Deep-sea vent ε-proteobacterial genomes provide insights into emergence of pathogens

Satoshi Nakagawa†, Yoshihiro Takaki§, Shigeru Shimamura§, Anna-Louise Reysenbach¶, Ken Takai†, and Koki Horikoshi†§

†Subground Animalcule Retrieval (SUGAR) Program and ¤Extremophiles Research Program, Extremobiosphere Research Center (XBR), Japan Agency for Marine–Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka 237-0061, Japan; and ¶Department of Biology, Portland State University, Portland, OR 97201

Edited by Rita R. Colwell, University of Maryland, College Park, MD, and approved June 7, 2007 (received for review January 30, 2007)

Deep-sea vents are the light-independent, highly productive ecosystems driven primarily by chemolithoautotrophic microorganisms, in particular by ε-Proteobacteria phylogenetically related to important pathogens. We analyzed genomes of two deep-sea vent ε-Proteobacteria strains, Sulfito- rum sp. NBC37-1 and Nitratiruptor sp. SB155-2, which provide insights not only into their unusual niche on the seafloor, but also into the origins of virulence in their pathogenic relatives, Helicobacter and Campylobacter species. The deep-sea vent ε-proteobacterial genomes encode for multiple systems for respiration, sensing and responding to environment, and detoxifying heavy metals, reflecting their adaptation to the deep-sea vent environment. Although they are nonpathogenic, both deep-sea vent ε-Proteobacteria share many virulence genes with pathogenic ε-Proteobacteria, including genes for virulence factor MviN, hemolysin, invasion antigen CIAd, and the N-linked glycosylation gene cluster. In addition, some virulence determinants (such as the H2 uptake hydrogenase) and genomic plasticity of the pathogenic descendents are to appear to have roots in deep-sea vent ε-Proteobacteria. These provide ecological advantages for hydrothermal vent ε-Proteobacteria who thrive in their deep-sea habitat and are essential for both the efficient colonization and persistent infections of their pathogenic relatives. Our comparative genomic analysis suggests that there are previously unrecognized evolutionary links between important human/animal pathogens and their nonpathogenic, symbiotic, chemolithoautotrophic deep-sea relatives.

Deep-sea hydrothermal vents are areas on the sea floor of high biological productivity fueled primarily by chemosynthesis. Most of the invertebrates thrive in this hostile environment through their relationship with chemolithoautotrophic proteobacterial symbionts. It has become evident that by far the most prevalent microorganisms at deep-sea vents belong to the ε-Proteobacteria (1, 2). They often account for >90% of total rRNA in various hydrothermal habitats as both epi- or endosymbionts of hydrothermal vent invertebrates (3–6) and free-living organisms associated with actively venting sulfide deposits and in areas where vent fluids and seawater mix [supporting information (SI) Fig. 5] (7–9). Until recently, the metabolic role of these key players in the deep-sea vent ecosystem has remained unknown because of their resistance to cultivation. However, in a recent breakthrough, pure cultures of deep-sea vent ε-Proteobacteria provided evidence that the majority of these microbes were mesophilic to thermophilic (capable of growth at 4°C to 70°C) chemolithoautotrophs capable of oxidation of hydrogen and sulfur compounds coupled with the reduction of oxygen, nitrate, and sulfur compounds (9–11).

These deep-sea vent ε-Proteobacteria diverge before their pathogenic relatives in small-subunit rRNA gene trees, and thus comparative genomics can provide insights into the origins and evolution of pathogenesis in the ε-Proteobacteria such as Helicobacter and Campylobacter species. It is estimated that >50% of the human population is infected by Helicobacter pylori, which causes gastric ulcer and cancer (12). In addition, Campylobacter jejuni is the leading cause of food-borne diarrhea (13). Complete genome sequences have been determined for three Helicobacter species (five different strains in total) (12, 14–17), three strains of Campylobacter jejuni (13, 18) and an additional five Campylobacter strains (18). Complete genome sequences and comparative genomic analysis of these pathogenic ε-Proteobacteria have provided important insights into their virulence factors, habitat specificities, and mechanisms of initial colonization, as well as into the significant genome plasticity (15–20). Wolinella succinogenes DSM1740† is the only nonpathogenic, host (cattle)-adapted ε-proteobacterium for which a complete genome sequence is available (21).

