Revised MS of CMIC-D-24-01876R1

Title: Unexpected Diversity in Gene Clusters Encoding Formate Hydrogenlyase Complex Machinery in *Vibrionaceae* Correlated to Fermentative Hydrogen Production

Author names and affiliations

Tomoo Sawabe[1*], Yuito Umeki[1], Ramesh Kumar Natarajan[2], Chunqi Jiang[1,3], Fabiano Thompson[4], Sayaka Mino[1*]

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

[2] National Institute for Interdisciplinary Science and Technology (CSIR), Thiruvananthapuram, Kerala, India

[3] Atmosphere and Ocean Research Institute, The University of Tokyo, Chiba, Japan

[4] Laboratory of Microbiology, Institute of Biology, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

***Corresponding Author**

Tomoo Sawabe and Sayaka Mino, 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

Declarations

Funding: This work was supported by the MEXT Kaken JP16H04976, JP19H03041 and JSPS Bilateral Research Project between Brazil and Japan 120200101.

Conflicts of interest/Competing interests: TS is an associate editor for Current Microbiology. Authors except TS declare no competing interests.

Ethics approval (Not applicable)

Consent to participate (Not applicable)

Consent for publication (Not applicable)

Availability of data and material: Genome sequence of *Vibrio tritonius* strain MA35 was deposited DBJ/GenBank/ENA under accession number GCA_038695265.1.

Code availability (Not applicable)

Author Contributions

Conceptualization: TS; Research and Data analysis: YU, CJ and TS; Writing - original draft preparation: TS; Writing -review and editing: YU, NRK, JC, SM, FT and TS; Resources: NRK, JC and TS. All authors contributed to the article and approved the submitted version.

Running title

Unexpected Diversity of FHL Gene Cluster in H_2-Producing Vibrios

Abstract

An entire Hyf-type formate hydrogenlyase complex (Hyf-FHL) gene cluster was first discovered in a marine *Vibrio* species, *Vibrio tritonius* isolated from the digestive tract of the sea hare *Aplysia kurodai* [1]. The bacterium is also the first marine bacterium in which hydrogen production ability exceeds that of *Escherichia coli* under saline conditions [1-2]. However, we were still unable to answer the evolutionary question as to why only minor groups of vibrios could maintain the FHL gene clusters and hydrogen (gas) production ability. Here, we set up comparative genomics and fermentative hydrogen production profiling using all 16 currently known *Vibrionaceae* species, which maintain FHL gene clusters and/or gas production, including 12 *Vibrio* and 4 *Photobacterium* species. Whole genome comparison using complete genome sequences revealed unexpected diversity of FHL gene clusters, at least, with two new types of FHL gene clusters. Additional fermentative hydrogen profiling and structure modeling of FHLs showed formate detoxification as a part of formate and pH homeostasis could be one of the selective pressures in the evolution of FHL gene clusters responsible for high hydrogen production in vibrios.

Keywords: vibrio, hydrogen, genome, physiology, formate, marine

Introduction

Fermentative hydrogen gas production is an atypical trait in the family *Vibrionaceae* [1-7]. With the recent revolution in genome sequencing, the list of hydrogen-producing marine vibrios has currently been updated to include 13 *Vibrionaceae* species; *Vibrio gazogenes*, *Vibrio spartinae*, *Vibrio ruber*, *Vibrio rhizosphaerae*, *Vibrio mangrovi*, *Vibrio aerogenes*, *Vibrio quintilis* (Gazogenes clade), *Vibrio porteresiae*, *Vibrio tritonius* (Porteresiae clade), *Vibrio furnissii* (Fluvialis clade), *Vibrio aphrogenes* (Rumoiensis clade), *Vibrio mytili* (Harveyi clade) and *V. plantisponsor* (Diazotrophicus clade) [3-6, 8-22]. All Gazogenes and Porteresiae clade species possess a gene set for Hyf-FHL consisting of 20 to 21 genes encoding FHL, formate dehydrogenase, FhlA activator protein and hydrogenase maturation (Hyp) proteins, however, only

two species, one in the Fluvialis clade and another in the Rumoiensis clade, possess this gene set [4-6, 23]. Interestingly, *V. furnissi* in the Fluvialis clade possesses a Hyc-like FHL gene cluster, which lacks three transmembrane protein genes, *hyfDEF*, similar to the Hyc-FHL gene cluster of *E. coli* [5-6]. More interestingly, *V. aerogenes* in the Gazogenes clade possesses a separated FHL gene cluster [6] and formate dehydrogenase (FDH-H) of *V. furnissii* and *V. aphrogenes* is a selenoprotein, which is the same type as in *E. coli*, using the UGA stop codon as the codon for selenocysteine (U) [4,24].

Furthermore, the structure of Hyc-FHL in *E. coli* has recently been determined using cryo-electron microscopy, which clearly shows it is a member of the complex I-like respiratory complexes [25-26]. However, the part of the membrane arm HycC of Hyc-FHL is 180°C rotated to the membrane anchor subunit (HycD). Such a structural uniqueness of the membrane arm rotation has already been observed in the membrane-bound [NiFe] hydrogenase complex (MBH) and the membrane-bound sulfane sulfur reductase complex (MBS) [25-26]. As it is proposed that complex I has evolved from an ancestral [NiFe] hydrogenase and *m*ultiple *r*esistance and *p*H (Mrp) complex, Hyf-FHL could be the most important molecules to elucidate the evolution of complex I family molecules and its related gene arrangements because of more structural resemblance to the complex I with the longer membrane arm than those of Hyc-FHL, MBH and MBS. Therefore, evolutionary questions on how the FHL gene cluster diverged, maintained and functioned to affect the atypical gas producing phenotype in the family *Vibrionaceae* are still under debates [6,25].

Among the *Vibrionaceae* species, *V. porteresiae* and *V. tritonius* in the Porteresiae clade show particularly high hydrogen production ability, with a maximum hydrogen production of 1.7 mol H_2/mol mannitol, exceeding that of *E. coli* even under saline conditions [2,6]. Compared to the hydrogen production efficiency of around 0.5 mol H_2/mol mannitol in the Gazogenes clade species [2,5-6], the hydrogen production ability of the Porteresiae clade species is high, however, genome structural differences are found in the presence of the different type of nickel transport-related genes such as the hydrogenase/urease accessory protein gene (*hupE*/*ureJ*) encoding a transmembrane nickel permease, *nikR-nikABCDE* encoding nickel-responsive regulators and nickel ABC transporters, and nickel/cobalt transporter NiCoT permease genes [6]. It remains difficult to answer the question why the Porteresiae clade species maintain high hydrogen production ability using the partial data set used in previous studies [6].

In this study, we compare all available hydrogen producing vibrios focusing on both genomic and

physiological comparisons, in particular, elucidating (1) structural diversity of the FHL complex gene cluster based on the complete genome sequences of 16 major hydrogen-producing and/or FHL gene cluster retaining marine vibrios species, and (2) fermentative hydrogen production profiling using the same strain set. Here, we report for the first time a strong positive correlation between high ability of formate re-import and hydrogen production in hydrogen-producing vibrios.

