



Title	Vibrio Clade 3.0: New Vibrionaceae Evolutionary Units Using Genome-Based Approach
Author(s)	Jiang, Chunqi; Tanaka, Mami; Nishikawa, Sayo et al.
Citation	Current microbiology, 79(1), 10 https://doi.org/10.1007/s00284-021-02725-0
Issue Date	2022-01
Doc URL	https://hdl.handle.net/2115/87847
Rights	This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/s00284-021-02725-0
Type	journal article
File Information	20211104_CMIC_Manuscript_Rev2.pdf



1 **Title**

2 *Vibrio* Clade 3.0: New *Vibrionaceae* evolutionary units using genome-based approach

3

4 **Author names and affiliations**

5 Chunqi Jiang^a, Mami Tanaka^a, Sayo Nishikawa^a, Sayaka Mino^a, Jesús L. Romalde^b, Fabiano
6 L. Thompson^c, Bruno Gomez-Gil^d, Tomoo Sawabe^a

7 ^aLaboratory of Microbiology, Faculty of Fisheries Sciences, Hokkaido University, Hakodate,
8 Japan

9 ^bDepartamento de Microbiología y Parasitología, CRETUS & CIBUS-Facultad de Biología,
10 Universidade de Santiago de Compostela, Campus Vida, Santiago de Compostela, España

11 ^cInstitute of Biology and SAGE-COPPE, Federal University of Rio de Janeiro (UFRJ), Rio
12 de Janeiro, Brazil

13 ^dCIAD, AC, Mazatlan Unit for Aquaculture and Environmental Management, Mazatlán,
14 México

15

16 **Corresponding Author**

17 Tomoo Sawabe, Laboratory of Microbiology, Faculty of Fisheries Sciences, Hokkaido
18 University, Hakodate, Japan.

19 E-mail: sawabe@fish.hokudai.ac.jp; telephone/fax number: +81-138-40-5569

20

21 **Compliance with Ethical Standards**

22 ● **Funding:** This work was partly supported by KAKEN 19H03041.

23 ● **Competing Interests:** The authors declare no competing financial interests.

24

25 **Declarations**

26 ● **Funding**

27 This work was partly supported by KAKEN 19H03041.

28 ● **Conflicts of interest/Competing interests**

29 The authors declare no competing financial interests.

30 ● **Ethics approval (Not applicable)**

31 ● **Consent to participate (Not applicable)**

32 ● **Consent for publication (Not applicable)**

33 ● **Availability of data and material**

34 The whole genome sequence data obtained in this study was deposited at

35 DDBJ/EMBL/GenBank under BioProject Accession: PRJDB11924.

36 ● **Code availability (Not applicable)**

37

38 **Author Contributions**

39 ● Chunqi Jiang conceived, designed and performed the experiments, analyzed the data,
40 visualized the data, drafted and reviewed the manuscript.

41 ● Mami Tanaka and Sayo Nishikawa performed the experiments, reviewed the manuscript.

42 ● Sayaka Mino, Jesús L. Romalde, Fabiano L. Thompson and Bruno Gomez-Gil analyzed
43 the data, reviewed the manuscript.

44 ● Tomoo Sawabe conceived and designed the experiments, reviewed the manuscript.

45

46 **ABSTRACT**

47 Currently, over 190 species in family *Vibrionaceae*, including not-yet-cultured taxa,
48 have been described and classified into over nine genera, in which the number of species has
49 doubled compared to the previous vibrio evolutionary update (Vibrio Clade 2.0) (Sawabe et
50 al., 2014). In this study, “Vibrio Clade 3.0”, the second update of the molecular phylogenetic
51 analysis was performed based on nucleotide sequences of eight housekeeping genes (8-
52 HKGs) retrieved from genome sequences including 22 newly determined genomes. A total of
53 51 distinct clades were observed, of which 21 clades are newly described. We further
54 evaluated the delineation powers of the clade classification based on nucleotide sequences of
55 34 single-copy genes and 11 ribosomal protein genes (11-RPGs) retrieved from core genome
56 sequences, however, the delineation power of 8-HKGs is still high and that gene set can be
57 reliably used for the classification and identification of *Vibrionaceae*. Furthermore, the 11-
58 RPGs set proved to be useful in identifying uncultured species among metagenome-
59 assembled genome (MAG) and/or single-cell genome assembled genome (SAG) pools. This
60 study expands the awareness of the diversity and evolutionary history of the family
61 *Vibrionaceae* and accelerates the taxonomic applications in classifying as not-yet-cultured
62 taxa among MAGs and SAGs.

63

64 **Keywords:** *Vibrionaceae*, multilocus sequence analysis, housekeeping gene, single-copy
65 gene, ribosomal protein gene, genome taxonomy, metagenome-assembled genome

66

67 INTRODUCTION

68 The family *Vibrionaceae* is a monophyletic group of Gram staining-negative facultative
69 anaerobic bacteria forming curved rods that occur naturally in marine and brackish water
70 systems. To date, over 190 species have been described, consisting of over 9 genera.
71 Members of *Vibrionaceae* are important bacteria for marine mineral cycles, geochemistry,
72 pathogenicity, evolution, ecology and systematics; it has been an excellent model for testing
73 modern methodologies and techniques in bacterial systematics and ecology for better
74 establishing “Genomic Taxonomy” [1–3]. Genomic taxonomy is defined on the basis of
75 comprehensive comparative genomics methods including various genome indices, e.g.,
76 average nucleotide identity (ANI), digital DNA-DNA hybridization (dDDH) [4], average
77 amino acid identity (AAI), and multilocus sequence analysis (MLSA) [5]. Among those,
78 MLSA has become one of the most accurate methods in identification of not only species but
79 also strains [6–12].

80 The first broad research on reconstructing the evolutionary history of 78 *Vibrionaceae*
81 type strains by MLSA using nine genes (*ftsZ*, *gapA*, *gyrB*, *mreB*, *pyrH*, *recA*, *rpoA*, *topA*, and
82 16S rRNA) was performed by Sawabe et al. (2007), and 14 monophyletic clades were
83 described (“Vibrio Clade 1.0”) [6]. Subsequently, the molecular phylogeny was updated to 96
84 species including 10 genome sequenced strains based on eight housekeeping genes used in
85 2007 eliminating the 16S rRNA gene due to its low resolution, and eight clades were newly
86 identified (“Vibrio Clade 2.0”). It also concluded that the “8-HKGs MLSA” demonstrated
87 enough delineation power for species description, and should be used as the default method
88 before alternative approaches are applied [13]. However, due to the difficulties in developing
89 universal primers for amplifying MLSA genes and the lack of genome sequences of type
90 strains in *Vibrionaceae*, new clades had yet to be identified. Recent rapid progress in genome

91 sequencing projects in vibrios provides an ideal opportunity to update the recent evolutionary
92 units of *Vibrionaceae* and has resulted in a massive increase in novel lineages.

93 Pan-genome was defined to be entire genome repertoire of a given group, and a set of
94 genes shared by all genomes was named core-genome [14–16]. Since the first pan-genome
95 analysis on pathogenic *Streptococcus agalactiae* [17], pan-genome studies have been
96 successfully applied to a variety of biological research [18]. In particular, the currently most
97 used multifunctional program, Anvi'o, which is capable of combing both the pangenomics
98 and phylogenomics for single-copy genes in core-genome to investigate the relationships
99 between a given group of draft/complete genomes [19, 20], provides new insights in
100 exploring the phylogeny and taxonomy of bacteria, such as *Arenibacter* [21], *Vibrio* [22], and
101 *Salmonella* [23]. In addition to the pangenome analyses, both metagenome assembled
102 genome (MAG) and single-cell assembled genome (SAG) are a recently emerging
103 methodology in many branches of modern microbiology. Using this approach, a new view of
104 the tree of life in three domains was reconstructed using ribosomal protein sequences [24],
105 and then this picture was expanded with the concatenation of 120 ubiquitous single-copy
106 protein genes [25]. All these genome-based approaches provide a wide range of
107 phylogenetically informative sequences that can be used to classify or identify species, and
108 finally suggest many novel not-yet-cultured microbes [24, 25]. Furthermore, it became
109 practical to characterize not-yet-cultured *Vibrionaceae* by metagenomic sequencing
110 approaches, especially the ones in a symbiotic relationship to angular fishes [26]. It has
111 accelerated the improvements in related research in diversity, evolution and ecological
112 symbiotic relationships between bacteria and fishes [26–29]. Previously, the taxa in
113 metagenomes could be easily classified into family or higher levels but classifications in
114 species/strain level remains a challenge [30], accurate identification of *Vibrionaceae*

115 species/strains could push forward in understanding the diversity, evolution, and phylogeny
116 of bacteria.

117 The aims of this study are; 1) to update the knowledge in the diversity and evolution of
118 the family *Vibrionaceae*, 2) to examine the delineation power of 8-HKGs against SCGs and
119 recently developed gene sets such as ribosomal protein genes (RPGs), and 3) to evaluate the
120 RPGs set as a potential new approach for species identification/classification of uncultured
121 microbial MAG/SAG. Twenty-one new clades are delineated in this study. This study
122 updates the most recent *Vibrionaceae* phylogeny to help us better understand and explore the
123 diversity and evolution of *Vibrionaceae* species, and it proposes a potential new approach to
124 the identification and classification of new bacterial species candidates using MAG/SAG.

