ORIGINAL ARTICLE

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

Sayaka Mino[1], Satoshi Nakagawa[2,3], Hiroko Makita[3], Tomohiro Toki[4], Junichi Miyazaki[3], Stefan M Sievert[5], Martin F Polz[6], Fumio Inagaki[7,8], Anne Godfroy[9], Shingo Kato[10], Hiromi Watanabe[11], Takuro Nunoura[12], Koichi Nakamura[13], Hiroyuki Imachi[3], Tomo-o Watsuji[3], Shigeaki Kojima[14], Ken Takai[3] and Tomoo Sawabe[1]

[1]*Laboratory of Microbiology, Faculty of Fisheries Sciences, Hokkaido University, Hakodate, Japan;* [2]*Laboratory of Marine Environmental Microbiology, Division of Applied Biosciences, Graduate School of Agriculture, Kyoto University, Kyoto, Japan;* [3]*Department of Subsurface Geobiology Analysis and Research (D-SUGAR), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokosuka, Japan;* [4]*Department of Chemistry, Biology, and Marine Science, Faculty of Science, University of the Ryukyus, Nishihara, Japan;* [5]*Department of Biology, Woods Hole Oceanographic Institution, Woods Hole, MA, USA;* [6]*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA;* [7]*Kochi Institute for Core Sample Research, JAMSTEC, Nankoku, Japan;* [8]*Research and Development Center for Ocean Drilling Science (ODS), JAMSTEC, Yokohama, Japan;* [9]*Ifremer, UMR6197, Laboratoire de Microbiologie des Environnements Extrêmes, Plouzané, France;* [10]*Japan Collection of Microorganisms, RIKEN BioResource Center, Tsukuba, Japan;* [11]*Department of Marine Biodiversity Research, JAMSTEC, Yokosuka, Japan;* [12]*Research and Development Center for Marine Biosciences, JAMSTEC, Yokosuka, Japan;* [13]*National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan and* [14]*Atmosphere and Ocean Research Institute, The University of Tokyo, Kashiwa, Japan*

Rich animal and microbial communities have been found at deep-sea hydrothermal vents. Although the biogeography of vent macrofauna is well understood, the corresponding knowledge about vent microbial biogeography is lacking. Here, we apply the multilocus sequence analysis (MLSA) to assess the genetic variation of 109 *Sulfurimonas* strains with $\geqslant$ 98% 16S rRNA gene sequence similarity, which were isolated from four different geographical regions (Okinawa Trough (OT), Mariana Volcanic Arc and Trough (MVAT), Central Indian Ridge (CIR) and Mid-Atlantic Ridge (MAR)). Sequence typing based on 11 protein-coding genes revealed high genetic variation, including some allele types that are widespread within regions, resulting in 102 nucleotide sequence types (STs). This genetic variation was predominantly due to mutation rather than recombination. Phylogenetic analysis of the 11 concatenated genes showed a clear geographical isolation corresponding to the hydrothermal regions they originated from, suggesting limited dispersal. Genetic differentiation among *Sulfurimonas* populations was primarily influenced by geographical distance rather than gas composition of vent fluid or habitat, although *in situ* environmental conditions of each microhabitat could not be examined. Nevertheless, *Sulfurimonas* may possess a higher dispersal capability compared with deep-sea hydrothermal vent thermophiles. This is the first report on MLSA of deep-sea hydrothermal vent *Epsilonproteobacteria*, which is indicative of allopatric speciation.
The ISME Journal (2017) **11,** 909–919; doi:10.1038/ismej.2016.178; published online 3 January 2017

Introduction

Biogeography is a fundamental principle in ecology, but whether it applies to microorganisms is controversial.

Although unlimited dispersal due to large cell numbers and small cell size has long been generalized (Baas-Becking, 1934), microbial communities with an unexpectedly high degree of spatial complexity have been reported from some ecosystems (Martiny *et al.*, 2006; Hanson *et al.*, 2012). Additionally, by using higher resolution techniques, microbial geographical separation has been assessed for some microbes at the subspecies level. For instance, allopatric speciation was detected in a single species of a hyperthermophilic

archaeon inhabiting terrestrial hot springs (Whitaker *et al.*, 2003). Extremophiles from terrestrial environments requiring specific growth conditions (e.g., high temperature, low pH, high salinity) have served as good models for detailed investigations on the microbial dispersal capability (Papke *et al.*, 2003; Whitaker *et al.*, 2003). However, only a few biogeographic studies have been reported on marine extremophiles.

Deep-sea hydrothermal vents are one of the most extreme environments where macrofauna thrive, enabled by symbiotic relationships with microorganisms. Although deep-sea hydrothermal vents are connected through the ocean via the currents, deep-sea vent macrofauna exhibit a clear geographical isolation (Rogers *et al.*, 2012). In contrast, similar microbial community members, including *Epsilonproteobacteria*, *Aquificales*, *Thermococcales*, *Methanococcales*, *Archaeoglobales* and 'deep-sea hydrothermal vent euryarchaeota 2', have been identified in geographically separated hydrothermal regions (Takai *et al.*, 2006; Callac *et al.*, 2013), suggesting that their populations are globally mixed. However, recent studies using the multilocus sequence analysis (MLSA) (Maiden, 2006) have successfully captured the endemism of such cosmopolitan deep-sea vent thermophiles (Flores *et al.*, 2012; Mino *et al.*, 2013; Price *et al.*, 2015). For example, the dispersal range of bacteria belonging to the widespread hydrogen/sulphur-oxidizing thermophilic genus *Persephonella* (order *Aquificales*) was found to be < 2500 km (Mino *et al.*, 2013), possibly reflecting their small population size. However, at present little is known about the dispersal capability of dominant deep-sea vent mesophiles, and if their large population sizes would lead to increased dispersal ranges as has been seen in other studies (Finlay and Fenchel, 2004).

Members of the genus *Sulfurimonas* (class *Epsilonproteobacteria*) represent ubiquitous and dominant mesophilic bacteria in deep-sea hydrothermal environments around the world. They are strictly chemolithoautotrophic, fastidious, metabolically versatile sulphur and/or hydrogen oxidizers inhabiting various hydrothermal habitats, including plumes, sediments, chimneys and diffuse-flow vent fluids (Nakagawa *et al.*, 2005; Campbell *et al.*, 2006). In addition to deep-sea vents, they have also been shown to dominate in shallow hydrothermal vents (Zhang *et al.*, 2012), cold seeps (Niemann *et al.*, 2013) and the pelagic redoxcline (Grote *et al.*, 2007, 2012), suggesting high dispersal ability of this group of bacteria. However, because of the difficulty in obtaining strains from geographically separated environments, the dispersal capability of *Sulfurimonas* has not yet been evaluated. We previously developed an MLSA scheme and applied it to 49 *Sulfurimonas* strains mainly isolated from the Western Pacific (Nakagawa, 2011). In the present study, we performed MLSA with 60 additional *Sulfurimonas* strains isolated from the Western Pacific and the CIR. This study revealed for the first time endemicity of *Sulfurimonas* populations at deep-sea hydrothermal vents, representing an important step towards a better understanding of population genetics of deep-sea hydrothermal vent microorganisms.

