

# Deep-sea vent $\epsilon$ -proteobacterial genomes provide insights into emergence of pathogens

Satoshi Nakagawa<sup>†‡</sup>, Yoshihiro Takaki<sup>§</sup>, Shigeru Shimamura<sup>§</sup>, Anna-Louise Reysenbach<sup>¶</sup>, Ken Takai<sup>†</sup>, and Koki Horikoshi<sup>†§</sup>

<sup>†</sup>Subground Animalcule Retrieval (SUGAR) Program and <sup>§</sup>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 <sup>¶</sup>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 25, 2007)

Deep-sea vents are the light-independent, highly productive ecosystems driven primarily by chemolithoautotrophic microorganisms, in particular by  $\epsilon$ -Proteobacteria phylogenetically related to important pathogens. We analyzed genomes of two deep-sea vent  $\epsilon$ -Proteobacteria strains, *Sulfurovum* 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  $\epsilon$ -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  $\epsilon$ -Proteobacteria share many virulence genes with pathogenic  $\epsilon$ -Proteobacteria, including genes for virulence factor MviN, hemolysin, invasion antigen CiaB, and the *N*-linked glycosylation gene cluster. In addition, some virulence determinants (such as the H<sub>2</sub>-uptake hydrogenase) and genomic plasticity of the pathogenic descendants appear to have roots in deep-sea vent  $\epsilon$ -Proteobacteria. These provide ecological advantages for hydrothermal vent  $\epsilon$ -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.

$\epsilon$ -Proteobacteria | comparative microbial genomics | deep-sea hydrothermal vent | pathogenesis | symbiosis

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  $\epsilon$ -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  $\epsilon$ -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  $\epsilon$ -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  $\epsilon$ -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  $\epsilon$ -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<sup>T</sup> is the only nonpathogenic, host (cattle)-adapted  $\epsilon$ -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  $\epsilon$ -Proteobacteria, *Sulfurovum* 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 Iheya North hydrothermal field, Japan (9).

## Results and Discussion

**Bacterial Strains.** Both  $\epsilon$ -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 with those of *Sulfurovum lithotrophicum* 42BKT (a strict sulfur-oxidizer) (22) and *Nitratiruptor tergaricus* MI55-1 (a strict hydrogen-oxidizer) (23), respectively (Fig. 1). DNA–DNA hybridization value between strain NBC37-1 and *S. lithotrophicum* 42BKT was  $\approx$ 42%. These results indicated that strains NBC37-1 and SB155-2 are new species within the genera *Sulfurovum* 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

Author contributions: S.N. and Y.T. designed research; S.N., Y.T., and S.S. performed research; S.N., Y.T., and S.S. analyzed data; and S.N., A.-L.R., K.T., and K.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Abbreviation: NLG, *N*-linked glycosylation.

Data deposition: Genome sequences of *Sulfurovum* sp. NBC37-1 and *Nitratiruptor* sp. SB155-2 have been deposited in DDBJ/EMBL/GenBank databases under the project accession nos. AP009179 and AP009178, respectively.

<sup>†</sup>To whom correspondence should be addressed. E-mail: nakagawas@jamstec.go.jp.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0700687104/DC1](http://www.pnas.org/cgi/content/full/0700687104/DC1).

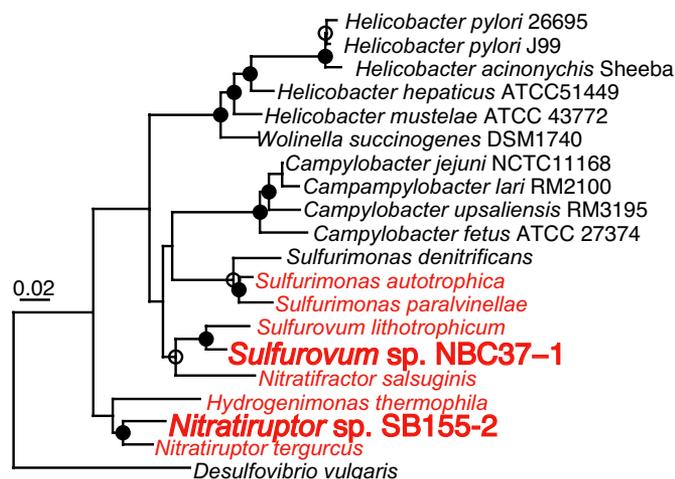
© 2007 by The National Academy of Sciences of the USA