Materials and Methods

Bacterial Strains and Culture Conditions

A total of 17 strains was examined in this study (Table S1). All strains were grown on ZoBell 2216E medium at 25°C with minor modifications according to Matsumura *et al*. [6] except *Vibrio quintilis* at 33°C [16], *Photobacterium carnosum* at 15°C [19].

Complete Genome Sequences

Complete genomes of bacterial strains have been sequenced previously [23]. Accession numbers of genome assembly are indicated in Table S1. In brief, after purification of genomic DNA, libraries were prepared using the Rapid Barcoding Sequencing Kit (SQK-RBK004, Oxford Nanopore Technologies, Oxford, UK), and sequenced using the MinION device installed with MinKNOW ver. 19.05 and Flow Cell (R9.4.1) (Oxford Nanopore Technology). The basecalling was performed using Guppy v3.4.4 (Oxford Nanopore Technology). Flye version 2.8.3 [27], Racon v1.4.20 [28], Medaka v1.0.1 (Oxford Nanopore Technology) and/or Unicycler version 0.4.7 [29] were used to obtain the complete genome with a hybrid assembly approach using Illumina and Nanopore reads.

Genome Analyses

DFAST [30], BLAST [31] and InterProScan [32] were used to annotate complete genome sequences. The Artemis Comparison Tool [33-34] was used to visualize rearrangements. In silico MolecularCloning (In Silico Biology Inc., Yokohama, Japan) and anvi'o v5.5.0 [35] were used for gene structure comparison and pan-genomic analysis, respectively. ClustalX [36] was used to align the four house-keeping gene (*ftsZ*, *mreB*, *pyrH* and *topA*) sequences and seven FHL gene (*hyfABCGHIJ*) sequences. Molecular networks and phylogenetic trees were reconstructed using SplitsTree version 4 [37] and MEGAX [38], respectively.

Hydrogen Production Profiling

Hydrogen production profiling was performed according to Matsumura *et al*. [6] with minor

modifications. In brief, ZoBell 2216E liquid medium was used as the basal medium for hydrogen production by adding 2-morpholinoethanesulfonic acid (MES) (Dojindo, Kumamoto, Japan) to a final concentration of 0.1 M (pH 6.0) and adding filter-sterilized D-glucose (Glc) to a final concentration of 3% (w/v). A total of 1 mL preculture was inoculated to 99 mL of the fermentation medium in a reactor (100 mL medium bottle, Corning, USA) sealed with silicon equipped with a pH electrode (FermProbe pH electrodes, Broadly-James Corp., Branford, USA) and a gas collection aluminum bag (GL Sciences, Tokyo). The pH electrode was connected to a pH controller system (DJ-1023P, ABLE, Tokyo, Japan), which was able to pump an alkaline (5 M NaOH) solution. The reactor was maintained at optimum temperatures for each strain with stirring (480 rpm) using a magnetic stirrer (RO 5 power IKAMAG, IKA, Staufen, Germany). At 24-hour intervals, culture medium and gas were collected to determine turbidity (OD_{620}), organic acid production, hydrogen production and Glc consumption.

The profiling was repeated twice for each strain. Turbidity (OD_{620}) was measured using a plate reader (Infinite F200, Tecan, Switzerland). Organic acids and Glc were determined by high-performance liquid chromatography (SHIMADZU, Kyoto, Japan). For organic acids, the detector was CDD-10A vp (SHIMADZU) and the column was Shim-Pack SCR-102H (SHIMADZU) at an analytical flow rate of 0.8 mL/min for 40 minutes. For Glc measurement, the detector was set to RID-10A (SHIMADZU) and the Shodex Asahipak NH2P-50 4E column (Showa Denko, Tokyo, Japan) was selected. Hydrogen production was determined by the amount of gas collected in an aluminum bag (GL Sciences, Tokyo, Japan) using the water displacement method and the ratio of hydrogen in the gas was determined using gas chromatography GC-2014 (SHIMADZU) equipped with a ShinCarbon ST column (Chemical Industries Ltd., Kyoto, Japan).

Statistics

Bayesian statistics to evaluate correlation between hydrogen production and formate re-import was performed using JASP 0.18.3.0 (https://jasp-stats.org/).

Structural Prediction of Vibrio FHL Complex

Structure of vibrio FHL complex was predicted using AlphaFold-multimer 2.3.2 [39] with options to uses a best relax mode, a PDB database release till 1 August 2024 and three predictions for each model implemented in the supercomputing system in National Institute of Genetics (Mishima, Japan). To reduce the time for prediction, structures of soluble arm with a membrane anchor and membrane arm were predicted, separately, and superposed at the membrane anchor unit. Predicted

structures were visualized and analyzed using ChimeraX version 1.8 [40]. In preliminary analyses using the *E. coli* Hyc-FHL PDB data (7Z0S), root mean square deviation (RMSD) of each subunit between the 7Z0S and a predicted structure were measured within 0.5 to 0.8 Å.

Results

Gap-Closed Complete Genome Sequences Comparison Reveal Unexpected Diverse Structures in Vibrio FHL Complex Gene Clusters

All strains except *V. palustris* and *V. zhugei*, which are reported to be non-gas-producing species [17-18], possessed the FHL complex gene cluster (Fig. 1). When comparing the franking region of the FHL gene cluster among four species belonging to the Porteresiae clade (*V. porteresiae*, *V. tritonius*, *V. palustris* and *V. zhugei*), conserved homologous regions in both upstream and downstream were observed in all species (Fig. S1). Neither a FHL gene cluster or any hydrogenase genes including [FeFe] hydrogenase were found in *V. plantisponsor* in the Diazotrophicus clade, nevertheless, this species is known to be gas-producing [15]. Core gene assessments revealed all strains possessed the formate dehydrogenase related (*fdhD*) gene and the major formate channel (*focA*) gene (Fig. S2). The *fdhF* gene responsible for FDH-H was conserved on the genomes of all hydrogen producing vibrios, in more precisely, in their FHL gene cluster (Fig. 1). No formate dehydrogenase-O (FDH-O) and FDH-N homologues genes [24,41] were not found in the genomes of the Porteresiae and Gazogenes clade species except in *V. quintilis*.

In addition to the three previously reported FHL gene clusters on Chromosome (Chr.) 1 including 1) the Porteresiae-type typical hyf-FHL with an entire *hypFCDEAB* gene cluster present in the Porteresiae clade species, *V. porteresiae* and *V. tritonius*, 2) the Gazogenes-type typical hyf-FHL gene cluster with *hupE* present in *V. gazogenes*, *V. mangrovi*, *V. rhizosphaerae*, *V. ruber* and *V. spartinae* and 3) Furnisii-type hyc-like FHL with *hyp* [6], several more atypical FHL gene clusters were observed (Figs. 1 and S3).