## Materials and methods

Sample collections
Hydrothermal samples were obtained with *R/V Natsushima*, *R/V Kaiyo* and *ROV Hyper-Dolphin*, *R/V Yokosuka* and *DSV Shinkai 6500*, or *R/V Atlante* and *ROV Victor* during cruises to the Okinawa Trough (OT) (NT02-06Leg2, NT03-09, NT05-03Leg1, YK07-04Leg2, NT11-19 and KY14-01), to the Mariana Volcanic Arc and Trough (MVAT) (NT05-18, YK10-10 and YK10-11), to the Central Indian Ridge (CIR) (YK05-16Leg2, YK09-13Leg2 and YK13-02) and to the Mid-Atlantic Ridge (MAR) (EXOMAR) (Supplementary Table S1). After recovery on-board, the samples were processed as described in Mino *et al.* (2013). Samples were taken using the ROV/DSV manipulator and fluid/sediment/ animal samplers. Briefly, each of chimney samples was sectioned immediately onboard ship (e.g., into the exterior surface and the inside parts), and slurried with the sterilized seawater in the presence or absence of 0.05% (w v^{-1}) neutralized sodium sulfide in 100 ml glass bottles (Schott Glaswerke, Mainz, Germany). Bottles were then tightly sealed with butyl rubber caps under a gas phase of 100% N_2 (0.2 MPa). Similarly, fluid, sediment and biological samples were prepared anaerobically in 30 ml glass bottles. Samples were stored at 4 °C until use.

Site descriptions
The hydrothermal samples were collected from a total of 15 hydrothermal fields/sites in four geographical regions: the OT, the MVAT, the CIR and the MAR (Figure 1). OT is a backarc basin, and three active hydrothermal fields (Iheya North, Izena Hole and Hatoma Knoll) were investigated in the present study. The Iheya North hydrothermal field includes three hydrothermal vent sites (Original, Natsu and Aki). MVAT is an arc-backarc system, stretching in parallel ~ 1300 km from north to south. In the North Mariana Volcanic Arc area (NMVA), two seamounts on hydrothermal systems (NW Eifuku and Daikoku) were studied, and three active vent sites (Archaean, Pika and Urashima) were investigated in the South Mariana Trough (SMT). In CIR, the samples were obtained from a basalt-hosted system (Solitaire field) and an ultramafic-rock-associated system (Kairei field). In MAR, samples from two basalt-hosted systems (Lucky Strike and TAG fields) and an ultramafic-rock-hosted system (Rainbow field) were used.

Habitat description
Microbial habitats in deep-sea hydrothermal environments were classified into four types as shown in Supplementary Table S1. Hydrothermal sediments
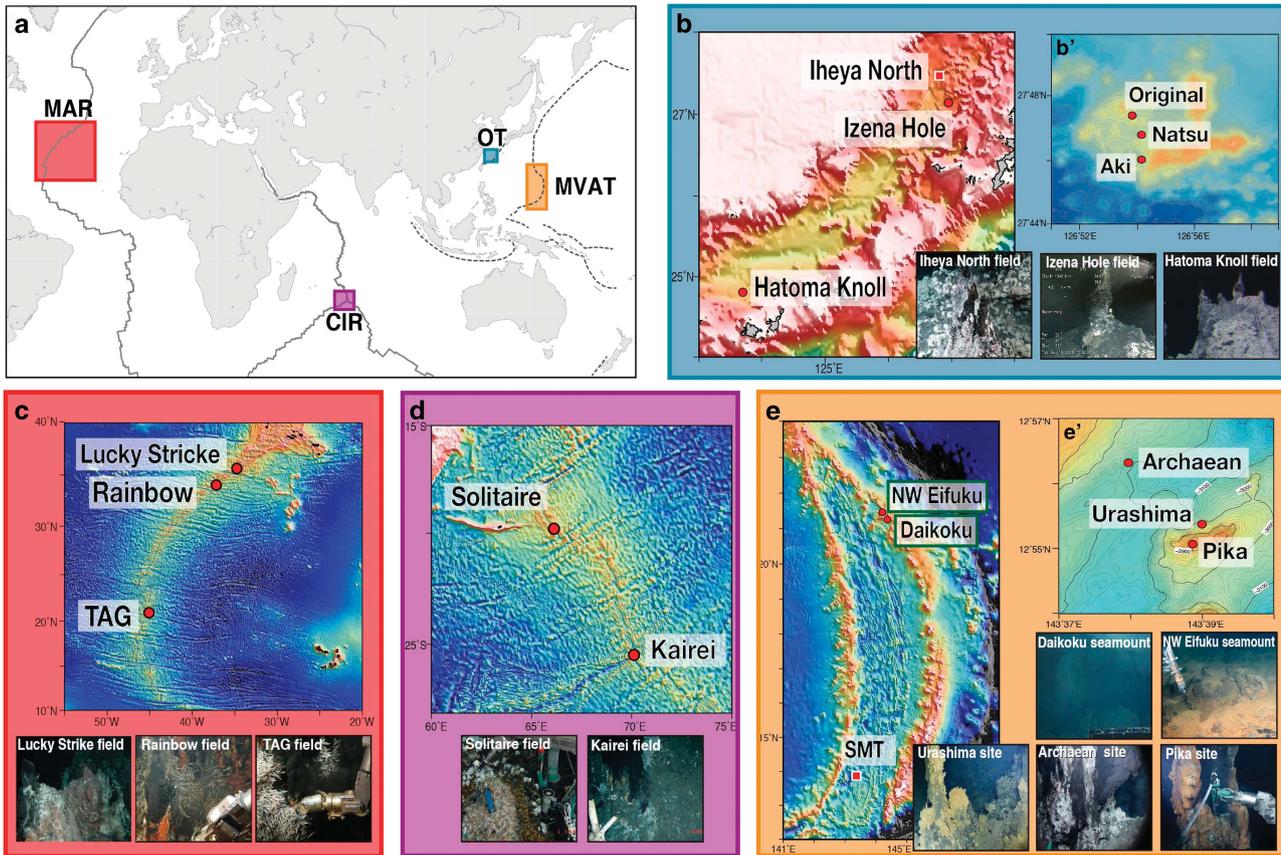
Figure 1 Maps and pictures of sampling locations. (**a**) World map. Solid and dotted lines indicate the mid-ocean ridge and the subduction zone, respectively. Zoom-in maps of the OT (**b** and **b'**), the MAR (**c**), the CIR (**d**) and the MVAT (**e**). Sites in OT and the SMT are shown in (**b'**) and (**e'**), respectively. OT, Iheya North field including Original (27°47′N, 126°53′E, ~1000 m in depth), Natsu (27°47′N, 126°54′E, ~1100 m in depth), and Aki (27°47′N, 126°54′E, ~1100 m in depth) sites, Izena Hole (27°16′N, 127°05′E, ~1500 m in depth) and Hatoma Knoll (24°51′N, 123°50′E, ~1500 m in depth) fields; MAR, Lucky Strike (37°18′N, 32°16′W, ~1700 m in depth), Rainbow (36°14′N, 33°54′W, ~2300 m in depth), TAG (26°08′N, 44°50′W, ~3400 m in depth) fields; CIR, Kairei (25°19′S, 70°02′E, ~2500 m in depth) and Solitaire (19°32′S, 65°51′E, ~2600 m in depth) fields; MVAT, NW Eifuku (21°29′N, 144°02′E, ~1600 m in depth), Daikoku (21°19′N, 144°11′E, ~400 m in depth) fields and SMT field including Archaean (12°56′N, 143°38′E, ~3000 m in depth), Pika (12°55′N, 143°39′E, ~2800 m in depth) and Urashima (12°55′N, 143°39′E, ~2900 m in depth) sites. Hydrothermal regions are indicated by colors as follows: blue, OT; red, MAR; purple, CIR; orange, MVAT.