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

Characteristics	<i>Sulfurovum</i> sp. NBC37-1	<i>Nitratiruptor</i> sp. SB155-2
Energy source	H <sub>2</sub> , S <sup>2-</sup> , S <sup>0</sup> , S <sub>2</sub> O <sub>3</sub> <sup>2-</sup>	H <sub>2</sub> , S <sup>2-</sup> , S <sup>0</sup> , S <sub>2</sub> O <sub>3</sub> <sup>2-</sup>
Electron acceptor	O <sub>2</sub> , NO <sub>3</sub> <sup>-</sup>	O <sub>2</sub> , NO <sub>3</sub> <sup>-</sup>
Temperature range for growth, °C (optimal)	10–37 (33)	37–65 (55)
Chromosome		
Size, bp	2,562,277	1,877,931
G+C content, mol %	43.8	39.7
Coding sequence (CDS)		
Total no.	2,466	1,857
CDSs with putative function, %	49.4	60.2
Percent of coding region	90.1	95.1
No. of <i>rrn</i> operons	3	3
No. of tRNA genes	44	45

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  $\epsilon$ -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.

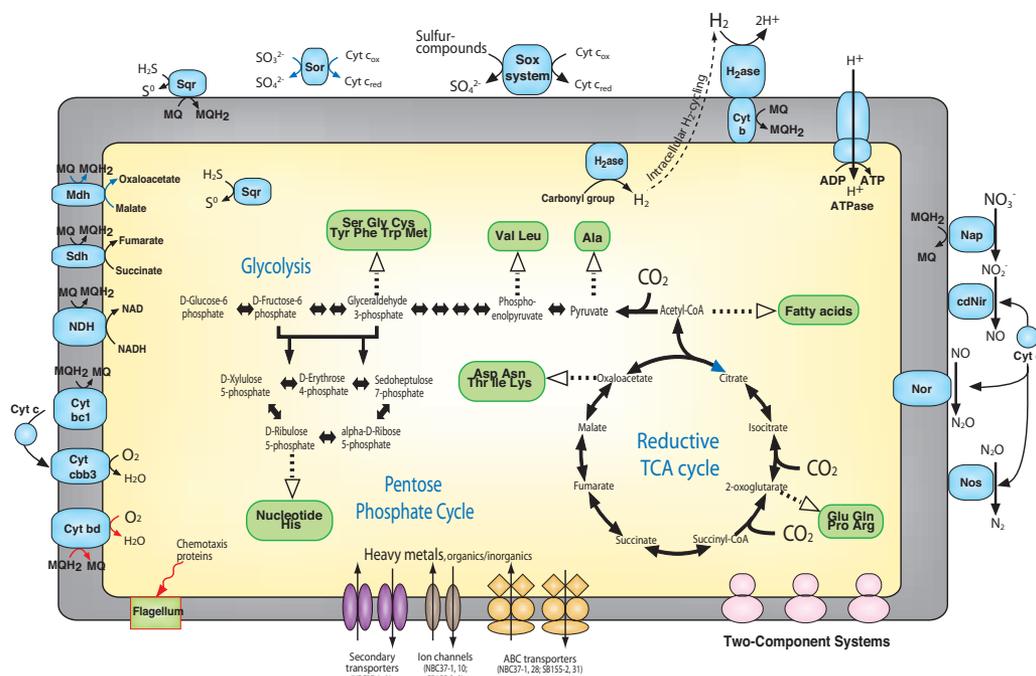
**Genome Features.** Strains NBC37-1 and SB155-2 each have a single circular chromosome (2,562,277 and 1,877,931 bp, respectively) with the GC content of 43.8% and 39.7% and represent the largest and smallest genomes among nonpathogenic, sequenced  $\epsilon$ -Proteobacteria (2.1–2.2 Mb (21) (A. Copeland, 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  $\epsilon$ -proteobacterial pathogens, e.g., *Helicobacter hepaticus* ATCC51449 (1.8 Mb) (14).



**Fig. 1.** Phylogenetic tree of 16S rRNA gene sequences. The deep-sea vent  $\epsilon$ -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.

