Title

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

Author names and affiliations

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

[a]Laboratory of Microbiology, Faculty of Fisheries Sciences, Hokkaido University, Hakodate, Japan

[b]Departamento de Microbiología y Parasitología, CRETUS & CIBUS-Facultad de Biología, Universidade de Santiago de Compostela, Campus Vida, Santiago de Compostela, España

[c]Institute of Biology and SAGE-COPPE, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

[d]CIAD, AC, Mazatlan Unit for Aquaculture and Environmental Management, Mazatlán, México

Corresponding Author

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

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

Compliance with Ethical Standards

- **Funding:** This work was partly supported by KAKEN 19H03041.
- **Competing Interests:** The authors declare no competing financial interests.

Declarations

- **Funding**

  This work was partly supported by KAKEN 19H03041.

- **Conflicts of interest/Competing interests**

  The authors declare no competing financial interests.

- **Ethics approval (Not applicable)**
- **Consent to participate (Not applicable)**
- **Consent for publication (Not applicable)**
- **Availability of data and material**

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

- **Code availability (Not applicable)**

Author Contributions

- Chunqi Jiang conceived, designed and performed the experiments, analyzed the data, visualized the data, drafted and reviewed the manuscript.
- Mami Tanaka and Sayo Nishikawa performed the experiments, reviewed the manuscript.
- Sayaka Mino, Jesús L. Romalde, Fabiano L. Thompson and Bruno Gomez-Gil analyzed the data, reviewed the manuscript.
- Tomoo Sawabe conceived and designed the experiments, reviewed the manuscript.

ABSTRACT

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

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

INTRODUCTION

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

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

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

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

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

Ribosomal Protein Genes (RPGs)

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

Sequence Analysis

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

RESULTS

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

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

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

Clade Based Genome Features

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

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

Single-Copy Genes (SCGs) MLSA

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

Ribosomal Protein Genes (RPGs) MLSA

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

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

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

Identification of MAGs Using 11-RPGs

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

DISCUSSION

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

Possible Clade Boundaries

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

Possible Changes in The Future

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

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

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

Misidentification of Species/Strains

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

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

Delineation Power of 8-HKGs MLSA

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

Availability of a Potential RPG Set for Metagenomes

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

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

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

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

Conclusion

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

ACKNOWLEDGMENTS

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

DATA AVAILABILITY

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

Author Contributions

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

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

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

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

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

1. Thompson FL, Iida T, Swings J (2004) Biodiversity of Vibrios. Microbiol Mol Biol Rev 68:403–431. https://doi.org/10.1128/MMBR.68.3.403-431.2004
2. Thompson CC, Vicente ACP, Souza RC, et al (2009) Genomic taxonomy of vibrios. BMC Evol Biol 9:1–16. https://doi.org/10.1186/1471-2148-9-258
3. Gomez-Gil B, Thompson CC, Matsumura Y, et al (2014) The Famlily *Vibrionaceae*. in: The Prokaryotes. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 659–747
4. Auch, A.F., von Jan, M., Klenk, HP. et al (2010) Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand in Genomic Sci 2 117–134. https://doi.org/10.4056/sigs.531120

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

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

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

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

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

demarcation of prokaryotes. Int J Syst Evol Microbiol 64:346–351. https://doi.org/10.1099/ijs.0.059774-0

55. Ghosh A, Bhadury P (2019) *Vibrio chemaguriensis* sp. nov., from Sundarbans, Bay of Bengal. Curr Microbiol 76:1118–1127. https://doi.org/10.1007/s00284-019-01731-7

56. Tan L, Gómez-Betancur I, Guo S, et al (2020) Complete genome of *Vibrio neocaledonicus* CGJ02-2, an active compounds producing bacterium isolated from South China Sea. Curr Microbiol 77:2665–2673. https://doi.org/10.1007/s00284-020-02047-7