(HSs) were influenced by diffuse venting. Mixing zone fluids (MFs) were collected from near vent orifice or in animal colonies or hydrothermal plumes. Chimney structures (CSs) include both exterior surface and the inside parts of active chimneys. Animal body parts (ANs) include setae of galatheid crabs, a polychaete nest and a shell of gastropod. The number of strains isolated from each habitat in each region was as follows: 35, 27, 3 and 5 strains from CS, MF HS and AN, respectively, in OT; 10, 2, 1 and 1 strains from CS, MF, HS and AN, respectively, in MVAT; 18 and 2 strains from CS and MF, respectively, in CIR; 5 strains from CS in MAR.

Strains and sequences for MLSA
Cultivation and isolation of *Sulfurimonas* strains were performed under 25 or 33 °C. Strains sharing ⩾98% 16S rRNA gene sequence similarity determined by ARB were selected for MLSA. The number of isolates is different for each area since different number of

hydrothermal samples were collected per region (Supplementary Table S1). Eleven protein-coding genes (*atpA*, *dnaK*, *gyrB*, *napA*, *pheS*, *metG*, *glyA*, *tkt*, *rplA*, *feoB* and *valS*) for MLSA were amplified and sequenced by primers shown in Supplementary Table S2. Details of strain isolation and sequencing were provided in Supplementary Methods.

Sequence typing
A number was assigned to each distinct allele within a locus. The combination of alleles at each MLSA gene defined its nucleotide ST. The relatedness between STs was shown as a dendrogram constructed by START2 (Jolley *et al.*, 2001) with the unweighted pair group cluster method with arithmetic averages.

Phylogenetic analysis based on protein-coding genes
Each MLSA locus, as well as the concatenated sequence of all 11 loci, was used for phylogenetic

analysis based on maximum likelihood (ML). ML trees were constructed using raxmlGUI ver 1.31 (Silvestro and Michalak, 2012). Because not all gene loci evolve at the same rate, GTR+I+Γ or GTR+Γ models were applied according to results from the Akaike information criterion analysis in jModelTest ver 2.1.5 (Darriba et al., 2012). For ML trees, 100 bootstrap replicates were performed. The tree data were visualized using FigTree ver 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Population differences between regions, habitats and sampling years
Genetic differences between populations were evaluated with pairwise F_{ST} values estimated by analysis of molecular variance using Arlequin ver 3.5 software (Excoffier and Lischer, 2010). The 109 strains were grouped according to differences in (1) geographical region (OT, SMT, NMVA, CIR and MAR) and (2) habitat (CS, MF, AN and HS). Additionally, Iheya North isolates were grouped by sampling years (2002, 2011 and 2014), and F_{ST} values were determined. In addition, the permutational multivariate analysis of variance test with 999 permutations using the genetic distance matrix was performed by the R package (Oksanen et al., 2013) to assess the impact of habitat, region and year differences.

Geochemical analysis
To consider the physicochemical conditions of the *Sulfurimonas* populations' habitats, the physical and chemical properties of the endmember hydrothermal fluids at each of the hydrothermal fields/sites were analysed. The chemical compositions of the SMT vent fluids were analysed as described previously (Toki et al., 2015). Concentrations of H_2 (Archaean, Pika and Urashima sites), CO_2 and CH_4 (Archaean site) in the endmember fluids were calculated by extrapolation of $Mg = 0$ using a linear regression of the water chemistry data. The maximum endmember H_2 was used for further analysis. The chemical compositions of the endmember vent fluids from other hydrothermal fields were taken from previous studies (Sakai et al., 1990a, b; Charlou et al., 2002; Gallant and Von Damm, 2006; Lupton et al., 2006; McCollom, 2007; Kawagucci et al., 2011; Nakamura et al., 2012; Mino et al., 2013; Kawagucci, 2015). Drastic change in the physicochemical composition of vent fluids have not been observed in the hydrothermal fields studied (Supplementary Table S3).

Effects of environmental differences and geographical distances on population differentiation
To determine the major factors influencing genetic diversification, we assessed the effects of the geographical distance and the physicochemical environment of the habitat. The geographical distance between hydrothermal fields/sites in the same region was

calculated as the straight-line sea route distance using latitude and longitude data. For the estimation of geographical distances across the regions, the averaged shortest sea route distances were calculated using online tools, SEA-DISTANCES.ORG (www.sea-distances.org), Portworld Distance Calculator (www.portworld.com/map) and SeaRates.com (www.searates.com). The environmental divergence between hydrothermal fields was calculated as the Euclidean distance using the gas compositions (H_2, H_2S and CH_4) of the endmember vent fluids with XLSTAT 2016.2 (Addinsoft, New York, NY, USA). The genetic differentiation was calculated using MEGA as the value of the pairwise distance. Correlations of genetic distance with the chemistry and the geographical distance were examined by the Mantel test using XLSTAT. Strains originating from the Solitaire field were removed in this analysis because of the lack of endmember gas composition data.

We compared the impact of geographical distance on the genetic variation of either deep-sea hydrothermal vent thermophiles (*Aciduliprofundum* and *Persephonella*) (Flores et al., 2012; Mino et al., 2013) or deep-sea hydrothermal vent mesophile (*Sulfurimonas*). Protein-coding gene sequences of thermophiles were obtained from the NCBI database, and were concatenated for each strain. For normalization, genetic distances (π) estimated from protein-coding gene sequences were divided by the dissimilarity of 16S rRNA gene sequences within the taxon.

Analysis of mutation and recombination
The relative frequency of recombination in comparison with point mutation (ρ/θ) (Milkman and Bridges, 1990) was calculated by ClonalFrame (Didelot and Falush, 2007). Because the ρ/θ value ignores the sequence length and nucleotide diversity of imported fragments, it contains no information about the actual impact of recombination on evolutionary changes (Vos and Didelot, 2009). We therefore also used ClonalFrame to calculate the ratio of nucleotide changes resulting from recombination compared with those resulting from mutation (r/m). The program was run with 100 000 Markov chain Monte Carlo iterations with a thinning interval of 100 after an initial burn-in phase of 50 000 iterations. SplitsTree ver 4.11.3 (Huson, 2006) was used to visualize the mutation and recombination events and to estimate the pairwise homoplasy index (Bruen et al., 2006) for each MLSA locus and concatenated sequence. For the OT strains, graphic evidence of recombination events that occur between different genes (intergenic recombination) and between alleles of the same gene (intragenic recombination) was shown by a matrix of informative sites provided by the SITES program (Hey and Wakeley, 1997).