**Carbon and Energy Metabolism.** Whereas strain SB155-2 has the small genome, both deep-sea vent  $\epsilon$ -proteobacterial genomes do encode for all of the genes of the reductive tricarboxylic acid cycle for CO<sub>2</sub> 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  $\epsilon$ -Proteobacteria are not organized in a single region. The scattered organization of *sox* genes was reported for a strictly autotrophic  $\gamma$ -proteobacterium, *Thiomicrospira crunogena* (SI Fig. 7) (26). In addition to *sox*, the strains have both cytoplasmic and periplasmic sulfide-quinone oxidoreductases that catalyze the oxidation of sulfide to elemental sulfur, probably contributing to filamentous sulfur formation by deep-sea vent  $\epsilon$ -Proteobacteria (27). Furthermore, consistent with our previous enzyme assay (28), strain NBC37-1 has two copies of sulfite: 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 a unique *Sor* or a different system for sulfite oxidation. Both deep-sea vent  $\epsilon$ -Proteobacteria have no dissimilatory sulfite reductase, adenylylsulfate reductase or adenylylsulfate:phosphate adenylyltransferase.

**Response to Environment.** Deep-sea vent  $\epsilon$ -Proteobacteria do not tolerate high oxygen concentrations as other  $\epsilon$ -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. 2.** Central metabolism and solute transport in the deep-sea vent  $\epsilon$ -Proteobacteria. Pathways for which no predictable enzymes were found in strains SB155-2 and NBC37-1 genomes are shown in blue and red arrows, respectively. Numbers of transport machineries are shown for both strains. The KEGG database was used for the reconstruction of metabolic pathways. Cyt, cytochrome; H<sub>2</sub>ase, hydrogenase; Sqr, sulfide-quinone oxidoreductases; Nap, periplasmic nitrate reductase; cdNir, cytochrome *cd*; nitrite reductase; Nor, nitric oxide reductase; Nos, nitrous oxide reductase; Mdh, malate dehydrogenase; Sdh, succinate dehydrogenase.

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  $\epsilon$ -Proteobacteria genomes (SI Figs. 6 and 9).

**Virulence Gene Homologs.** It has been shown that pathogenic  $\epsilon$ -Proteobacteria have acquired virulence genes horizontally from evolutionarily distant sources (12, 13). Both deep-sea vent  $\epsilon$ -Proteobacteria genomes lack orthologs of virulence genes of pathogenic  $\epsilon$ -Proteobacteria, such as type IV secretion pathway and *cag* pathogenicity island genes. However, both deep-sea vent  $\epsilon$ -Proteobacteria share many virulence genes that were identified in pathogenic  $\epsilon$ -Proteobacteria, including genes for virulence factor MviN, hemolysin, invasion antigen CiaB, and lytic murein transglycosylase (SI Table 2). One of the most remarkable virulence genes in deep-sea vent  $\epsilon$ -Proteobacteria is the *N*-linked glycosylation (NLG) gene cluster. This cluster includes a *pglB/stt3* ortholog encoding for an essential oligosaccharyl-transferase with a conserved catalytic motif (WWDYG) (Fig. 3A) (18, 31). The glycosylation by PglB/Stt3 is the most frequent protein-modification system of secretory and membrane protein in Eukarya but was reported exclusively in *Campylobacteriales* 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 PglB/Stt3 phylogenetic tree, deep-sea vent  $\epsilon$ -Proteobacteria represent the deepest branching Bacteria (SI Fig. 10), therefore, bacterial NLG probably has arisen in deep-sea  $\epsilon$ -Proteobacteria to maintain a symbiotic relationship with hydrothermal vent invertebrates. Establishing a symbiotic relationship (whether endosymbiotic or episymbiotic) with an ani-

mal 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  $\epsilon$ -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 H<sub>2</sub>-uptake hydrogenase encoded outside the pathogenicity island, which is essential for its efficient initial colonization (34, 35). Interestingly, these pathogenic  $\epsilon$ -Proteobacteria have only a single H<sub>2</sub>-uptake hydrogenase (36) and do not have the suite of hydrogenases that the deep-sea vent  $\epsilon$ -Proteobacteria have. Strain NBC37-1 has four different hydrogenases (two of H<sub>2</sub>-uptake type, one H<sub>2</sub>-sensing type, and one H<sub>2</sub>-evolving type), and strain SB155-2 has three (one each of H<sub>2</sub>-uptake type, H<sub>2</sub>-sensing type, and H<sub>2</sub>-evolving type) (SI Fig. 11 and SI Table 2). Despite their genomic plasticity (described below), the  $\epsilon$ -Proteobacteria have well-conserved H<sub>2</sub>-uptake hydrogenase gene clusters (Fig. 3B), offering an explanation for the presence of high-affinity hydrogenase in heterotrophic pathogens (35). Because the concentration of H<sub>2</sub> in stomachs ( $\approx 100 \mu\text{M}$ ) (34) is much higher than in habitats of the deep-sea vent  $\epsilon$ -Proteobacteria, which are generally less than  $10 \mu\text{M}$ , it is likely that pathogenic  $\epsilon$ -Proteobacteria lost H<sub>2</sub>-sensing and -evolving hydrogenases in the course of adaptation to constantly H<sub>2</sub>-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 chemolithoautotrophic deep-sea vent  $\epsilon$ -Proteobacteria have adapted to their pathogenic lifestyle and no longer need to fix CO<sub>2</sub>.