125

126 **MATERIALS AND METHODS**

127 *Bacterial Strains*

128 To increase the number of complete genome sequences of *Vibrionaceae*, 27 type strains
129 (see **Table S1**) including 4 *Photobacterium* and 23 *Vibrio* species were cultured on ZoBell
130 2216E agar and broth at 25°C and used for genome sequencing.

131 *DNA Extraction and Whole Genome Sequencing*

132 Genome sequences for the 27 type strains were newly determined/updated according to
133 previously described methods using a hybrid assembly of Nanopore long and Illumina short
134 reads [31]. Sequencing library for MinION (Oxford Nanopore Technologies, Oxford, UK)
135 was prepared using Rapid Barcoding Kit (SQK-RBK004). MinION reads were basecalled by
136 Guppy 1.1. Demultiplexing and adaptor trimming of the reads were performed using
137 Deepbinner 0.2.0 [32]. Paired end DNA libraries were prepared using Nextera XT and were
138 sequenced with the Illumina MiniSeq platform (150-300 bp length) following the
139 manufacturer's instructions. Removal of adaptor sequences were performed using the

140 platanus trim function in *Platanus_B* [33]. Most of the complete genomes were assembled
141 with Unicycler 0.4.7 or 0.4.8 [34] using both long and short reads. For *V. ezurae* JCM
142 21522^T, draft assembly was created by Flye 2.8.3 [35] with genomeSize=5 m using MinION
143 long reads, then sequences were corrected with Racon 1.4.20 [36] and Medaka 1.0.1 (Oxford
144 Nanopore Technologies Ltd., <https://github.com/nanoporetech/medaka>), finally polished by
145 Pilon 1.24 [37] using Illumina short reads.

146 ***Data Collection***

147 A total 163 draft genomes of type or reference strains in *Vibrionaceae* and one complete
148 genome of *Escherichia coli* K-12 (ASM584v2) were retrieved from the National Center for
149 Biotechnology Information (NCBI) and GenBank database (Release 238, 15 June 2020) to
150 update *Vibrionaceae* phylogeny in this study. All the genome sequences used in the analysis
151 are listed in **Supplementary Table S1**.

152 ***Eight Housekeeping Gene Sequences (8-HKGs)***

153 MLSA was performed accordingly to Sawabe et al. (2013). The nucleotide sequences of
154 the eight housekeeping genes (*ftsZ*, *gapA*, *gyrB*, *mreB*, *pyrH*, *recA*, *rpoA*, and *topA*) were
155 retrieved from genome sequences using in silico MolecularCloning ver. 7 (In Silico Biology,
156 Inc., Yokohama, Japan). The domains used to reconstruct the phylogenetic trees were regions
157 of the *ftsZ*, *gapA*, *gyrB*, *mreB*, *pyrH*, *recA*, *rpoA*, and *topA* genes: positions 1-948, 1-960, 13-
158 1650, 1-1038, 31-663, 1-906, 1-990, and 1-2073, respectively (*V. cholerae* ATCC 14035^T
159 (ASM62164v1) numbering).

160 ***Pan-Genome Analysis and Single-Copy Genes (SCGs)***

161 Pangenomic analyses were performed using a total of 192 genome sequences (191
162 *Vibrionaceae* and *E. coli*) and the Anvi'o program ver. 6.2 [19-20]. Briefly, each of the
163 genome sequence fasta file was converted into an anvi'o contigs database (anvi-gen-contigs-
164 database), the contigs databases were decorated with hits from HMM models (anvi-run-

165 hmms) and annotated with functions from the NCBI's Clusters of Orthologous Groups (anvi-
166 run-ncbi-cogs). An anvi'o genomes storage database was created (anvi-gen-genomes-storage)
167 and its pangenome was analyzed (anvi-pan-genome) with the help of NCBI's blastp v. 2.7.1+
168 and MUSCLE [38]. Core single-copy genes were filtered (anvi-get-sequences-for-gene-
169 clusters) by default settings from the clusters generated by MCL [39] and extracted in fasta
170 files (anvi-get-sequences-for-gene-clusters) for further analysis. Gene information is listed in
171 Supplementary **Table S2**.

172 ***Ribosomal Protein Genes (RPGs)***

173 Ribosomal protein gene sets used in Hug et al. (2016) and Park et al. (2018) were
174 retrieved from the same genome dataset using in silico MolecularCloning ver. 7 (In Silico
175 Biology, Inc.) with the exception of three species (*Candidatus Photodesmus blepharus*,
176 *Candidatus Photodesmus katoptron*, and *Vibrio gallaecicus*) due to the lack of certain genes.
177 The region of each gene used in this study is listed in Supplementary **Table S3**.

178 ***Sequence Analysis***

179 The sequences of different gene sets were aligned using MACSE v2.03 [40] or
180 MUSCLE [38] and edited using MEGA-X v10.1.8 [41]. Split decomposition analysis using
181 the concatenated sequence was performed using SplitsTree 4.14.8 with a neighbor net
182 drawing and a Jukes-Cantor correction. The concatenated sequences were also used for a
183 phylogenetic analysis using Maximum Likelihood (ML) method with 500 bootstraps by
184 MEGA-X v10.1.8 [42, 43]. Average Nucleotide Identity (ANI) values were estimated using
185 FastANI [44], PyANI [45] and Orthologous Average Nucleotide Identity Tool version 0.93.1
186 [46]. In silico DNA-DNA hybridization (DDH) values were estimated using a Genome-to-
187 Genome Distance Calculator 2.1 (GGDC) [47, 48]. Average Amino Acid Identity (AAI)
188 values were estimated by AAI calculator, Kostas lab [49].

189

190 RESULTS

191 *Vibrio Clade 3.0: Updated MLSA Network Based on Typical 8-HKGs*

192 8-HKGs MLSA using 191 *Vibrionaceae* strains including 27 newly determined/updated
193 genome sequences revealed a total of 51 distinct clades with 21 newly defined clades, with
194 the other twelve clades remaining unchanged as previously described (**Fig. 1**, Table 1). A
195 description of *Vibrio* Clade 3.0 was obtained in a supplemental document. The robustness of
196 these clades was strong enough to indicate their monophyly in the Maximum Likelihood
197 phylogenetic analyses (**Fig. S1a, S2a**). On the basis of genome analysis, most of the clades
198 shared >71.9% intra-ANI, >18.7% intra-*in silico* DDH, and >66.6% intra-AAI (Table 1), the
199 highest ANI (98.9%) and DDH (90.1%) were observed between *V. qinghaiensis* Q67^T and *V.*
200 *ordalii* ATCC 33509^T, and the highest AAI (98.6%) was observed between *V.*
201 *neocaledonicus* CGJ02-2 and *V. alginolyticus* ATCC 17749^T.

202 In addition, according to the heatmaps of ANI and AAI matrices (**Fig. S3, S4**), AAI
203 heatmap showed higher clade-based hierarchical clustering. The AAI boundaries of clades
204 could be clearly inferred to be around 70%, while the ANI boundaries remained
205 undetermined. Meanwhile, ANI matrices of all genomes for preliminary screening using
206 FastANI and PyANI indicated that 17 species/strains pairs showed values above the species
207 boundary (95%), further confirmation was examined using OrthoANI between these pairs.
208 According to the OrthoANI, 14 of the 17 preliminary pairs must be re-evaluated for further
209 identification (Table 2).

210 *Clade Based Genome Features*

211 The *Vibrionaceae* clades except singletons had the mean 4.7 ± 0.4 Mb genome size and
212 $44.5 \pm 0.8\%$ GC content. Each clade had similar genome size and GC content, but *vibrio*
213 clades overlapped with each other more than other clades in other genera (**Fig. 2**). Even
214 though members of the Hollisae clade comes from different genera (*Enterovibrio* and

215 *Grimontia*), they shared similar size and content. In addition, the *Nigripulchritudo* had the
216 biggest genome (6.3 Mb), while the Halotolerans had the highest GC content (50.6%).

217 ***Single-Copy Genes (SCGs) MLSA***

218 The same 192 genomes were used for pan-genome analysis. A total of 95,334 gene
219 clusters with 844,099 genes were defined in the pangenome, in which 403 gene clusters with
220 82,829 genes were recognized in the core-genome, and 34 single-copy genes were identified
221 by default settings (listed in **Table S2**). The 34 single-copy genes (34-SCGs) were extracted
222 and concatenated from the core-genome for the phylogenetic analysis. According to the split
223 network constructed by concatenated 34-SCGs (**Fig. 3a**), 50 of 51 clades (98%) in 8-HKGs
224 MLSA were congruent despite differing positions, except the singleton *Ganghwense* clade,
225 which was assigned into the neighboring *Rosenbergii* clade. This was consistent with the
226 results in the phylogenetic tree using the same data (**Fig. 4a, S1b**), *P. ganghwense* was
227 grouped in the *Rosenbergii* clade. *Scophthalmi* clade lost the monophyly. Although the
228 network topology base on 34-SCGs MLSA was similar to that of 8-HKGs, it showed
229 comparatively lower gene resolution than 8-HKGs MLSA (**Fig. S5a**).