An atypical Gazogenes-hyf gene cluster, which had previously been discovered in *V. aerogenes* was also confirmed in *V. quintilis* (Fig. 1). The atypical FHL gene cluster was separated into two regions in *hyfABCDEFGHIJ-hycH* and *hydN-fdhF-fhlA-fdhD-hupE-hypFCDEAB* on Chr. 1 (Fig. S3), and these gene structures were shared between those two species (Fig. 1). The *hyfABCDEFGHIJ-hycH* was located approximately 513 kb and 584 kb downstream of *hydN-fdhF-fhlA-fdhD-hupE-hypFCDEAB* gene cluster in *V. aerogenes* and *V. quintilis*, respectively. Loci of the FHL gene clusters were related to dynamic genome re-arrangements of Chr. 1 in the Gazogenes

clade species (Figs. S3 and S4).

More diverse FHL gene clusters were discovered in *Photobacterium* species, and those were likely to be shuffled (Figs. 1 and S3). In *P. swingsii* and *P. sanguinicancri* in the Jeanii clade [21-23], *hypAB* genes were located upstream of the FHL gene clusters and the rest of *hypFCDE* genes were separated by the core of FHL gene clusters (Figs. 1 and S3). More interestingly, these FHL gene clusters were found on Chr. 2 (Figs. 1 and S3). *P. carnosum* and *P. toruni* in the Phosphoreum clade [19-20,23] possessed a hyc-FHL-like gene cluster, which lacks *hyfDEF* transmembrane genes on Chr. 2 (Figs. 1 and S3). A separated *hypBCDEA* gene cluster was found on Chr. 1 (Figs. 1 and S3). These *Photobacterium* HypA on Chr. 2 (341 bp) showed a rather high amino acid sequence identity of 56-57% to *V. gazogenes* HypA, while the HypA on Chr. 1 showed only 29% identity. Similarly, the HypB on Chr. 2 showed 57-59% identity to *V. gazogenes* HypB, while the HypB on Chr. 1 showed only 29% identity.

Surprisingly, an extended genome survey on how those FHL gene cluster are widespread in the genus *Photobacterium* using a draft genome sequence collection of *Photobacterium* type strains established by Jiang *et al*. [23] revealed some evolutionary laws and more discovery for FHL gene cluster diversity (Fig. S5): (1) all Phosphoreum clade species except *P. angustum* possessed the Phosphoreum-type FHL gene cluster, but *P. angustum* might possess a typical Gazogenes-type FHL gene cluster. All Phosphoreum clade species also might possess *hyd2*-related uptake hydrogenase gene clusters on those genomes. The separated *hypBCDEA* cluster in the Phosphoreum-type FHL was adjacently located to those uptake hydrogenase genes. (2) *P. jeanii* also possessed a FHL gene cluster, so all the Jeanii clade species possessed the Jeanii-type of the FHL gene cluster. (3) *P. indicum* in the Profundum clade might possess the Fluvialis type FHL gene cluster, but this is not the case in other members of the Profundum clade such as *P. frigidiphilum*, *P. lipolyticum* and *P. profundum*. (4) Only two species, *P. chitinilyticum* and *P. marinum*, in the Proteolyticum clade, possessed duplicated FHL gene clusters, one related to the Fluvialis-type FHL and the other a shuffled FHL gene cluster which have never been observed before. (5) *P. damselae* subsp. *damselae* possessed the Fluvialis-type related FHL gene cluster but *hyp* gene clusters were re-shaped.

As the FHL pathway has genetic link amongst *hyf*, *hyp*, *fdhF*, *focA* and *pfl* (pyruvate formate lyase and the related protein genes) genes [24,41], we further assessed the status of co-localization of *hyf*, *hyp*, *fdhF*, *focA* and *pfl* genes. *focA-pflBA* clusters, at least, were observed in *P. toruni*, *P.*

carnosum and *P. phosphoreum*. *pflB* gene was not clustered neither to *focA* and *pflA* in the genus *Vibrio*: e.g. *pflB*, *pflA* and *focA* at around 2,286,997, 2,291,660 and 1,375,018 nucleotide positions on Chr. 1 in *V. tritonius*, respectively, and *pflB* and *pflA* at around 1,585,309 and 158,227on Chr. 1 in *V. aerogenes*, respectively, and *focA* and 869,729 on Chr. 2 in *V. aerogenes*. No confidential evidence in lack of *hypF* gene or presence of *hypF* gene alone separated from *hyp* gene clusters in *Vibrionaceae* type strain genome collection was obtained.

Assessment of Fermentative Hydrogen Production Profile

As unexpectedly diverse FHL gene clusters were discovered in the hydrogen-producing *Vibrionaceae*, we performed fermentative hydrogen production profiling on the same strain set as the genome comparison to assess whether the dynamic rearrangements of the FHL gene cluster are related to hydrogen production ability (Figs. 2 and S6-S7).

No hydrogen production was observed in *V. palustris*, *V. zhugei*, *V. plantisponsor* and *P. carnosum* (Figs. S6-S7). *P. carnosum* possessed the atypical FHL gene cluster related to that of *P. toruni* but no hydrogen produced. Hydrogen production was high in *V. porteresiae* and *V. tritonius*, moderate in Gazogenes clade species and low in *P. swingsii* and *P. sanguinicancri* (Figs. S6-S7). The trends in hydrogen production were high in the Porteresiae clade species and less than half in the Gazogenes clades species (Figs. 2 and S6-S7). The hydrogen production of the three *Photobacterium* species was lower than that of *V. aerogenes*, which was the lowest hydrogen producer in the Gazogenes clade (Figs. 2 and S6-S7).

Formate production was high in the Porteresiae clades species and medium to low in the other *Vibrionaceae* species (Figs. 2 and S7BDF). Notably, formate re-import was observed in higher hydrogen-producing species. A strong positive correlation between hydrogen production and formate re-import was observed (Pearson's r was 0.969 supported with BF_{10} 3.646×10^{17}. Lower 95%; CI 0.930, Upper 95%; CI 0.984) (Fig. 3). Higher hydrogen production was related to higher formate re-import in the Porteresiae clade species (Fig. 3).

Inferring The Evolution of FHL Gene Cluster in *Vibrionaceae*

Molecular networks revealed that common ancestry of house-keeping genes and FHL genes were likely to be matched (Fig. 4AB). FHL gene cluster type was related to the clade in *Vibrionaceae*. Comparison of maximum likelihood tree topology also suggests divergence of FHL genes along with the *Vibrionaceae* speciation (Fig. S8). The entire set of the vibrio FHL gene cluster, consisting

of *hyfABCDEFGHIJ*, *fdhF*, *fhlA* and *hypFCDEAB* was found in both the Gazogenes and Porteresiae clades, and *hupE* gene was also observed in the *Photobacterium* species, nevertheless FHL gene clusters were shuffled (Fig. S8).