**Genomic Plasticity.** Both deep-sea vent  $\epsilon$ -Proteobacteria and their pathogenic relatives lack many DNA-repair genes (Fig. 4),



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 *Sulfurovum* 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 Clusters of Orthologous Groups (COG) database (48). 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 SIG-NALP (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* ( $\delta$ -proteobacterium) served as outgroup. Phylogenetic trees based on amino acid sequences of PglB/Stt3 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.

- Campbell BJ, Engel AS, Porter ML, Takai K (2006) *Nat Rev Microbiol* 4:458–468.
- Takai K, Nakagawa S, Reysenbach A-L, Hoek J (2006) in *Geophysical Monograph Series*, eds Christie D, Fisher C, Lee S-M, Givens S (American Geophysical Union, Washington, DC), pp 185–213.
- Polz MF, Cavanaugh CM (1995) *Proc Natl Acad Sci USA* 92:7232–7236.
- Cary SC, Cottrell MT, Stein JL, Camacho F, Desbruyères D (1997) *Appl Environ Microbiol* 63:1124–1130.
- Urakawa H, Dubilier N, Fujiwara Y, Cunningham DE, Kojima S, Stahl DA (2005) *Environ Microbiol* 7:750–754.
- Suzuki Y, Sasaki T, Suzuki M, Nogi Y, Miwa T, Takai K, Neelson KH, Horikoshi K (2005) *Appl Environ Microbiol* 71:5440–5450.
- Longnecker K, Reysenbach AL (2001) *FEMS Microbiol Ecol* 35:287–293.
- Nakagawa S, Takai K, Inagaki F, Chiba H, Ishibashi J, Kataoka S, Hirayama H, Nunoura T, Horikoshi K, Sako Y (2005) *FEMS Microbiol Ecol* 54:141–155.
- Nakagawa S, Takai K, Inagaki F, Hirayama H, Nunoura T, Horikoshi K, Sako Y (2005) *Environ Microbiol* 7:1619–1632.
- Takai K, Inagaki F, Nakagawa S, Hirayama H, Nunoura T, Sako Y, Neelson KH, Horikoshi K (2003) *FEMS Microbiol Lett* 218:167–174.
- Nakagawa S, Takai K (2006) *Methods Microbiol* 35:55–91.
- Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA, et al. (1997) *Nature* 388:539–547.
- Parkhill J, Wren BW, Mungall K, Ketley JM, Churcher C, Basham D, Chillingworth T, Davies RM, Feltwell T, Holroyd S, et al. (2000) *Nature* 403:665–668.
- Suerbaum S, Josenhans C, Sterzenbach T, Drescher B, Brandt P, Bell M, Droge M, Fartmann B, Fischer HP, Ge Z, et al. (2003) *Proc Natl Acad Sci USA* 100:7901–7906.
- Oh JD, Kling-Bäckhed H, Giannakis M, Xu J, Fulton RS, Fulton LA, Cordum HS, Wang C, Elliott G, Edwards J, et al. (2006) *Proc Natl Acad Sci USA* 103:9999–10004.
- Alm RA, Ling LS, Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, deJonge BL, et al. (1999) *Nature* 397:176–180.
- Eppinger M, Baar C, Linz B, Raddatz G, Lanz C, Keller H, Morelli G, Gressmann H, Achtman M, Schuster SC (2006) *PLoS Genet* 2:e120.
- Fouts DE, Mongodin EF, Mandrell RE, Miller WG, Rasko DA, Ravel J, Brinkac LM, DeBoy RT, Parker CT, Daugherty SC, et al. (2005) *PLoS Biol* 3:e15.
- Eppinger M, Baar C, Raddatz G, Huson DH, Schuster SC (2004) *Nat Rev Microbiol* 2:872–885.
- Marais A, Mendz GL, Hazell SL, Mégraud F (1999) *Microbiol Mol Biol Rev* 63:642–674.
