ORIGINAL ARTICLE

Heterogeneous composition of key metabolic gene clusters in a vent mussel symbiont population

Tetsuro Ikuta[1,8], Yoshihiro Takaki[2,8], Yukiko Nagai[1], Shigeru Shimamura[2], Miwako Tsuda[2], Shinsuke Kawagucci[2], Yui Aoki[1], Koji Inoue[3], Morimi Teruya[4], Kazuhito Satou[5], Kuniko Teruya[5], Makiko Shimoji[5], Hinako Tamotsu[5], Takashi Hirano[5,6], Tadashi Maruyama[7] and Takao Yoshida[1]

[1]Department of Marine Biodiversity Research, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka, Kanagawa, Japan; [2]Department of Subsurface Geobiological Analysis and Research, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka, Kanagawa, Japan; [3]Department of Marine Bioscience, Atmosphere and Ocean Research Institute, The University of Tokyo, Kashiwa, Chiba, Japan; [4]Okinawa Industrial Technology Center, 12-2 Suzaki, Uruma, Okinawa, Japan; [5]Okinawa Institute of Advanced Sciences (OIAS), 5-1 Suzaki, Uruma, Okinawa, Japan; [6]Okinawa Science and Technology Promotion Center (OSTC), 112-18 Asahimachi, Naha, Okinawa, Japan and [7]Research and Development Center for Marine Biosciences, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka, Kanagawa, Japan

Chemosynthetic symbiosis is one of the successful systems for adapting to a wide range of habitats including extreme environments, and the metabolic capabilities of symbionts enable host organisms to expand their habitat ranges. However, our understanding of the adaptive strategies that enable symbiotic organisms to expand their habitats is still fragmentary. Here, we report that a single-ribotype endosymbiont population in an individual of the host vent mussel, *Bathymodiolus septemdierum* has heterogeneous genomes with regard to the composition of key metabolic gene clusters for hydrogen oxidation and nitrate reduction. The host individual harbours heterogeneous symbiont subpopulations that either possess or lack the gene clusters encoding hydrogenase or nitrate reductase. The proportions of the different symbiont subpopulations in a host appeared to vary with the environment or with the host's development. Furthermore, the symbiont subpopulations were distributed in patches to form a mosaic pattern in the gill. Genomic heterogeneity in an endosymbiont population may enable differential utilization of diverse substrates and confer metabolic flexibility. Our findings open a new chapter in our understanding of how symbiotic organisms alter their metabolic capabilities and expand their range of habitats.
The ISME Journal (2016) **10,** 990–1001; doi:10.1038/ismej.2015.176; published online 29 September 2015

Introduction

A wide variety of animals have acquired the ability to live on inorganic carbon sources by establishing symbioses with chemoautotrophic bacteria in various different habitats from shallow water to deep sea, such as hydrothermal vents. In the deep-sea vent symbioses, the symbiotic bacteria use the chemical energy of reduced sulphur, methane or hydrogen from vent fluid (Dubilier *et al.*, 2008; Petersen *et al.*, 2011). These bacteria also require sources for assimilation such as carbon dioxide, methane,

ammonia, nitrate and so on. The metabolic ability of the symbionts is largely affected by the geochemical conditions, whereas the physicochemical features of the hydrothermal vent environment are highly variable (Zielinski *et al.*, 2011), and the availability of metabolic substrates is unpredictable. Due to the variable nature of hydrothermal vent environments, some host animals have multiple phylogenetically and physiologically distinct endosymbiont species allowing them to use a variety of energy sources, that is, conferring metabolic flexibility, and to occupy various habitats (Fisher *et al.*, 1993; Distel *et al.*, 1995; Woyke *et al.*, 2006; Dubilier *et al.*, 2008; Robidart *et al.*, 2008; Beinart *et al.*, 2012). On the other hand, intra-species subtype variations with single nucleotide substitutions have been reported in some bacterial endosymbionts (DeChaine *et al.*, 2006; Vrijenhoek *et al.*, 2007; Dubilier *et al.*, 2008). However, until recently, genomic diversity on the level of gene composition within a

Correspondence: T Yoshida, Department of Marine Biodiversity Research, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka, Kanagawa 237-0061, Japan.
E-mail: tyoshida@jamstec.go.jp
[8]These authors contributed equally to this work.

single symbiont species has been considerably under-estimated, and we have known little about the ecological importance of such genomic variations.

Here, we report a new and unique genome heterogeneity in a single thioautotrophic endosymbiont population in the deep-sea vent mussel, *Bathymodiolus septemdierum*. Whole-genome sequencing showed metabolically heterogeneous genomes in a *B. septemdierum* endosymbiont (BSEPE) population with a single type of 16S ribosomal gene in a single host individual. The symbiont population was composed of several subpopulations that possessed or lacked gene clusters for key metabolic enzymes such as hydrogenase and nitrate reductase. To confirm the generality of the heterogeneity in the BSEPE genome and to examine the proportions of the symbiont subpopulations harbouring each gene cluster within an individual host, we performed quantitative polymerase chain reaction (qPCR) in multiple hosts from two different vent sites. To investigate the distributions of the symbiont subpopulations in the host gill, we conducted *in situ* hybridization (ISH) analyses. In addition, to test whether the genomic variants corresponding to the symbiont subpopulations exist in the habitat, we performed PCR using DNA extracted from seawater near the *Bathymodiolus* mussel colony as a template. On the basis of our findings, we discuss possible scenarios for producing the heterogeneous symbiont subpopulations.

Materials and methods

Sample collection and preparation
B. septemdierum was collected from hydrothermal vents on the Myojin Knoll, at depths of 1228 m (Dive #1126), 1249 m (Dive #1127), 1303 m (Dive #1284) and 1278 m (Dive #1288), during the cruises NT10-08 (11–18 May 2010) and NT11-09 (15–26 June 2011), and from Suiyo Seamount at depths of 1386 m (Dive #0679, Dive #0680 and Dive #1286) and 1384 m (Dive #1287) m during the cruises NT07-08 (15–17 May 2007) and NT11-09 (15–26 June 2011) (see also Supplementary Table S1) with the ROV *Hyper Dolphin*, operated by the R/V *Natsushima* of the Japan Agency of Marine-Earth Science and Technology. The collected mussels were either frozen immediately in liquid nitrogen or dissected on-board. The gills were cut out using a disposable scalpel and used for DNA extraction or fixed as follows. For ISH analysis, the gills were fixed on-board in 4% paraformaldehyde in $1 \times$ PBS for 16 h at 4 °C, followed by stepwise dehydration in an ethanol series. For RNA extraction, the gills were placed into RNAlater (Qiagen, Venlo, Netherlands), incubated for 16 h at 4 °C, and stored at −80 °C.

