Endemicity of the cosmopolitan mesophilic chemolithoautotroph *Sulfurimonas* at deep-sea hydrothermal vents

Mino et al.

Supplementary Methods

Enrichment, isolation and selection of strains

Serial dilution of the samples was performed at 25°C and 33°C with 3 ml of MMJHS medium (Takai *et al*., 2003), and the purity of isolates was confirmed as described by Mino *et al.* (2013). Extraction of the genomic DNA and amplification of the 16S rRNA gene of the isolates were performed as described by Mino *et al.* (2013). Almost complete sequences of the 16S rRNA gene were assembled using Sequencher ver 4.8 (Gene Codes Corporation, Ann Arbor, MI, USA). To determine the phylogenetic positions of isolates, sequences were aligned using MacQIIME (http://www.wernerlab.org/software /macqiime) in combination with PyNAST (Caporaso *et al.*, 2010) and were compiled using the ARB software ver 03.08.22 (Ludwig *et al.*, 2004).

Designing the primers and sequencing

11 protein-coding genes were selected for MLSA based on the MLST database of pathogenic epsilonproteobacterium *Helicobacter pylori* and *Campylobacter jejuni* (http://www.mlst.net): *tkt*, *atpA*, *dnaK*, *napA*, *metG*, *gyrB*, *glyA*, *feoB*, *valS*, *rplA*, and *pheS*. Consensus primers (Supplementary Table S2) were designed on the basis of publicly available genes encoded in the complete epsilonproteobacterial genomes deposited in the NCBI/DDBJ/EMBL databases: *Sulfurimonas autotrophica* DSM 16294 (CP002205) (Sikorski *et al.*, 2010), *Sulfurovum* sp. NBC37-1 (AP009179) (Nakagawa *et al.*, 2007), *Nitratiruptor* sp. SB155-2 (AP009178) (Nakagawa *et al.*, 2007), *Nautilia profundicola* AmH (CP001279) (Campbell *et al.*, 2009), *H.*

pylori SMJ180 (CP002073) (Kersulyte *et al.*, 2000), *H. pylori* 26695 (AE000511) (Tomb *et al.*, 1997), *H. pylori* P12 (CP001217) (Fischer *et al.*, 2010), and *C. jejuni* NCTC 11168 (AL111168) (Parkhill *et al.*, 2000). ClustalX ver 2.0 was used for the alignment of nucleotide sequences (Larkin *et al.*, 2007). PCR using LA *Taq* polymerase with GC buffer (Takara Bio, Otsu, Japan) was performed with the following thermal profile: 96°C for 2 min followed by 32 cycles of denaturation at 96°C for 20 sec, primer annealing for 45 sec at 52 or 55°C, and extension at 72°C for 2 min. For samples that showed lower amplification, three additional PCR cycles were added to obtain sufficient product. Sequences were assembled using Sequencher ver 4.8. All assembled sequences were aligned in ClustalX and were translated to amino acids using the online program EMBOSS-Transeq (http://www.ebi.ac.uk/emboss/transeq/). Sequences obtained in this study have been deposited in DDBJ/EMBL/GenBank under Accession No. AB489258-AB489845, AB697196-AB697375, and LC73183-LC074357.

Suitability of the genes for MLSA

MLSA is based on housekeeping genes, which are subject to purifying selection, slow evolution, and variation within gene sequences (Maiden, 2006; Margos *et al*., 2008). The suitability for MLSA of each gene was evaluated as described by Mino *et al.* (2013). Briefly, the nucleotide similarity, amino acid similarity, and non-synonymous and synonymous substitution ratio (*Ka/Ks*) were determined using the software program SWAAP ver 1.0.3 (Pride, 2000). The nucleotide diversity per site, the number of polymorphic sites, G+C content, haplotype diversity, and Tajima's D based on the total number of mutations were calculated using DnaSP ver 5 (Librado & Rozas, 2009). The *Z*-test of purifying selection was

performed using the MEGA ver 5.05 software (Tamura *et al.*, 2011).

Population structure

The linkage model of STRUCTURE was applied, which accounts for the correlation between nearby nucleotide sites arising in admixed populations (Falush *et al.*, 2003). Sequence data (from the 11 protein-coding genes) were formatted using xmfa2struct (http://www.xavierdidelot.xtreemhost.com/clonalframe.htm). Four replicate MCMC runs were performed for each value of *K* ranging from 1 to 6 with 100,000 burn-in and 200,000 sampling iterations. The number of clusters (ancestral populations) was estimated according to ΔK values (Evanno *et al.*, 2005) by the program STRUCTURE Harvester (Earl & vonHoldt, 2011). The programs CLUMPP ver 1.1.2 (Jakobsson & Rosenberg, 2007) and DISTRUCT ver 1.1 (Rosenberg, 2003) were used to produce graphics. For the STRUCTURE analysis, strains that had the same ST were considered as one strain.

