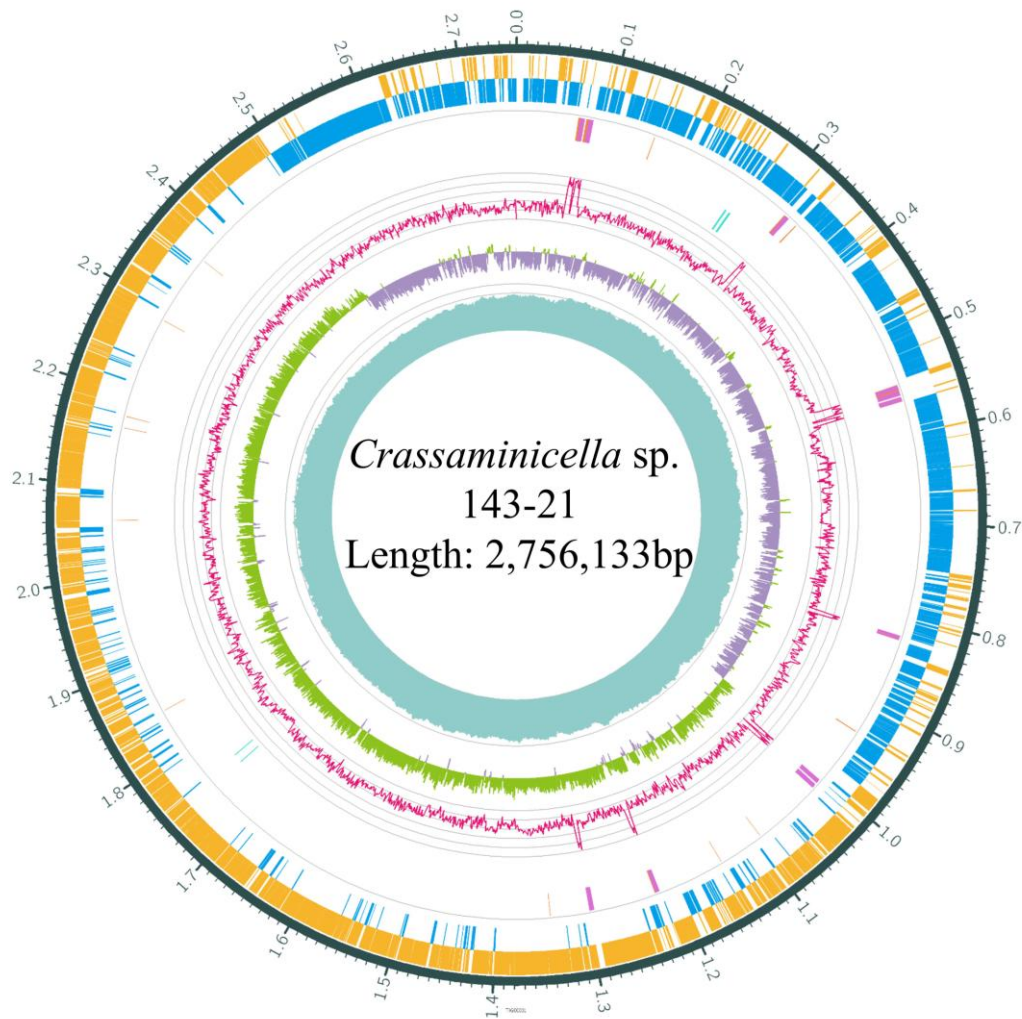
106 Table 1. General features and genome sequencing project information for
 107 *Crassaminicella* sp. 143-21 according to MIGS recommendations

Item	Description
MIGS data	
Project_name	<i>Crassaminicella</i> sp. 143-21 genome sequencing and assembly
NCBI BioProject	Accession: PRJNA743153
NCBI BioSample	Accession: SAMN19999416