Inferring the ancestral population structure
The Bayesian analysis tool STRUCTURE ver 2.3 (Pritchard et al., 2000) was used to determine the

ancestral lineage among our strains. STRUCTURE assumes that the observed data are derived from K ancestral lineages. Details of analysis were described in Supplementary Methods.

Results

Characterization of protein-coding gene sequences in closely related *Sulfurimonas* strains

The properties of the 11 protein-coding genes selected for MLSA are summarized in Supplementary Table S4. Sequenced lengths of the protein-coding genes ranged from 129 to 879 bp, resulting in a 6720-bp-long concatenated sequence. Similarities of nucleotide and amino-acid sequences at MLSA loci varied from 89.6% to 93.4% and from 94.0% to 99.6%, respectively. The nucleotide diversity (π) ranged from 0.066 to 0.104, showing different evolutionary rates among the 11 protein-coding genes. The ratios of Ka/Ks for all 11 MLSA loci were below 1, indicating that all loci were under negative selection pressure. This result was supported by significant Z-test values. Tajima's D values were not significantly different from 0 for all MLSA loci, suggesting no significant departure from the neutral model with purifying selection. These results confirmed that the genes selected in this study are suitable for MLSA.

Sequence typing and relationship between STs

The number of allele types per MLSA locus was between 39 and 61 (Supplementary Table S4), and concatenated sequences were classified into 102 unique STs (Supplementary Figure S1). Results of the sequence typing indicate weak geographical isolation between vent fields or sites within a geographical region. Five STs (ST8, 12, 19, 94 and 99) were shared by two to four strains that originated from the same hydrothermal field, for example, strain M-4 and strain M-21 from the Iheya North field shared ST19. Moreover, one or more alleles in the MLSA loci were shared among strains isolated from different fields in the same hydrothermal region, but not across the different regions, except between NMVA and SMT in MVAT. For example, allele no. 11 of *glyA* was found in strains M-94 and M-95 from the Hatoma Knoll field, strain M-63 from the Izena Hole field and strains M-143 and M-184 from the Iheya North field. These OT strains originated from various hydrothermal habitats, that is, chimneys and vent fluids. Two allele types (allele no. 5 of *rplA* and *napA*) were shared by strains isolated from the Kairei and Solitaire fields in CIR, and one allele type (allele no. 17 of *gyrB*) was shared by strains obtained from the TAG and Lucky Strike fields in MAR. In addition to allele types shared by strains from different fields in the same geographical region, ST12 and allele types (e.g., allele no. 8 of *gyrB*) were shared by strains from different habitats (Supplementary Figure S1).

Table 1 F_{ST} values between populations grouped by geographical origin[a]

	OT (n = 70)	SMT (n = 12)	NMVA (n = 2)	CIR (n = 20)	MAR (n = 5)
OT	0.000				
SMT	**0.634**	0.000			
NMVA	**0.658**	0.687	0.000		
CIR	**0.706**	**0.753**	**0.526**	0.000	
MAR	**0.737**	**0.746**	**0.652**	**0.607**	0.000

Abbreviations: CIR, Central Indian Ridge; MAR, Mid-Atlantic Ridge; NMVA, North Mariana Volcanic Arc; OT, Okinawa Trough; SMT, South Mariana Trough.
The numbers in parentheses show the number of strains isolated from each region.
[a]Values are shown in bold where $P < 0.05$.

Differentiation associated with differences in province, habitat and sampling year

To assess factors affecting genetic differentiation among strains, F_{ST} values were estimated according to geographical region, habitat (isolation source) and sampling year. The F_{ST} values imply the existence of genetically isolated populations in different regions (Table 1). Although the F_{ST} value separating strains originating from SMT and NMVA was not statistically significant, probably because of the small number of strains from NMVA, the F_{ST} values separating strains from the other geographical regions were highly significant (ranging from 0.526 to 0.753). The analysis of molecular variance analysis also supported the existence of geographically distinct populations (70.1% of the variance between populations, $P < 0.001$). However, the F_{ST} values between strains grouped by habitat and sampling year were mostly low and not significant (Supplementary Tables S5 and S6). Permutational multivariate analysis of variance tests showed that the geographical region significantly explained 85% ($P < 0.001$) of the genetic variation in *Sulfurimonas* population, whereas habitat and year explained 0.3% ($P > 0.1$) and 0.2% ($P > 0.1$) of the genetic variation, respectively (Supplementary Table S7) Although the number of strains from HS and AN were small, the effect of habitat remained insignificant even when HS and AN strains were eliminated ($P > 0.1$; 0.26% of the genetic variation).

Geographical distribution influenced by distance and environmental factors

The ML tree using the 11 protein-coding genes of the 109 strains showed that most of the sequences clustered into four groups according to the geographical region (OT, SMT, CIR and MAR) (Figure 2). Within the four groups, there was no clear cluster corresponding to specific fields/sites/habitats. A significant correlation coefficient was obtained by applying a Mantel test including all pairwise comparisons of genetic and geographical distance between strains ($r = 0.786$, $P < 0.0001$) (Figure 3).
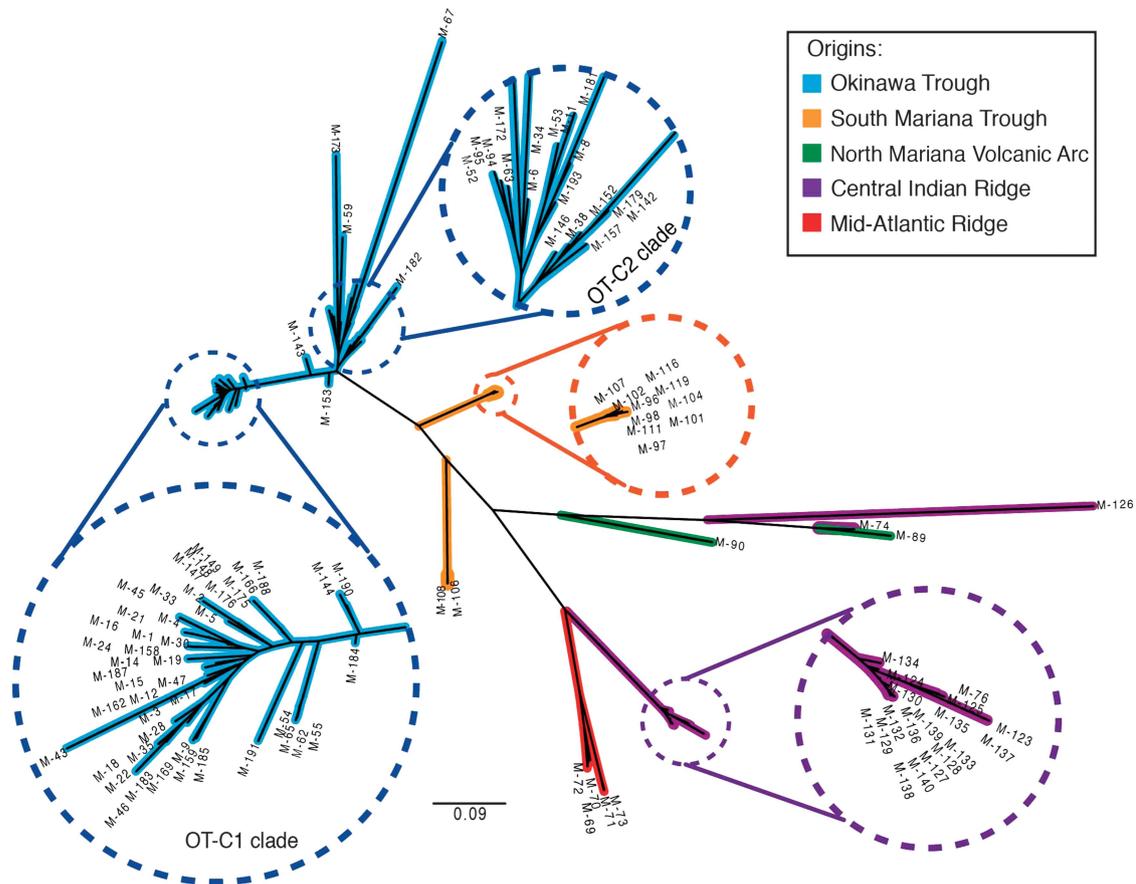
Figure 2 Maximum-likelihood tree based on 11 protein-coding gene sequences of 109 *Sulfurimonas* strains. Blue, OT; orange, SMT; green, NMVA; purple, CIR; red, MAR.