Supplementary Tables and Figures

Supplementary Table S1. Summary of cruise information

Research area	Cruise ID	Ship	Equipment	Affiliation	Cruise period	Sample type (No. of strains)
Okinawa Trough	NT02-06Leg2	R/V Natsushima	ROV Hyper-Dolphin	JAMSTEC	21 Apr - 4 May 2002	CS (8), MF[a] (15), HS[a] (3)
	NT03-09	R/V Natsushima	ROV Hyper-Dolphin	JAMSTEC	12 Aug - 6 Sep 2003	CS (2), MF[b] (7)
	NT05-03Leg1	R/V Natsushima	ROV Hyper-Dolphin	JAMSTEC	14 Apr - 20 Apr 2005	MF[c] (3)
	YK07-04Leg2	R/V Yokosuka	DSV Shinkai 6500	JAMSTEC	12 Mar - 15 Mar 2007	CS (2)
	NT11-19	R/V Natsushima	ROV Hyper-Dolphin	JAMSTEC	24 Sep - 26 Sep 2011	CS (6)
	KY14-01	R/V Kaiyo	ROV Hyper-Dolphin	JAMSTEC	8 Jan - 31 Jan 2014	CS (17), MF[d] (2), AN[d] (5)
Mariana Volcanic Arc and Trough	NT05-18	R/V Natsushima	ROV Hyper-Dolphin	JAMSTEC	22 Oct -7 Nov 2005	MF[e] (1), HS[e] (1)
	YK10-10	R/V Yokosuka	DSV Shinkai 6500	JAMSTEC	17 Aug - 30 Aug 2010	CS (10), MF[f] (1), AN[f] (1)
	YK10-11	R/V Yokosuka	DSV Shinkai 6500	JAMSTEC	2 Sep - 14 Sep 2010	
Central Indian Ridge	YK05-16Leg2	R/V Yokosuka	DSV Shinkai 6500	JAMSTEC	4 Feb - 22 Mar 2006	CS (2)
	YK09-13Leg2	R/V Yokosuka	DSV Shinkai 6500	JAMSTEC	2 Nov - 10 Dec 2009	CS (4)
	YK13-02	R/V Yokosuka	DSV Shinkai 6500	JAMSTEC	6 Feb - 25 Feb 2013	CS (12), MF[g] (2)
Mid-Atlantic Ridge	EXOMAR	R/V Atlante	ROV Victor	Ifremer	3 Aug - 28 Aug 2005	CS (5)

a, MF samples were taken from 2-10 m away from the vent orifice. HS samples were surface layer (0-7 cm) of sediments (Takai *et al.*, 2003; Nakagawa *et al.*, 2005).

b, MF samples were taken from 6-10 m away from the vent orifice or from immediately below the flange structure.

c, MF samples were taken from 3 m away from vent orifice or from animal (*Paralvinella* and *Lamellibrachia*) colonies.

d, MF samples were taken from the *Shinkaia* colonies, and AN samples were taken from a tube of *Lamellibrachia* and setae of *S. crosnieri.*

e, MF sample was taken from highly turbid plume. Vent orifice could not be explored. HS sample was iron-rich mat associated with diffuse venting.

f, MF sample were taken from the *Alviniconcha* colony, and AN sample was a piece of the *Alviniconcha* shell.

g, MF samples were taken from a scaly-foot colony.

Supplementary Table S2. Primer sets to amplify the 11 protein-coding genes for MLSA

Gene	Forward primer (5'→3')	Reverse primer (5'→3')
atpA (ATP synthase, A subunit)	CAAYGTTATGGCWGGAGA	ACGTCACCWGCYTGWGTTTC
dnaK (Hsp 70 chaperon protein)	AGAAAAGCGACAAAAGADGC	GCNACMACTTCATCAGGRTT
gyrB (DNA gyrase, B subunit)	GGWYTKCAYGGBGTBGGBGT	CVACRTCVGCATCGGTCAT
	CYATHGATGAAGCDATGGC*	CVACRTCSGCATCGGTCAT*
napA (nitrate reductase, large subunit)	TTCTGYGGTACDGGHTGTGG	GRTTCATACCCATYGTCCA
pheS (phenylalanyl-tRNA synthetase subunit alpha)	CCNGCAMGWGATATGCARGA	CCRAANGCATAWCCRCTTACAT
metG (methionyl-tRNA synthetase)	ACNGGHACNGAYGARCAYGG	GCHGGCCARTAVAYNGCATG
glyA (serine hydroxymethyltransferase)	GCNGTYATGGARGCBATGGG	CTTCTTGTCTCTCCDGGAAC
tkt (transketolase)	TGTHTNTGYGGTGAYGGAG	TCTTCNCCHACRCCGATRCT
	CATGCBACAGGDCTWATCTA*	
rplA (50S ribosomal protein B)	CCDAGACAYGCDGAYCARATG	GGCATHADRCCTTTDGGBCC
feoB (ferrous iron transport protein B)	GARGCRVAAAAAGAGGGCAT	GGCATYTCCATAACAAAMGG
	GGRCAGCCMAATGTMGGWAA*	AARCTYTKSCCGTGCAG*
valS (valyl-tRNA synthetase)	CATTTTRCTGATGGAAGCGG	AGGAATTTGATGWCCCCACC

*Extra primers were designed for the amplification of some strains.