Modeling of Representative FHL Complex of Vibrios

Three dimensional (3D) structures of soluble arm consisting of HyAGHI and FDH-H and membrane arm consisting of HyfBCDEF of vibrio FHL complex were first predicted (Fig. S9CDEFG). The overall structure of Hyf-FHL of *V. tritonius* strain AM2 was reconstructed based on L-shaped complex I of *Thermus thermophiles* (4HEA) (Fig. S9A) with a long Helix HL at the side of membrane arm [25] by superposed at the membrane anchor HyfC. Three transmembrane proteins HyfBDF were predicted to be 180°C rotated compared to that of complex I based on the membrane anchor (HyfC) position (Fig. S9ABCD). Reconstructed overall structures, active sites and a putative path of electron flow and proton paths of Hyf-FHL of representative vibrios, *V. gazogenes*, *V. aerogenes* and *P. swingsii*, were unlikely to be different in the overall structure of *V. tritonius* (Fig. S9CEFG).

The active site molybdopterin guanine dinucleotide (Mo-bis-PGD) cofactor (C8/11/15/42 quadrad) and the active site selenocysteine SeCys(U)140/H141/R333 at the *E. coli* FdhF position (SeCys140 was Cys in FHL models of *V. tritonius*, *V. gazogenes* and *V. aerogenes*) [24] and an iron-sulfur cluster relay [26] were conserved in the predicted structure of vibrios. In addition to the terminal cysteine (C531) and the conserved β1-β2 loop glutamate (E193) in at the HycE, the canopy residues R478/C529/S283 and proton paths extend from the active site of hydrogenase to the bulk via conserved residues (R478/D529/H284/E437/R395) at the HycE position were conserved in HyfG of representative vibrio Hyf-FHL models [26]. Putative substrate proton paths predicted in Hyc-FHL of *E. coli* were also conserved in those vibrio Hyf-FHL models located in the E-channel Y102/R108/E130/E138/D196/E199/E201/E203/E206/E280 at HycD position in HyfC and E135/L208/H222/A230/S234/K239/H291/E294/T292/H328/H332/K336/D354/E391/K433/C430 except Q285K and R353K at HycC position in HyfB [25-26].

Discussion

Evolutionary history and functional constraints of Hyf-FHL complex have been discussed based on the structural similarity of complex I since the first proposal of structural model of Hyf-FHL of *E. coli* [42]. However, due to the faint activity of Hyf-FHL (FHL-2) in *E. coli* [43], evolutionary

and structural elucidation has been less successful until the discovery of hydrogen producing *V. tritonius* possessing the Hyf-FHL gene cluster as the sole FHL machinery [1,6] and a series of molecular genetics and biochemical studies on FHL-2-related complexes of *Trabulsiella guamensis* [44] and *Pectobacterium atrosepticum* [45]. Here, we update knowledge on diversity of Hyf-FHL gene clusters using all available hydrogen producing *Vibrionaceae* species based on genome comparison using gap-closed complete genomes and comparative fermentative hydrogen production profiling. These analyses unveil unexpected diversity of FHL gene clusters and relationships between gene cluster and function in hydrogen production ability. This data also provides insights into the maintenance of typical Gazogenes type Hyf-FHL complex gene clusters that might show a common ancestry with hydrogen producing vibrios. During speciation, niche separation and changes of ecological function, particularly in the needs of detoxification of formate as a part of formate and pH homeostasis in certain micro-environments, where active carbohydrate metabolisms under anaerobic conditions proceeded, such as the gut of herbivorous sea hare or mangrove sediments and rhizosphere [1,13,46-48], as a selective pressure. A small number of *Vibrionaceae* species still retain the functional Hyf-FHL gene cluster, which expresses to show gas producing phenotypes in *Vibrionaceae* until today.

Comparison of tree topology between house-keeping genes and genes of interest is a simple approach to evaluating divergence or convergence of the genes of interest [49]. Network and tree topologies of FHL genes were unlikely to be different from that of house-keeping gene phylogeny of *Vibrionaceae* species with FHL gene clusters (Figs. 4 and S8), suggesting divergence in FHL gene clusters (Figs. 4AB and S8). It should be emphasized that observed divergence of FHL gene clusters in not only hydrogen producing *Vibrio* but also in those of *Photobacterium* species suggests that the common ancestor of *Vibrionaceae* might have possessed the functional FHL gene clusters, but currently those genes have been lost in most of the species.

Unexpectedly, at least, two more types of FHL gene clusters were discovered in this study (Fig. 1). In total, six types of FHL gene clusters could be described in *Vibrionaceae* [6]. The first type, called Porteresiae-hyf, is retained in *V. tritonius* and *V. porteresiae* in the Porteresiae clade. These types of gene cluster consist of 21 genes (*hyfABCDEFGHIJ*, *hycH*, *hydN*, *fdhF*, *fhlA*, *fdhD* and *hypFCDEAB*) with approximately 24 kb in length on Chr. 1 in those vibrios (Fig. 1) [5-6]. *V. zhugei* and *V. palustris*, which are recently described members of the Porteresiae clade [23], lost the gene clusters (Figs. 1 and S8). The second type, designated as typical Gazogenes-hyf, is retained in *V. gazogenes*, *V. spartinae*, *V. ruber*, *V. rhizosphaerae* and *V. mangrovi* in the Gazogenes clade (Fig.

1). In the typical Gazogenes-hyf, *hupE* gene is located upstream of *hyp* genes (Fig. 1). The length of this gene cluster is ca. 24 kb on Chr. 1 (Fig. 1) [5-6]. The third type, designated as atypical Gazogenes-hyf, formed separated gene clusters after *hydN* on Chr. 1 (Fig. 1). Those atypical gene clusters are reported in *V. aerogenes* [6] and further confirmed in *V. quintilis* in this study (Fig. 1). The fourth type, designated as Jeanii-hyf, forms rearranged and separated gene clusters of typical Gazogenes-hyf, in particular, in *hyp* gene cluster (Figs. 1 and S5). The Jeanii-hyf is located on Chr. 2. The fifth type, designated as Furnisii-hyf, shows lack of *hyfDEF* from typical Gazogenes-hyf, previously reported [6]. The Furnissi-hyf was previously assigned as hyc-FHL, but molecular network clearly shows lower relation to *E. coli* Hyc (Fig. 4). The sixth type, designated as Phosphoreum-hyf, is atypical Furnisii-hyf, in which *hypBCDEA* is separated and located on Chr. 1 (Figs. 1 and S5). The atypical Gazogenes-hyf is confirmed (third type), and the Jeanii-hyf (fourth type) and Phosphoreum-hyf (sixth type) are newly observed in this study. Surprisingly, a further gene search suggested a more diverse world of FHL gene clusters in the genus *Photobacterium* (Fig. S5), which may trigger complete genome reconstruction and evolutionary studies for all *Photobacterium* species in future.

Interestingly, diverse FHL gene clusters are likely to be explained along with *Vibrionaceae* speciation (Fig. 4B). One possible hypothesis is maintenance of typical Gazogenes-hyf gene cluster in the ancestry of *Vibrionaceae*. During speciation, niche separation and adaptation, the FHL gene clusters have evolved into diverse types of gene clusters and are maintained in a portion of current *Vibrionaceae* species. Porteresiae-hyf gene cluster might be caused by lack of *hupE* of the typical Gazogenes-hyf, maintained in *V. tritonius* and *V. porteresiae*, but completely replaced by other genes in *V. zhugei* and *V. palustris*.