A relatively strong correlation was detected at a scale $<10\,000$ km ($r=0.840$, $P<0.0001$), possibly due to small discrepancies between shortest sea routes and actual dispersal pathways. Strain pairs at large geographic distances, that is, pairs of MAR strains and others (Figures 3d–f), were less distant genetically from those at intermediate geographical distances ($P<0.001$, Kruskal–Wallis test followed by Dunn's *post hoc* test). This could be due to the small number of MAR strains analyzed. In contrast, we found no significant correlation between the population genetic distance and the physicochemical variance of the vent fluids, such as the maximum temperature, pH and chemical composition of endmember fluids, including the concentration of H_2 (Mantel $r=-0.04$, $P=0.450$) (Supplementary Table S8).

Two major clades, that is, OT-C1 and OT-C2, were identified within OT, and contained 45 and 22 strains, respectively. These two clades, however, were not significantly correlated with the features of their environment such as isolation temperatures, origins (fields/sites), habitats and sampling years (Fisher test; $P>0.1$ in all cases).

By comparing *Sulfurimonas*, *Persephonella* and *Aciduliprofundum*, we found the difference in correlations between the geographical distance and the genetic
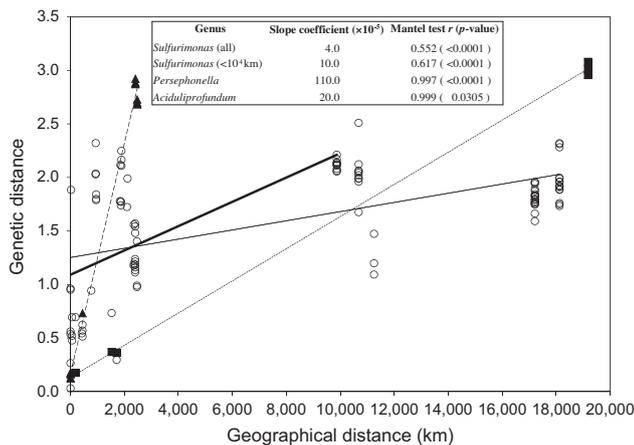
distance (Figure 4). The higher slope coefficient of *Persephonella* could be explained by their higher evolutionary rate. However, this is not likely considering the lack of many DNA repair genes in genomes of *Epsilonproteobacteria* (Nakagawa *et al.*, 2007). Thus, thermophilic *Persephonella* likely have less dispersal capability than mesophilic *Sulfurimonas*. *Aciduliprofundum* had a dispersal limitation as well, but its slope coefficient value was relatively low, possibly due to the small number of studied deep-sea hydrothermal sites/fields.

Evolution of population structure
The evolution of the population structure of strains related to *Sulfurimonas* strains was further analysed by STRUCTURE. The number of ancestral populations (K) needed to explain the current population structure was estimated to be 3 based on the distribution of ΔK. The MVAT and CIR strains formed genetically homogeneous groups with an average of $\geqslant 96.6\%$ and $\geqslant 87.7\%$ of genetic features descended from ancestral lineages 2 (green) and 3 (yellow), respectively (Figure 5). The MAR strains were admixtures composed of two ancestral lineages, indicating that they share common genetic
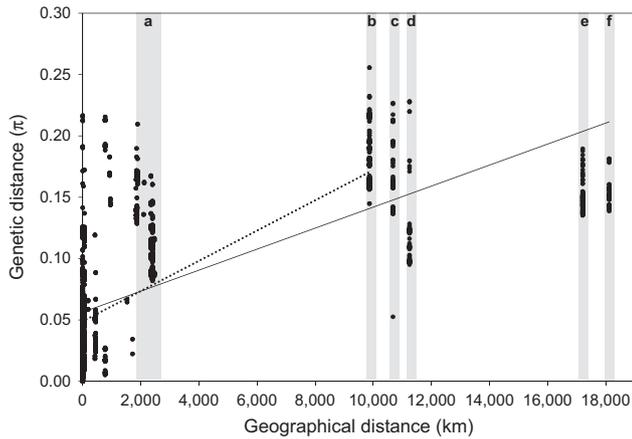
Figure 3 Relationship between genetic distance and geographical distance. Dots represent pairwise comparisons evaluated by MEGA. The black line is the regression line (solid line, all comparison; dotted line, comparison below 10 000 km). Gray areas indicate the comparison across the regions, (**a**) OT-MVAT, (**b**) OT-CIR, (**c**) MVAT-CIR, (**d**) CIR-MAR, (**e**) MAR-OT and (**f**) MAR-MVAT. Dots on the left side of area (**a**) indicate the comparison within the regions.

Genus	Slope coefficient ($\times 10^{-5}$)	Mantel test r (p-value)
Sulfurimonas (all)	4.0	0.552 (<0.0001)
Sulfurimonas (<10^4km)	10.0	0.617 (<0.0001)
Persephonella	110.0	0.997 (<0.0001)
Aciduliprofundum	20.0	0.999 (0.0305)

Figure 4 Comparison of the impact of geographical distance on genetic diversity of microorganisms living in deep-sea vent environments. Genetic distance is a measure of the genetic distance between groups per 100 nucleotides divided by evolution rate (MLSA loci divergence in entire population/16S rRNA divergence). Filled squares indicate *Aciduliprofundum* (Flores *et al.*, 2012). Filled triangles indicate *Persephonella* (Mino *et al.*, 2013). Open circles indicate deep-sea hydrothermal vent *Sulfurimonas* (this study). Black lines are the regression line (solid thin line, *Sulfurimonas* (all strains pairs); solid thick line, *Sulfurimonas* (strain pairs <10 000 km); dotted line, *Aciduliprofundum*; broken line, *Persephonella*].

features with both the SMT and CIR ancestral populations. The STRUCTURE analysis differentiated OT-C1 members from OT-C2 members. The genetic features of OT-C1 members probably derived from mostly ancestral lineage 1 (red), whereas those of OT-C2 members from ancestral lineages 1 and 2. Additionally, the STRUCTURE analysis allowed us to determine the F_{ST} values as indices of the divergence of each ancestral population from their common ancestral population (Falush *et al.*, 2003). The F_{ST} values of the ancestral lineage 1, 2 and 3

were 0.899, 0.168 and 0.889, respectively. This result indicates that the ancestral lineage 2 is the most ancestral of the three populations.