Genome sequencing and assembly
Genomic DNA of the symbiont was purified from the gill of a single individual of *B. septemdierum* from the Myojin Knoll as described previously (Kuwahara

et al., 2007). Briefly, the gill was homogenized, and tissue debris was removed by filtration through nylon filters. The bacterial cells were collected by centrifugation and were subjected to DNase digestion to eliminate host DNA. Symbiont DNA was then extracted from the pelleted bacterial cells and sequenced using 454 Titanium (Roche, Basel, Switzerland) with the 3-kb paired-end library. The 454 sequence data yielded 74.5 Mb over 282 803 reads with an average read length of 264 nt. *De novo* genomic assembly was performed using the GS De Novo Assembler version 2.3 (Newbler, Roche). Assembly of the data resulted in 267 contigs with length > 500 bp, totalling 1.5 Mb with G+C content of 36%, and ordered to 25 scaffolds. Together, all reads provided ~40 × coverage of the symbiont genome. Gaps in the genome were closed using PCR amplification and Sanger sequencing. To validate the genomic architecture and nucleotide sequence, the DNA samples were resequenced using PacBio RS (Pacific Biosciences, Menlo Park, CA, USA) and Illumina platforms (see Supplementary Methods for details). The genome sequence and annotation of BSEPE are available at GenBank/EMBL/DDBJ under accession number AP013042. The Illumina/PacBio reads for resequencing analysis are under accession number DRA002953

Construction of the pangenomic sequence of the *B. septemdierum* symbiont
The PCR results for gap filling suggested that there was structural heterogeneity in the BSEPE genome resulting from the presence or absence of a genomic fragment, or variation in nucleotide sequences with base substitutions and short indels. The architecture of regions with structural heterogeneity was investigated by amplifying PCR products covering the boundaries of both 5′ and 3′ sides of the heterogeneous region. In cases where gaps in the genome could not be filled by PCR because of the presence of more than two alternative architectures at the locus, cloning analysis by the TA cloning method was performed to obtain a single clone from PCR fragments. Finally, a pangenomic sequence of the *B. septemdierum* symbiont was constructed based on the longer gap sequences.

To validate the genomic architecture and nucleotide sequence, the DNA samples were resequenced using PacBio RS (Pacific Biosciences) and Illumina platforms. A 10-kb SMRT-bell library was constructed according to the manufacturer's protocol and sequenced with eight SMRT cells using C2 chemistry. The longer PacBio RS reads produced 367 131 reads totalling about 757 Mb, with a median of 1707 nt (~505 × coverage). To confirm the structural accuracy and to correct misassembled repeat regions, PacBio read mappings were performed by a BLASTN search against the pangenomic sequence. Next, paired-end Illumina sequencing (2 × 150 nt) generated 59 137 101 reads totalling ~ 8.87 Gb using

one lane of a HiSeq2000 system. These data were mapped to the pangenomic sequence to an average depth of 2065 using CLC genomics workbench (Qiagen) with the default parameter settings. The consensus sequence of the mapped reads was created using the 'make-consensus' module available in the AMOS package (Treangen et al., 2011).

Detection of DNA fragments from seawater

Seawater near the *Bathymodiolus* mussel colony on the Myojin Knoll was collected in the Dive #1645 dive of the ROV *Hyper Dolphin*, operated by the R/V *Natsushima* NT14-06 cruise of the Japan Agency of Marine-Earth Science and Technology. Microbes in ~ 1.2 l seawater were collected by filtration through a nitrocellulose membrane with a pore size of 0.22 μm (Millipore, Billerica, MA, USA). DNA was extracted from the collected microbes with DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's manual. PCR was conducted with the specific primer sets for the membrane-bound uptake hydrogenase (*hup*) and nitrate reductase (*nar*) gene clusters. For PCR amplification, *La Taq* DNA polymerase (Takara, Shiga, Japan) was used with an initial denaturation phase of 94 °C for 1 min, followed by 35 cycles of 96 °C for 20 s, 68 °C for 3 min and 72 °C for 4 min. Symbiont genomic DNA purified from the gill was used as a positive control. Template DNA was omitted in the negative control, and no amplification was observed. The primers used in this study were as follows: Hup_5'F, 5′-TATCATTCAACCTGCCATCCGCAG-3′; Hup_5'R, 5′-AAGAAGCCGATTTGGTGCTCAGAG-3′; Hup_3'F, 5′-TGGTAAACGCATAACCATAACGAC G-3′; Hup_3'R, 5′-CAATCAACACCGGAATGATGA TGG-3′; Nar_5'F, 5′-TTCACGCCAAGCGATGATTAT TCAG-3′; Nar_5'R, 5′-AATACTAATGTTTGGCTTGC CACACG-3′; Nar_3'F, 5′-GAATCGGATGCGAATG AGATTGATAG-3′; Nar_3'R, 5′-TCTAGGTTTCCAA TCGGTGTAGCG-3′. Predicted amplicon lengths for these PCR conditions were 1203 bp for Hup_5'F and Hup_5'R, 3287 bp for Hup_3'F and Hup_3'R, 2932 bp for Hup_5'F and Hup_3'R, 1215 bp for Nar_5'F and Nar_5'R, 1496 bp for Nar_3'F and Nar_3'F, and 1895 bp for Nar_5'F and Nar_3'R. The primers were designed to be specific to the BSEPE genome sequence, and primer specificity was verified by comparison with sequence data of other known organisms via a BLASTN search against the NCBI database (http://blast.ncbi.nlm.nih.gov/Blast. cgi). Furthermore, the sequence of each PCR product amplified with DNA from seawater was confirmed by DNA Sanger sequencing and found to match 99.42–100% to our BSEPE genome sequence.