Geographic location	Indian Ocean
Latitude and longitude	25.32° S, 70.04° E
Water depth	2,440 m
Collection date	02–2019
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 PromethION and MGI-SEQ 2000
Assembly method	Unicycler version 0.4.8
Coverage	837.0x
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	
Size (bp)	2,756,133
G+C content (%)	31.1
Number of predicted CDSs	2,427
Genes assigned to COG	1,991
Genes assigned to KEGG	1,355
Number of rRNA genes	11, 11, 11 (5S, 16S, 23S)
Number of tRNAs	121
Number of CRISPR Arrays	4

108 The resulting complete genome of *Crassaminicella* sp. 143-21 consists of a
109 single circular chromosome comprising 2,756,133 nucleotides with 31.1 % G+C
110 content (Fig. 1; Table 1). No plasmid was identified in the genome. A total of 2,427
111 protein-coding sequences (CDS) were predicted, which covered approximately 85.59 %
112 of the genome. The genome encodes 121 tRNA genes, 33 rRNA genes, and 4 loci of

113 CRISPR. Among the 2,427 CDSs, over half CDSs were annotated using the COG
114 (82.04 %), GO (64.56 %), and KEGG (55.83 %) databases. Upon COG classification,
115 1,991 genes were assigned to 23 functional categories (Fig S1), whereas the
116 remaining 436 genes could not be assigned to any category. The abundance of COG
117 category in strain 143-21, such as cell cycle control, cell division, chromosome
118 partitioning (D), Coenzyme transport and metabolism (H), translation, ribosomal
119 structure and biogenesis (J), replication , recombination and repair (L), and mobilome:
120 prophages, transposons (X), are higher than that in *Crassaminicella* sp. SY095 and *C.*
121 *profunda* Ra1766H (Fig S1). While other categories, including amino acid transport
122 and metabolism (E), carbohydrate transport and metabolism (G), inorganic ion
123 transport and metabolism (P), and secondary metabolites biosynthesis, transport and
124 catabolism (Q), are lower than the other two strains. These differences may reflect the
125 adaptation of strains to its local habitat.



126

127 Fig. 1. Schematic representation of the *Crassaminicella* sp. 143-21 genome. Labeling
 128 from the outside to the center is as follows: circle 1, genes on the forward strand;
 129 circle 2, genes on reverse strand; circle 3, RNA genes (tRNAs orange, rRNAs purple);
 130 circle 4, CRISPR arrays; circle 5, GC content; circle 6, GC skew; and circle 7,
 131 sequencing depth.

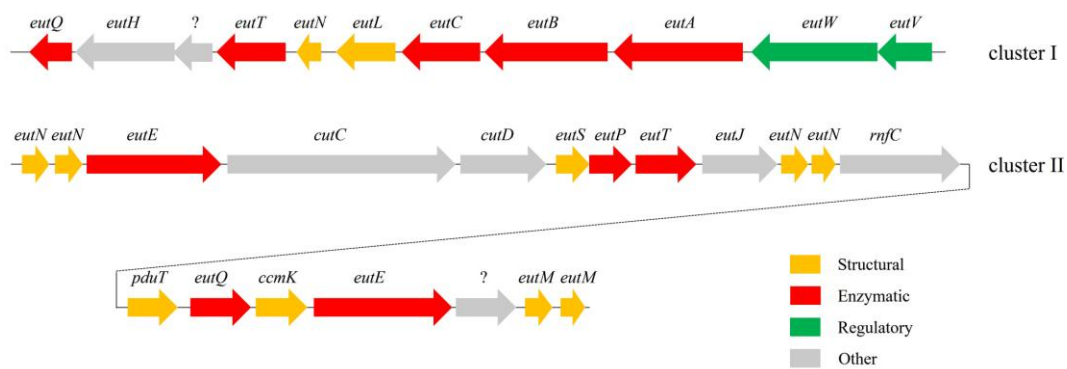
132 2.1 Organic matter catabolism and energy metabolism