How does FHL gene cluster diversity affect fermentative hydrogen production of those vibrios? *V. tritonius* and *V. porteresiae* showed the highest hydrogen production followed by *V. quintilis*, *V. spartinae*, *V. gazogenes*, *V. rhizosphaerae* and *V. ruber* (Figs. 2 and S7). Faint hydrogen production was observed in *V. aerogenes* and *V. mangrovi* (Fig. S7). Little or no hydrogen production was recorded in *Photobacterium* species examined in this study (Fig. S7). Nevertheless, there were two exceptions in *V. aerogenes* and *V. quintilis*, the separated gene cluster franking upstream of *hycH* and downstream of *hydN* region observed in typical and atypical Gazogenes-hyf cluster was unlikely to affect hydrogen production. However, in the *Photobacterium* FHL cluster, separation of *hyp* and/or presence of Chr. 2 affected FHL gene regulation, which caused detrimental effects on normal fermentative hydrogen production. Biosynthesis of active [NiFe] hydrogenase requires

several mature proteins [24,50], and in *E. coli*, the major steps in the formation of the large subunit precursor (pre-HycE) are known to involve the Hyp protein: 1) CN^- ligand (HypFE), (2) attachment of CN^- and CO ligands to Fe atoms (HypCD), (3) insertion of Fe-CO-$(CN^-)_2$ into pre-HycE (HypC), (4) insertion of Ni^{2+} into pre-HycE (HypAB), (5) C-terminal peptide degradation process of pre-HycE [24]. In *P. carnosum* and *P. toruni*, *hypFAB* and *hypBCDEA* were present on different chromosomes, and in *P. swingsii* and *P. sanguinicancri*, the presence of *hypAB* and *hypFCDE* were in opposite directions. The latter gene structure also made us consider plasmid-borne horizontal gene transfer of the FHL gene clusters. Moreover, all the Phosphoreum clade including *P. carnosum* and *P. toruni*, the separated *hypBCDEA* gene clusters were located on a downstream region of an uptake hydrogenase gene cluster (Fig. S5), which means effects of the uptake hydrogenase activity on the net hydrogen production in the Phosphoreum clade species must be re-evaluated in the future, nevertheless, faint formate re-import was observed in *P. carnosum* and *P. toruni* (Fig. S7F). We also consider that the origin of the uptake hydrogenase genes should be analyzed to understand how those *hyp* gene cluster shuffles are regulated on bacterial genomes too. We took a snapshot of the way in which FHL gene clusters rearranged and collapsed in *Vibrionaceae* and normal Hyp gene expression and translation could largely affect hydrogen production in *Vibrionaceae* species.

Interestingly, comprehensive hydrogen profiling using all available hydrogen producing vibrios first revealed a strong positive correlation supported by Bayesian statistics (BF_{10}=3.646x10^{17}) between hydrogen production and formate re-import (Fig. 2). Strains that possessed higher formate re-import produced higher hydrogen production. All species in the Porteresiae clade species could produce higher formate in glucose fermentation, but only *V. tritonius* and *V. porteresiae* performed high levels of formate re-import (Fig. S7). In Gazogenes clade species and *Photobacterium* species, low to moderate levels of formate production were observed in all species, but strains showing formate re-import could produce hydrogen (Figs. S6-S7). One of the primary factors affecting the high hydrogen production of *V. tritonius* and *V. porteresiae* in the Porteresiae clade is the ability of higher formate production and re-import than the other vibrios (Fig. S7). *V. tritonius* was isolated from gut of sea hare [1] and *V. porteresiae* was isolated as endophyte of mangrove-associated wild rice [13]. The gut of sea hare and mangrove rhizosphere are one of the representative marine micro-environments of active fermentation by rich carbohydrates under anaerobic conditions [46-48]. Those micro-environments are possible formate-rich environments, which affected the evolution of typical Gazogenes-type FHL gene cluster making them more functional in *V. tritonius* and *V. porteresiae*.

Formate is a key intermediate to enhance fermentative hydrogen production in *E. coli* at the molecular level [24,41,44-45]. A pyruvate formate lyase, PflB, and a formate channel, FocA, co-ordinates to exclude and re-import formate in *E. coli* under anaerobic conditions as a bacterial cellular formate and pH homeostasis, as so a genetic link as the *focA-pflAB* operon has been suggested in *E. coli* and the related bacteria [24]. However, *focA-pflB* operon was not observed on genomes of the hydrogen-producing vibrios belonging to the Porteresiae and Gazogenes clades. In *E. coli*, at least, *pflA* is expressed constitutive [41], which was also observed in *V. tritonius* [6]. The structure determination of the Hyf-FHL complex described below, FocA-PflB coordination and the single cell-based biochemical and physiological assays could also lead us to better understand high hydrogen production ability of the Porteresiae clade species. More interestingly, FDH-H is likely to be the sole major enzyme for formate detoxification in the Porteresiae and Gazogenes clade species due to lack of FDH-O and FDH-N homologes on those genomes [24,41]. These genotypes strengthen formate metabolism as a possible key factor in maintaining fermentative hydrogen production in specific groups of vibrios.

Finally, we obtained vibrio Hyf-FHL structure models for the first time in this study (Fig. S8). Hyf-FHL models from representative FHL gene clusters are likely to be similar conserving major paths for electron relay and proton (Fig. S9). Due to computing limitations, those models were still mosaic superposed at the membrane anchor. Beyond the Andrews *et al*. 's proposal of Hyf-type FHL [42], structural similarity of Hyf-FHL to Hyc-FHL and MBH and MBS, which major transmembrane subunits rotated 180°C compared to complex I [25-26]. Major differences in putative functional sites in those structural models of vibrio FHL compared to the Hyc-FHL of *E. coli* have not been observed. Proton paths proposed in MBS and complex I were also likely to be conserved in the vibrio Hyf-FHL. As proposed for Hyc-FHL structure of *E. coli*, proton paths to the hydrogenase HycEG might be established to the final proton donor glutamate on β1-β2 loop through HycD and HycC [25]. *in vitro* experiments using a purified Hyc-FHL of *E. coli* were unable to verify the hypothesis of an energy-conserving proton pump of Hyc-FHL [24]. Removal of HyfDEF from the Hyf-FHL complex of *P. atrosepticum* did not affect the physiological activity compared to that of wild type of the protein complex [45]. Considering the same function of antiporter-like transmembrane subunits HyfBDF as a proton translocation, HyfBDF could have an extra-function in formate detoxification transporting substrate protons to hydrogenase. In other words, three more transmembrane antiporter units compared to Hyc-FHL might contribute to high proton transports and hydrogen production in vibrios, which might lack FDH-O and -N, under

formate-rich conditions. Re-arrangements and loss of *hupE* gene in Hyf-FHL gene clusters in *Vibrionaceae* could be related to the level of formate detoxification during speciation and ecophysiological selection and/or adaptation of vibrios. Biochemical studies, structure determination and *in vitro* evolution studies of vibrios FHLs are further needed to clarify the formate detoxification hypothesis.