As an effort to address (i) the mechanistic reasons for their prevalence in deep-sea vents and (ii) whether deep-sea vent chemolithoautotrophs could be an evolutionary source of human/animal pathogens, we sequenced and analyzed complete genomes of two representative deep-sea vent ε-Proteobacteria, Sulfito- rum sp. NBC37-1 and Nitratiruptor sp. SB155-2. Both strains were isolated from in situ samplers deployed on the actively venting sulfide mound in the Iheyua North hydrothermal field, Japan (9).

Results and Discussion

Bacterial Strains. Both ε-proteobacterial strains are Gram-negative chemolithoautotrophs, capable of growth via hydrogen and sulfur-compound oxidation under microaerobic and anaerobic conditions (Table 1) (9). The 16S rRNA gene sequences of strains NBC37-1 and SB155-2 are 99.1% and 95.3% similar to those of Sulfito- rum lithotrophicum 42BKT (a strict sulfur-oxidizer) (22) and Nitratiruptor tergarcus MI55–1 (a strict hydrogen-oxidizer) (23), respectively (Fig. 1). DNA–DNA hybridization value between strain NBC37-1 and S. lithotrophicum 42BKT was ≈42%. These results indicated that strains NBC37-1 and SB155-2 are new species within the genera Sulfito- rum and Nitratiruptor, respectively. Although, given our sample collection, it is unclear whether these isolates are epibiotic symbionts or any other variation of symbionts, both strains have many close
relatives colonizing to hydrothermal vent animals such as polychaetes and shrimps (3, 4, 9). In addition, Sulfurovum sp. NBC37-1 is closely related to the endosymbiont of a hydrothermal vent gastropod Alviniconcha (97.1% similarity of 16S rRNA gene sequence) (6). Therefore, the ε-Proteobacteria sequenced in this study probably have some symbiotic relationship with vent animals. Many genomic features, such as small genome size and high coding density (described below), strongly support their symbiotic lifestyle.

**Genome Features.** Strains NBC37-1 and SB155-2 each have a single circular chromosome (2,562,277 and 1,877,931 bp, respectively) with the GC content of 43.8% and 39.7% and represent the largest and smallest genomes among nonpathogenic, sequenced ε-Proteobacteria (2.1–2.2 Mb (21) (A. Copeland, S. Lucas, A. Lapidus, K. Barry, J. C. Detter, T. Glavina, N. Hammon, S. Israni, S. Pitluck, P. Chain et al., http://genome.jgi-psf.org) (SI Fig. 6 and Table 1). One of the keys to surviving and thriving in the deep-sea vents is having multiple mechanisms to overcome broad fluctuations in the availabilities (both kinds and concentration) of energy sources, electron acceptors, and carbon sources, which might explain the larger genome of strain NBC37-1. The small genome size of strain SB155-2 is comparable to host-adapted ε-proteobacterial pathogens, e.g., Helicobacter hepaticus ATCC51449 (1.8 Mb) (14).

**Table 1. General features of the Sulfurovum sp. NBC37-1 and Nitratiruptor sp. SB155-2**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Sulfurovum sp. NBC37-1</th>
<th>Nitratiruptor sp. SB155-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy source</td>
<td>H₂, S偶, SO₄²⁻</td>
<td>H₂, S偶, SO₄²⁻</td>
</tr>
<tr>
<td>Electron acceptor</td>
<td>O₃, NO₂⁻</td>
<td>O₃, NO₂⁻</td>
</tr>
<tr>
<td>Temperature range for growth, °C (optimal)</td>
<td>10–37 (33)</td>
<td>37–65 (55)</td>
</tr>
<tr>
<td>Chromosome</td>
<td>single circular chromosome (2,562,277 and 1,877,931 bp, respectively)</td>
<td>single circular chromosome (2,562,277 and 1,877,931 bp, respectively)</td>
</tr>
<tr>
<td>G+C content, mol%</td>
<td>43.8</td>
<td>39.7</td>
</tr>
<tr>
<td>CDSs with putative function, %</td>
<td>49.4</td>
<td>60.2</td>
</tr>
<tr>
<td>Percent of coding region</td>
<td>90.1</td>
<td>95.1</td>
</tr>
<tr>
<td>No. of rrr operons</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>No. of tRNA genes</td>
<td>44</td>
<td>45</td>
</tr>
</tbody>
</table>