ISH

To amplify the genomic fragments corresponding to the loci targeted in ISH, we designed three primer sets for *Locus_D*, four for *Locus_H* and one for *Locus_N*. These primers produced 5.5–10.9 kb of PCR fragments covering the target locus as a whole (Figure 1). The PCR fragments were labelled with fluorescein or digoxigenin using nick translation kits from Enzo Life Sciences or Roche, respectively (see Supplementary Methods for details). For whole-mount *in situ* hybridization (WISH) for genomic loci, the fixed gill filaments of two host individuals from Myojin Knoll (Supplementary Table S1) were used. The hybridization reaction with the fluorescein labelled probe was carried out overnight at 37 °C. After incubation in anti-fluorescein-AP (Roche) in PBST with blocking, hybridization was detected using NBT/BCIP solution (Roche) in TMNT buffer (see Supplementary Methods for details). For fluorescent *in situ* hybridization (FISH), the fixed gills of three host mussels from Myojin Knoll, different individuals from those used for WISH (Supplementary Table S1), were embedded in paraffin and sliced with a microtome into 4-μm-thick sections. Two-colour FISH of the gill sections was performed using probes labelled with fluorescein or digoxigenin and TSA Plus Cyanine 3/Fluorescein System (Perkin Elmer, Waltham, MA, USA) (see Supplementary Methods for details).

qPCR

Four adult *B. septemdierum* specimens were collected from Myojin Knoll, and two adults and six juveniles were collected from Suiyo Seamount (Supplementary Table S1). The shell lengths of the adults and juveniles were 110–120 and 2.8–5.0 mm, respectively. Genomic DNAs from the symbionts were purified from the gills of the six adult individuals, as described previously (Kuwahara et al., 2007). Total DNA was extracted from each of six whole juveniles using a DNeasy Blood and Tissue Kit (Qiagen). The qPCR analysis was performed on an Applied Biosystems 7300 Real-Time PCR System (Life Technologies, Waltham, MA, USA) with the Power SYBR Green PCR Master Mix (Life Technologies), and the threshold cycle (Cq) was determined automatically for each gene. All reactions were carried out in triplicate, including a no template control, with PCR conditions of 95 °C for 10 min, 40 cycles of 95 °C for 15 s and 60 °C for 1 min, and a dissociation stage at the end of the run from 60 °C to 95 °C. A five-fold dilution series (from 1e−5 to 1.6e−8 pmol) of the PCR fragment encompassing the region targeted by each qPCR primer set was used as a template to produce a calibration curve for absolute quantification, and the values acquired with the calibration curves were computed reflecting PCR amplification efficiency (see Supplementary Methods and Supplementary Table S2 for details). Statistical analysis was performed using Microsoft Office Excel for Mac 2011 (Microsoft, Redmond, WA, USA), in which *P*-values were calculated via Mann–Whitney *U*-tests. All qPCR analyses followed the MIQE guidelines (Bustin et al., 2009).
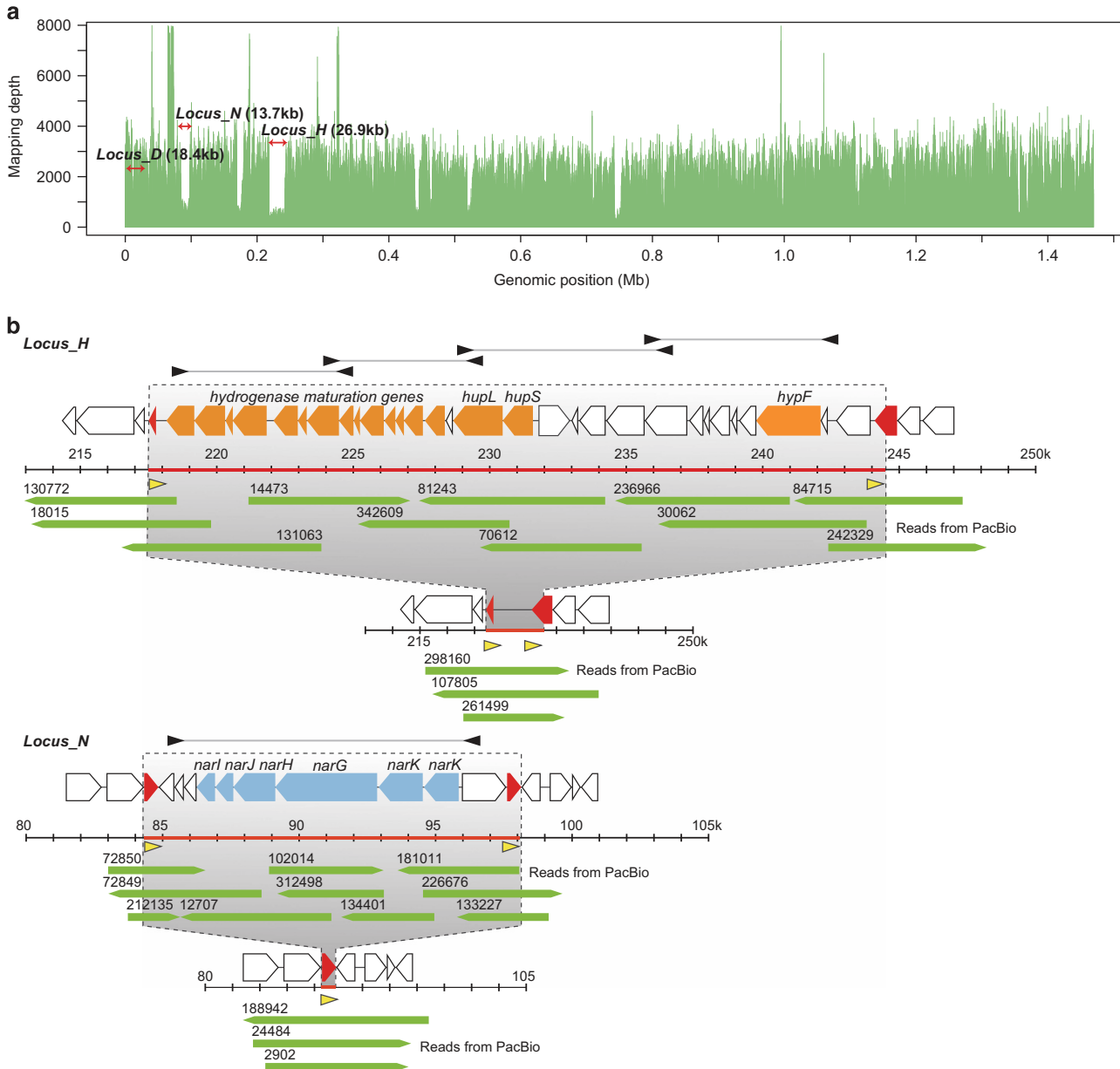
Figure 1 Loci with low mapping depth including the *hup* and *nar* gene clusters. (**a**) Illumina sequence redundancies mapped on their corresponding genomic positions. The mapping depth of some loci, represented by *Locus_H* and *Locus_N*, is lower than that of other genomic regions. (**b**) Predicted genomic architecture of *Locus_H* and *Locus_N*. Cases with or without each gene cluster (*hup* or *nar*) are shown. Genes related to the hydrogenase complex and nitrate reductase are indicated with orange and blue arrows, respectively. Duplicated genes (red arrows) are located at both ends of each cluster and form direct repeats (yellow arrowheads). Green arrows indicate PacBio reads. PCR primers used for template preparation of probes for ISH analysis are indicated with black arrowheads.