FIGURE CAPTIONS

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

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

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

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

Figure 5. Metagenome-assembled genomes (MAGs) identification using concatenated split 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
Photobacterium	*carnosum*	CECT 9394^T	Draft	BPPU01000001-BPPU01000011
Photobacterium	*sanguinicancri*	CECT 7579^T	Complete	AP024850-AP024851
Photobacterium	*swingsii*	CECT 7576^T	Complete	AP024852-AP024853
Photobacterium	*toruni*	CECT 9189^T	Draft	AP024854-AP024860
Vibrio	*aerogenes*	LMG 19650^T	Complete	AP024861-AP024863
Vibrio	***agarivorans***	**CECT 5085^T**	**Draft**	BLAT01000001-BLAT01000058
Vibrio	*breoganii*	CAIM 1829^T	Complete	AP024864-AP024865
Vibrio	*comitans*	LMG 23416^T	Complete	AP024866-AP024868
Vibrio	***ezurae***	**JCM 21522^T**	**Complete**	AP024869-AP024870
Vibrio	*gallicus*	LMG 21878^T	Complete	AP024871-AP024872
Vibrio	*gazogenes*	ATCC 29988^T	Complete	AP024873-AP024874
Vibrio	***halioticoli***	**IAM 14596^T**	**Complete**	AP024875-AP024877
Vibrio	*inusitatus*	LMG 23434^T	Complete	AP024878-AP024880
Vibrio	***ishigakensis***	**JCM 19231^T**	**Complete**	AP024881-AP024882
Vibrio	*mangrovi*	CECT 7927^T	Complete	AP024883-AP024884
Vibrio	*neonatus*	JCM 21521^T	Complete	AP024885-AP024886
Vibrio	*palustris*	CECT 9027^T	Complete	AP024887-AP024888
Vibrio	*pectenicida*	LMG 19642^T	Complete	AP024889-AP024892
Vibrio	*plantisponsor*	CECT 7581^T	Complete	AP024893-AP024894
Vibrio	*porteresiae*	MSSRF30^T	Complete	AP024895-AP024896
Vibrio	*quintilis*	CECT 7734^T	Draft	AP024897-AP024899
Vibrio	*rarus*	LMG 23674^T	Complete	AP024900-AP024902
Vibrio	*rhizosphaerae*	LMG 23790^T	Complete	AP024903-AP024904
Vibrio	*ruber*	LMG 23124^T	Complete	AP024905-AP024906
Vibrio	*spartinae*	CECT 9026^T	Complete	AP024907-AP024908
Vibrio	***superstes***	**JCM 21480^T**	**Complete**	AP024909-AP024910
Vibrio	*zhugei*	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	*Candidatus* Photodesmus	***Candidatus* Photodesmus blepharus** and ***Candidatus* Photodesmus katoptron**	2	1.1±0.1	33.2±3.4	73.1	18.7	66.6
Pacificus	*Enterovibrio*	*E. pacificus*	1	5.3	45.5	-	-	-
Hollisae	*Grimontia* /*Enterovibrio*	***E. baiacu***, ***E. calviensis***, *E. coralii*, ***E. nigricanis***, *E. norvegicus*, ***G. celer***, ***G. indica***, ***G. marina***, and *G. hollisae*	9	5.3±0.5	48.1±0.8	74.6-89.9	19.5-38.5	76.5-94.6
Marinum	*Paraphotobacterium*	***Paraphotobacterium marinum***	1	2.6	31.2	-	-	-
Aquae	*Photobacterium*	*P. aquae*	1	5.1	49.1	-	-	-
Halotolerans	*Photobacterium*	*P. halotolerans* and ***P. salinisoli***	2	4.7±0.0	50.6±0.5	91.1	42.5	94.6
Jeanii	*Photobacterium*	*P. jeanii*, *P. sagunicancri*, and *P. swingsii*	3	5.4±0.3	44.1±0.9	77.6-85.0	21.8-28.8	82.6-91.2
Proteolyticum	*Photobacterium*	*P. alginatilyticum*, ***P. chitinilyticum***, *P. marinum*, and *P. proteolyticum*	4	6.1±0.5	47.1±0.9	76.6-93.4	21.0-52.3	79.4-95.1
Costicola	*Salinivibrio*	*S. costicola.* subsp. *alcaliphilus*, *S. costicola.* subsp. *costicola*, *S. kushneri*, *S. proteolyticus*, *S. shamensis*, *S. siamensis*, and ***S. socompensis***	7	3.4±0.1	49.9±0.6	78.6-98.5	21.7-86.2	84.5-98.3
Occultus	*Thaumasiovibrio*	***T. occutus*** and ***T. subtropicus***	2	4.9±0.7	48.3±1.6	71.9	20.8	68.5
Aestivus	*Vibrio*	*V. aestivus* and ***V. mexicanus***	2	4.8±0.0	44.9±0.2	90.0	39.5	92.7
Albus	*Vibrio*	*V. albus*	1	4.3	46.1	-	-	-
Caribbeanicus	*Vibrio*	*V. caribbeanicus*	1	4.4	41.6	-	-	-
Fluvialis	*Vibrio*	*V. fluvialis* and *V. furnisii*	2	4.9±0.1	50.3±0.5	86.0	30.5	91.8
Maerlii	*Vibrio*	*V. maerlii*	1	4.8	44.5	-	-	-
Metschnikovii	*Vibrio*	*V. cincinnatiensis*, *V. fujianensis*, *V. injenensis*, and *V. metschnikovii*	4	3.7±0.1	43.8±0.3	76.5-91.9	20.2-45.7	80.5-95.3
Ostreicida	*Vibrio*	*V. ostreicida*	1	4.4	45.6	-	-	-
Pacinii	*Vibrio*	*V. pacinii* and *V. salilacus*	2	4.0±0.5	45.2±0.1	95.7	65.9	95.5
Sonorensis	*Vibrio*	*V. sonorensis*	1	4.8	44.4	-	-	-
Viridaestus	*Vibrio*	*V. viridaestus*	1	4.7	41.9	-	-	-
Xiamenensis	*Vibrio*	*V. xiamenensis*	1	5.5	46.4	-	-	-
Emended clades (18)								
Fischeri	*Aliivibrio*	***A. finisterrensis***, *A. fischeri*, *A. logei*, *A. sifiae*, *A. wodanis* and *A. salmonicida*	6	4.5±0.3	38.7±0.5	79.3-97.5	22.2-79.7	83.5-97.1
Ganghwense	*Photobacterium*	*P. ganghwense*	1	5.5	50.5	-	-	-