**Carbon and Energy Metabolism.** Whereas strain SB155-2 has the small genome, both deep-sea vent ε-proteobacterial genomes do encode for all of the genes of the reductive tricarboxylic acid cycle for CO₂ fixation (24) and provide the genetic basis for their versatile respiration, including gene clusters for hydrogenases (described below), sulfur-compounds oxidation system (sox) (25), and denitrification (sequential reduction of nitrate to dinitrogen) pathway (Fig. 2). In contrast to those of facultatively autotrophic sulfur-oxidizing bacteria, the sox genes of deep-sea vent ε-Proteobacteria are not organized in a single region. The scattered organization of sox genes was reported for a strictly autotrophic γ-proteobacterium, Thiomicrospira crunogena (SI Fig. 7) (26). In addition to sox, the strains have both cyttoplasmic and periplasmic sulfide-quinone oxidoreductases that catalyze the oxidation of sulfide to elemental sulfur, probably contributing to filamentous sulfur formation by deep-sea vent ε-Proteobacteria (27). Furthermore, consistent with our previous enzyme assay (28), strain NBC37-1 has two copies of sulfate:cytochrome c oxidoreductase (Sor). However, no predicted gene of strain SB155-2 showed significant similarity to any sor in the nonredundant protein database from the National Center for Biotechnology Information, suggesting that this strain has an unique Sor or a different system for sulfate oxidation. Both deep-sea vent ε-Proteobacteria have no dissimilatory sulfite reductase, adenylylsulfate reductase or adenylylsulfate:phosphate adenyltransferase.

**Response to Environment.** Deep-sea vent ε-Proteobacteria do not tolerate high oxygen concentrations as other ε-Proteobacteria do (9). Although both strains have no superoxide dismutase (SI Table 2), strain NBC37-1 has catalase. In addition, there are several genes in both genomes that encode for peroxidases and alkyl hydroperoxide reductases (ahp) that provide these microaerobic bacteria with resistance to oxidative stress (SI Table 2). It was recently demonstrated that the Ahp was the major antioxidant enzyme in the endosymbiont of tubeworm Riftia pachyptila inhabiting deep-sea vents (29). Furthermore, fitting to their metal-rich niche, both strains contain a wide array of mineral transport systems including detoxification mechanisms of heavy metals such as arsenate, cadmium, and copper (SI Table 2). Both strains have a large number of two-component signal transduction systems and PAS-GGDEF systems that are used commonly by prokaryotes to sense and respond to environmental signals (SI Table 2) (30). There are almost no flagellar or chemotaxis genes in the strain NBC37-1 genome, whereas the flagellated strain SB155-2 has 35 flagellar genes, 12 chemotaxis (che) genes, and 8 methyl-accepting chemotaxis (mcp) genes (SI Fig. 8), which are organized mostly in a single region with atypical G+C content (SI Fig. 6 and SI Table 3). These genes exhibit the highest similarities to those of phylogenetically

---

**Fig. 1.** Phylogenetic tree of 16S rRNA gene sequences. The deep-sea vent ε-Proteobacteria are shown in red. Branch points conserved with bootstrap values of 75% (filled circles) and bootstrap value of 50 to 74% (open circles) are indicated. Scale bar represents the expected number of changes per position.

**Nakagawa et al.**
distant Bacteria, e.g., *Aquifex aeolicus*, suggesting that this cluster is a genomic island possibly acquired via horizontal gene transfer. In addition to the potential chemotaxis island in strain SB155-2, we identified several functionally unknown genomic islands (up to 37,176 bp) in both deep-sea vent *γ*-Proteobacteria genomes (SI Figs. 6 and 9).