68 Supplementary Table S3. Physicochemical compositions of endmember hydrothermal fluids

	Okinawa Trough							Central Indian Ridge			Mid-Atlantic Ridge						
	Iheya North (NBC)			Izena Hole		Hatoma Knoll		Kairei			TAG		Lucky Strike		Rainbow		
Sampling year	2011	2007	2002	2003–2011	1989	2007	2000	2006	2001	2000	1994	1990	2008	1994	2008	2006	1997
Max temp (°C)	309	309	304	320	320	321	325	360	365	360	321	360–366	163–324	324	191–370	365	365
pH	4.65	5	4.8	4.7	4.7	no data	5.2	3.4	3.4	3.5	2.8	3.35	3.6–3.9	3.7	3.0–3.4	3	2.8
Cl (mmol/kg)	599	557	544	590	550	no data	381	623	620	642	659	636	no data	472	no data	no data	750
K (mmol/kg)	73	72.4	72.3	72	72	no data	55	13.8	15.2	14.3	18	no data	no data	24.8	no data	no data	20.4
Ca (mmol/kg)	21.1	21.9	16.1	22	22	no data	17	27.2	31.3	30	26	no data	no data	36.7	no data	no data	66.6
Mn (μmol/kg)	694	658	619	340	110	26.2	483	828	857	840	1	no data	no data	0.446	no data	no data	2.25
H_2S (mmol/kg)	3.6	4.5	no data	5.6	12.6	no data	13.5	4.81	3.9	4	6.7	2.5–3.5	2.4–3.4	3	1.8–3.3	no data	1.2
H_2 (mmol/kg)	0.1	0.23	no data	0.06	0.05	0.3	0.2	3.71	7.89	no data	0.15	no data	0.025–0.071	0.0767	12.3–16.9	12.9	16
CH_4 (mmol/kg)	2.5	3.7	no data	4.9	7.6	no data	80	0.123	0.17	0.082	0	0.14–0.15	0.124–0.147	0.4	1.9–2.3	1.65	2.5
CO_2 (mmol/kg)	63	227	no data	156	209	322	315	9.5	4.8	no data	0.124	no data	35–133	28.4	21–25	17	16
Ref	Kawagucci et al., 2015	Kawagucci et al., 2011		Ishibashi et al, 2014	Sakai et al., 1990	Kawagucci et al., 2011	Kawagucci et al., 2015	Takai & Nakamura, 2010	Gallant & Von Damm, 2007	Gamo et al., 2001	McCollom, 2007	Charlou et al, 1996	Flores et al, 2011	Charlou et al., 2002	Flores et al., 2011	Konn et al., 2015	McCollom, 2007

69

70

71 Supplementary Table S4. Properties of the 11 protein-coding genes for MLSA

Gene	Seq length (bp)	ANI [a] (SD)	AAI [b] (SD)	G+C	Snt[c]	π[d]	Ka/Ks (SD)	Z-test[e]	Tajima's D	No. of allele	Hd [f] (SD)	jModelTest	PHI test *P*-value[g]
gyrB	879	91.2 (5.1)	98.0 (1.5)	0.36	301	0.09	0.03 (0.083)	**13.12**	-0.14	59	0.97 (0.01)	GTR+G+I	**0.000**
tkt	615	90.7 (7.9)	96.1 (5.2)	0.41	331	0.09	0.05 (0.113)	**9.54**	-0.13	52	0.97 (0.01)	GTR+G	0.581
rplA	129	93.4 (4.8)	99.6 (0.9)	0.42	52	0.07	0.01 (0.039)	**4.39**	-0.42	39	0.92 (0.02)	GTR+G+I	n.d.[h]
feoB	714	89.6 (6.6)	94.0 (4.0)	0.36	302	0.10	0.06 (0.055)	**8.77**	-0.41	61	0.90 (0.20)	GTR+G+I	**0.001**
atpA	702	91.3 (5.9)	98.2 (2.7)	0.43	285	0.08	0.04 (0.037)	**8.54**	-1.04	47	0.96 (0.01)	GTR+G+I	**0.000**
pheS	366	90.1 (6.4)	97.7 (1.7)	0.41	116	0.10	0.03 (0.034)	**7.59**	0.30	53	0.98 (0.01)	GTR+G	0.818
metG	555	91.0 (7.6)	96.4 (5.5)	0.38	279	0.09	0.05 (0.123)	**7.12**	-0.69	51	0.95 (0.01)	GTR+G+I	0.207
napA	660	91.4 (5.5)	97.7 (2.3)	0.38	244	0.09	0.04 (0.118)	**11.43**	-0.30	54	0.96 (0.01)	GTR+G+I	**0.000**
glyA	645	91.5 (5.8)	98.4 (1.8)	0.43	250	0.09	0.02 (0.108)	**11.95**	-0.80	44	0.96 (0.01)	GTR+G+I	0.476
valS	492	90.9 (5.8)	97.7 (3.2)	0.39	205	0.09	0.03 (0.060)	**10.69**	-0.90	51	0.96 (0.01)	GTR+G+I	0.492
dnaK	513	92.5 (4.4)	97.5 (2.0)	0.39	142	0.08	0.05 (0.030)	**8.16**	0.90	50	0.97 (0.01)	GTR+G+I	0.188

a, Average nucleotide sequence identity.

b, Average amino acid sequence identity.

c, Number of polymorphic nucleotide sites.

d, Nucleotide diversity per site.

e, Statistically significant results are shown in bold.

f, Haplotype diversity.

g, Number of informative sites is small for the PHI test.