Phosphoreum	*Photobacterium*	***P. andalusiense***, *P. angustum, P. aquimaris, P. carnosum, P. iliopscarium, P. kishitanii,* ***P. malacitanum***, *P. piscicola, P. leiognathi subsp. leiognathi, P. leiognathi.* subsp. *mandapamensis, P. phosphoreum, P. toruni, and P. damselae.* subsp. *damselae*	13	4.6±0.2	39.5±0.7	75.7-97.6	20.5-78.2	76.7-97.8
Profundum	*Photobacterium*	***P. frigidiphilum***, *P. indicum, P. lipolyticum,* and *P. profumdum*	4	5.9±0.7	42.7±2.1	76.7-95.7	21.7-63.8	79.0-95.8
Rosenbergii	*Photobacterium*	***P. aphoticum***, *P. gaetbulicola, P. lutimaris, P. rosenbergii,* and ***P. sanctipauli***	5	5.9±0.4	48.5±1.1	75.5-83.0	21.2-26.3	76.5-89.5
Anguillarum	*Vibrio*	*V. aestuarianus, V. anguillarum, V. ordalii,* and ***V. qinghaiensis***	4	4.0±0.4	44.1±1.0	78.7-98.9	21.5-90.1	84.1-98.5
Cholerae	*Vibrio*	*V. cholerae,* ***V. metoecus***, *V. mimicus,* and ***V. parillis***	4	4.1±0.2	46.8±0.5	85.0-88.2	29.1-35.2	91.4-93.1
Diazotrophicus	*Vibrio*	*V. diazotrophicus* and ***V. plantisponsor***	2	4.6±0.2	43.7±0.3	94.2	56.7	96.5
Harveyi	*Vibrio*	***V. alfacsensis***, *V. alginolyticus, V. azureus, V. campbellii, V. chemaguriensis,* ***V. diabolicus***, *V. harveyi,* ***V. hyugaensis***, *V. jasicida, V. mytili, V. natriegens,* ***V. neocaledonicus***, *V. owensii, V. parahaemolyticus, V. rotiferianus, and V. sagamiensis*	16	5.2±0.5	44.4±1.2	75.2-98.4	20.9-85.7	76.1-98.6
Marisflavi	*Vibrio*	*V. marisflavi*	1	4.7	42.2	-	-	-
Mediterranei	*Vibrio*	*V. barjaei,* ***V. hangzhouensis***, *V. maritimus, V. mediterranei, V. thalassae, and V. variabilis*	6	5.4±0.3	45.3±1.2	76.7-95.0	20.6-61.3	81.1-97.2
Nereis	*Vibrio*	***V. hepatarius*** and *V. nereis*	2	4.5±0.5	45.8±0.6	79.3	22.3	85.4
Orientalis	*Vibrio*	***V. aquaticus***, ***V. atypicus***, ***V. bivalvicida***, *V. brasiliensis,* ***V. europaeus***, ***V. galatheae***, *V. orientalis,* ***V. ouci***, *V. tubiashii, V. sinaloensis* and ***V. xuii***	11	5.0±0.6	45.0±0.6	75.8-94.2	20.1-56.1	80.4-95.8
Pectenicida	*Vibrio*	*V. pectenicida*	1	4.5	41.3	-	-	-
Porteresiae	*Vibrio*	***V. palustris***, *V. porteresiae, V. tritonius, and* ***V. zhugei***	4	4.6±0.9	44.4±0.9	72.5-84.8	19.2-28.4	71.8-91.1
Scophthalmi	*Vibrio*	*V. ichthyoenteri,* ***V. renipiscarius***, *V. scophthalmi,* and ***V. sinensis***	4	4.8±0.4	44.2±1.2	74.9-84.4	21.3-28.4	76.6-89.9