**Virulence Gene Homologs.** It has been shown that pathogenic *γ*-Proteobacteria have acquired virulence genes horizontally from evolutionarily distant sources (12, 13). Both deep-sea vent *γ*-Proteobacteria genomes lack orthologs of virulence genes of pathogenic *γ*-Proteobacteria, such as type IV secretion pathway and *cag* pathogenicity island genes. However, both deep-sea vent *γ*-Proteobacteria share many virulence genes that were identified in pathogenic *γ*-Proteobacteria, including genes for virulence factor MviN, hemolysin, invasion antigen CiaB, and lysine decarboxylase (SI Table 2). One of the most remarkable virulence genes in deep-sea vent *γ*-Proteobacteria is the *N*-linked glycosylation (NLG) gene cluster. This cluster includes a pgIB/stt3 ortholog encoding for an essential dolicholipidyltransferase with a conserved catalytic motif (WWDYG) (Fig. 3A) (18, 31). The glycosylation by PgIB/Stt3 is the most frequent protein-modification system of secretory and membrane protein in Eukarya but was reported exclusively in *Campylobacter* in Bacteria (31). Although little is known about the origin of bacterial NLG, glycosylated proteins play important roles in microbial escape from the host immune system in both symbiotic and pathogenic host–bacterial interaction (32). A *Campylobacter* NLG mutant has reduced ability in adhesion and invasion into host tissue (33). In the PgIB/Stt3 phylogenetic tree, deep-sea vent *γ*-Proteobacteria represent the deepest branching Bacteria (SI Fig. 10), therefore, bacterial NLG probably has arisen in deep-sea *γ*-Proteobacteria to maintain a symbiotic relationship with hydrothermal vent invertebrates. Establishing a symbiotic relationship (whether endosymbiotic or epibiotic) with an animal should provide more opportunity in interacting with other microbes, including pathogens to the host, thus leading to acquiring virulence genes in pathogenic descendants.

**Hydrogenase.** It is increasingly recognized that pathogenic *γ*-Proteobacteria have virulence determinants that are not classified as virulence genes in general but do play important roles in virulence. For example, *Helicobacter* species have a *H2*-uptake hydrogenase encoded outside the pathogenicity island, which is essential for its efficient initial colonization (34, 35). Interestingly, these pathogenic *γ*-Proteobacteria have only a single *H2*-uptake hydrogenase (36) and do not have the suite of hydrogenases that the deep-sea vent *γ*-Proteobacteria have. Strain NBC37-1 has four different hydrogenases (two of *H2*-uptake type, one *H2*-sensing type, and one *H2*-evolving type), and strain SB155-2 has three (one each of *H2*-uptake type, *H2*-sensing type, and *H2*-evolving type) (SI Fig. 11 and SI Table 2). Despite their genomic plasticity (described below), the *γ*-Proteobacteria have well-conserved *H2*-uptake hydrogenase gene clusters (Fig. 3B), offering an explanation for the presence of high-affinity hydrogenase in heterotrophic pathogens (35). Because the concentration of *H2* in stomachs (∼100 μM) (34) is much higher than in habitats of the deep-sea vent *γ*-Proteobacteria, which are generally less than 10 μM, it is likely that pathogenic *γ*-Proteobacteria lost *H2*-sensing and -evolving hydrogenases in the course of adaptation to constantly *H2*-rich environments. Likewise, the presence of several genes specific to the reductive tricarboxylic acid cycle in *Helicobacter, Campylobacter,* and *Wolinella* (12, 13, 18, 19, 21) further suggests that these heterotrophic relatives of chemolithotrophic deep-sea vent *γ*-Proteobacteria have adapted to their pathogenic lifestyle and no longer need to fix *CO2*.

**Genomic Plasticity.** Both deep-sea vent *γ*-Proteobacteria and their pathogenic relatives lack many DNA-repair genes (Fig. 4),...
leading to frequent recombination, mutation, gene loss, and horizontal gene transfer (19, 37–39). Such genomic plasticity is not advantageous for organisms living in stable environments (40). Nevertheless, genomic plasticity increases microdiversity and enables H. pylori to persist in infections (37, 39, 41). Similarly, it is likely that the high microdiversity reported for deep-sea vent e-Proteobacteria (9) results from the genomic plasticity of e-Proteobacteria and confers a competitive advantage enabling this lineage to thrive in ever changing steep physical–chemical gradients. The genomic plasticity may explain not only the dominance but also the endemism of many e-Proteobacteria to very specific habitats.

Conclusions
Genomes of deep-sea vent e-Proteobacteria give insights into their versatile energy metabolism, and into an enhanced ability to sense and respond to conditions outside cells. These reflect their adaptation to the deep-sea vent environment by responding to geochemical gradients and through associations with deep-sea invertebrates. Our comparative genomic analysis shows that deep-sea vent chemooautotrophy has provided the core of virulence for important human/animal pathogens. As additional genome sequences become available, the complex forces that have molded this physiologically diverse group of Bacteria will be further resolved. That autotrophy at deep-sea vents and pathogenesis in the acid gut have common genomic roots is evidence for the role genomic plasticity has played in the diversification of microbial life.