230 ***Ribosomal Protein Genes (RPGs) MLSA***

231 Two RPGs sets, 16-RPGs from Hug et al. (2016) and 11-RPGs from Park et al. (2018),
232 were examined for MLSA analysis among the 188 *Vibrionaceae* species. Three species (*V.*
233 *gallaecicus* and two *Candidatus* *Photodesmus* species) were removed due to the lack of
234 certain genes.

235 Since these two sets had a similar network (**Fig. 3b, S6**) and gene resolution (98.6% and
236 98.5% respectively, **Fig. S5a**), the 11-RPGs set is recommended as a subject for further
237 research, thus time and effort has been saved in collecting these genes of interest. According
238 to the split network base on the concatenated 11-RPGs nucleotide sequences (**Fig. 3b**), all
239 clades in 8-HKGs MLSA were congruent. Member exchange occurred between two clades:

240 Proteolyticum and Rosenbergi clade. *P. marinum*, which belong to Proteolyticum in 8-
241 HKGs, was clustered in Rosenbergi, and *P. sanctipauli*, which belong to Rosenbergi in 8-
242 HKGs, was clustered in Proteolyticum. It was even worse in the phylogenetic tree
243 constructed using the same data, the members of Proteolyticum, Rosenbergi, and Aquae
244 clades in 8-HKGs interfered with each other here that we could not classify the clades among
245 them (Fig. 4b, S2).

246 *Identification of MAGs Using 11-RPGs*

247 A total of 19 *Vibrionaceae* MAGs from different BioProjects were collected to test
248 identification using 11-RPGs (Table 3). Finally, 10 MAGs were used for the 11-RPGs MLSA
249 after checking the presence of 11-RPGs. Results on the basis of 11-RPGs MLSA using 188
250 *Vibrionaceae* and 10 *Vibrionaceae* MAGs (Fig. 5, Table 3) showed that 7 MAGs (MAG3 -
251 MAG9) were classified into known clades, in which 5 MAGs (MAG5 - MAG9) were further
252 identified as know species based on ANI (MAG6; *V. casei*, MAG7; *V. litoralis*, and MAG5,
253 MAG8, MAG9; *V. campbellii*). MAG3 and MAG4 were estimated to belong to
254 Diazotrophicus and Nereis clade, respectively, but they showed lower ANI below the species
255 boundary (<95%) against any known clade members, which means they are likely to be new
256 species. In addition, three MAGs could not be classified into any known vibrio clades,
257 MAG1 might share common ancestry with Rumoiensis clade species but MAG1 had a distant
258 long branch, and MAG2 was located close to Fischeri clade, but was not clustered with any
259 vibrio species. Moreover, MAG10 was primarily labeled as *Vibrionaceae* MAG but it could
260 be placed to *Aeromonadaceae* (Table 3 and Fig. 5). In further blast, MAG10 was affiliated to
261 be relative of *Tolumonas auensis* DSM 9187^T with the 11 RPG sequence similarity 86.1-
262 94.6%.

263

264 **DISCUSSION**

265 As of October 2020, over 5,000 genomes have been described in *Vibrionaceae*, but only
266 164 genomes with less than 30 complete genomes of *Vibrionaceae* type strains were
267 available in public databases. Twenty-seven genomes (22 complete) of type strains in
268 *Vibrionaceae* have been added or updated in this study (Table S1), which covers nearly 15%
269 of described *Vibrionaceae* species. This could achieve the second update of vibrio
270 evolutionary units and we propose it be “Vibrio Clade 3.0”. The second update of vibrio
271 clades is described in the supplemental document.

272 **Possible Clade Boundaries**

273 In this study, a total of 51 clades (including 17 orphan clades) were described in the
274 family *Vibrionaceae*, which is almost twice the number described in 2013, specifically the
275 increase of singletons (from 4 to 17), showed greater diversity of this family. Members of a
276 clade shared at least 71.9% ANI, 18.7% DDH, and 66.6% AAI (Table 1), which may suggest
277 clear boundaries in classifying clade members in the future. However, in order to improve the
278 accuracy of these boundaries, we would like to suggest excluding the two-member clades
279 since they are more likely to be unstable. Integrating with the heatmaps of ANI and AAI
280 (Fig. S3, S4), the AAI boundary for *Vibrionaceae* clades classification could be 69.5-71.8%,
281 but an apparent ANI clade boundary was not found.

282 **Possible Changes in The Future**

283 According to the 8-HKGs MLSA in this study, some clades were likely to be split into
284 several different branches, indicating different evolutionary directions and potential new
285 clades in the future (Fig. 1, S7). For example, in the Mediterranei clade, two major branches
286 could be found: branch with 1) *V. thalassae*, *V. mediterranei*, and *V. barjaei*, and 2) *V.*
287 *variabilis*, *V. maritimus*, and *V. hangzhouensis*. The same could be found in the Porteressiae
288 clade: branch with 1) *V. palustris* and *V. zhugei*, and 2) *V. porteresiae* and *V. tritonius*.
289 Furthermore, there are still some individual orphan-like species showing distant relationships

290 with other members in a clade, usually occurring at the edge of a clade. Examples include *V.*
291 *gallicus* in the Halioticoli clade, *V. sinaloensis* in the Orientalis clade, *V. fortis* and *V.*
292 *profundi* in the Splendidus clade. All these candidates possess singleton potential or could
293 form a new clade with newly included members. Meanwhile, the classification of clades in
294 the genus *Photobacterium* seemed troublesome since the results differed between different
295 gene sets, particularly between the Rosenbergii and Proteolyticum clades (**Fig. 4**), which may
296 require further studies.

297 We also consider not-yet validly published genus/species in future studies, e.g. genera
298 "*Corallibacterium*" [50] and newly described "*Veronia*" [51]. "*Corallibacterium*" strains
299 were not included in this study due to the lack of genomes, but "*Veronia pacificus*" is a later
300 synonym of *Enterovibrio pacificus*, which was determined to be an orphan clade in this
301 study.

302 ***Misidentification of Species/Strains***

303 Some species/strains in clades were found to be closely related and shared long branches
304 in the split network tree (**Fig. 1**). According to the ANI calculation, they may have been
305 misidentified thus need further confirmation (Table 2). First confirmation was performed
306 using different ANI calculators, those (14 pairs) who reach the boundary value (95%) for
307 species delineation [46, 52] came to the second confirmation by *in silico* DDH checkup; and
308 the final eight pairs were examined by AAI calculation [53, 54]. As shown in Table 2, all the
309 final 8 pairs showed values over the boundaries. However, half of them were subspecies
310 (*Salinivibrio costicola*, *Photobacterium damsela* and *P. leiognathi* subspecies) and one pair
311 (*Aliivibrio logei*-*A. salmonicida*) were not type strains, so are not discussed in this study.
312 Finally, we re-identified 3 species which were likely to have been previously misidentified.
313 *V. chemaguriensis* was newly classified in 2019 from the Harveyi clade, showed 45%
314 GGDC, 92% ANI and 96.2% AAI against to genome of *V. alginolyticus* ATCC 17749^T [55],

315 however, with the new member of *V. diabolicus* was classified into the Harveyi clade, the
316 genome of *V. chemaguriensis* Iso1^T showed 98.1% ANI, 83.1% DDH and 98.4% AAI against
317 *V. diabolicus* CNCMI-1629^T, which indicates they are the same species; another member of
318 this clade, *V. neocaledonicus* CGJ02-2, showed 98.5% ANI, 85.7% DDH and 98.5% AAI
319 against *V. alginolyticus* ATCC 17749^T, was identified as *V. alginolyticus* ATCC 17749^T,
320 consistent with the recent research [56] and the luminescent bacterium *V. qinghaiensis* Q67^T
321 in Anguillarum was identified as *V. ordalii* ATCC 33509^T (98.9% ANI, 90.1% DDH and
322 98.5% AAI). These results highlight the contribution of 8-HKGs MLSA in identifying the
323 new and previously misidentified species/strains in *Vibrionaceae*, which could contribute to
324 further elucidation of species-level ecology more appropriately.

325 ***Delineation Power of 8-HKGs MLSA***

326 Compared to the 8-HKGs, newly examined 34-SCGs and 11-RPGs were determined to
327 be lower resolution (94.9% and 98.5%, separately) (**Fig. S5a**), which caused issues in
328 identifying closely related species/clades. Most of the incongruences were observed in the
329 genus *Photobacterium* (**Fig. 4**), which suggests that both the 34-SCGs and 11-RPGs may not
330 be the best gene set for the identification of *Photobacterium* species. These results indicated
331 that the 8-HKGs MLSA is still an effective and reliable tool in the identification of
332 *Vibrionaceae*. Meanwhile, by means of evaluation of the 8-HKGs individually, six of the
333 eight housekeeping genes showed relatively higher gene resolution (**Fig. S5b**), which
334 indicates that it is highly possible that we could use the 6-HKGs set for future research.