Results

Unique heterogeneity in the *Bathymodiolus septemdierum* symbiont genome

B. septemdierum, a dominant species at hydrothermal vents in the Izu-Ogasawara Arc, Japan, harbours a thiotrophic bacterium (Fujiwara *et al.*, 2000), presumably acquired from the environment (Won *et al.*, 2003). The BSEPE is harboured in the gill epithelial cells called bacteriocytes (Fujiwara *et al.*, 2000), although the mechanism and timing of infection remains unclear. Our assembly of BSEPE genomic sequences yielded a circular chromosome of 1 469 434 bp containing one ribosomal RNA operon and 1471 protein-coding genes (Table 1, Supplementary Figure S1 and Supplementary Table S1). We detected no polymorphisms in the ribosomal gene operon (16S+23S) by 454 sequencing with an average read depth of 40 and Illumina sequencing with an average read depth of >1000. These data indicate that one individual *B. septemdierum* host harbours a single symbiont ribotype with no polymorphisms. This is in line with previous report of highly similar to identical ribotypes for nineteen 16S

Table 1 Genomic features of the *B. septemdierum* symbiont, the related thioautotrophic symbionts and free-living bacteria

Symbol	BSEPE	eofBa	SUP05OMZ	SUP05GB1	SUP05GB2	Rma	Vok	eofSc
Organism	Endosymbiont of *Bathymodiolus septemdierum* str. Myojin knoll	Endosymbiont of *Bathymodiolus sp.*	Saanich Inlet OMZ SUP05	Guaymas Basin SUP05 GB-1	Guaymas Basin SUP05 GB-2	*Candidatus* Ruthia magnifica str. Cm	*Candidatus* Vesicomyosocius okutanii HA	Scaly-foot gastropod endosymbiont
Host	*Bathymodiolus septemdierum*	*Bathymodiolus sp.*	Free-living	Free-living	Free-living	*Calyptogena magnifica*	*Calyptogena okutanii*	Scaly foot gastropod
Accession number	AP013042	CAEB00000000	GG729945-GG729951, GG729953-GG729965, GG729967-GG730040	AJXC01000000[a]	AJXC01000000[a]	CP000488	AP009247	AP012978
Genome size	1 469 434	>619 851	>1 164 035	>1 246 425	>1 238 377	1 160 782	1 022 154	2 597 759
GC content	0.387	0.384	0.393	0.395	0.387	0.340	0.316	0.651
State	Complete	Draft	Draft	Draft	Draft	Complete	Complete	Complete
Contigs	1	78	90	100	120	1	1	1
Ave. redundancy	40	14.5	7.25	13.07	12.59	—	—	65
Protein coding sequences	1471	636	1282	1535	1568	976	937	2249
Pseudogene	52	—	23	—	—	100	3	152
rRNA operon	1	—	1	1	1	1	1	1
tRNA genes	35	11	29	34	34	36	35	40
Coverage of genes	0.924	0.889	0.878	0.933	0.929	0.799	0.852	0.826
Orthologs[b]	—	509	948	1066	1055	901	877	1013
Maximum likelihood distance (Mode)	—	0.18	0.144	0.162	0.14	0.404	0.398	0.707
Reference	—	Petersen et al. (2011)	Walsh et al. (2009)	Anantharaman et al. (2013)	Anantharaman et al. (2013)	Newton et al. (2007)	Kuwahara et al. (2007)	Nakagawa et al. (2014)

[a]Accession numbers of individual genomic contigs are shown in the Supplementary Methods section. [b]Number of orthologs between BSEPE and the reference genome.
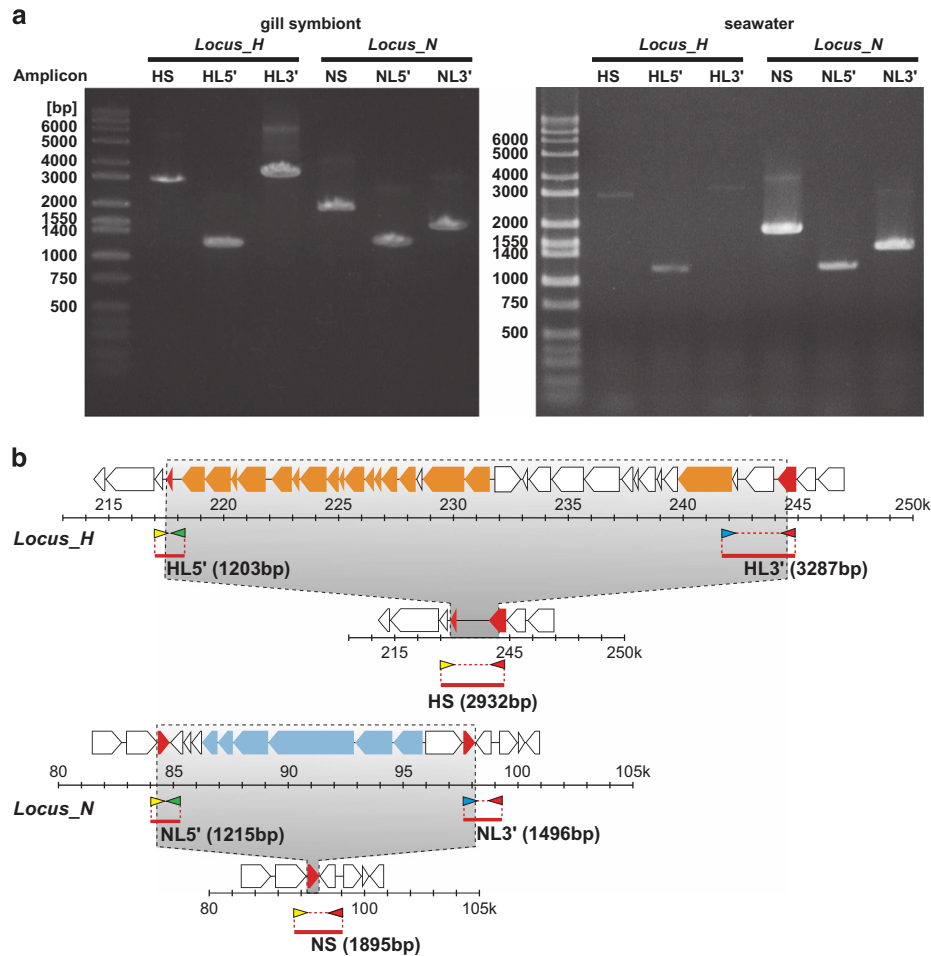
Figure 2 Detection of DNA fragments that possess or lack the symbiont's *hup* and *nar* gene clusters, from seawater near a *Bathymodiolus* mussel colony. (**a**) Left, result of PCR using symbiont genomic DNA purified from the gill. Right, result of PCR using DNA extracted from seawater. Detection of amplicon HS or NS indicates lack of the *hup* or *nar* gene cluster, respectively, and detection of amplicon HL5′/ HL3′ or NL5′/ NL3′ indicates presence of each cluster. To detect the presence of the *hup* or *nar* gene cluster, both the 5′ and 3′ terminal ends of each gene cluster were amplified. In our PCR conditions using Hup_5′F or Nar_5′F primers (represented by yellow arrowheads in (**b**) and Hup_3′R or Nar_3'R (represented by red arrowheads in (**b**), an amplicon containing the *hup* or *nar* gene cluster was not detected because estimated lengths of these amplicons from our bioinformatic data were too long at 27 865 and 15 178 bp, respectively. (**b**) Genomic positions of amplicons and specific primers. Genomic architectures of *Locus_H* and *Locus_N* with or without each gene cluster (*hup* or *nar*) are shown with the original genomic positions. The *hup* and *nar* gene clusters are indicated in orange and blue, respectively. Duplicated genes are indicated in red. Primer sets of amplicons HS and NS are indicated with yellow (Hup_5′F or Nar_5′F primers) and red (Hup_3′R or Nar_3′R primers) arrowheads. Primer sets of amplicon HL5′ and NL5′ are indicated with yellow and green (Hup_5′R or Nar_5′R primers) arrowheads, and primer sets of amplicon HL3′ and NL3′ are indicated with blue (Hup_3′F or Nar_3′F primers) and red arrowheads. Lengths of amplicons estimated from our bioinformatic data are shown in the respective parentheses.