Materials and Methods

Bacterial Strains. The deep-sea vent e-Proteobacteria strains were isolated from a 30-m-tall sulfide mound in the Ihey North field, Japan (water depth, 1,000 m) (SI Fig. 5) (9). Strains NBC37-1 and SB155-2 were grown without shaking in MMJHS medium at 33 and 55°C, respectively. MMJHS medium contained per liter of synthetic seawater: 1 g of NaHCO₃, Na₂S₄O₇·H₂O and NaNO₃, 30 g of S⁰ and 10 ml of vitamin solution (10, 11). The medium was prepared under a H₂/CO₂ (80:20) gas phase (300 kPa). The pH of the medium was adjusted to 6.7. Genomic DNA was obtained from late-exponential growth-phase cultures. To clarify the phylogenetic relationship between strain NBC37-1 and S. lithotrophicum 42BKT, we performed DNA–DNA hybridization following the method by Ezaki et al. (42). Briefly, RNase-treated and heat-denatured DNA was adsorbed to microplates (Maxisorb; Nalge Nunc, Rochester, NY). Probe DNA was labeled with photobiotin (Sigma, St. Louis, MO), Salmon sperm DNA (Nacalai Tesque, Kyoto, Japan) was used as a control. After sonication and denaturation of probe DNA, DNA–DNA hybridization was performed at 30°C for 3 h. Streptavidin-conjugated β-galactosidase (GIBCO BRL, Gaithersburg, MD) and 4-methylumbelliferyl-β-D-galactopyranoside (Sigma) were used as an enzyme and a fluorogenic substrate, respectively. The reaction product was quantified with a microplate reader (Wallac 1420 ARVO sx; PerkinElmer, Boston, MA) at wavelengths of 360 nm for excitation and 450 nm for emission.

Sequencing and Assembly. The complete genomes were sequenced by using the whole-genome shotgun method as described (43), from libraries of 1.5–2.2 kb and 3–5 kb insert size. Sequences were assembled by using Phred/Phrap (P. Green and B. Ewing, University of Washington, Seattle, WA), producing an average of 8-fold coverage across each genome. Gaps were closed by editing, walking library clones, and linking assemblies by PCR.

Nakagawa et al.
The sequence assemblies were confirmed by pulsed-field gel electrophoresis and Southern blotting analysis. The low sequence quality region was searched with Phrap and improved by sequencing. Genome sequences of Sulfurvom sp. NBC37-1 and Nitratiruptor sp. SB155-2 are deposited under DDBJ/EMBL/GenBank accession nos. AP009179 and AP009178, respectively.

**Sequence Analysis and Annotation.** We identified ORFs using Glimmer (44) and GeneMarkS (45). Regions <30 aa were eliminated. Additional ORFs >50 aa were searched from intergenic regions. The ORFs were compared with the nonredundant protein database of the NCBI by using BLASTP (46). The protein domain families were searched against the PFAM database (47). The annotated functions were classified based on the KEGG database (49). The metabolic pathway was reconstructed based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (49).

Signal peptides in proteins were predicted by using the SIGNALP (50), and membrane-spanning domains were predicted by using SOSUI (51). Transfer RNAs were predicted by tRNAscan-SE (52). By using cumulative GC profile method (53), we detected genomic segments that are different in GC content from the rest of the genome. We identified a segment as horizontally transferred genomic island when it has genomic island-specific features (53), such as biased codon usage, the presence of mobile genes (e.g., integrases and transposases), and the presence of repeat elements and a tRNA locus at junctions.

**Phylogenetic Tree Constructions.** A phylogenetic tree of 16S rRNA gene sequences was constructed by using ARB (54). Alignments were manually checked, and unambiguously aligned nucleotide positions were used. Distances were estimated with the Jukes–Cantor correction. Bootstrapping (100 resamplings) was performed with PAUP* 4.0. Desulfovibrio vulgaris (δ-proteobacterium) served as outgroup. Phylogenetic trees based on amino acid sequences of PglB/Sst3 and hydrogenase (SI Figs. 10 and 11) was created by using ClustalX (55).

We thank the crew of the RV Natsushima and the DSROV Hyper Dolphin team for their assistance and K. Uematsu for technical help with TEM. A.-L.R. was supported by National Science Foundation Grant OCE-0242038.