Conclusion

Using all available hydrogen producing and FHL gene cluster maintaining vibrios, a snapshot of FHL gene dynamic evolution has been observed. As we proposed "formate detoxification hypothesis" in maintaining Hyf-FHL in vibrios, this dynamic evolution might have been driven by environmental formate detoxification as a result of bacterial intercellular formate and pH homeostasis, which further evolves the Porteresiae-type gene clusters to cope with high fermentative hydrogen production under micro-environments where active fermentation is progressing such as the gut of herbivorous animals and mangrove sediments and rhizospheres. Moreover, more diverse FHL gene clusters in particular group of the genus *Photobacterium* including duplicated FHL gene clusters might be present. Those results might help us to further elucidate how complex I family complex including *E. coli* FHL has evolved.

Acknowledgements

We would like to thank Ms. Nishikawa (Hokkaido University) for her technical support. We also would like to thank National Institute of Genetics (NIG) in the use of AlphaFold 2.3.2: Computations were partially performed on the NIG supercomputer at ROIS NIG. This work was partly supported by the MEXT Kaken JP16H04976, JP19H03041 and JSPS Bilateral Research Project between Brazil and Japan 120200101.

References

1. 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 4:414. https://doi.org/10.3389/fmicb.2013.00414.

2. Matsumura Y, Sato K, Al-saari H, Nakagawa S, Sawabe T (2014) Enhanced hydrogen production by a newly described heterotrophic marine bacterium, *Vibrio tritonius* strain AM2, using seaweed as the feedstock. Int J Hydrog Energy 39:7270-7277. https://doi.org/10.1016/j.ijhydene.2014.02.164.

3. Gomez-Gil B, Thompson CC, Vicente A, Matsumura Y, Sawabe T, Iida T, Christen R, Sawabe T (2014). "Family *Vibrionaceae*". In: Rosenberg E et al. (eds) The Prokaryotes, 4th edn. Springer, New York, pp. 659-747.

4. Tanaka M, Endo S, Kotake F et al (2017) *Vibrio aphrogenes* sp. nov., in the Rumoiensis clade isolated from a seaweed. PLoS One. 12:e0180053. https://doi.org/10.1371/journal.pone.0180053.

5. Matsumura Y, Al-saari H, Mino M et al (2015) Identification of a gene cluster responsible for hydrogen evolution in *Vibrio tritonius* strain AM2 with transcriptional analyses. Int J Hydrog Energy 40:9137-9146. https://doi.org/10.1016/j.ijhydene.2015.05.137.

6. Matsumura Y, Sato K, Jiang C et al (2022) Comparative physiology and genomics of hydrogen-producing vibrios. Curr Microbiol 79:360. 10.1007/s00284-022-03065-3.

7. Sato Y, Mino S, Thompson F et al (2024) Core transcriptome of hydrogen producing marine vibrios reveals contribution of glycolysis in their efficient hydrogen production. Curr Microbiol 81, 230. https://doi.org/10.1007/s00284-024-03764-z.

8. Harwood CS (1978) *Beneckea gazogenes* sp. nov., a red, facultatively anaerobic marine bacterium. Curr Microbiol 1:233-238. https://doi.org/10.1007/BF02602849.

9. Shieh WY, Chen AL, Chiu HH (2000) *Vibrio aerogenes* sp. nov., a facultatively anaerobic marine bacterium that ferments glucose with gas production. Int J Syst Evol Microbiol 50:321-329. https://doi.org/10.1099/00207713-50-1-321.

10. Shieh WY, Chen YW, Chaw SM, Chiu HH (2003) *Vibrio ruber* sp. nov., a red, facultatively anaerobic, marine bacterium isolated from sea water. Int J Syst Evol Microbiol 53:479-484. https://doi.org/10.1099/ijs.0.02307-0.

11. Kumar NR, Sudha N (2007) *Vibrio rhizosphaerae* sp. nov., a red-pigmented bacterium that antagonizes phytopathogenic bacteria. Int J Syst Evol Microbiol 57:2241-2246. https://doi.org/10.1099/ijs.0.65017-0.

12. Pujalte MJ, Ortigosa M, Urdaci MC, Garay E, Grimont PAD (1993) *Vibrio mytili* sp. nov., from Mussels. Int J Syst Bacteriol 43:358-362. https://doi.org/10.1099/00207713-43-2-358.

13. Rameshkumar N, Fukui Y, Sawabe T, Nair S (2008) *Vibrio porteresiae* sp. nov., a diazotrophic bacterium isolated from a mangrove-associated wild rice (*Porteresia coarctata* Tateoka) Int J Syst Evol Microbiol 58:1608-1615. https://doi.org/10.1099/ijs.0.65604-0.

14. Rameshkumar N, Sproer C, Lang E, Nair S (2010). *Vibrio mangrovi* sp. nov., a diazotrophic bacterium isolated from mangrove-associated wild rice (*Poteresia coarctata* Tateoka). FEMS Microbiol Lett 307: 35–40. 10.1111/j.1574-6968.2010.01958.x.

15. Rameshkumar N, Gomez-Gil B, Spröer C et al (2011) *Vibrio plantisponsor* sp. nov., a diazotrophic bacterium isolated from a mangrove associated wild rice (*Porteresia coarctata Tateoka*). Syst Appl Microbiol 34: 487–493. https://doi.org/10.1016/j.syapm.2011.02.005.

16. Lucena T, Ruvira MA, Arahal DR et al (2012) *Vibrio aestivus* sp. nov. and *Vibrio quintilis* sp. nov., related to Marisflavi and Gazogenes clades, respectively. Syst Appl Microbiol 35:427–431. 10.1016/j.syapm.2012.08.002. https://doi.org/10.1016/j.syapm.2012.08.002.

17. Lucena T, Arahal DR, Ruvira MA et al (2017) *Vibrio palustris* sp. nov. and *Vibrio spartinae* sp. nov., two novel members of the Gazogenes clade, isolated from salt-marsh plants (*Arthrocnemum macrostachyum* and *Spartina maritima*). Int J Syst Evol Microbiol 67:3506–3512. https://doi.org/10.1099/ijsem.0.002155.

18. Guo Z, Li W, Wang Y et al (2019) *Vibrio zhugei* sp. nov., a moderately halophilic bacterium isolated from pickling sauce. Int J Syst Evol Microbiol 69:1313–1319. https://doi.org/10.1099/ijsem.0.003308.

19. Hilgarth M, Fuertes S, Ehrmann M, Vogel RF (2018) *Photobacterium carnosum* sp. nov., isolated from spoiled modified atmosphere packaged poultry meat. Syst Appl Microbiol 41:44–50. https://doi.org/10.1016/j.syapm.2017.11.002.

20. Labella AM, Arahal DR, Lucena T et al (2017). *Photobacterium toruni* sp. nov., a bacterium isolated from diseased farmed fish. Int J Syst Evol Microbiol 67:4518–4525. ttps://doi.org/10.1099/ijsem.0.002325.