133 Strain 143-21 is chemoheterotrophic that can grow on amino acids, tryptone and
 134 saccharides. Correspondingly, its genome encoded for several transport systems that
 135 might participate in the uptake of peptides, amino acids, glucose, ribose, fructose, and
 136 raffinose/stachyose/melibiose. We further analyzed the utilization pattern of carbon

137 source with the Carbohydrate-Active Enzymes (CAZy) database. The result showed
138 that strain 143-21 genome contains 35 CAZymes including 18 glycosyltransferases
139 (GTs), 10 glycoside hydrolases (GHs), 6 Carbohydrate esterases (CEs), and 1
140 carbohydrate-binding modules (CBM). GTs could catalyze glycosyl group transfer to
141 a nucleophilic group and form a glycoside, which may be involved in the amino sugar
142 and nucleotide sugar metabolism. Among GHs, genes related to degradation of
143 N-acetylglucosamine, glycogen, cellobiose, maltose, and fructose were found. We
144 speculated that these genes could be helpful for the survival of strain 143-21 in
145 deep-sea vent habitats by reinforcing acquisition of nutrients from the extreme
146 environment. Strain 143-21 contained a great number of genes that encoded complete
147 glycolysis/gluconeogenesis pathway and pentose phosphate pathway. In addition, 12
148 families of peptidases and 10 kinds of aminotransferases were annotated, which digest
149 peptides to produce a variety of amino acids and further metabolism.

150 In the genome of strain 143-21, thirty genes encoding ethanolamine
151 utilization-related proteins were observed in two clusters (Fig. 2). Gene cluster I
152 (KVH43_01845-01895) contains the key genes of ethanolamine utilization, *eutB* and
153 *eutC*, which encoding the two subunits of the ethanolamine ammonia lyase and breaks
154 ethanolamine down into the gases acetaldehyde and ammonia [26]. In addition, some
155 accessory genes and regulatory genes can be found in this cluster. Gene cluster II
156 (KVH43_06710-06800) mainly contains genes encoding the bacterial
157 microcompartments, where the catabolic reactions occurred efficiently [26].
158 Interestingly, such gene clusters were also presented in the strain SY095 genome,

159 while absent in the genome of strain Ra1766H. All microbial cells are made of a
 160 bilayer of phospholipids, which contain the membrane lipid
 161 phosphatidylethanolamine. Ethanolamine can be produced when phosphodiesterases
 162 break down phosphatidylethanolamine, and is considered a valuable source of carbon
 163 and/or nitrogen for bacteria capable of its catabolism [26]. Deep-sea vent ecosystem is
 164 the vast reservoir of microorganisms, it is plausible that cell membrane decomposition
 165 will accompany the release of ethanolamine. Thus, possessing ethanolamine
 166 utilization genes provide substantial competitive advantages for microbes surviving in
 167 such extreme environment.



168
 169 Fig. 2. The putative ethanolamine utilization gene clusters and gene organization of
 170 the strain 143-21.

171 Genomic analysis also predicted an electron transport complex encoding gene
 172 cluster, which is assumed to encode a membrane-bound enzyme complex with six
 173 subunits (RnfCDGEAB). Rnf complex is a respiratory enzyme that oxidizes reduced
 174 ferredoxin and reduces NAD, coupled to ion (Na^+) transport across the cytoplasmic
 175 membrane and generate an electrochemical gradient for ATP synthesis [27, 28].
 176 Moreover, ten genes encoding F₀F₁ ATP synthase were observed in strain 143-21
 177 genome. Intriguingly, we also identified similar gene clusters encoding Rnf complex

178 and F₀F₁ ATP synthase in the genomes of strain SY095 and strain Ra1766H. It is
179 proposed that Rnf complex coupled with F₀F₁ ATP synthase form a simple
180 respiratory chain to increase the ATP yield, which could help *Crassaminicella* species
181 to survive in the deep-sea hydrothermal vent environment.