rRNA sequences of *Bathymodiolus* symbionts in 11 host individuals (two host species collected from four different vent sites) (Duperron *et al.*, 2006). In a phylogenetic tree based on 16S rRNA genes, the thiotrophic endosymbionts of *Bathymodiolus* (including BSEPE) form a robust monophyletic clade with *Calyptogena* clam endosymbionts and their free-living relatives SUP05 (Supplementary Figure S2).

The mapping depth of short reads obtained from next-generation sequencing of the genome of BSEPE was obviously lower in some genomic loci than in others (Figure 1a and Supplementary Table S3). The largest of these loci (*Locus_H*) occupied 26.9 kb

and included a cluster of genes encoding key enzymes for hydrogen oxidation (Figures 1a and b and Supplementary Table S3). In the BSEPE genome, we found a membrane-bound uptake hydrogenase (*hup*) (Vignais and Billoud, 2007) gene cluster consisting of the key structural genes *hupL* and *hupS* and other accessory genes (Figure 1b and Supplementary Table S3), as has been reported for other bacteria (Hansel *et al.*, 2001; Brito *et al.*, 2005). The second-largest locus with a low mapping depth (*Locus_N*, 13.7 kb) included a cluster of genes that encode enzymes involved in dissimilatory nitrate reduction, nitrate reductase (Nar) (Figures 1a and b and Supplementary Table S3). In the BSEPE genome,

narG, -H, -J and *-I* are present in this cluster (Figure 1b and Supplementary Table S3), as in some other bacteria (Blasco *et al.*, 1990; Hartig *et al.*, 1999). Our PCR analysis using primer sets covering these two loci (Figure 2) and the long sequencing reads obtained from a PacBio RS sequencer showed the occurrence of two genomic architectures: one with and the other without these gene clusters (Figure 1b). In our genome sequence data, we did not observe any 16S rRNA sequence likely derived from the close relatives of the symbiont. Furthermore, in BLASTN searches with *Locus_H* and *Locus_N* (Figure 1 and Supplementary Table S3), no homologous sequences (coverage >60% and identity >80%) were found in sequence data of other known organisms. These results indicated the presence of genomic hetero-geneity in BSEPE within a single host mussel, that is, the BSEPE genome is not uniform, but rather is heterogeneous within a symbiont population with a single type of 16S ribosomal gene. Genomic regions that include the *hup* or *nar* gene clusters are absent in some symbiont subpopulations in a population living in one host mussel.

Composition and distribution of the symbiont subpopulations

To confirm the generality of this finding, we performed qPCR in multiple host individuals from two different vent sites, Myojin Knoll and Suiyo Seamount. In adult specimens (A1–4 from Myojin Knoll and A5 and 6 from Suiyo Seamount in Figure 3 and Supplementary Table S1), the relative amount of *hupL* and *narG* DNA compared with a reference single-copy gene (*dnaA*) present with average map-ping depth (2486 of mapping depth) (Figure 1a) was low (4.28–24.17% for *hupL* and 27.22–61.24% for *narG* in Figures 3a and b, respectively), except for *narG* in one specimen from Suiyo Seamount (98.96%, A6 in Figure 3b). Therefore, genome heterogeneity in BSEPE was common in two differ-ent *B. septemdierum* colonies from geographically separated sites. Furthermore, the relative amount of *hupL* or *narG* compared with *dnaA* in juveniles (from Suiyo Seamount) was higher than that in adult specimens (from Myojin Knoll and Suiyo Seamount) (P-value = 0.016 in *hupL* and 0.037 in *narG*, Figure 3 and Supplementary Table S1). Also, the relative amount of *narG* compared with *dnaA* in specimens from Suiyo Seamount (including adults and juve-niles) was higher than that in specimens from Myojin Knoll (adults only) (P-value = 0.007 in *narG*, Figure 3 and Supplementary Table S1). In addition, the sums of the proportions of *hupL* to *dnaA* and *narG* to *dnaA* in some specimens were >100% (Figure 3), indicating that *Locus_H* and *Locus_N* coexist in some symbiont genomes.