SD, Standard deviation.

Supplementary Table S5. F_{ST} values between strains grouped by habitat. The numbers in parentheses show the number of strains originating from each habitat type

	CS (n=66)	MF (n=31)	AN (n=8)	HS (n=4)
CS	0.000			
MF	**0.097**	0.000		
AN	**0.074**	-0.008	0.000	
HS	0.023	-0.006	-0.012	0.000

Values are shown in bold where $p<0.05$.

Supplementary Table S6. F_{ST} values between strains grouped by sampling year. The strains isolated from the Iheya North hydrothermal field were used for this estimation. The numbers in parentheses show the number of strains obtained at each sampling year

	2002 (n=26)	2005 (n=3)	2011 (n=6)	2014 (n=24)
2002	0.000			
2005	-0.057	0.000		
2011	-0.012	-0.069	0.000	
2014	**0.079**	0.072	0.000	0.000

Values are shown in bold where $p<0.05$.

80 Supplementary Table S7. Chemical properties of end member vent fluid in different hydrothermal sites/fields

Venting site	Hydrothermal system	Ref	Tmax[j] °C	pH	Cl^- mmol/kg	K mmol/kg	Ca mmol/kg	Mn µmol/kg	H_2S mmol/kg	H_2 mmol/kg	CH_4 mmol/kg	CO_2 mmol/kg	Depth m
OT	BAB								**10.1**	**0.16**	**30.4**	**250**	
Iheya North (NBC)		a	309	5.0	557	72.4	21.9	$6.58x10^2$	4.5	0.23	3.7	227	1,070
Izena Hole (JADE)		b	320	4.7	550	72	22.0	$1.10x10^2$	12.4	0.05	7.6	209	1,450
Hatoma Knoll		c	325	5.2	381	55	17	$4.83x10^2$	13.5	0.2	80	315	1,473
SMT	BAB								**6.3**	**0.037**	**$1.3x10^{-2}$**	**26**	
Pika		d	322	3.0	469	31.6	37.8	$1.14x10^3$	7.0	~0.03	$7.20x10^{-3}$	33.7	2,830
Urashima		d	280	3.0	623	37.0	31.9	$2.22x10^3$	2.4	~0.03	$1.08x10^{-2}$	23.2	
Archaean		d	318	3.0	401	33.0	15.8	$1.28x10^3$	9.6	~0.05	~0.02	~20	
NMVA	IA								**30.0**	**0.2**	**0.2**	**800**	
NW Eifuku		e	108	ND	ND	ND	ND	ND	30.0	<0.2	<0.2	800	1,604
CIR	MOR								**3.9**	**7.89**	**0.17**	**4.8**	
Kairei		f	365	3.4	620	15.2	31.3	$8.57x10^2$	3.9	7.89	0.17	4.8	2,460
Solitaire		g	296	4.8	~500					0.46	~0.05	~8	2,745
MAR	MOR								**3.6**	**5.41**	**1.01**	**15.8**	
TAG		h	321	2.8	659	18	26	$1.00x10^3$	6.7	0.15	0.12	2.9	3,436
Lucky Strike		i	324	3.7	472	24.8	36.7	$4.46x10^2$	3.0	$7.67x10^{-2}$	0.40	28.4	1,740
Rainbow		h	365	2.8	750	20.4	66.6	$2.25x10^3$	1.2	16	2.5	16	2,320

a, Kawagucci *et al.* (2011)

b, Sakai *et al.* (1990a, b)

c, Kawagucci *et al.* (2015)

d, Mino *et al.* (2013), Toki *et al.* (2015) and this study.

e, Lupton *et al.*, (2006)

f, Gallant and Von Damm (2006)

g, Nakamura *et al.* (2012)

h, McCollom (2007)

i, Charlou *et al.* (2002)

j, Maximum temperature

ND, no data

Bold values are average for each vent field.

Supplementary Table S8. Outputs of PERMANOVA analysis by the R package.