335

336 ***Availability of a Potential RPG Set for Metagenomes***

337 The success of identifications of MAGs proved the potential of the 11-RPGs set in
338 classifying or identifying (new) species/clades using metagenomes. However, there were still
339 several MAGs that could not be used for identification due to the lack of certain ribosomal

340 protein genes. The reason seems to depend on genome completeness and complete regions of
341 the genome. Most of the MAGs, which had all 11-RPGs hits, showed over 90.9%
342 completeness (checked by DFAST, <https://dfast.ddbj.nig.ac.jp/>), except for MAG5 and
343 MAG10, while other MAGs, which had only eight or even none hits showed lower
344 completeness (average 61.3%). Unfortunately, we did not find any available SAGs data to
345 test for the identification using the 11-RPGs set, but its application was believed as well. As a
346 result, it is necessary to obtain more complete genomes or decrease the number of genes for
347 analysis, which would be needed in future research. More suitable and reliable ribosomal
348 gene sets should be evaluated since S17 was deleted in the 11-RPGs set but had the highest
349 gene resolution (**Fig. S5c**). More attempts are needed for the identification of new species,
350 and its application for other bacterial families to be demonstrated as well. Moreover,
351 evaluation of 8-HKGs or its reduced set must be performed to study further population
352 studies using MAGs/SAGs data set because only 8-HKGs or its reduced set has high
353 delineation powers in differentiate populations [7].

354 This MLSA approach established in this study could help to further understanding of
355 not-yet-cultured vibrios. MAGs used in this study recovered from various materials
356 (bioreactor, invertebrates, and environments) and diverse environments (surface seawater,
357 sediment, and hydrothermal vent), and the new species candidates were found in the
358 sediment-related samples (MAG1, and MAG2), and marine phytoplankton exometabolite
359 enrichments (MAG3, and MAG4). In particular, MAG2 was reconstructed from a
360 Foraminifera, *Globobulimina* sp., obtained from sediment in Gullmar Fjord (Sweden), and
361 estimated to belong in not only a new genus but also a new clade (**Fig. 5**). Unveiling new
362 species candidates from marine environmental MAGs/SAGs are not surprising considering
363 the pace in new species description in recent decades, but those from marine protozoans

364 associated taxa might be a new frontier to expand *Vibrionaceae* ecology, or more specifically
365 to reveal unexpected host-microbe associations.

366 **Conclusion**

367 The new phylogenetic analysis using 191 genomes for family *Vibrionaceae* was updated
368 with a total of 51 distinct clades including 21 newly defined ones. According to the
369 comparison with two new approaches, the 8-HKGs MLSA is still an effective and reliable
370 tool for delineating new species, monophyletic groups/clades in *Vibrionaceae*. Using the
371 dataset in this study, 96.1% ANI may be the boundary for species delineation, 69.5-71.8%
372 AAI may be the boundary for *Vibrionaceae* clade delineation. The success of identification in
373 *Vibrionaceae* using MAGs showed the potential of the 11-RPGs set in classifying or
374 identifying species candidates in MAG or SAG applications. This is the most comprehensive
375 study to date of the family *Vibrionaceae* with both the most genera and the most species
376 described. However, more efforts on acquiring genomes for the remaining *Vibrionaceae*
377 species will improve and perfect the phylogenetic analysis to better illustrate the ecological
378 diversity, evolutionary history, and host-microbes interactions of the family *Vibrionaceae*.
379 Finally, this methodology could be a universal tool, so it could apply to any bacterial taxa
380 after evaluating the HKGs and RPGs set suitability.

381

382

383 **ACKNOWLEDGMENTS**

384 We sincerely thank Tomomi Shimizu for technical assistance. This work was partly
385 supported by KAKEN 19H03041.

386

387 **DATA AVAILABILITY**

388 The whole genome sequence data obtained in this study was deposited at
389 DDBJ/EMBL/GenBank under BioProject Accession: PRJDB11924.

390

391 **Author Contributions**

392 Chunqi Jiang conceived, designed and performed the experiments, analyzed the data,
393 visualized the data, drafted and reviewed the manuscript.

394 Mami Tanaka and Sayo Nishikawa performed the experiments, reviewed the manuscript.

395 Sayaka Mino, Jesús L. Romalde, Fabiano L. Thompson and Bruno Gomez-Gil analyzed the
396 data, reviewed the manuscript.

397 Tomoo Sawabe conceived and designed the experiments, reviewed the manuscript.

398

399 **COMPETING INTERESTS**

400 The authors declare no competing interests.

401

402 **REFERENCES**

403 1. Thompson FL, Iida T, Swings J (2004) Biodiversity of Vibrios. *Microbiol Mol Biol Rev*
404 68:403–431. <https://doi.org/10.1128/MMBR.68.3.403-431.2004>

405 2. Thompson CC, Vicente ACP, Souza RC, et al (2009) Genomic taxonomy of vibrios.
406 *BMC Evol Biol* 9:1–16. <https://doi.org/10.1186/1471-2148-9-258>

407 3. Gomez-Gil B, Thompson CC, Matsumura Y, et al (2014) The Family *Vibrionaceae*. in:
408 *The Prokaryotes*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 659–747

409 4. Auch, A.F., von Jan, M., Klenk, HP. et al (2010) Digital DNA-DNA hybridization for
410 microbial species delineation by means of genome-to-genome sequence comparison.

411 *Stand in Genomic Sci* 2 117–134. <https://doi.org/10.4056/sigs.531120>

- 412 5. Thompson CC, Chimetto L, Edwards RA, et al (2013) Microbial genomic taxonomy.
413 BMC Genomics 14:. <https://doi.org/10.1186/1471-2164-14-913>
- 414 6. Sawabe T, Kita-Tsukamoto K, Thompson FL (2007) Inferring the evolutionary history
415 of vibrios by means of multilocus sequence analysis. J Bacteriol 189:7932–7936.
416 <https://doi.org/10.1128/JB.00693-07>
- 417 7. Thompson FL, Gomez-Gil B, Vasconcelos ATR, Sawabe T (2007) Multilocus sequence
418 analysis reveals that *Vibrio harveyi* and *V. campbellii* are distinct species. Appl Environ
419 Microbiol 73:4279–4285. <https://doi.org/10.1128/AEM.00020-07>
- 420 8. López-Hermoso C, de la Haba RR, Sánchez-Porro C, et al (2017) Assessment of
421 multilocus sequence analysis as a valuable tool for the classification of the genus
422 *Salinivibrio*. Front Microbiol 8:1–14. <https://doi.org/10.3389/fmicb.2017.01107>
- 423 9. Young JM, Park D-C, Shearman HM, Fargier E (2008) A multilocus sequence analysis
424 of the genus *Xanthomonas*. Syst Appl Microbiol 31:366–377.
425 <https://doi.org/10.1016/j.syapm.2008.06.004>
- 426 10. Glaeser SP, Kämpfer P (2015) Multilocus sequence analysis (MLSA) in prokaryotic
427 taxonomy. Syst. Appl. Microbiol. 38:237–245
- 428 11. Cano-Gomez A, Høj L, Owens L, Andreakis N (2011) Multilocus sequence analysis
429 provides basis for fast and reliable identification of *Vibrio harveyi*-related species and
430 reveals previous misidentification of important marine pathogens. Syst Appl Microbiol
431 34:561–565. <https://doi.org/10.1016/j.syapm.2011.09.001>
- 432 12. Guo Y, Zheng W, Rong X, Huang Y (2008) A multilocus phylogeny of the
433 *Streptomyces griseus* 16S rRNA gene clade: use of multilocus sequence analysis for
434 streptomycete systematics. Int J Syst Evol Microbiol 58:149–159.
435 <https://doi.org/10.1099/ijs.0.65224-0>

- 436 13. Sawabe T, Ogura Y, Matsumura Y, et al (2013) Updating the *Vibrio* clades defined by
437 multilocus sequence phylogeny: Proposal of eight new clades, and the description of
438 *Vibrio tritonius* sp. nov. *Front Microbiol* 4:1–14.
439 <https://doi.org/10.3389/fmicb.2013.00414>
- 440 14. Medini D, Donati C, Tettelin H, et al (2005) The microbial pan-genome. *Curr Opin*
441 *Genet Dev* 15:589–594. <https://doi.org/10.1016/j.gde.2005.09.006>
- 442 15. Tettelin H, Riley D, Cattuto C, Medini D (2008) Comparative genomics: the bacterial
443 pan-genome. *Curr Opin Microbiol* 11:472–477.
444 <https://doi.org/10.1016/j.mib.2008.09.006>
- 445 16. Vernikos G, Medini D, Riley DR, Tettelin H (2015) Ten years of pan-genome analyses.
446 *Curr Opin Microbiol* 23:148–154. <https://doi.org/10.1016/j.mib.2014.11.016>
- 447 17. Tettelin H, Masignani V, Cieslewicz MJ, et al (2005) Genome analysis of multiple
448 pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial “pan-
449 genome.” *Proc Natl Acad Sci* 102:13950–13955.
450 <https://doi.org/10.1073/pnas.0506758102>
- 451 18. Kim Y, Gu C, Kim HU, Lee SY (2020) Current status of pan-genome analysis for
452 pathogenic bacteria. *Curr Opin Biotechnol* 63:54–62.
453 <https://doi.org/10.1016/j.copbio.2019.12.001>
- 454 19. Eren AM, Esen ÖC, Quince C, et al (2015) Anvi'o: an advanced analysis and
455 visualization platform for ‘omics data. *PeerJ* 3:e1319. <https://doi.org/10.7717/peerj.1319>
- 456 20. Delmont TO, Eren AM (2018) Linking pangenomes and metagenomes: the
457 *Prochlorococcus* metapangenome. *PeerJ* 6:e4320. <https://doi.org/10.7717/peerj.4320>
- 458 21. Sidhu C, Saini MK, Srinivas Tanuku NR, Pinnaka AK (2019) *Arenibacter amylolyticus*
459 sp. nov., an amylase-producing bacterium of the family *Flavobacteriaceae* isolated from