To confirm the occurrence of genomic heteroge-neity in the symbiont harboured in the gill, and to investigate the distribution of symbiotic bacteria possessing the *hup* and *nar* gene clusters in the gill,
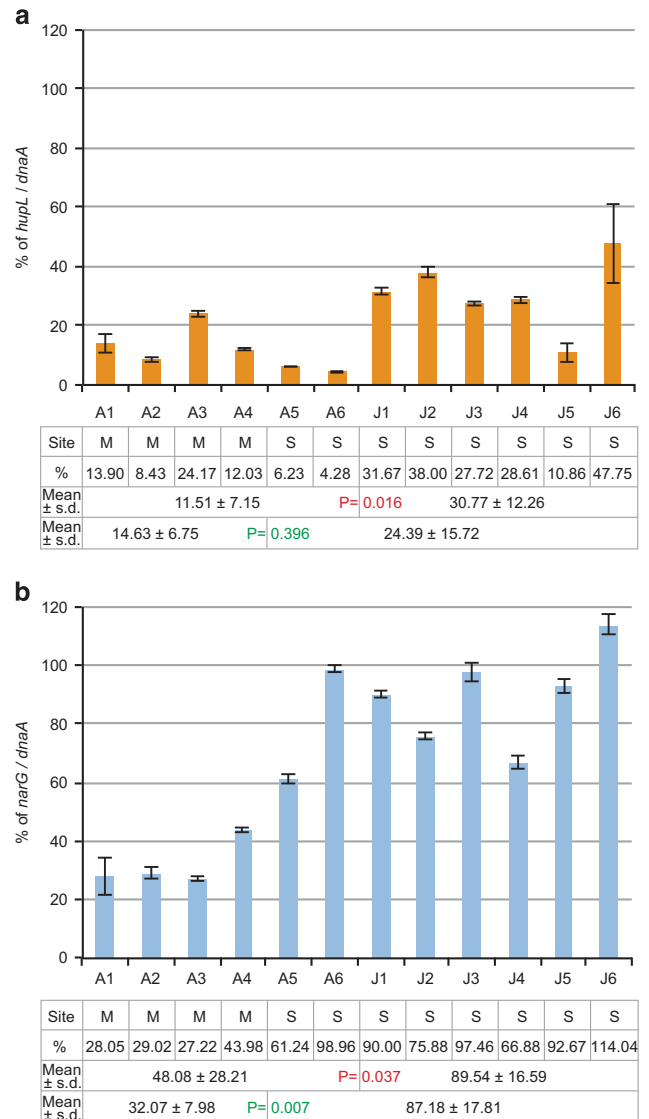

a

Site	M	M	M	M	S	S	S	S	S	S	S	S
	A1	A2	A3	A4	A5	A6	J1	J2	J3	J4	J5	J6
%	13.90	8.43	24.17	12.03	6.23	4.28	31.67	38.00	27.72	28.61	10.86	47.75
Mean ± s.d.	11.51 ± 7.15				P= 0.016		30.77 ± 12.26					
Mean ± s.d.	14.63 ± 6.75			P= 0.396			24.39 ± 15.72					

b

Site	M	M	M	M	S	S	S	S	S	S	S	S
	A1	A2	A3	A4	A5	A6	J1	J2	J3	J4	J5	J6
%	28.05	29.02	27.22	43.98	61.24	98.96	90.00	75.88	97.46	66.88	92.67	114.04
Mean ± s.d.	48.08 ± 28.21				P= 0.037		89.54 ± 16.59					
Mean ± s.d.	32.07 ± 7.98			P= 0.007			87.18 ± 17.81					

Figure 3 The proportions of the subpopulation harbouring *hupL* or *narG*. (**a**, **b**) Proportion (%) of symbiont subpopulations with *hupL* (**a**) or *narG* (**b**) in each host individual. The amounts of DNA of *hupL* and *narG* genes relative to DNA of the reference single-copy gene, *dnaA* on a locus with average mapping depth, were measured by qPCR in six adult individuals (A1–A6) and six juveniles (J1–J6). Error bars indicate the s.d. for three experimental replicates. The table below the chart indicates the vent sites (M, Myojin Knoll and S, Suiyo Seamount), proportions of *hupL* or *narG* to *dnaA*, mean proportions ± s.d. in adults and juveniles, and mean proportions ± s.d. in two vent sites. *P*-values by Mann–Whitney *U*-test between adults and juveniles, and between Myojin Knoll and Suiyo Seamount are shown in red and green, respectively.

we conducted ISH analysis for BSEPE genomic DNA. We used probes for *Locus_H*, *Locus_N* and *Locus_D* (which includes the reference *dnaA* gene) (Figure 1a). *Locus_H* and *Locus_N* were detected in patches in the whole-mount gill filaments and the gill sections, whereas *Locus_D* was detected in all the bacteriocytes (Figure 4 and Supplementary Table S1). The proportions of the area of fluores-cence signals for *Locus_H* and *Locus_N* compared
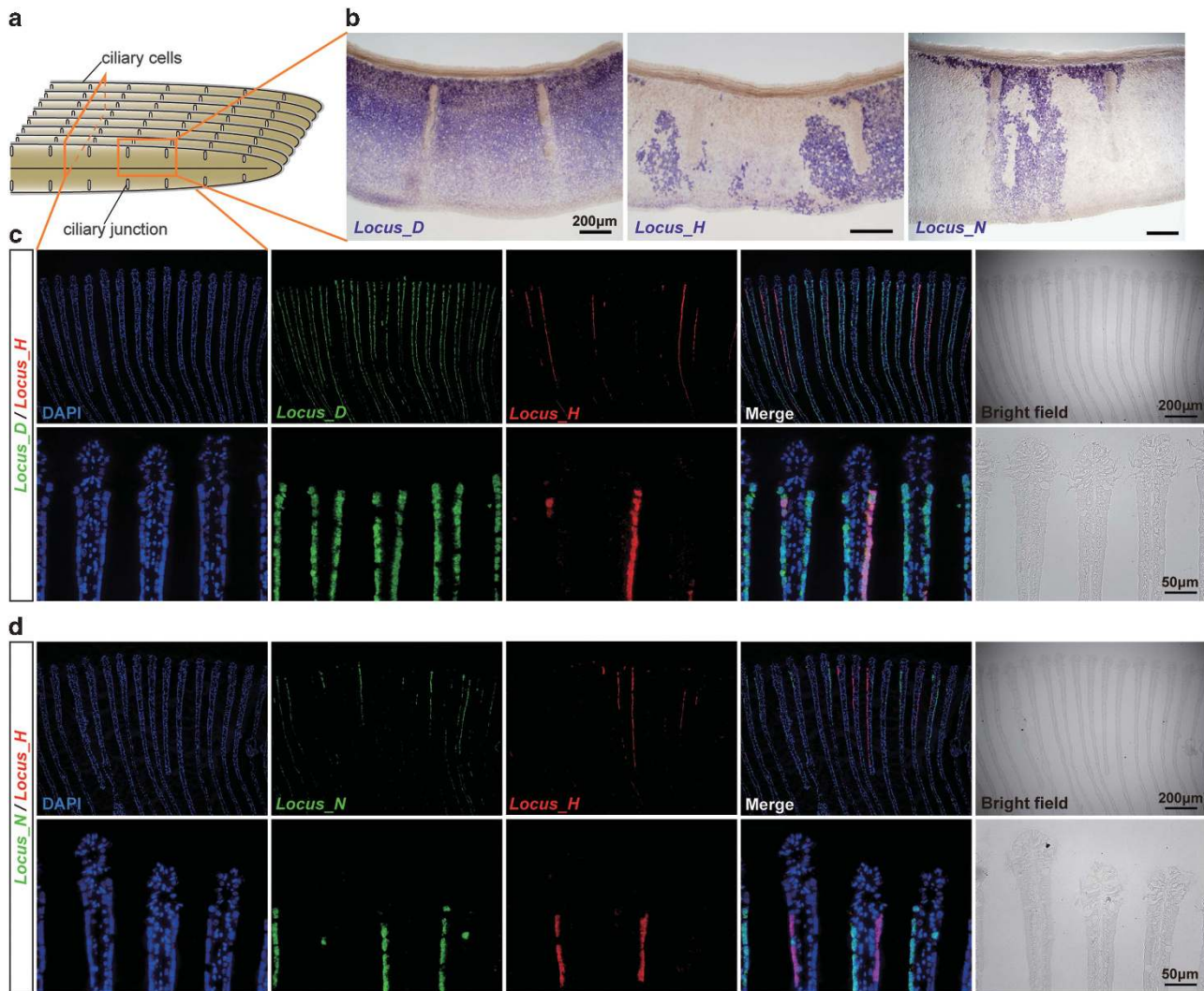
Figure 4 Distribution of symbiont subpopulations harbouring *Locus_H* or *Locus_N* in the gill. (**a**) Schematic drawing of the gill filaments. (**b**) Whole-mount *in situ* hybridization was performed for *Locus_D*, *Locus_H* and *Locus_N* (purple-blue signals). (**c**, **d**) Cross sections of gill filaments stained using fluorescence *in situ* hybridization and counterstained with DAPI staining (blue). (**c**) Double detection of *Locus_D* (green) and *Locus_H* (red). Lower panels show higher magnification. (**d**) Double detection of *Locus_N* (green) and *Locus_H* (red). Lower panels show higher magnification. For all panels in (**b–d**), frontal is to the top.