	Df	Sum Sqs	MeanSqs	F. Model	R^2	Pr (>F)	
Region	4	29.049	7.2623	151.911	0.85427	0.001	***
Habitat	3	0.095	0.0315	0.660	0.00278	0.638	
Year	1	0.080	0.0802	1.677	0.00236	0.173	
Residuals	100	4.781	0.0478		0.14059		
Total	108	34.005			1.00000		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

92

93 **Supplementary Figure legends**

94 Supplementary Figure S1. A dendrogram of the 109 strains of *Sulfurimonas* constructed by
95 UPGMA cluster analysis using MLSA allele data. The symbol color and shape indicates
96 sampling year and habitat, respectively. NMVA, the North Mariana Volcanic Arc (green); OT,
97 the Okinawa Trough (blue); CIR, the Central Indian Ridge (purple); SMT, the South Mariana
98 Trough (orange); MAR, Mid-Atlantic Ridge (red). E, NW Eifuku; D, Daikoku; IN, Iheya
99 North Original site; AS, Iheya North Aki site; NS, Iheya North Natsu site; IZ, Izena Hole; HK,
100 Hatoma Knoll; S, Solitaire; K, Kairei; A, Archaean site; U, Urashima; P, Pika; TAG, TAG
101 field; LS, Lucky Strike; R, Rainbow.