- 460 marine water in India. *Int J Syst Evol Microbiol* 71:.
461 <https://doi.org/10.1099/ijsem.0.004664>
- 462 22. Tanaka M, Kumakura D, Mino S, et al (2020) Genomic characterization of closely
463 related species in the Rumoiensis clade infers ecogenomic signatures to non-marine
464 environments. *Environ Microbiol* 22:3205–3217. [https://doi.org/10.1111/1462-
465 2920.15062](https://doi.org/10.1111/1462-2920.15062)
- 466 23. Aguirre-Sanchez JR, Ibarra-Rodriguez JR, Vega-Lopez IF, et al (2021) Genomic
467 signatures of adaptation to natural settings in non-typhoidal *Salmonella enterica*
468 Serovars Saintpaul, Thompson and Weltevreden. *Infect Genet Evol* 90:104771.
469 <https://doi.org/10.1016/j.meegid.2021.104771>
- 470 24. Hug LA, Baker BJ, Anantharaman K, et al (2016) A new view of the tree of life. *Nat*
471 *Microbiol* 1:16048. <https://doi.org/10.1038/nmicrobiol.2016.48>
- 472 25. Parks DH, Chuvochina M, Waite DW, et al (2018) A standardized bacterial taxonomy
473 based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 36:996.
474 <https://doi.org/10.1038/nbt.4229>
- 475 26. Hendry TA, Freed LL, Fader D, et al (2018) Ongoing transposon-mediated genome
476 reduction in the luminous bacterial symbionts of deep-sea ceratioid anglerfishes. *MBio*
477 9: <https://doi.org/10.1128/mBio.01033-18>
- 478 27. Baker LJ, Freed LL, Easson CG, et al (2019) Diverse deep-sea anglerfishes share a
479 genetically reduced luminous symbiont that is acquired from the environment. *Elife* 8:.
480 <https://doi.org/10.7554/eLife.47606>
- 481 28. Collins FWJ, Walsh CJ, Gomez-Sala B, et al (2021) The microbiome of deep-sea fish
482 reveals new microbial species and a sparsity of antibiotic resistance genes. *Gut Microbes*
483 13:1–13. <https://doi.org/10.1080/19490976.2021.1921924>

- 484 29. Gould AL, Fritts-Penniman A, Gaisiner A (2021) Museum genomics illuminate the high
485 specificity of a bioluminescent symbiosis for a genus of reef fish. *Front Ecol Evol* 9:.
486 <https://doi.org/10.3389/fevo.2021.630207>
- 487 30. Tu Q, He Z, Zhou J (2014) Strain/species identification in metagenomes using genome-
488 specific markers. *Nucleic Acids Res* 42:e67–e67. <https://doi.org/10.1093/nar/gku138>
- 489 31. Tanaka M, Hongyu B, Jiang C, et al (2020) *Vibrio taketomensis* sp. nov. by genome
490 taxonomy. *Syst Appl Microbiol* 43:.. <https://doi.org/10.1016/j.syapm.2019.126048>
- 491 32. Wick RR, Judd LM, Holt KE (2018) Deepbiner: Demultiplexing barcoded Oxford
492 Nanopore reads with deep convolutional neural networks. *PLOS Comput Biol*
493 14:e1006583. <https://doi.org/10.1371/journal.pcbi.1006583>
- 494 33. Kajitani R, Yoshimura D, Ogura Y, et al (2020) Platanus_B: an accurate de novo
495 assembler for bacterial genomes using an iterative error-removal process. *DNA Res* 27:..
496 <https://doi.org/10.1093/dnares/dsaa014>
- 497 34. Wick RR, Judd LM, Gorrie CL, Holt KE (2017) Unicycler: Resolving bacterial genome
498 assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595.
499 <https://doi.org/10.1371/journal.pcbi.1005595>
- 500 35. Kolmogorov M, Yuan J, Lin Y, Pevzner PA (2019) Assembly of long, error-prone reads
501 using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019->
502 0072-8
- 503 36. Vaser R, Sović I, Nagarajan N, Šikić M (2017) Fast and accurate de novo genome
504 assembly from long uncorrected reads. *Genome Res* 27:737–746.
505 <https://doi.org/10.1101/gr.214270.116>
- 506 37. Walker BJ, Abeel T, Shea T, et al (2014) Pilon: An integrated tool for comprehensive
507 microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963.
508 <https://doi.org/10.1371/journal.pone.0112963>

- 509 38. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high
510 throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>
- 511 39. Van Dongen S, Abreu-Goodger C (2012) Using MCL to Extract Clusters from
512 Networks. pp 281–295
- 513 40. Ranwez V, Harispe S, Delsuc F, Douzery EJP (2011) MACSE: Multiple alignment of
514 coding sequences accounting for frameshifts and stop codons. *PLoS One* 6:e22594.
515 <https://doi.org/10.1371/journal.pone.0022594>
- 516 41. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular evolutionary genetics
517 analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874.
518 <https://doi.org/10.1093/molbev/msw054>
- 519 42. Nei M, Kumar S (2000) *Molecular evolution and phylogenetics*. Oxford university press
- 520 43. Kumar S, Stecher G, Li M, et al (2018) MEGA X: Molecular evolutionary genetics
521 analysis across computing platforms. *Mol Biol Evol* 35:1547–1549.
522 <https://doi.org/10.1093/molbev/msy096>
- 523 44. Jain C, Rodriguez-R LM, Phillippy AM, et al (2018) High throughput ANI analysis of
524 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9:5114.
525 <https://doi.org/10.1038/s41467-018-07641-9>
- 526 45. Pritchard L, Glover RH, Humphris S, et al (2016) Genomics and taxonomy in
527 diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods*
528 8:12–24. <https://doi.org/10.1039/C5AY02550H>
- 529 46. Lee I, Ouk Kim Y, Park S-C, Chun J (2016) OrthoANI: An improved algorithm and
530 software for calculating average nucleotide identity. *Int J Syst Evol Microbiol* 66:1100–
531 1103. <https://doi.org/10.1099/ijsem.0.000760>

- 532 47. Meier-Kolthoff JP, Göker M, Spröer C, Klenk HP (2013) When should a DDH
533 experiment be mandatory in microbial taxonomy? Arch Microbiol 195:413–418.
534 <https://doi.org/10.1007/s00203-013-0888-4>
- 535 48. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M (2013) Genome sequence-based
536 species delimitation with confidence intervals and improved distance functions. BMC
537 Bioinformatics 14:60. <https://doi.org/10.1186/1471-2105-14-60>
- 538 49. Rodriguez-R LM, Konstantinidis KT (2014) Bypassing cultivation to identify bacterial
539 species. Microbe Mag 9:111–118. <https://doi.org/10.1128/microbe.9.111.1>
- 540 50. Hettiarachchi SA, Lee S-J, Lee Y, et al (2018) *Corallibacterium pacifica* gen. nov., sp.
541 nov., a novel bacterium of the family *Vibrionaceae* isolated from Hard Coral. Curr
542 Microbiol 75:835–841. <https://doi.org/10.1007/s00284-018-1455-7>
- 543 51. Gomez-Gil B, González-Castillo A, Aguilar-Méndez MJ, et al (2021) *Veronia*
544 *nyctiphanis* gen. nov., sp. nov., Isolated from the Stomach of the Euphausiid
545 *Nyctiphanes simplex* (Hansen, 1911) in the Gulf of California, and Reclassification of
546 *Enterovibrio pacificus* as *Veronia pacifica* comb. nov. Curr Microbiol 78:3782–3790.
547 <https://doi.org/10.1007/s00284-021-02627-1>
- 548 52. Richter M, Rosselló-Móra R (2009) Shifting the genomic gold standard for the
549 prokaryotic species definition. Proc Natl Acad Sci 106:19126–19131.
550 <https://doi.org/10.1073/pnas.0906412106>
- 551 53. Goris J, Konstantinidis KT, Klappenbach JA, et al (2007) DNA–DNA hybridization
552 values and their relationship to whole-genome sequence similarities. Int J Syst Evol
553 Microbiol 57:81–91. <https://doi.org/10.1099/ijs.0.64483-0>
- 554 54. Kim M, Oh H-S, Park S-C, Chun J (2014) Towards a taxonomic coherence between
555 average nucleotide identity and 16S rRNA gene sequence similarity for species

- 556 demarcation of prokaryotes. *Int J Syst Evol Microbiol* 64:346–351.
557 <https://doi.org/10.1099/ijms.0.059774-0>
- 558 55. Ghosh A, Bhadury P (2019) *Vibrio chemaguriensis* sp. nov., from Sundarbans, Bay of
559 Bengal. *Curr Microbiol* 76:1118–1127. <https://doi.org/10.1007/s00284-019-01731-7>
- 560 56. Tan L, Gómez-Betancur I, Guo S, et al (2020) Complete genome of *Vibrio*
561 *neocaledonicus* CGJ02-2, an active compounds producing bacterium isolated from
562 South China Sea. *Curr Microbiol* 77:2665–2673. [https://doi.org/10.1007/s00284-020-](https://doi.org/10.1007/s00284-020-02047-7)
563 [02047-7](https://doi.org/10.1007/s00284-020-02047-7)
- 564

565 **FIGURE CAPTIONS**

566 Figure 1. The updated concatenated split network based on multilocus sequence analysis
567 (MLSA) of eight housekeeping genes (8-HKGs) retrieved from 191 *Vibrionaceae*
568 species. The *ftsZ*, *gapA*, *gyrB*, *mreB*, *pyrH*, *recA*, *rpoA*, and *topA* gene sequences were
569 concatenated and the tree was reconstructed using the SplitsTree4 ver. 4.14.8. Clades
570 indicated by red, green, and blue represent the “new”, “emended”, and “un-changed”
571 clades, respectively.