Mutation and recombination
Homologous recombination and point mutations are significant evolutionary forces in bacteria (Tang *et al.*, 2012). The ρ/θ value estimated by ClonalFrame ranged from 0.06 to 0.28, indicating that recombination occurred less frequently than mutation. However, the relative effect of recombination per point mutation (r/m) was much >1, suggesting that rare recombination events had an important role in the diversification of the *Sulfurimonas* strains (Table 2). Extensive recombination events are supported by bushy network structures in a SplitsTree analysis (Supplementary Figure S2). The pairwise homoplasy index test also statistically supported the presence of recombination events in the *gyrB*, *feoB*, *atpA* and *napA* sequences ($P < 0.05$). In addition, inter- and intragenic recombinations within the OT strains were detected by the SITES program. The program found 740 informative sites in 6720 bp. The graphic evidence of the inter- and intragenic recombination is shown in Supplementary Figure S3.

Discussion

Recombination and mutation contribute to genetic diversity in both prokaryotes and eukaryotes, and the rate of occurrence varies widely depending on microbial taxa and ecological characteristics (Vos and Didelot, 2009). The ratio ($r/m = 2.4$–3.7) in *Sulfurimonas* supported the idea that marine/aquatic species have high (mostly $\geqslant 2$) r/m ratios (Vos and Didelot, 2009), and these values were similar to those of *Campylobacter*, *Mycoplasma*, *Haemophilus*, *Pseudomonas* and *Halorubrum*, which, with the exception of halophilic *Halorubrum*, are commensal or opportunistic pathogens (Vos and Didelot, 2009). Comparison of ρ/θ and r/m values among *Epsilonproteobacteria* shows that those of *Sulfurimonas* were comparable to those of *Campylobacter*, but were different from those of *Helicobacter*, which has high recombination rates possibly to establish persistent infection in specific hosts (Robinson *et al.*, 2005). Recombination appears to occur less frequently in *Sulfurimonas* and *Campylobacter*, potentially promoting the separation of daughter clusters (Sheppard *et al.*, 2014), which allow these microbes to colonize a wide range of environments/hosts.

Frequent dispersal of *Sulfurimonas* between sites separated by small distances (2–5 km) was demonstrated in this study by the high degree of genetic homogeneity, as previously shown in thermophilic *Persephonella* (Mino *et al.*, 2013). Additionally, in mesophilic *Sulfurimonas*, identical allele types were found in hydrothermal fields separated by medium distances (< 800 km), for example, Iheya
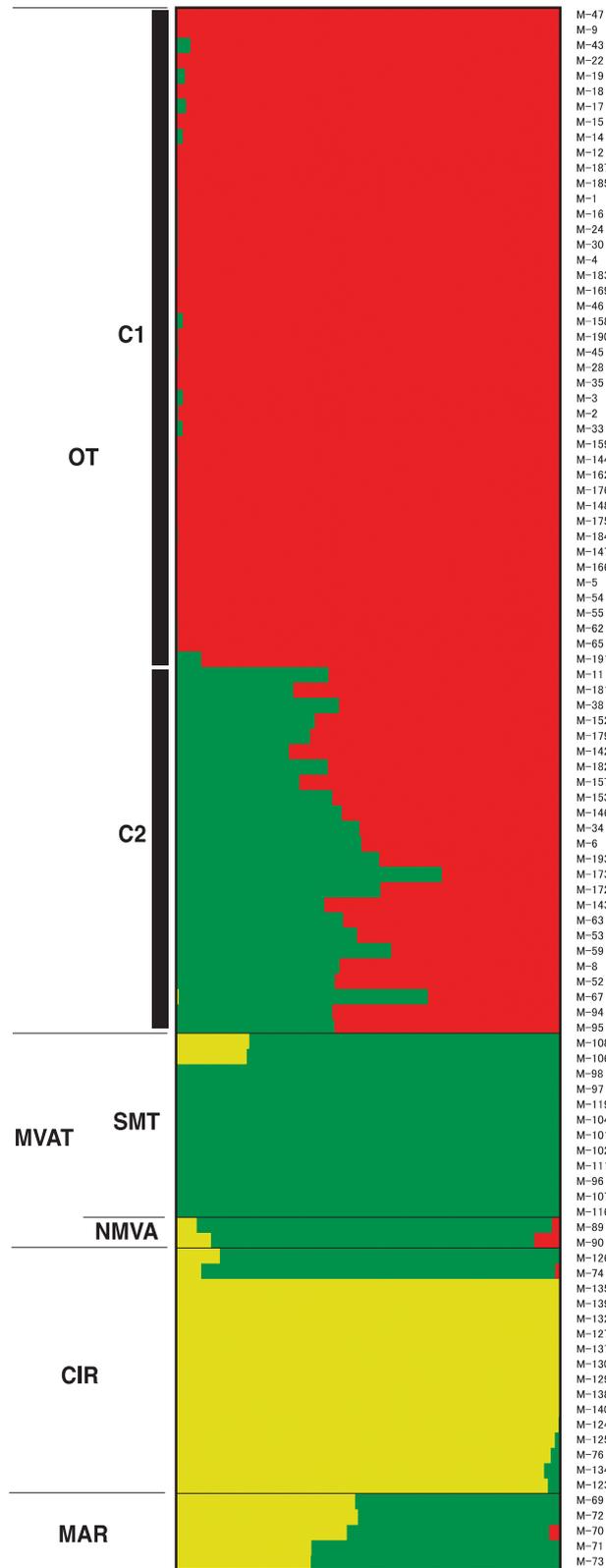
Figure 5 *Sulfurimonas* population structure at $K = 3$. Each horizontal line represents an individual strain. The three ancestral lineages are shown in red (lineage 1), green (lineage 2) and yellow (lineage 3).