182 **2.2 Environmental adaptation**

183 Genomic analysis showed that strain 143-21 possesses a series of genes related
184 to high pressure and high temperature adaptation. Firstly, according to the COG
185 categorization, strain 143-21 genome contains a large proportion of genes (10.20 %)
186 related to translation, ribosomal structure and biogenesis. It has been reported that
187 increasing the synthesis of ribosome subunits and tRNA synthase could help vent
188 microorganism to counteract the effects of HHP on the ribosome complex [29].
189 Secondly, strain 143-21 genome contains 9.84 % of total COG related to signal
190 transduction mechanisms, which is 7.76 % in *Crassaminicella* sp. SY095 and 10.88 %
191 in *C. profunda* Ra1766H (Fig S1). For strain 143-21, twelve histidine kinase sensor
192 and eleven diguanylate cyclase were identified. Histidine kinase acts as a sensor in a
193 two-component signal transduction system, diguanylate cyclase are involved in the
194 synthesis of global second messenger cyclic diguanosine monophosphate, which
195 involved in many aspects, such as biofilm formation, motility, and resistance [30]. As
196 reported, two-component regulatory systems are crucial for strains to sense changes of
197 surrounding environment [31]. Thus, it is reasonable to suppose that possessing
198 multiple signal transduction systems plays a role in microbial survival in deep-sea
199 hydrothermal vent environment. Thirdly, strain 143-21 can form an endospore at late

200 stage of growth. Searching its genome, we discovered that a total of 73 genes are
201 involved in sporulation initiation, sporulation regulation, DNA translocation,
202 engulfment, cortex synthesis, coat formation and assembly. *spo0A* is the key gene for
203 sporulation initiation, which was activated after phosphorylated by histidine kinases.
204 Once Spo0A was phosphorylated, many genes will participate in the sporulation
205 regulation, such as *spoIIAA*, *spoIIAB*, *spoIIIE*, *spoIIIGA*, *spoIIM*, *spoIIR*, *sigE-H*,
206 *spoIIIA-H*, and *spoIVA-B*. At spore assembly stage, SpoVM functions as a landmark
207 protein. In addition, there are also 21 genes associated with spore germination.
208 Forming an endospore under extreme conditions and spore germination in suitable
209 environments may help strain 143-21 surviving in the changing deep-sea
210 hydrothermal vent environment. Fourthly, strain 143-21 genome harbor 89
211 transposases, which may facilitate genomic rearrangements and gene duplication. A
212 high proportion of transposases confer strain a strong genetic plasticity to deal with
213 changing vent environment. In addition, the strain 143-21 genome contains four
214 CRISPR arrays comprising a total of 78 spacer sequences, significantly more than the
215 spacers identified in strain SY095 (28) and strain Ra1766H^T (53). Variation in
216 CRISPR array sequences among different *Crassaminicella* species is probably an
217 adaptation to the local habitats. In conclusion, the genomic analysis has revealed the
218 genetic basis of strain 143-21 and provided insight into the adaptation strategies used
219 by *Crassaminicella* spp. in the deep-sea hydrothermal vent.

220 2.3. Genome sequence accession numbers

221 The complete genome sequence of *Crassaminicella* sp. 143-21 was deposited at

222 GenBank database under the accession number CP078093. The strain has been
223 submitted to Marine Culture Collection of China, MCCC (<http://www.mccc.org.cn/>)
224 with accession number MCCC 1K06400.

225 **Acknowledgements**

226 This research was funded by grant 2018YFC0310704, 2018YFC0309904 and
227 2016YFC0304905-08 from the National Key R&D Program of China, grants NSFC
228 91751202 and 41806174 from the National Natural Science Foundation of China,
229 grant ZDKJ2019011 from the Major Scientific and Technological Projects of Hainan
230 Province, grant Y9719105 for Institute of Deep-sea Technology Innovation, Chinese
231 Academy of Sciences and a grant for LIA-MagMC from the Centre National de la
232 Recherche Scientifique. We are very grateful to all the participants of the TS10 cruise
233 of R/V *Tan Suo Yi Hao* and the pilots of the *Shen Hai Yong Shi* for their professional
234 service during this cruise.