- Baar C, Eppinger M, Raddatz G, Simon J, Lanz C, Klimmek O, Nandakumar R, Gross R, Rosinus A, Keller H, et al. (2003) *Proc Natl Acad Sci USA* 100:11690–11695.
- Inagaki F, Takai K, Neelson KH, Horikoshi K (2004) *Int J Syst Evol Microbiol* 54:1477–1482.
- Nakagawa S, Takai K, Inagaki F, Horikoshi K, Sako Y (2005) *Int J Syst Evol Microbiol* 55:925–933.
- Hügler M, Wirsén CO, Fuchs G, Taylor CD, Sievert SM (2005) *J Bacteriol* 187:3020–3027.
- Friedrich CG, Bardischewsky F, Rother D, Quentmeier A, Fischer J (2005) *Curr Opin Microbiol* 8:253–259.
- Scott KM, Sievert SM, Abril FN, Ball LA, Barrett CJ, Blake RA, Boller AJ, Chain PS, Clark JA, Davis CR, et al. (2006) *PLoS Biol* 4:e383.
- Taylor CD, Wirsén CO (1997) *Science* 277:1483–1485.
- Takai K, Campbell BJ, Cary SC, Suzuki M, Oida H, Nunoura T, Hirayama H, Nakagawa S, Suzuki Y, Inagaki F, et al. (2005) *Appl Environ Microbiol* 71:7310–7320.
- Markert S, Arndt C, Felbeck H, Becher D, Sievert SM, Hügler M, Albrecht D, Robidart J, Bench S, Feldman RA, et al. (2007) *Science* 315:247–250.
- Galperin MY, Nikolskaya AN, Koonin EV (2001) *FEMS Microbiol Lett* 203:11–21.
- Szymanski CM, Wren BW (2005) *Nat Rev Microbiol* 3:225–237.
- Hooper LV, Gordon JI (2001) *Glycobiology* 11:1R–10R.
- Szymanski CM, Burr DH, Guerry P (2002) *Infect Immun* 70:2242–2244.
- Olson JW, Maier RJ (2002) *Science* 298:1788–1790.
- Maier RJ, Olson J, Olczak A (2003) *J Bacteriol* 185:2680–2682.
- Vignais PM, Billoud B, Meyer J (2001) *FEMS Microbiol Rev* 25:455–501.
- Monack DM, Mueller A, Falkow S (2004) *Nat Rev Microbiol* 2:747–765.
- Kraft C, Suerbaum S (2005) *Int J Med Microbiol* 295:299–305.
- Kang J, Blaser MJ (2006) *Nat Rev Microbiol* 4:826–836.
- Tamames J (2001) *Genome Biol* 2:RESEARCH0020.
- Robinson K, Loughlin MF, Potter R, Jenks PJ (2005) *J Infect Dis* 191:579–587.
- Ezaki T, Hashimoto Y, Yabuuchi E (1989) *Int J Syst Bacteriol* 39:224–229.
- Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JA, et al. (1995) *Science* 269:496–512.
- Salzberg SL, Delcher AL, Kasif S, White O (1998) *Nucleic Acids Res* 26:544–548.
- Besemer J, Borodovsky M (2005) *Nucleic Acids Res* 33:W451–454.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) *Nucleic Acids Res* 25:3389–3402.
- Bateman A, Birney E, Durbin R, Eddy SR, Finn RD, Sonnhammer EL (1999) *Nucleic Acids Res* 27:260–262.
- Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) *Nucleic Acids Res* 28:33–36.
- Kanehisa M, Goto S, Kawashima S, Nakaya A (2002) *Nucleic Acids Res* 30:42–46.
- Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) *Int J Neural Syst* 8:581–599.
- Hirokawa T, Boon-Chieng S, Mitaku S (1998) *Bioinformatics* 14:378–379.
- Lowe TM, Eddy SR (1997) *Nucleic Acids Res* 25:955–964.
- Zhang R, Zhang CT (2004) *Bioinformatics* 20:612–622.
- Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadukumar B, Buchner A, Lai T, Steppi S, Jobb G, et al. (2004) *Nucleic Acids Res* 32:1363–1371.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) *Nucleic Acids Res* 25:4876–4882.