with that of *Locus_D* were comparable to the results obtained by our qPCR (Figure 3 and Supplementary Figure S3A, B). Double staining of *Locus_H* and *Locus_N* showed that symbiont subpopulations containing these loci did not make up the entire population of the gill symbiont (Figure 4d). Therefore, symbiotic bacteria that lack both *Locus_H* and *Locus_N* are common throughout the bacteriocytes.

We conclude that each *B. septemdierum* individual harbours heterogeneous subpopulations of the sulphur-oxidizing symbiont with different sets of genes encoding key metabolic enzymes such as Hup and Nar. There are at least four subpopulations: subpopulation I with *hup* genes but without *nar* genes, subpopulation II with *nar* genes but without *hup* genes, subpopulation III without either gene and subpopulation IV with both genes (Figure 5). In addition, the finding that some other loci had

low-mapping depths suggests the presence of other subpopulations with different genetic compositions (Supplementary Table S3 and Supplementary Figure S4). Thus, the genome sequence of BSEPE we report here is a pan-genome from a number of subpopulations with a single ribotype.

Coexisting genomic variants corresponding to the symbiont subpopulations in the surrounding seawater
If *B. septemdierum* acquires its symbionts from the environment, similarly to other *Bathymodiolus* species (Won *et al.*, 2003), the heterogeneous symbiont subpopulations within an individual host may reflect coexisting genomic variants corresponding to subpopulations I–IV (or others) in the habitat. We performed a PCR analysis of seawater near a *Bathymodiolus* mussel colony and detected DNA fragments that either possessed or lacked the symbiont's
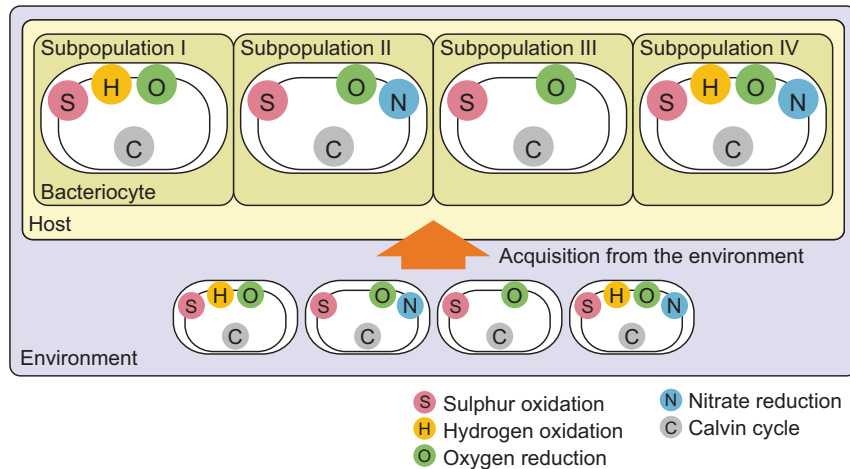
Figure 5 Model for production of heterogeneous symbiont subpopulations in a host individual. Genomic variants of the symbiont corresponding to subpopulations I–IV (or more) exist in the environment and are acquired from the environment in a host individual and localized in patches in the gill by unknown mechanisms.

hup and *nar* gene clusters (Figure 2). Therefore, symbiont variations corresponding to subpopulations I–IV likely exist in the surrounding seawater. The amount of DNA recovered from seawater near the *Bathymodiolus* mussel colony was too small to conduct qPCR using it; therefore, the relative proportion of the symbiont variations in the environment remained unclear.

Discussion

Possible scenarios for producing the heterogeneous subpopulations

Here we have shown that a symbiont population was composed of several subpopulations possessing one of the heterogeneous genomes, each of which had or lacked gene clusters for key metabolic enzymes such as hydrogenase and nitrate reductase. Petersen et al. (2011) reported that hydrogen could be used as an energy source by thiotrophic symbionts of *Bathymodiolus* mussels from the Mid-Atlantic Ridge. Nitrate respiration has been reported in some symbiotic bacteria (Hentschel and Felbeck, 1993; Hentschel *et al.*, 1996), and it has been recently shown that nitrate reductase is also used for nitrate assimilation by chemoautotrophic symbioses (Liao *et al.*, 2014).