235 **Declaration of Competing Interest**

236 The authors declare that they have no known competing financial interests or
237 personal relationships that could have appeared to influence the work reported in this
238 paper.

239 **References**

- 240 [1] W. Martin, J. Baross, D. Kelley, M.J. Russell, Hydrothermal vents and the origin of life, *Nat*
241 *Rev Microbiol*, 6 (2008) 805-814.
- 242 [2] G.J. Dick, The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped
243 locally, *Nat Rev Microbiol*, 17 (2019) 271-283.
- 244 [3] S. Nakagawa, K. Takai, Deep-sea vent chemoautotrophs: diversity, biochemistry and
245 ecological significance, *FEMS Microbiol Ecol*, 65 (2008) 1-14.
- 246 [4] C.S. Fortunato, B. Larson, D.A. Butterfield, J.A. Huber, Spatially distinct, temporally stable
247 microbial populations mediate biogeochemical cycling at and below the seafloor in hydrothermal

248 vent fluids, *Environmental Microbiology*, 20 (2018) 769-784.

249 [5] S.M. Sievert, C. Vetriani, Chemoautotrophy at Deep-Sea Vents Past, Present, and Future,
250 *Oceanography*, 25 (2012) 218-233.

251 [6] W. Wang, Z. Li, L. Zeng, C. Dong, Z. Shao, The oxidation of hydrocarbons by diverse
252 heterotrophic and mixotrophic bacteria that inhabit deep-sea hydrothermal ecosystems, *Isme J*, 14
253 (2020) 1994-2006.

254 [7] B.N. Orcutt, J.B. Sylvan, N.J. Knab, K.J. Edwards, Microbial ecology of the dark ocean above,
255 at, and below the seafloor, *Microbiol Mol Biol Rev*, 75 (2011) 361-422.

256 [8] G.B. Slobodkina, T.V. Kolganova, T.P. Tourova, N.A. Kostrikina, C. Jeanthon, E.A.
257 Bonch-Osmolovskaya, A.I. Slobodkin, *Clostridium tepidiprofundum* sp. nov., a moderately
258 thermophilic bacterium from a deep-sea hydrothermal vent, *Int J Syst Evol Microbiol*, 58 (2008)
259 852-855.

260 [9] H.L. Cao, Y. Wang, O.O. Lee, X. Zeng, Z.Z. Shao, P.Y. Qian, Microbial Sulfur Cycle in Two
261 Hydrothermal Chimneys on the Southwest Indian Ridge, *Mbio*, 5 (2014).

262 [10] R.K. Sinha, K.P. Krishnan, F.A. Thomas, M.B. Binish, M. Mohan, P.J. Kurian, Polyphasic
263 approach revealed complex bacterial community structure and function in deep sea sediment of
264 ultra-slow spreading Southwest Indian Ridge, *Ecological Indicators*, 96 (2019) 40-51.

265 [11] R. Lakhil, N. Pradel, A. Postec, B. Ollivier, J.L. Cayol, A. Godfroy, M.L. Fardeau, G. Gales,
266 *Crassaminicella profunda* gen. nov., sp. nov., an anaerobic marine bacterium isolated from
267 deep-sea sediments, *Int J Syst Evol Microbiol*, 65 (2015) 3097-3102.

268 [12] X.G. Li, W.J. Zhang, X.Q. Qi, L.F. Wu, Genome analysis of *Crassaminicella* sp. SY095, an
269 anaerobic mesophilic marine bacterium isolated from a deep-sea hydrothermal vent on the
270 Southwest Indian Ridge, *Mar Genomics*, 52 (2020) 100733.

271 [13] R.R. Wick, L.M. Judd, C.L. Gorrie, K.E. Holt, Unicycler: Resolving bacterial genome
272 assemblies from short and long sequencing reads, *PLoS Comput Biol*, 13 (2017) e1005595.