North–Izena Hole, 61 km; Iheya North–Hatoma Knoll, 530 km; Izena Hole–Hatoma Knoll, 513 km; Kairei-Solitaire, 773 km. These results are suggestive of gene flow within the same hydrothermal region as it has been reported for deep-sea vent macrofauna (Nakamura *et al.*, 2012; Beedessee *et al.*, 2013). For the Mariana region, the data suggest that NMVA and SMT (separated by 926 km) possibly harbour genetically distinct *Sulfurimonas* populations. The biogeographical barrier between SMT and NMVA, which was previously reported for vent macrofauna (Kojima and Watanabe, 2015), separated microbial *Sulfurimonas* populations as well, suggesting that deep-sea hydrothermal vent mesophilic bacteria and macrofauna have similar dispersal capabilities or possibly disperse together at times. At the large scale across geographical regions (>1000 km), a clear geographical separation of *Sulfurimonas* populations emerged, implying that they are not mixed globally, most likely due to physical barriers that limit their distribution. Recently, the dispersal barrier between OT and the Mariana region has been demonstrated by macrofaunal biophysical model and ocean current data at a depth of 1000 m (Mitarai *et al.*, 2016), suggesting that deep currents and not surface currents affect the dispersal of *Sulfurimonas*. Since meso- to macro-scale bottom water currents remain unknown for the field sites investigated, we cannot estimate realistic dispersal pathway at present. Although geographical isolation at a regional scale emerged from the analysis, the STRUCTURE analysis detected genetic admixture (e.g., between OT and SMT), implying ancient connectivity between geographical regions. In the future, microbial population genetic studies might help to better understand the tectonic history of deep-sea hydrothermal fields.

Microbial geographical distribution is influenced by both geographical distance and local environmental conditions (Martiny *et al.*, 2006). Microbial community structures in the vicinity of deep-sea hydrothermal vents are thought to be controlled by the geochemistry of hydrothermal fluids (Takai and Nakamura, 2011), but environmental differences, that is, gas composition of vent fluid endmember and habitat, could not be strongly linked to the observed genetic variation among *Sulfurimonas* strains. This suggests that the measured vent fluid chemistry does not have a significant role in the genetic differentiation within or between geographical regions. The metabolic versatility of *Epsilonproteobacteria* might help them to adapt to the varying environmental conditions by inducing a shift in metabolism that takes advantage of available energy sources (Nakagawa *et al.*, 2005; Campbell *et al.*, 2006). However, *in situ* chemical conditions of each microhabitat could not be examined here. Additionally, endmember geochemical compositions of vent fluids can vary even within a single hydrothermal site due to phase separation and phase segregation

Table 2 The impact of recombination relative to mutation (r/m) and recombination frequency relative to mutation (ρ/θ)

Class or species	Region	Number of NST	Number of locus	r/m (95% CI[a])	ρ/θ (95% CI[a])	Reference
Epsilonproteobacteria	OT	70	11	2.7 (2.2–3.2)	0.25 (0.20–0.31)	This study
Sulfurimonas sp.	SMT	12	11	1.9 (1.0–3.2)	0.11 (0.04–0.22)	This study
	NMVA	2	11	ND	ND	This study
	CIR	20	11	3.1 (1.8–4.6)	0.06 (0.03–0.10)	This study
	MAR	5	11	3.0 (1.6–5.2)	0.24 (0.11–0.45)	This study
	All	102	11	3.1 (2.4–3.7)	0.28 (0.21–0.34)	This study
Helicobacter pylori		117	8	13. 6 (12.2–15.5)	ND	Vos and Didelot (2009)
		ND	16[b]	ND	4.0[c]	Pérez-Losada *et al.* (2006)
Campylobacter jejuni		110	7	2.2 (1.7–2.8)	ND	Vos and Didelot (2009)
		ND	7	ND	0.1[c]	Pérez-Losada *et al.* (2006)

Abbreviations: CIR, Central Indian Ridge; MAR, Mid-Atlantic Ridge; MLSA, multilocus sequence analysis; ND, not determined; NMVA, North Mariana Volcanic Arc; OT, Okinawa Trough; SMT, South Mariana Trough.
[a]Credibility interval.
[b]Some MLSA loci were overlapped.
[c]Average value of each gene calculated by LDhat package (McVean *et al.*, 2002).

(Massoth *et al.*, 1989). If all physicochemical data were taken *in situ* at the same time as when the isolations were done, it might be a better explanatory variable than geographical distance.

The geographical distance between sites significantly correlated with the genetic variation among *Sulfurimonas* strains, as has been previously demonstrated for other extremophiles (Whitaker *et al.*, 2003; Mino *et al.*, 2013). As we hypothesized, thermophiles tend to have less dispersal capability than the mesophilic *Sulfurimonas*, although it has to be kept in mind that genes selected for MLSA were different in both cases. Based on the influence of geographical distance on genetic variation of *Sulfurimonas*, we hypothesize that a scale <10 000 km is relevant for characterizing the dispersal capability of dominant deep-sea vent mesophiles.

In conclusion, by using MLSA we clearly demonstrate a genetic variation corresponding to hydrothermal regions, suggesting that *Sulfurimonas* is undergoing allopatric speciation as a consequence of the genetic drift occurring in geographically separated deep-sea hydrothermal regions. The genetic variability of *Sulfurimonas* may represent an advantage to adapt to deep-sea vent environments, possibly supporting the cosmopolitan distribution of *Sulfurimonas*. However, large population sizes of microbial species might increase microbial dispersal capability (Martiny *et al.*, 2006). In addition to further MLSA, whole-genome studies of cultivated and non-cultivated organisms, for example, through single-cell genomics, would lead to obtain a more comprehensive understanding of population genetics of deep-sea hydrothermal vent microbes.

Conflict of Interest

The authors declare no conflict of interest.

References

Baas-Becking LGM. (1934). *Geobiologie of inleiding tot de milieukunde*. Van Stockkum & Zoon: The Hague, The Netherlands.

Beedessee G, Watanabe H, Ogura T, Nemoto S, Yahagi T, Nakagawa S *et al.* (2013). High connectivity of animal populations in deep-sea hydrothermal vent fields in the Central Indian Ridge relevant to its geological setting. *PLoS One* **8**: e81570.

Bruen TC, Philippe H, Bryant D. (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**: 2665–2681.

Callac N, Rommevaux-Jestin C, Rouxel O, Lesongeur F, Liorzou C, Bollinger C *et al.* (2013). Microbial colonization of basaltic glasses in hydrothermal organic-rich sediments at Guaymas Basin. *Front Microbiol* **4**: 250.

Campbell BJ, Engel A, Porter M, Takai K. (2006). The versatile ε-proteobacteria: key players in sulfidic habitats. *Nat Rev Microbiol* **4**: 458–468.

Charlou J, Donval J, Fouquet Y, Jean-Baptiste P, Holm N. (2002). Geochemistry of high H_2 and CH_4 vent fluids issuing from ultramafic rocks at the Rainbow hydrothermal field (36º14′N, MAR). *Chem Geol* **191**: 345–359.

918

Darriba D, Taboada G, Doallo R, Posada D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**: 772.

Didelot X, Falush D. (2007). Inference of bacterial micro-evolution using multilocus sequence data. *Genetics* **175**: 1251–1266.

Excoffier L, Lischer HE. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* **10**: 564–567.

Falush D, Kraft C, Taylor NS, Correa P, Fox JG, Achtman M *et al.* (2001). Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. *Proc Natl Acad Sci USA* **98**: 15056–15061.

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.

Finlay JF, Fenchel T. (2004). Cosmopolitan metapopulations of free-living microbial eukaryotes. *Protist* **155**: 237–244.

Flores GE, Wagner ID, Liu Y, Reysenbach AL. (2012). Distribution, abundance, and diversity patterns of the thermoacidophilic 'deep-sea hydrothermal vent euryarchaeota 2'. *Front Microbiol* **3**: 47.

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.