572 Figure 2. Clade-based genome size and GC content relatedness of all 191 *Vibrionaceae*
573 species used in this study. Colors indicate different clades except singletons which are
574 all black. The size of shapes indicates the number of species in Clades. Error bars
575 indicate the standard deviation. Data were visualized with ggplot2 ver. 3.3.3.

576 Figure 3. Concatenated split network based on nucleotide sequences of a) 34 single-copy core
577 genes (34-SCGs) retrieved from 191 *Vibrionaceae* species pan-genome, b) 11 ribosomal
578 protein genes (11-RPGs) retrieved from 188 *Vibrionaceae* species. Gene sequences were
579 concatenated, and the tree was reconstructed using the SplitsTree4 ver. 4.14.8. Clade
580 underlined was not conserved compared to 8-HKGs MLSA. Colour was set as same as
581 Figure 1.

582 Figure 4. The rooted Maximum Likelihood bootstrap consensus tree based on a) 34-SCGs vs
583 8-HKGs, b) 11-RPGs vs 8-HKGs; each clade is marked by different colours (colour was
584 set as same as Figure 1, black branches are the singletons), shadowed clades with labels
585 were incongruent compared to 8-HKGs. The trees with all leaves labeled are available in
586 Supplemental Figure S4 and S5, respectively.

587 Figure 5. Metagenome-assembled genomes (MAGs) identification using concatenated split
588 network based on nucleotide sequences of 11-RPGs retrieved from 188 *Vibrionaceae*

589 species and 10 MAGs. Gene sequences were concatenated, and the tree was
590 reconstructed using the SplitsTree4 ver. 4.14.8.

591 Figure S1. The rooted Maximum Likelihood bootstrap consensus tree based on a) 8
592 housekeeping genes, b) 34 single-copy genes. Branches of each clade are marked by
593 different colors (color was set as same as Figure 2).

594 Figure S2. The rooted Maximum Likelihood bootstrap consensus tree based on a) 8
595 housekeeping genes, b) 11 ribosomal protein genes. Branches of each clade are marked
596 by different colors (color was set as same as Figure 2).

597 Figure S3. Heatmap representation based on the average nucleotide identity (ANI) distance
598 matrix. Dendrogram shows the hierarchical clustering. Species ID is listed in Table S1.
599 Heatmap was visualized with ComplexHeatmap ver. 2.2.0.

600 Figure S4. Heatmap representation based on the average amino identity (AAI) distance
601 matrix. Dendrogram shows the hierarchical clustering. Species ID is listed in Table S1.
602 Clades (over 2 members) were labeled with the same color as Figure 2, asterisks indicate
603 the split clades. Heatmap was visualized with ComplexHeatmap ver. 2.2.0.

604 Figure S5. Gene similarity of a) concatenated gene sets, b) eight housekeeping genes, and c)
605 16 ribosomal protein genes (bold indicates 11-RPGs set), circles represent different
606 species, red square indicates the median value, lower gene similarity means higher gene
607 resolution.

608 Figure S6. Concatenated split network based on nucleotide sequences of 16 ribosomal protein
609 genes retrieved from 188 *Vibrionaceae* species. Gene sequences were concatenated, and
610 the tree was reconstructed using the SplitsTree4 ver. 4.14.8.
611 Color was set as same as Figure 1.

612 Figure S7. Focused concatenated split networks for each clade (except for singletons) based
613 on 8-HKGs with *E. coli* and *V. cholerae* as outgroups.

614 Table 1. Newly obtained/updated genome sequences in this study

Genus	Species	Strains	Status	Accession
<i>Photobacterium</i>	<i>carosum</i>	CECT 9394 ^T	Draft	BPPU01000001-BPPU01000011
<i>Photobacterium</i>	<i>sanguinancrri</i>	CECT 7579 ^T	Complete	AP024850-AP024851
<i>Photobacterium</i>	<i>swingsii</i>	CECT 7576 ^T	Complete	AP024852-AP024853
<i>Photobacterium</i>	<i>toruni</i>	CECT 9189 ^T	Draft	AP024854-AP024860
<i>Vibrio</i>	<i>aerogenes</i>	LMG 19650 ^T	Complete	AP024861-AP024863
<i>Vibrio</i>	<i>agarivorans</i>	CECT 5085^T	Draft	BLAT01000001-BLAT01000058
<i>Vibrio</i>	<i>breoganii</i>	CAIM 1829 ^T	Complete	AP024864-AP024865
<i>Vibrio</i>	<i>comitans</i>	LMG 23416 ^T	Complete	AP024866-AP024868
<i>Vibrio</i>	<i>ezuræ</i>	JCM 21522^T	Complete	AP024869-AP024870
<i>Vibrio</i>	<i>gallicus</i>	LMG 21878 ^T	Complete	AP024871-AP024872
<i>Vibrio</i>	<i>gazogenes</i>	ATCC 29988 ^T	Complete	AP024873-AP024874
<i>Vibrio</i>	<i>haliotocoli</i>	IAM 14596^T	Complete	AP024875-AP024877
<i>Vibrio</i>	<i>inusitatus</i>	LMG 23434 ^T	Complete	AP024878-AP024880
<i>Vibrio</i>	<i>ishigakensis</i>	JCM 19231^T	Complete	AP024881-AP024882
<i>Vibrio</i>	<i>mangrovi</i>	CECT 7927 ^T	Complete	AP024883-AP024884
<i>Vibrio</i>	<i>neonatus</i>	JCM 21521 ^T	Complete	AP024885-AP024886
<i>Vibrio</i>	<i>palustris</i>	CECT 9027 ^T	Complete	AP024887-AP024888
<i>Vibrio</i>	<i>pectenicida</i>	LMG 19642 ^T	Complete	AP024889-AP024892
<i>Vibrio</i>	<i>plantisponsor</i>	CECT 7581 ^T	Complete	AP024893-AP024894
<i>Vibrio</i>	<i>porteresiae</i>	MSSRF30 ^T	Complete	AP024895-AP024896
<i>Vibrio</i>	<i>quintilis</i>	CECT 7734 ^T	Draft	AP024897-AP024899
<i>Vibrio</i>	<i>rarus</i>	LMG 23674 ^T	Complete	AP024900-AP024902
<i>Vibrio</i>	<i>rhizosphaerae</i>	LMG 23790 ^T	Complete	AP024903-AP024904
<i>Vibrio</i>	<i>ruber</i>	LMG 23124 ^T	Complete	AP024905-AP024906
<i>Vibrio</i>	<i>spartinae</i>	CECT 9026 ^T	Complete	AP024907-AP024908
<i>Vibrio</i>	<i>superstes</i>	JCM 21480^T	Complete	AP024909-AP024910
<i>Vibrio</i>	<i>zhugei</i>	KCTC 62784 ^T	Complete	AP024911-AP024912

615 Bold indicates newly updated genomes

616

617

618 Table 2. Update of clades description using 8 housekeeping genes for multilocus sequence analysis (MLSA) in
 619 the family *Vibrionaceae*