Splendidus	*Vibrio*	*V. atlanticus, V. celticus, V. chagasii, V. fortis,* ***V. coralliirubri***, *V. crassostreae, V. gigantis, V. cyclitrophicus, V. echinoideorum, V. lentus, V. gallaecicus, V. kanaloae,* ***V. profundi,*** *V. splendidus, V. tasmaeniensis,* and *V. toranzoniae*	16	5.2±0.4	43.9±0.8	76.1-94.8	21.2-59.4	78.7-96.8
Vulnificus	*Vibrio*	***V. cidicii***, *V. navarrensis*, and *V. vulnificus*	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	*Photobacterium*	*P. damselae* subsp. *damselae* and *P. damselae.* subsp. *piscicida*	2	4.7±0.5	40.9±0.1	97.0	74.2	97.0
Agarivorans	*Vibrio*	*V. agarivorans* and *V. astriarenae*	2	4.8±0.1	45.5±0.2	79.3	22.8	86.5
Coralliilyticus	*Vibrio*	*V. coralliilyticus* and *V. neptunius*	2	5.4±0.3	45.3±0.5	87.6	34.2	91.7
F10	*Vibrio*	*Vibrio* genomesp. F10	1	3.8	44.1	-	-	-
F6	*Vibrio*	*Vibrio* genomesp. F6	1	4.2	42.3	-	-	-
Gazogenes	*Vibrio*	*V. aerogenes, V. gazogenes, V. mangrovi, V. quintilis, V. ruber, V. rhizosphaerae,* and *V. spartinae*	7	5.1±0.5	45.7±0.3	72.1-92.0	19.8-45.7	69.5-93.2
Halioticoli	*Vibrio*	*V. breoganii, V. commitans, V. ezurae, V. gallicus, V. halioticoli, V. inusitatus, V. ishigakensis, V. neonatus, V. rarus,* and *V. superstes*	10	4.2±0.4	43.9±1.1	73.2-93.6	19.6-45.5	74.9-96.3
Nigripulchritudo	*Vibrio*	*V. nigripulchritudo* and *V. penaeicida*	2	6.3±0.3	44.9±1.3	77.1	21.5	84.0
Ponticus	*Vibrio*	*V. panuliri, V. ponticus, V. rhodolitus,* and *V. taketomensis*	4	4.6±0.3	44.9±0.4	77.6-85.0	22.2-30.3	83.0-89.6
Proteolyticus	*Vibrio*	*V. proteolyticus*	1	4.7	50.0	-	-	-
Rumoiensis	*Vibrio*	*V. algivorus, V. aphrogenes, V. casei, V. gangliei, V. litoralis,* and *V. rumoiensis*	6	3.8±0.3	41.8±0.9	76.5-88.5	21.6-35.5	79.9-92.6
Tapetis	*Vibrio*	*V. tapetis*	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	***V. qinghaiensis* Q67^T**	*V. ordalii* ATCC 33509^T	Anguillarum
2	98.3	98.4	98.5	86.2	98.3	*S. costicola* subsp. *alcaliphilus* DSM 16359^T	*S. costicola* subsp. *costicola* LMG 11651^T	Costicola
3	98.3	98.4	98.5	85.7	98.6	***V. neocaledonicus* CGJ02-2**	*V. alginolyticus* ATCC 17749^T	Harveyi
4	98.1	98.1	98.1	83.1	98.4	***V. chemaguriensis* Iso1^T**	*V. diabolicus* CNCM I-1629^T	Harveyi
5	97.2	97.6	97.5	79.7	97.1	*A. logei* 1S159	*A. salmonicida* LFI1238	Fischeri
6	97.5	97.5	97.6	78.2	97.8	*P. damselae* subsp. *damselae* ATCC 33539^T	*P. angustum* ATCC 25915^T	Phosphoreum
7	96.1	97.0	97.0	74.2	97.0	*P. damselae* subsp. *piscicida* 91-197	*P. damselae* NCTC 11647^T	Damselae
8	96.8	96.8	96.9	73.2	97.8	*P. leiognathi* subsp. *mandapamensis* svers.1.1	*P. leiognathi subsp. leiognathi* ATCC 25521^T	Phosphoreum