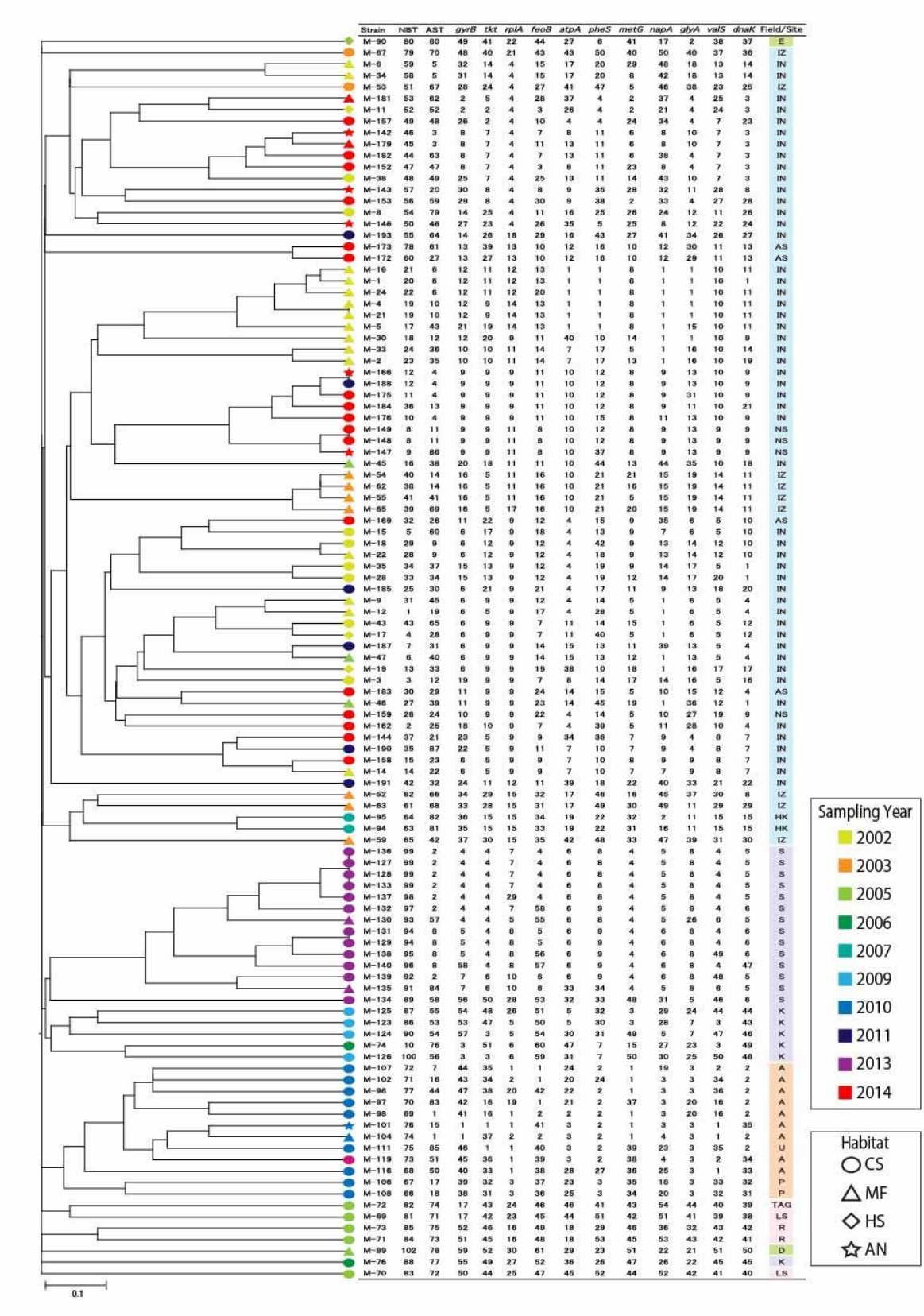
102

103 Supplementary Figure S2. A bushy network structure of *Sulfurimonas* populations inferred
104 from MLSA loci.

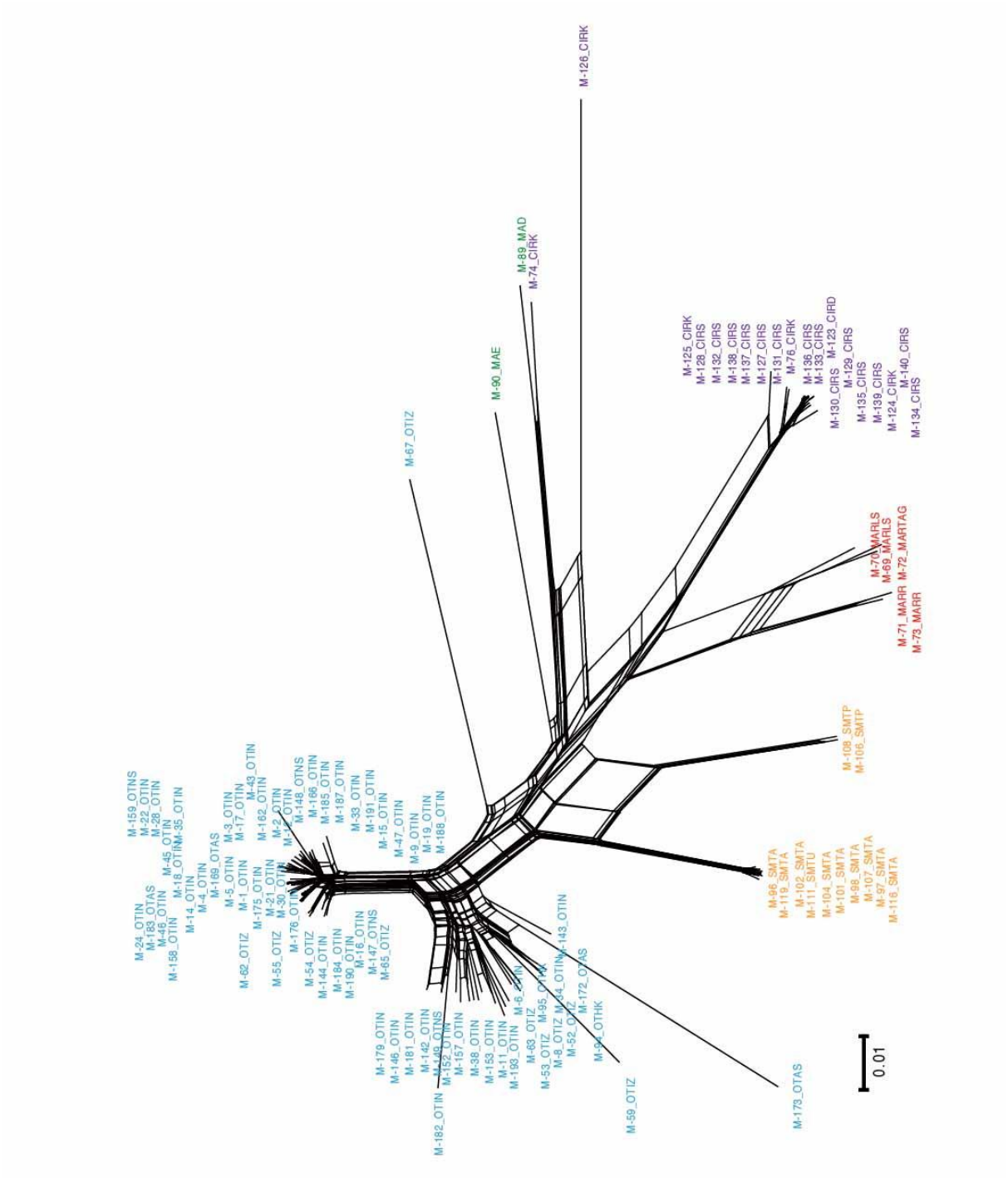
105

106 Supplementary Figure S3. Matrix of informative sites among the OT strains. Blue represents
107 incompatible sites, indicating the occurrence of recombination.