Clades	Genus	Species	Number of species	Genome size (Mb) Mean±SD	GC content (%) Mean±SD	ANI (%)	DDH (%)	AAI (%)
New clades (21)								
Photodesmus	<i>Candidatus Photodesmus</i>	<i>Candidatus Photodesmus blepharus</i> and <i>Candidatus Photodesmus katoptron</i>	2	1.1±0.1	33.2±3.4	73.1	18.7	66.6
Pacificus	<i>Enterovibrio</i>	<i>E. pacificus</i>	1	5.3	45.5	-	-	-
Hollisae	<i>Grimontia</i> / <i>Enterovibrio</i>	<i>E. baiacu</i>, <i>E. calviensis</i>, <i>E. coralii</i>, <i>E. nigricanis</i>, <i>E. norvegicus</i>, <i>G. celer</i>, <i>G. indica</i>, <i>G. marina</i>, and <i>G. hollisae</i>	9	5.3±0.5	48.1±0.8	74.6-89.9	19.5-38.5	76.5-94.6
Marinum	<i>Paraphotobacterium</i>	<i>Paraphotobacterium marinum</i>	1	2.6	31.2	-	-	-
Aquae	<i>Photobacterium</i>	<i>P. aquae</i>	1	5.1	49.1	-	-	-
Halotolerans	<i>Photobacterium</i>	<i>P. halotolerans</i> and <i>P. salinisoli</i>	2	4.7±0.0	50.6±0.5	91.1	42.5	94.6
Jeanii	<i>Photobacterium</i>	<i>P. jeanii</i> , <i>P. sagunicancrini</i> , and <i>P. swingsii</i>	3	5.4±0.3	44.1±0.9	77.6-85.0	21.8-28.8	82.6-91.2
Proteolyticum	<i>Photobacterium</i>	<i>P. alginatilyticum</i> , <i>P. chitinilyticum</i> , <i>P. marinum</i> , and <i>P. proteolyticum</i>	4	6.1±0.5	47.1±0.9	76.6-93.4	21.0-52.3	79.4-95.1
Costicola	<i>Salinivibrio</i>	<i>S. costicola</i> subsp. <i>alcaliphilus</i> , <i>S. costicola</i> subsp. <i>costicola</i> , <i>S. kushneri</i> , <i>S. proteolyticus</i> , <i>S. shanensis</i> , <i>S. siamensis</i> , and <i>S. socompensis</i>	7	3.4±0.1	49.9±0.6	78.6-98.5	21.7-86.2	84.5-98.3
Occultus	<i>Thaumasiovibrio</i>	<i>T. occultus</i> and <i>T. subtropicus</i>	2	4.9±0.7	48.3±1.6	71.9	20.8	68.5
Aestivus	<i>Vibrio</i>	<i>V. aestivus</i> and <i>V. mexicanus</i>	2	4.8±0.0	44.9±0.2	90.0	39.5	92.7
Albus	<i>Vibrio</i>	<i>V. albus</i>	1	4.3	46.1	-	-	-
Caribbeanicus	<i>Vibrio</i>	<i>V. caribbeanicus</i>	1	4.4	41.6	-	-	-
Fluvialis	<i>Vibrio</i>	<i>V. fluvialis</i> and <i>V. furnisii</i>	2	4.9±0.1	50.3±0.5	86.0	30.5	91.8
Maerlii	<i>Vibrio</i>	<i>V. maerlii</i>	1	4.8	44.5	-	-	-
Metschnikovii	<i>Vibrio</i>	<i>V. cincinnatiensis</i> , <i>V. fujianensis</i> , <i>V. injenensis</i> , and <i>V. metschnikovii</i>	4	3.7±0.1	43.8±0.3	76.5-91.9	20.2-45.7	80.5-95.3
Ostreicida	<i>Vibrio</i>	<i>V. ostreicida</i>	1	4.4	45.6	-	-	-
Pacini	<i>Vibrio</i>	<i>V. pacinii</i> and <i>V. salilacus</i>	2	4.0±0.5	45.2±0.1	95.7	65.9	95.5
Sonorensis	<i>Vibrio</i>	<i>V. sonorensis</i>	1	4.8	44.4	-	-	-
Viridaestus	<i>Vibrio</i>	<i>V. viridaestus</i>	1	4.7	41.9	-	-	-
Xiamenensis	<i>Vibrio</i>	<i>V. xiamenensis</i>	1	5.5	46.4	-	-	-
Emended clades (18)								
Fischeri	<i>Aliivibrio</i>	<i>A. finisterrensis</i>, <i>A. fischeri</i>, <i>A. logei</i>, <i>A. sijtae</i>, <i>A. wodanis</i> and <i>A. salmonicida</i>	6	4.5±0.3	38.7±0.5	79.3-97.5	22.2-79.7	83.5-97.1
Ganghwense	<i>Photobacterium</i>	<i>P. ganghwense</i>	1	5.5	50.5	-	-	-

Phosphoreum	<i>Photobacterium</i>	<i>P. andalusiense</i> , <i>P. angustum</i> , <i>P. aquimaris</i> , <i>P. carnosum</i> , <i>P. iliopscarium</i> , <i>P. kishitanii</i> , <i>P. malactanum</i> , <i>P. piscicola</i> , <i>P. leiognathi</i> subsp. <i>leiognathi</i> , <i>P. leiognathi</i> . subsp. <i>mandapamensis</i> , <i>P. phosphoreum</i> , <i>P. toruni</i> , and <i>P. damselae</i> . subsp. <i>damselae</i>	13	4.6±0.2	39.5±0.7	75.7- 97.6	20.5- 78.2	76.7- 97.8
Profundum	<i>Photobacterium</i>	<i>P. frigidiphilum</i> , <i>P. indicum</i> , <i>P. lipolyticum</i> , and <i>P. profundum</i>	4	5.9±0.7	42.7±2.1	76.7- 95.7	21.7- 63.8	79.0- 95.8
Rosenbergii	<i>Photobacterium</i>	<i>P. aphoticum</i> , <i>P. gaetbulicola</i> , <i>P. lutimaris</i> , <i>P. rosenbergii</i> , and <i>P. sanctipauli</i>	5	5.9±0.4	48.5±1.1	75.5- 83.0	21.2- 26.3	76.5- 89.5
Anguillarum	<i>Vibrio</i>	<i>V. aestuarianus</i> , <i>V. anguillarum</i> , <i>V. ordalii</i> , and <i>V. qinghaiensis</i>	4	4.0±0.4	44.1±1.0	78.7- 98.9	21.5- 90.1	84.1- 98.5
Cholerae	<i>Vibrio</i>	<i>V. cholerae</i> , <i>V. metoecus</i> , <i>V. mimicus</i> , and <i>V. parillis</i>	4	4.1±0.2	46.8±0.5	85.0- 88.2	29.1- 35.2	91.4- 93.1
Diazotrophicus	<i>Vibrio</i>	<i>V. diazotrophicus</i> and <i>V. plantisponsor</i>	2	4.6±0.2	43.7±0.3	94.2	56.7	96.5
Harveyi	<i>Vibrio</i>	<i>V. alfacensis</i> , <i>V. alginolyticus</i> , <i>V. azureus</i> , <i>V. campbellii</i> , <i>V. chemaguriensis</i> , <i>V. diabolicus</i> , <i>V. harveyi</i> , <i>V. hyugaensis</i> , <i>V. jasicida</i> , <i>V. mytili</i> , <i>V. natriegens</i> , <i>V. neocaledonicus</i> , <i>V. owensii</i> , <i>V. parahaemolyticus</i> , <i>V. rotiferianus</i> , and <i>V.</i> <i>sagamiensis</i>	16	5.2±0.5	44.4±1.2	75.2- 98.4	20.9- 85.7	76.1- 98.6
Marisflavi	<i>Vibrio</i>	<i>V. marisflavi</i>	1	4.7	42.2	-	-	-
Mediterranei	<i>Vibrio</i>	<i>V. barjaei</i> , <i>V. hangzhouensis</i> , <i>V. maritimus</i> , <i>V. mediterranei</i> , <i>V. thalassae</i> , and <i>V. variabilis</i>	6	5.4±0.3	45.3±1.2	76.7- 95.0	20.6- 61.3	81.1- 97.2
Nereis	<i>Vibrio</i>	<i>V. hepatarius</i> and <i>V. nereis</i>	2	4.5±0.5	45.8±0.6	79.3	22.3	85.4
Orientalis	<i>Vibrio</i>	<i>V. aquaticus</i> , <i>V. atypicus</i> , <i>V. bivalvicida</i> , <i>V. brasiliensis</i> , <i>V. europaeus</i> , <i>V. galathea</i> , <i>V. orientalis</i> , <i>V. ouci</i> , <i>V. tubiashii</i> , <i>V. sinaloensis</i> and <i>V. xuii</i>	11	5.0±0.6	45.0±0.6	75.8- 94.2	20.1- 56.1	80.4- 95.8
Pectenida	<i>Vibrio</i>	<i>V. pectenida</i>	1	4.5	41.3	-	-	-
Porteresiae	<i>Vibrio</i>	<i>V. palustris</i> , <i>V. porteresiae</i> , <i>V. tritonius</i> , and <i>V. zhugei</i>	4	4.6±0.9	44.4±0.9	72.5- 84.8	19.2- 28.4	71.8- 91.1
Scophthalmi	<i>Vibrio</i>	<i>V. ichthyenteri</i> , <i>V. renipiscarius</i> , <i>V. scophthalmi</i> , and <i>V. sinensis</i>	4	4.8±0.4	44.2±1.2	74.9- 84.4	21.3- 28.4	76.6- 89.9
Splendidus	<i>Vibrio</i>	<i>V. atlanticus</i> , <i>V. celticus</i> , <i>V. chagasii</i> , <i>V. fortis</i> , <i>V. corallirubri</i> , <i>V. crassostreae</i> , <i>V. gigantis</i> , <i>V. cyclitrophicus</i> , <i>V. echinoideorum</i> , <i>V. lentus</i> , <i>V. gallaecicus</i> , <i>V. kanaloae</i> , <i>V. profundus</i> , <i>V. splendidus</i> , <i>V. tasmaniensis</i> , and <i>V. toranzoniae</i>	16	5.2±0.4	43.9±0.8	76.1- 94.8	21.2- 59.4	78.7- 96.8
Vulnificus	<i>Vibrio</i>	<i>V. cidicii</i> , <i>V. navarrensis</i> , and <i>V. vulnificus</i>	3	4.7±0.3	47.7±0.9	77.1- 95.4	21.1- 62.8	79.8- 94.8