9	96.2	96.0	96.1	67.1	-	*V. ordalii* ATCC 33509^T	*V. anguillarum* DSM 21597^T	Anguillarum
10	95.9	95.8	95.9	66.2	-	*V. qinghaiensis* Q67^T	*V. anguillarum* DSM 21597^T	Anguillarum
11	95.5	95.8	95.7	65.9	-	*V. salilacus*	*V. pacinii*	Pacinii
12	95.5	95.4	95.7	63.8	-	*P. frigidiphilum* JCM 12947^T	*P. profundum* SS9	Profundum
13	95.4	95.3	95.4	62.8	-	*V. cidicii* 2756-81^T	*V. navarrensis* ATCC 51183^T	Vulnificus
14	95.3	95.2	95.3	62.9	-	*P. andalusiense* CECT9192^T	*P. aquimaris* LC2-065^T	Phosphoreum
15	95.2	95.0	95.0	-	-	*V. barjaei* 3062^T	*V. mediterranei* NBRC 15635^T	Mediterranei
16	95.3	94.6	94.8	-	-	*V. coralliirubri* corallo1^T	*V. celticus* Rd8.15^T	Splendidus
17	95.4	94.5	94.8	-	-	*V. toranzoniae* CECT 7225^T	*V. kanaloae* CCUG 56968^T	Splendidus
	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		11-RPGs hits	MAG completeness (%)	Belonged clade	The closest species based on 11-RPGs	ANI against the closest species (%)
MAG1	**ASM234268v1**	***Vibrio* sp.**	**UBA2441**	**11**	**95.5**	**New**	***V. albus***	**71.5**
MAG2	**ASM335412v1**	***Vibrionaceae***	**Glo_1**	**11**	**95.5**	**New**	***Aliivibrio fischeri***	**71.5**
MAG3	**ASM1180648v1**	***Vibrio diazotrophicus***	**HF9B**	**11**	**98.7**	**Diazotrophicus**	***V. diazotrophicus***	**90.3**
MAG4	**ASM1180657v1**	***Vibrio hepatarius***	**HF70**	**11**	**100.0**	**Nereis**	***V. hepatarius***	**92.6**
MAG5	ASM337041v1	*Vibrio campbellii*	HD9-110m-PIT-SAG10	11	29.1	Harveyi	*V. campbellii*	98.4
MAG6	ASM351374v1	*Vibrio* sp.	UBA10714	11	94.1	Rumoiensis	*V. casei*	100.0
MAG7	ASM352238v1	*Vibrio* sp.	UBA10737	11	90.9	Rumoiensis	*V. litoralis*	97.4
MAG8	ASM1180653v1	*Vibrio campbellii*	HF17	11	99.2	Harveyi	*V. campbellii*	96.3
MAG9	ASM1337316v1	*Vibrio campbellii*	HF-Din11	11	100.0	Harveyi	*V. campbellii*	96.3
MAG10	ASM325067v1	*Vibrio diazotrophicus*	SZUA-511	11	30.8	*Aeromonadaceae*	*Tolumonas auensis*	69.4
MAG11	ASM187415v1	*Vibrio* sp.	MedPE-SWchi	1	70.8	NT	NT	-
MAG12	ASM233904v1	*Vibrio metoecus*	UBA1833	8	60.9	NT	NT	-
MAG13	ASM234273v1	*Vibrio* sp.	UBA2437	0	90.0	NT	NT	-
MAG14	ASM324725v1	*Vibrio diazotrophicus*	SZUA-363	0	43.0	NT	NT	-
MAG15	ASM337038v1	*Vibrio campbellii*	HD9-110m-PIT-SAG09	0	11.2	NT	NT	-
MAG16	ASM345089v1	*Vibrio* sp.	UBA12222	0	34.1	NT	NT	-
MAG17	ASM347639v1	*Vibrio* sp.	UBA9541	8	64.8	NT	NT	-
MAG18	ASM348339v1	*Vibrio* sp.	UBA8383	8	80.1	NT	NT	-
MAG19	ASM1337330v1	*Vibrionaceae*	HF-Dia40	0	96.8	NT	NT	-

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

626

627