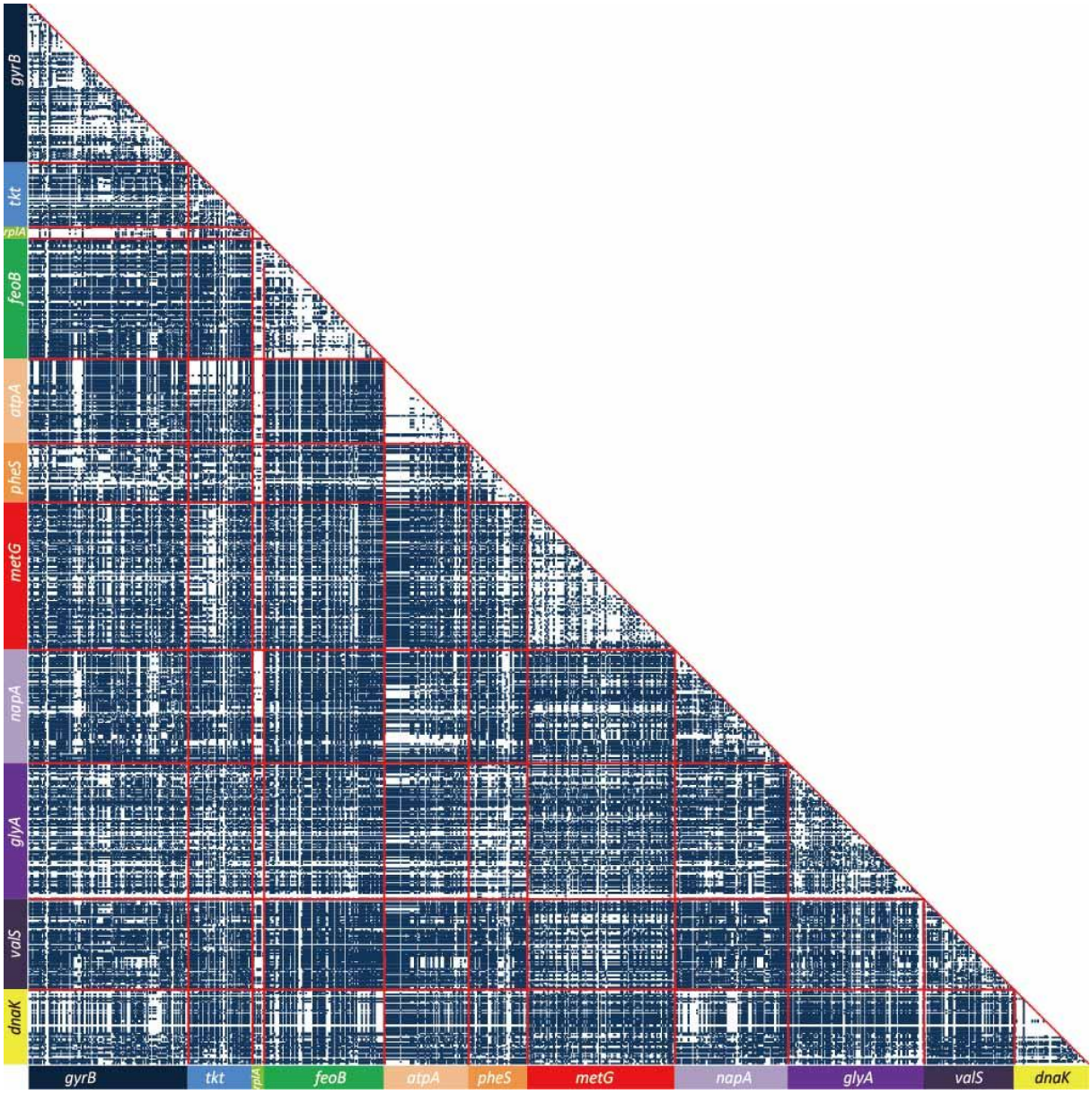
108



Supplementary Figure S1.



Supplementary Figure S2.



Supplementary Figure S3.

Supplementary references

Campbell B, Smith J, Hanson T, Klotz M, Stein L, Lee C *et al.* (2009). Adaptations to submarine hydrothermal environments exemplified by the genome of *Nautilia profundicola*. *PLoS Genet* **5:** e1000362.

Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. (2010). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26:** 266-267.

Charlou JL, Donval JP, Jean-Baptiste P, Dapoigny A, Rona PA. (1996). Gases and helium isotopes in high temperature solutions sampled before and after ODP Leg 158 drilling at TAG Hydrothermal Field (26°N, MAR). *Geophys Res Lett* **23:** 3491-3494.

Earl DA, vonHoldt BM. (2011). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* **4:** 359-361.

Evanno G, Regnaut S, Goudet J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14:** 2611-2620.

Falush D, Stephens M, Pritchard J. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164:** 1567-1587.

Fischer W, Windhager L, Rohrer S, Zeiller M, Karnholz A, Hoffmann R *et al.* (2010). Strain-specific genes of *Helicobacter pylori*: genome evolution driven by a novel type IV secretion system and genomic island transfer. *Nucleic Acids Res* **38:** 6089-6101.

Flores GE, Campbell JH, Kirshtein JD, Meneghin J, Podar M, Steinberg JI et al. (2011). Microbial community structure of hydrothermal deposits from geochemically different vent fields along the Mid-Atlantic Ridge. *Environ Microbiol* **13:** 2158-2171.

Gallant R, Von Damm K. (2006). Geochemical controls on hydrothermal fluids from the Kairei and Edmond Vent Fields, 23°-25°S, Central Indian Ridge. *Geochem Geophys Geosyst* **7:** Q06018. doi:06010.01029/02005GC001067.

Gamo T, Chiba H, Yamanaka T, Okudaira T, Hashimoto J, Tsuchida S et al. (2001) Chemical characteristics of newly discovered black smoker fluids and associated hydrothermal plumes at the Rodriguez Triple Junction, Central Indian Ridge. *Earth Planet Sci Lett* **193:** 371-379.