Unchanged clades (12)								
Damselae	<i>Photobacterium</i>	<i>P. damsela</i> subsp. <i>damsela</i> and <i>P. damsela</i> subsp. <i>piscicida</i>	2	4.7±0.5	40.9±0.1	97.0	74.2	97.0
Agarivorans	<i>Vibrio</i>	<i>V. agarivorans</i> and <i>V. astriarenae</i>	2	4.8±0.1	45.5±0.2	79.3	22.8	86.5
Coralliilyticus	<i>Vibrio</i>	<i>V. coralliilyticus</i> and <i>V. neptunius</i>	2	5.4±0.3	45.3±0.5	87.6	34.2	91.7
F10	<i>Vibrio</i>	<i>Vibrio</i> genomesp. F10	1	3.8	44.1	-	-	-
F6	<i>Vibrio</i>	<i>Vibrio</i> genomesp. F6	1	4.2	42.3	-	-	-
Gazogenes	<i>Vibrio</i>	<i>V. aerogenes</i> , <i>V. gazogenes</i> , <i>V. mangrovi</i> , <i>V. quintilis</i> , <i>V. ruber</i> , <i>V. rhizosphaerae</i> , and <i>V. spartinae</i>	7	5.1±0.5	45.7±0.3	72.1-92.0	19.8-45.7	69.5-93.2
Halioticoli	<i>Vibrio</i>	<i>V. breoganii</i> , <i>V. commitans</i> , <i>V. ezurae</i> , <i>V. gallicus</i> , <i>V. halioticoli</i> , <i>V. inusitatus</i> , <i>V. ishigakensis</i> , <i>V. neonatus</i> , <i>V. rarus</i> , and <i>V. superstes</i>	10	4.2±0.4	43.9±1.1	73.2-93.6	19.6-45.5	74.9-96.3
Nigripulchritudo	<i>Vibrio</i>	<i>V. nigripulchritudo</i> and <i>V. penaeicida</i>	2	6.3±0.3	44.9±1.3	77.1	21.5	84.0
Ponticus	<i>Vibrio</i>	<i>V. panuliri</i> , <i>V. ponticus</i> , <i>V. rhodolitus</i> , and <i>V. taketomensis</i>	4	4.6±0.3	44.9±0.4	77.6-85.0	22.2-30.3	83.0-89.6
Proteolyticus	<i>Vibrio</i>	<i>V. proteolyticus</i>	1	4.7	50.0	-	-	-
Rumoiensis	<i>Vibrio</i>	<i>V. algivorus</i> , <i>V. aphrogenes</i> , <i>V. casei</i> , <i>V. gangliei</i> , <i>V. litoralis</i> , and <i>V. rumoiensis</i>	6	3.8±0.3	41.8±0.9	76.5-88.5	21.6-35.5	79.9-92.6
Tapetis	<i>Vibrio</i>	<i>V. tapetis</i>	1	5.7	43.7	-	-	-

620 Bold indicates new/emended species, underlined indicates likely mistaken species

621

622 Table 3. List of strains/species pair showing a value above species boundary

No.	FastANI (%)	PyANI (%)	OrthoANI (%)	DDH (%)	AAI (%)	Paired strains/species (bold indicates the likely misidentified species)	Clade
1	98.8	98.7	98.9	90.1	98.5	<i>V. qinghaiensis</i> Q67 ^T / <i>V. ordalii</i> ATCC 33509 ^T	Anguillarum
2	98.3	98.4	98.5	86.2	98.3	<i>S. costicola</i> subsp. <i>alcaliphilus</i> DSM 16359 ^T / <i>S. costicola</i> subsp. <i>costicola</i> LMG 11651 ^T	Costicola
3	98.3	98.4	98.5	85.7	98.6	<i>V. neocaledonicus</i> CGJ02-2 / <i>V. alginolyticus</i> ATCC 17749 ^T	Harveyi
4	98.1	98.1	98.1	83.1	98.4	<i>V. chemaguriensis</i> Iso1 ^T / <i>V. diabolicus</i> CNCM I-1629 ^T	Harveyi
5	97.2	97.6	97.5	79.7	97.1	<i>A. logei</i> 1S159 / <i>A. salmonicida</i> LFI1238	Fischeri
6	97.5	97.5	97.6	78.2	97.8	<i>P. damsela</i> subsp. <i>damsela</i> ATCC 33539 ^T / <i>P. angustum</i> ATCC 25915 ^T	Phosphoreum
7	96.1	97.0	97.0	74.2	97.0	<i>P. damsela</i> subsp. <i>piscicida</i> 91-197 / <i>P. damsela</i> NCTC 11647 ^T	Damsela
8	96.8	96.8	96.9	73.2	97.8	<i>P. leiognathi</i> subsp. <i>mandapamensis</i> svers.1.1 / <i>P. leiognathi</i> subsp. <i>leiognathi</i> ATCC 25521 ^T	Phosphoreum
9	96.2	96.0	96.1	67.1	-	<i>V. ordalii</i> ATCC 33509 ^T / <i>V. anguillarum</i> DSM 21597 ^T	Anguillarum
10	95.9	95.8	95.9	66.2	-	<i>V. qinghaiensis</i> Q67 ^T / <i>V. anguillarum</i> DSM 21597 ^T	Anguillarum
11	95.5	95.8	95.7	65.9	-	<i>V. salilacus</i> / <i>V. pacinii</i>	Pacini
12	95.5	95.4	95.7	63.8	-	<i>P. frigidophilum</i> JCM 12947 ^T / <i>P. profundum</i> SS9	Profundum
13	95.4	95.3	95.4	62.8	-	<i>V. cidicii</i> 2756-81 ^T / <i>V. navarrensis</i> ATCC 51183 ^T	Vulnificus
14	95.3	95.2	95.3	62.9	-	<i>P. andalusiense</i> CECT9192 ^T / <i>P. aquimaris</i> LC2-065 ^T	Phosphoreum
15	95.2	95.0	95.0	-	-	<i>V. barjaei</i> 3062 ^T / <i>V. mediterranei</i> NBRC 15635 ^T	Mediterranei
16	95.3	94.6	94.8	-	-	<i>V. coralliirubri</i> corallo1 ^T / <i>V. celticus</i> Rd8.15 ^T	Splendidus
17	95.4	94.5	94.8	-	-	<i>V. toranzoniae</i> CECT 7225 ^T / <i>V. kanaloae</i> CCUG 56968 ^T	Splendidus
95	95	95	95	70	95	Threshold for same species	

623

624 Table 4. A metagenome-assembled genomes (MAGs) survey using 11 ribosomal protein genes (11-RPGs)
625 scheme

ID	Assembly accession	Primary MAG name	UBA	11-RPGs hits	MAG completeness (%)	Belonged clade	The closest species based on 11-RPGs	ANI against the closest species (%)
MAG1	ASM234268v1	<i>Vibrio</i> sp.	UBA2441	11	95.5	New	<i>V. albus</i>	71.5
MAG2	ASM335412v1	<i>Vibrionaceae</i>	Glo_1	11	95.5	New	<i>Allivibrio fischeri</i>	71.5
MAG3	ASM1180648v1	<i>Vibrio diazotrophicus</i>	HF9B	11	98.7	Diazotrophicus	<i>V. diazotrophicus</i>	90.3
MAG4	ASM1180657v1	<i>Vibrio hepatarius</i>	HF70	11	100.0	Nereis	<i>V. hepatarius</i>	92.6
MAG5	ASM337041v1	<i>Vibrio campbellii</i>	HD9-110m-PIT-SAG10	11	29.1	Harveyi	<i>V. campbellii</i>	98.4
MAG6	ASM351374v1	<i>Vibrio</i> sp.	UBA10714	11	94.1	Rumoiensis	<i>V. casei</i>	100.0
MAG7	ASM352238v1	<i>Vibrio</i> sp.	UBA10737	11	90.9	Rumoiensis	<i>V. litoralis</i>	97.4
MAG8	ASM1180653v1	<i>Vibrio campbellii</i>	HF17	11	99.2	Harveyi	<i>V. campbellii</i>	96.3
MAG9	ASM1337316v1	<i>Vibrio campbellii</i>	HF-Din11	11	100.0	Harveyi	<i>V. campbellii</i>	96.3
MAG10	ASM325067v1	<i>Vibrio diazotrophicus</i>	SZUA-511	11	30.8	<i>Aeromonadaceae</i>	<i>Tolomonas auensis</i>	69.4
MAG11	ASM187415v1	<i>Vibrio</i> sp.	MedPE-SWchi	1	70.8	NT	NT	-
MAG12	ASM233904v1	<i>Vibrio metoecus</i>	UBA1833	8	60.9	NT	NT	-
MAG13	ASM234273v1	<i>Vibrio</i> sp.	UBA2437	0	90.0	NT	NT	-
MAG14	ASM324725v1	<i>Vibrio diazotrophicus</i>	SZUA-363	0	43.0	NT	NT	-
MAG15	ASM337038v1	<i>Vibrio campbellii</i>	HD9-110m-PIT-SAG09	0	11.2	NT	NT	-
MAG16	ASM345089v1	<i>Vibrio</i> sp.	UBA12222	0	34.1	NT	NT	-
MAG17	ASM347639v1	<i>Vibrio</i> sp.	UBA9541	8	64.8	NT	NT	-
MAG18	ASM348339v1	<i>Vibrio</i> sp.	UBA8383	8	80.1	NT	NT	-
MAG19	ASM1337330v1	<i>Vibrionaceae</i>	HF-Dia40	0	96.8	NT	NT	-

Possible new species indicated by bold. NT: not tested.

626

627