Grote J, Labrenz M, Pfeiffer B, Jost G, Jürgens K. (2007). Quantitative distributions of *Epsilonproteobacteria* and a *Sulfurimonas* subgroup in pelagic redoxclines of the central Baltic Sea. *Appl Environ Microbiol* **73**: 7155–7161.

Grote J, Schott T, Bruckner CG, Glöckner FO, Jost G, Teeling H *et al.* (2012). Genome and physiology of a model *Epsilonproteobacterium* responsible for sulfide detoxification in marine oxygen depletion zones. *Proc Natl Acad Sci USA* **109**: 506–510.

Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506.

Hey J, Wakeley J. (1997). A coalescent estimator of the population recombination rate. *Genetics* **145**: 833–846.

Huson B. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**: 254–267.

Jolley KA, Feil EJ, Chan MS, Maiden MC. (2001). Sequence type analysis and recombinational tests (START). *Bioinformatics* **17**: 1230–1231.

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: New York, NY, USA.

Kojima S, Watanabe H. (2015). Vent fauna in the Mariana Trough. In: Ishibashi J-I, Okino K, Sunamura M (eds). *Subseafloor Biosphere Linked to Hydrothermal Systems: TAIGA Concept*. Cold Spring Harbor: New York, NY, USA.

Lupton J, Butterfield D, Lilley M, Evans L, Nakamura K, Chadwick W *et al.* (2006). Submarine venting of liquid carbon dioxide on a Mariana Arc volcano. *Geochem Geophys Geosyst* **7**: Q08007.

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

Martiny J, Bohannan B, Brown J, Colwell R, Fuhrman J, Green J *et al.* (2006). Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4**: 102–112.

Massoth GJ, Butterfield DA, Lupton JE, McDuff RE, Lilley MD, Jonasson IR. (1989). Submarine venting of phase-separated hydrothermal fluids at Axial Volcano, Juan de Fuca Ridge. *Nature* **340**: 702–705.

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

McVean G, Awadalla P, Fearnhead P, Awadalla P, Fearnhead P, Fearnhead P. (2002). A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**: 1231–1241.

Milkman R, Bridges MM. (1990). Molecular evolution of the Escherichia coli chromosome. III. Clonal frames. *Genetics* **126**: 505–517.

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.

Mitarai S, Watanabe H, Nakajima Y, Shchepetkin AF, McWilliams JC. (2016). Quantifying dispersal from hydrothermal vent fields in the western Pacific Ocean. *Proc Natl Acad Sci USA* **113**: 2976–2981 pii: 201518395.

Nakagawa S, Takai K, Inagaki F, Hirayama H, Nunoura T, Horikoshi K *et al.* (2005). Distribution, phylogenetic diversity and physiological characteristics of epsilon-*Proteobacteria* in a deep-sea hydrothermal field. *Environ Microbiol* **7**: 1619–1632.

Nakagawa S, Takaki Y, Shimamura S, Reysenbach AL, Takai K, Horikoshi K. (2007). Deep-sea vent epsilon-proteobacterial genomes provide insights into emergence of pathogens. *Proc Natl Acad Sci USA* **104**: 12156–12150.

Nakagawa S. (2011). Genetic analysis and preservation of chemolithoautotrophic bacteria dominating deep-sea vents. *IFO Res Commun* **25**: 31–39 (in Japanese with English abstract).

Nakamura K, Watanabe H, Miyazaki J, Takai K, Kawagucci S, Noguchi T *et al.* (2012). Discovery of new hydrothermal activity and chemosynthetic fauna on the Central Indian Ridge at 18°–20°S. *PLoS One* **7**: e32965.

Niemann H, Linke P, Knittel K, MacPherson E, Boetius A, Brückmann W *et al.* (2013). Methane-carbon flow into the benthic food web at cold seeps—a case study from the Costa Rica subduction zone. *PLoS One* **8**: e74894.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB *et al.* (2013). vegan: Community Ecology Package. R package version 2.4-1 [online]. Available at: http://CRAN.R-project.org/package=vegan.

Papke RT, Ramsing NB, Bateson MM, Ward DM. (2003). Geographical isolation in hot spring cyanobacteria. *Environ. Microbiol* **5**: 650–659.

Pérez-Losada M, Browne EB, Madsen A, Wirth T, Viscidi RP, Crandall KA. (2006). Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. *Infect Genet Evol* **6**: 97–112.

Price MT, Fullerton H, Moyer CL. (2015). Biogeography and evolution of *Thermococcus* isolates from hydrothermal vent systems of the Pacific. *Front Microbiol* **6**: 968.

Pritchard J, Stephens M, Donnelly P. (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Robinson K, Loughlin MF, Potter R, Jenks PJ. (2005). Host adaptation and immune modulation are mediated by homologous recombination in *Helicobacter pylori*. *J Infect Dis* **191**: 579–587.

Rogers AD, Tyler PA, Connelly DP, Copley JT, James R, Larter RD *et al.* (2012). The discovery of new deep-sea hydrothermal vent communities in the southern ocean and implications for biogeography. *PLoS Biol* **10**: e1001234.

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.

Sheppard SK, Cheng L, Méric G, de Haan CP, Llarena AK, Marttinen P *et al.* (2014). Cryptic ecology among host generalist *Campylobacter jejuni* in domestic animals. *Mol Ecol* **23**: 2442–2451.

Silvestro D, Michalak I. (2012). raxmlGUI: a graphical front-end for RAxML. *Org Divers Evol* **12**: 335–337.

Takai K, Nakamura K. (2011). Archaeal diversity and community development in deep-sea hydrothermal vents. *Curr Opin Microbiol* **14**: 282–291.

Takai K, Nakagawa S, Reysenbach AL, Hoek J. (2006). Microbial ecology of mid-ocean ridges and back-arc basins. In: Christie DM, Fisher CR, Lee SM, Givens S (eds). *Back-Arc Spreading Systems: Geological, Biological, Chemical and Physical Interactions*, vol 166. Geophysics Monograph Series AGU: Washington, DC, USA.

Tang J, Bromfield ES, Rodrigue N, Cloutier S, Tambong JT. (2012). Microevolution of symbiotic *Bradyrhizobium* populations associated with soybeans in east North America. *Ecol Evol* **2**: 2943–2961.

Toki T, Ishibashi J, Noguchi T, Tawata M, Tsunogai U, Yamanaka T *et al.* (2015). Chemical and isotopic compositions of hydrothermal fluids at Snail, Archaean, Pika, and Urashima sites in the Southern Mariana Trough. In: Ishibashi J-i, Okino K, Sunamura M (eds). *Subseafloor Biosphere Linked to Hydrothermal Systems: TAIGA Concept*. Cold Spring Harbor: New York, NY, USA.

Vos M, Didelot X. (2009). A comparison of homologous recombination rates in bacteria and archaea. *ISME J* **3**: 199–208.

Whitaker RJ, Grogan DW, Taylor JW. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301**: 976–978.

Zhang Y, Zhao Z, Chen CT, Tang K, Su J, Jiao N. (2012). Sulfur metabolizing microbes dominate microbial communities in Andesite-hosted shallow-sea hydrothermal systems. *PLoS One* **7**: e44593.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)