Ishibashi J, Noguchi T, Toki T, Miyabe S, Yamagami S, Onishi Y et al. (2014). Diversity of fluid geochemistry affected by processes during fluid upwelling in active hydrothermal fields in the Izena Hole, the middle Okinawa Trough back-arc basin. *Geochemical J* **48:** 357-369.

Jakobsson M, Rosenberg NA. (2007). CLUMPP: a cluster matching and permutation program

for dealing with label switching and multimodality in analysis of population structure. *Bioinformatic*s 23: 1801-1806.

Kawagucci S, Chiba H, Shibash J, Yamanaka T, Toki T, Muramatsu Y et al. (2011). Hydrothermal fluid geochemistry at the Iheya North field in the mid-Okinawa Trough: Implication for origin of methane in subseafloor fluid circulation systems. *Geochem J* **45:** 109-124.

Kawagucci S. (2015). Fluid geochemistry of high-temperature hydrothermal fields in the Okinawa Trough. In: Ishibashi J-i, Okino K, Sunamura M (eds). Subseafloor biosphere linked to global hydrothermal systems: TAIGA Concept, Cold Spring Harbor Laboratory: Cold Spring Harbor, NY.

Kersulyte D, Mukhopadhyay A, Velapatiño B, Su W, Pan Z, Garcia C *et al.* (2000). Differences in genotypes of *Helicobacter pylori* from different human populations. *J Bacteriol* **182:** 3210-3218.

Konn C, Charlou JL, Holm NG, Mousis O. (2015). The production of methane, hydrogen, and organic compounds in ultramafic-hosted hydrothermal vents of the Mid-Atlantic Ridge. *Astrobiology* 15: 381-399.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947-2948.

Librado P, Rozas J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:** 1451-1452.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar B *et al.* (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* 32: 1363-1371.

McCollom TM. (2007). Geochemical constraints on sources of metabolic energy for chemolithoautotrophy in ultramagic-hosted deep-sea hydrothermal systems. *Astrobiology* **7:** 933-950.

Maiden MC. (2006). Multilocus sequence typing of bacteria. *Annu Rev Microbiol* **60:** 561-588.

Margos G, Gatewood AG, Aanensen DM, Hanincová K, Terekhova D, Vollmer SA et al. (2008). MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*. *Proc Natl Acad Sci U S A* **105:** 8730-8735.

Mino S, Makita H, Toki T, Miyazaki J, Kato S, Watanabe H *et al.* (2013). Biogeography of *Persephonella* in deep-sea hydrothermal vents of the Western Pacific. *Front Microbiol* **4:** 107. doi: 10.3389/fmicb.2013.00107.

Nakagawa S, Takaki Y, Shimamura S, Reysenbach A, Takai K, Horikoshi K. (2007). Deep-sea

vent ε-proteobacterial genomes provide insights into emergence of pathogens. *Proc Natl Acad Sci U S A* **104:** 12146-12150.

Parkhill J, Wren B, Mungall K, Ketley J, Churcher C, Basham D *et al.* (2000). The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* **403:** 665-668.

Pride DT. (2000). SWAAP Version 1.0.0 - Sliding Windows Alignment Analysis Program: A Tool for Analyzing Patterns of Substitutions and Similarity in Multiple Alignments. Distributed by the author.

Rosenberg NA. (2003). DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes* **4:** 137-138.

Sakai H, Gamo T, Kim E, Shitashima K, Yanagisawa F, Tsutsumi M et al. (1990a). Unique chemistry of the hydrothermal solution in the mid-Okinawa Trough backarc basin. *Geophys Res Lett* **17:** 2133-2136.

Sakai H, Gamo T, Kim ES, Tsutsumi M, Tanaka T, Ishibashi J et al. (1990b). Venting of carbon-dioxide rich fluid and hydrate formation in mid-Okinawa Trough backarc basin. *Science* **248:** 1093-1096.

Sikorski J, Munk C, Lapidus A, Ngatchou Djao O, Lucas S, Glavina Del Rio T *et al.* (2010). Complete genome sequence of *Sulfurimonas autotrophica* type strain (OK10). *Stand Genomic Sci* **3:** 194-202.

Takai K, Inagaki F, Nakagawa S, Hirayama H, Nunoura T, Sako Y *et al.* (2003). Isolation and phylogenetic diversity of mem- bers of previously uncultivated epsilon-Proteobacteria in deep-sea hydrothermal fields. *FEMS Microbiol Lett* **218:** 167-174.

Takai K and Nakamura K. (2010). Compositional, physiological and metabolic variability in microbial communities associated with geochemically diverse, deep-sea hydrothermal vent fluids. In *Geomicrobiology: Molecular & Environmental Perspective*, L. Barton, M. Mendl and A. Loy (eds.), Springer, NY., pp. 251-283.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28:** 2731-2739.

Tomb J, White O, Kerlavage A, Clayton R, Sutton G, Fleischmann R *et al.* (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388:** 539-547.