273 [14] B.J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C.A. Cuomo, Q.
274 Zeng, J. Wortman, S.K. Young, A.M. Earl, Pilon: an integrated tool for comprehensive microbial
275 variant detection and genome assembly improvement, *PLoS One*, 9 (2014) e112963.

276 [15] D. Hyatt, G.L. Chen, P.F. Locascio, M.L. Land, F.W. Larimer, L.J. Hauser, Prodigal:
277 prokaryotic gene recognition and translation initiation site identification, *BMC Bioinformatics*, 11
278 (2010) 119.

279 [16] M. Galperin, K. Makarova, Y. Wolf, E. Koonin, Expanded microbial genome coverage and
280 improved protein family annotation in the COG database. *Nucleic Acids Res* 43: D261–D269,
281 2015.

282 [17] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K.
283 Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis,
284 J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, G.O. Consortium, Gene
285 Ontology: tool for the unification of biology, *Nat Genet*, 25 (2000) 25-29.

286 [18] Y. Yin, X. Mao, J. Yang, X. Chen, F. Mao, Y. Xu, dbCAN: a web resource for automated
287 carbohydrate-active enzyme annotation, *Nucleic Acids Res*, 40 (2012) W445-451.

288 [19] V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The
289 carbohydrate-active enzymes database (CAZy) in 2013, *Nucleic acids research*, 42 (2014)
290 D490-D495.

291 [20] M. Kanehisa, S. Goto, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, Data, information,

292 knowledge and principle: back to metabolism in KEGG, *Nucleic Acids Res*, 42 (2014) D199-205.
293 [21] K. Lagesen, P. Hallin, E.A. Rodland, H.H. Staerfeldt, T. Rognes, D.W. Ussery, RNAmmer:
294 consistent and rapid annotation of ribosomal RNA genes, *Nucleic Acids Res*, 35 (2007)
295 3100-3108.
296 [22] T.M. Lowe, P.P. Chan, tRNAscan-SE On-line: integrating search and context for analysis of
297 transfer RNA genes, *Nucleic Acids Res*, 44 (2016) W54-57.
298 [23] C. Bland, T.L. Ramsey, F. Sabree, M. Lowe, K. Brown, N.C. Kyrpides, P. Hugenholtz,
299 CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced
300 palindromic repeats, *Bmc Bioinformatics*, 8 (2007).
301 [24] C.M. Hudson, B.Y. Lau, K.P. Williams, Islander: a database of precisely mapped genomic
302 islands in tRNA and tmRNA genes, *Nucleic Acids Research*, 43 (2015) D48-D53.
303 [25] M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, M.A.
304 Marra, Circos: an information aesthetic for comparative genomics, *Genome Res*, 19 (2009)
305 1639-1645.
306 [26] K.G. Kaval, D.A. Garsin, Ethanolamine Utilization in Bacteria, *Mbio*, 9 (2018).
307 [27] M. Kuhns, D. Trifunovic, H. Huber, V. Muller, The Rnf complex is a Na(+)coupled
308 respiratory enzyme in a fermenting bacterium, *Thermotoga maritima*, *Commun Biol*, 3 (2020).
309 [28] Q. Li, L. Li, T. Rejtar, D.J. Lessner, B.L. Karger, J.G. Ferry, Electron transport in the pathway
310 of acetate conversion to methane in the marine archaeon *Methanosarcina acetivorans*, *J Bacteriol*,
311 188 (2006) 702-710.
312 [29] G. Michoud, M. Jebbar, High hydrostatic pressure adaptive strategies in an obligate
313 piezophile *Pyrococcus yayanosii*, *Sci Rep*, 6 (2016) 27289.
314 [30] U. Römling, M.Y. Galperin, M. Gomelsky, Cyclic di-GMP: the First 25 Years of a Universal
315 Bacterial Second Messenger, *Microbiology and Molecular Biology Reviews*, 77 (2013) 1-52.
316 [31] E.S. Boyd, Archaea on the move, *Environ Microbiol Rep*, 7 (2015) 385-387.