21. Gomez-Gil B, Roque A, Rotllant G et al (2011). *Photobacterium swingsii* sp. nov., isolated from marine organisms. Int J Syst Evol Microbiol 61:315–319. https://doi.org/10.1099/ijs.0.019687-0.

22. Gomez-Gil B, Roque A, Rotllant G et al (2016). *Photobacterium sanguinicancri* sp. nov. isolated from marine animals. Anton van Leeuwen 109:817–825. https://doi.org/10.1007/s10482-016-0681-x.

23. Jiang C, Tanaka M, Nishikawa S, Mino S, Romalde JL, Thompson FL, Gomez-Gil B, Sawabe T (2022) Vibrio clade 3.0: new *Vibrionaceae* evolutionary units using genome-based approach. Cur Microbiol 79:10. https://doi.org/10.1007/s00284-021-02725-0.

24. Pinske C, Sawers RG (2016) Anaerobic formate and hydrogen metabolism, EcoSal Plus 2016. 10.1128/ecosalplus.ESP-0011-2016.

25. Yu H, Shut GJ, Haja DK et al (2021) Evolution of complex I–like respiratory complexes. J Biol Chem 296:100740. https://doi.org/10.1016/j.jbc.2021.100740.

26. Steinhilper R, Höff G, Heider J et al (2022) Structure of the membrane-bound formate hydrogenlyase complex from *Escherichia coli*. Nat Commun 13:5395. 10.1038/s41467-022-32831-x. https://doi.org/10.1038/s41467-022-32831-x.

568 27. Kolmogorov M, Yuan J, Lin Y, Pevzner PA (2019) Assembly of long, error-prone reads using
569 repeat graphs. Nat Biotechnol 37:540–546.
570
571 28. Vaser R, Sović I, Nagarajan N, Šikić M (2017) Fast and accurate de novo genome assembly
572 from long uncorrected reads. Genome Res 27:737-746. 10.1101/gr.214270.116.
573
574 29. Wick RR, Judd LM, Gorrie CL, Holt KE (2017) Unicycler: Resolving bacterial genome
575 assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595.
576 https://doi.org/10.1371/journal.pcbi.1005595.
577
578 30. Tanizawa Y, Fujisawa T, Nakamura Y (2017) DFAST: a flexible prokaryotic genome
579 annotation pipeline for faster genome publication. Bioinformatics 34:1037-1039.
580 https://doi.org/10.1093/bioinformatics/btx713.
581
582 31. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search
583 tool. J Mol Biol 215:403-410. https://doi.org/10.1016/S0022-2836(05)80360-2.
584
585 32. Jones P, Binns D, Chang HY et al (2014) InterProScan 5: Genome-scale protein function
586 classification. Bioinformatics 30:1236-1240. https://doi.org/10.1093/bioinformatics/btu031.
587
588 33. Rutherford K, Parkhill J, Crook J et al (2000) Artemis: Sequence visualization and
589 annotation. Bioinformatics 16:944-945. https://doi.org/10.1093/bioinformatics/16.10.944.
590
591 34. Abbott JC, Aanensen DM, Rutherford K et al (2005) WebACT-an online companion for the
592 Artemis Comparison Tool. Bioinformatics 21:3665-3666.
593 https://doi.org/10.1093/bioinformatics/bti601.
594
595 35. Eren AM, Esen ÖC, Quince C et al (2015) Anvi'o: An advanced analysis and visualization
596 platform for 'omics data. PeerJ 3: e1319. https://doi.org/10.7717/peerj.1319.
597
598 36. Larkin MA, Blackshields G, Brown NP et al (2007). Clustal W and Clustal X version 2.0.
599 Bioinformatics, 23, 2947-2948. 10.1093/bioinformatics/btm404.
600
601 37. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies.
602 Mol Biol Evol 23:254-267. https://doi.org/10.1093/molbev/msj030.

603

604 38. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular evolutionary
605 genetics analysis across computing platforms. Mol Biol Evol 35:1547-1549.
606 10.1093/molbev/msy096.
607

608 39. Liu J, Guo Z, Wu T et al (2023) Enhancing alphafold-multimer-based protein complex
609 structure prediction with MULTICOM in CASP15. Commun Biol 6:1140.
610 https://doi.org/10.1038/s42003-023-05525-3.
611

612 40. Meng EC, Goddard TD, Pettersen EF et al (2023) UCSF ChimeraX: Tools for structure
613 building and analysis. Protein Sci 32:e4792. https://doi.org/10.1002/pro.4792.
614

615 41. Kammel M, Pinske C, Sawers RG (2022) FocA and its central role in fine- tuning pH
616 homeostasis of enterobacterial formate metabolism. Microbiology 168:001253.
617 10.1099/mic.0.001253.
618

619 42. Andrews SC, Berks BC, McClay J et al (1997) A 12-cistron *Escherichia coli* operon (*hyf*)
620 encoding a putative proton-translocating formate hydrogenlyase system. Microbiology 143:3633-
621 3647. https://doi.org/10.1099/00221287-143-11-3633.
622

623 43. Self WT, Hasona A, Shanmugam, KT (2004) Expression and regulation of a silent operon, *hyf*,
624 coding for hydrogenase 4 isoenzyme in *Escherichia coli*. J Bacteriol 186:580-587.
625 https://doi.org/10.1128/jb.186.2.580-587.2004.
626

627 44. Lindenstrauß U, Pinske C (2019) Dissection of the hydrogen metabolism of the
628 enterobacterium *Trabulsiella guamensis*: identification of a formate-dependent and essential
629 formate hydrogenlyase complex exhibiting phylogenetic similarity to complex I. J Bacteriol
630 201:e00160-19. https://orcid.org/0000-0001-6248-2437.
631

632 45. Finney AJ, Lowden R, Fleszar M, Albareda M, Coulthurst SJ, Sargent F (2019) The plant
633 pathogen *Pectobacterium atrosepticum* contains a functional formate hydrogenlyase-2 complex.
634 Mol Microbiol 112:1440-1452. doi:10.1111/mmi.14370.
635
636

637 46. Tsuji A, Nishiyama N, Ohshima M et al (2014) Comprehensive enzymatic analysis of the
638 amylolytic system in the digestive fluid of the sea hare, *Aplysia kurodai*: Unique properties of two
639 α-amylases and two α-glucosidases. FEBS Openbio 4:560-570.
640 https://doi.org/10.1016/j.fob.2014.06.002.
641
642 47. Laux M, Ciapina LP, de Carvalho FM et al (2024) Living in mangroves: a syntrophic scenario
643 unveiling a resourceful microbiome. BMC Microbiol 24: 228. https://doi.org/10.1186/s12866-
644 024-03390-6.
645
646 48. Cai M, Duan C, Zhang X et al (2021) Genomic and transcriptomic dissection of *Theionarchaea*
647 in marine ecosystem. Sci China Life Sci 64. https://doi.org/10.1007/s11427-021-1996-x.
648
649 49. Martínez-Aquino A (2016) Phylogenetic framework for coevolutionary studies: a compass for
650 exploring jungles of tangled trees. Cur Zool 62:393–403. https://doi.org/10.1093/cz/zow018.
651
652 50. Casalot L, Rousset M (2001) Maturation of the [NiFe] hydrogenases. Trend Microbiol 9:228-
653 237. https://doi.org/10.1016/S0966-842X(01)02009-1.
654
655
656 **Figure legends**
657 **Fig. 1. Diversity of genomic architectures of FHL gene cluster among hydrogen-producing**
658 **marine vibrios.** Atypical Gazogenes-hyf type was confirmed in *V. quilitilis* in this study.
659 Phosphreum-hyf and Jeanii-hyf were newly discovered in this study.
660
661 **Fig. 2. Representative fermentative hydrogen production profiles of hydrogen-producing**
662 **vibrios.** (A) *V. porteresiae* MSSRF30^T, (B) *V. gazogenes* ATCC 29988^T and (C) *P. toruni* CECT
663 9189^T. Orange closed diamond: cell growth (OD620), blue closed circle: hydrogen production
664 (mmol), yellow closed triangle formate concentration (mM) and green closed square (glucose
665 concentration (mM). The data from duplicate samples of each strain are shown. See all
666 physiological data in Fig. S5.
667
668 **Fig. 3. Bayesian correlation between hydrogen production and formate re-import.** Pearson's
669 *r* was 0.969 supported with BF_{10} 3.646x1017. Lower 95% CI: 0.930, Upper 95% CI: 0.984. Red
670 closed circle indicates *V. tritonius* and *V. porteresiae*.

Fig. 4. Splitstree networks based on nucleotide sequences of four housekeeping genes and FHL genes (*hyfABCGHIJ*). (A) Network based on four housekeeping genes (*ftsZ*, *mreB*, *pyrH* and *topA*) (data from Jiang et al. (2022) [23]. (B) Network based on *hyfABCGHIJ*. *hyfABCGHIJ* genes were likely to be radiated before *Vibrio*, *Photobacterium* and *E. coli* radiations. Predicted common ancestry was supported in each *Vibrio* clade.

Supplementary figures

Fig. S1. A local rearrangement map of FHL gene cluster region among the Porteresiae clade species. From the top, *V. porteresiae*, *V. tritonius*, *V. palustris* and *V. zhugei* are compared. Red lines indicate homologous nucleotide sequence matches. Blue lines indicate reverse complement nucleotide sequence matches. Black encloses the FHL gene sequence, and yellow encloses homologous nucleotide sequence upstream and downstream of it.

Fig. S2. Pangenome of marine vibrio genomes used in this study. The colors represent the clades: orange represent Gazogenes clade, light blue represent Porteresiae clade, red represent Diazotrophicus clade, black represent Phosphoreum clade, and blue represent Jeanii clade. From the top to bottom; *V. gazogenes*, *V. spartinae*, *V. ruber*, *V. rhizosphaerae*, *V. mangrovi*, *V. aerogenes*, *V. quintilis*, *V. porteresiae*, *V. tritonius*, *V. palustris*, *V. zhugei*, *V. plantisponsor*, *P. carnosum*, *P. toruni*, *P. swingsii* and *P. sanguinicancri*. The red heatmap represents ANI; the darker color indicates closely related species.

Fig. S3. Loci of representative FHL gene clusters of *Vibrionaceae* genomes.

Fig. S4. Genome wide rearrangement map among the seven Gazogenes clade species. From the top, *V. gazogenes*, *V. spartinae*, *V. ruber*, *V. rhizosphaerae*, *V. mangrovi*, *V. aerogenes* and *V. quintilis* are compared. Red lines indicate homologous nucleotide sequence matches. Blue lines indicate reverse complement nucleotide sequence matches. Yellow lines indicate FHL gene sequence matches.

Fig. S5. Mining of FHL gene clusters among *Photobacterium* species. All Phosphoreum and Jeanii clade species possessed FHL gene clusters using draft genome data. In Phosphoreum calde species, genes responsible to uptake hydrogenases were found on a region franking to the separated *hypBCDEA* gene cluster in all Phosphoreum clade species except the *P. piscicola***

due to too fragmented genome. *P. chitinilyticum* and *P. marinum* in the Proteolyticum clade possessed a Furnissii-type FHL gene cluster and an incomplete second FHL-like gene cluster. * and ** indicates data from draft genome sequences. Colors on species names correspond to those of Fig. 1.

Fig. S6. Fermentative hydrogen production profiles of 17 vibrio strains. (A) *V. gazogenes* ATCC 29988^T, (B) *V. spartinae* CECT 9026^T, (C) *V. ruber* LMG 23124^T, (D) *V. rhizosphaerae* MSSRF3^T, (E) *V. mangrovi* CECT 7927^T, (F) *V. aerogenes* LMG 19650^T, (G) *V. quintilis* CECT 7734^T, (H) *V. porteresiae* MSSRF30^T, (I) *V. porteresiae* MSSRF31, (J) *V. tritonius* MA35, (K) *V. palustris* CECT 9027^T, (L) *V. zhugei* KCTC 62784^T, (M) *V. plantisponsor* CECT 7581^T, (N) *P. carnosum* CECT 9394^T, (O) *P. toruni* CECT 9189^T, (P) *P. swingsii* CECT 7576^T, (Q) *P. sanguinicancri* CECT 7579^T. The correspondence between each color, shape and profile is shown below. The data from duplicate samples of each strain are shown.

Fig. S7. Clade-by-clade comparison of hydrogen production. of formate concentration. Left (A, C and E): hydrogen production, right (B, D and F): formate production. A and B: Porteresiae clade, C and D: Gazogenes clade and E and F: the others (Diazotrophicus, Phosphoreum, Jeanii clades). High hydrogen production of the *V. tritonius* and *V. porteresiae* in the Porteresiae clade was suggested with $BF_{10} > 4$ based on Bayesian ANOVA test (n=2). Error bar: SD.

Fig. S8. Divergent evolution hypothesis of FHL gene structure in the genera *Vibrio* and *Photobacterium*. Left and right tree is based on ML tree (GTR + GI model) based on nucleotide sequences of four housekeeping genes and FHL genes, respectively. The same data set in Fig. 4 was used.

Fig. S9. 3D models of vibrio FHLs. AlphaFold-multimer 2.3.2 was used for structure predication. Membrane arm and soluble arm were superposed at HyfC. A & C: Complex I, B & D: *V. tritonius* AM2^T FHL, E: *V. gazogenes* ATCC 29988^T FHL, F: *V. aerogenes* LMG 19650^T FHL, G: *P. swingsii* CECT 7576^T FHL. In A and C, homologous subunits on the membrane arm were indicated with asterisks. N- and C-terminal regions were not trimmed in this prediction. As selenocysteine in the FDH-H of *P. swingsii* was replaced by cysteine for the prediction.

Table S1. List of strains and genome sequences used in this study