The different symbiont variants might have been produced by the gain or loss of gene clusters in specific symbiont populations. Recently, Kleiner et al. (2012) proposed that the ancestor of *Bathymodiolus* symbionts acquired *hupL* through lateral gene transfer. Our phylogenetic analysis of *hupL* gene products supports this idea. The *hupL* gene might have been laterally transferred into the lineage of the clade that includes *Bathymodiolus* symbionts and SUP05 (Supplementary Figure S5A). As the genomic synteny of the *hup* gene clusters was relatively well conserved in this clade (Supplementary Figure S6A), *hup* genes probably

moved intergenomically as a cluster. The most parsimonious explanation for the phylogenetic relationship between the core genomes of chemoautotrophic sulphur-oxidizing symbiotic and free-living bacteria and their hydrogen oxidation cluster (Supplementary Figure S6B) is that the shared ancestor of the *Bathymodiolus* symbionts and free-living SUP05, as well as the *Calyptogena* symbiont, already harboured the *hup* gene cluster, which might have been gained by lateral gene transfer into the ancestral lineage. The *hup* gene cluster was not found in one metagenomic datum of SUP05 (Walsh *et al.*, 2009) or *Calyptogena* symbiont lineages (Kuwahara *et al.*, 2007; Newton *et al.*, 2007). These situations can be accounted for by loss of *hup* genes from these lineages. Similarly, the *hup* gene cluster was probably lost from some BSEPE populations. Thus, at least for *hup* genes, the genomic variations in BSEPE could have been produced by the loss of gene clusters in specific symbiont populations. In contrast to HupL, NarG from BSEPE, a SUP05 and a *Calyptogena* symbiont did not form a monophyletic cluster (Supplementary Figure S5B). However, because only three NarG sequences are currently available in these taxa and their close relatives, the evolutionary history of *nar* genes is difficult to estimate in the lineage into BSEPE from this limited data set.

Our PCR data indicate that genomic variants of the symbiont also exist in the environment (Figure 2). Then, the genomic variants of the symbiont in the environment might be horizontally transmitted in a host individual likely being unable to discriminate among the multiple metabolic subtypes (Figure 5). In our qPCR, the proportion of the symbiont subpopulation harbouring *hup* or *nar* genes in juveniles (from Suiyo Seamount) was higher than that in adult specimens (from Myojin Knoll and Suiyo Seamount) (Figure 3 and Supplementary Table S1). Also, the qPCR results indicated that the proportion of the

symbiont subpopulation harbouring *nar* genes in specimens from Suiyo Seamount (including adults and juveniles) was higher than that in specimens from Myojin Knoll (adults only) (Figure 3 and Supplementary Table S1). These results suggest that the composition of symbiont subpopulations varies depending on the environment (difference within a vent site over time or between two different vent sites) or host developmental stage. Symbiont localization in *Bathymodiolus* mussels has been reported to shift from a wide range of epithelia in early life stages (⩽9 mm) to only the gills in later life stages (Wentrup *et al.*, 2013). Therefore, certain subpopulations may be selected in the process of establishing a symbiotic relationship in the gill epithelium during development. Alternatively, multiple infections of the symbiont over time may affect the composition of symbiont subpopulations. The cause of the variations in the composition of symbiont subpopulations remains to be determined.

It is also not clear how the patchy distribution patterns of subpopulations are produced, and each bacteriocyte contains one or more symbiont subpopulations. Further investigations into the infection and distribution of symbiont subpopulations throughout host ontogeny, the proliferation of bacteriocytes and symbionts, and single-cell genome sequencing will improve our understanding of these questions.

Genome heterogeneity may be beneficial for utilizing diverse energy substrates
The lack of hydrogenase genes may be associated with the geochemical and geographical characteristics of the studied fields. Three decades of investigation into global hydrothermal fields have revealed a clear relationship between H_2 concentration of high-temperature hydrothermal fluid and tectonic background (Nakamura and Takai, 2014). The H_2 concentration of the fluid venting at the Western Pacific Arc/Bac-Arc setting, including the Izu-Ogasawara Arc, is generally one or two orders of magnitude lower than at the Mid-Ocean Ridge setting, including the Mid-Atlantic Ridge and Eastern Pacific Rise, where the ability to use hydrogen as an energy source in symbioses is proposed to be widespread (Petersen *et al.*, 2011). Moreover, compilation of calculations of bioavailable energy based on fluid chemistry and vent-endemic microbiological investigations (Nakamura and Takai, 2014) indicates that the structure of chemolithotrophic microbial communities in hydrothermal environments is controlled primarily by the concentration of H_2 in the fluid. When the H_2 concentration is low, thiotrophic metabolism is favoured, whereas hydrogenotrophic metabolism is favoured when H_2 is high. Thus, in the Western Pacific region, where sulphur oxidation is more preferable, the BSEPE is likely to have a tendency to lose the hydrogenase. The loss of non-essential genes accompanying a size reduction in the bacterial genome is often attributed to minimization of the material costs of cellular replication, and to the genetic drift (Mira *et al.*, 2001). In experimental bacterial populations, it has been proposed that loss of specific gene(s) is beneficial and driven by selection (Lee and Marx, 2012).

One drawback to the loss of genes that encode non-essential metabolic pathways is reduced flexibility in occupying variable geochemical regimes. In the BSEPE genome, all of the genes necessary for thiotrophy and aerobic respiration occur on loci with average mapping depth, and thus are present in all symbiont subpopulations, whereas genes for hydrogen oxidation and nitrate respiration are not essential for autotrophic growth. However, in the expression analysis, two structural *hup* genes (*hupL* and -*S*) and four structural *nar* genes (*narG, -H, -I,* and -*J*) were transcribed in the host gill, suggesting that these genes are functional in the symbiosis (Supplementary Figure S3C and Supplementary Table S1), although actual activities of the enzymes need to be assayed to draw conclusion. In the hydrothermal vent environment occupied by *Bathymodiolus* mussels, sulphide and dissolved oxygen regimes appear to be unstable (Zielinski *et al.*, 2011); therefore, the retention of variation for energy acquisition potentially increases the opportunities to colonize different and variable geochemical conditions. Thus, *hup* (and possibly *nar*) genes may be conserved in specific subpopulations of BSEPE with functional activity, based on a subtle balance between the tendency to lose unnecessary genes and the tendency to retain flexibility to occupy a variable environment.

The genome heterogeneity in BSEPE may permit the host-symbiont association to utilize diverse metabolic substrates. In ocean microorganisms, metabolic capability is deeply correlated with the organism's local acclimation and niche acquisition (Kashtan *et al.*, 2014). It has been proposed that some host animals have physiologically distinct multiple endosymbiont species allowing them to use a variety of energy sources, which may confer metabolic flexibility and enable the organism to occupy a range of habitats (Fisher *et al.*, 1993; Distel *et al.*, 1995; Woyke *et al.*, 2006; Dubilier *et al.*, 2008; Robidart *et al.*, 2008; Beinart *et al.*, 2012). The genomic heterogeneity at the sub-species level we found here may also enable differential utilization of diverse substrates, although this model remains to be validated. In horizontally transmitted endosymbioses, the acquisition of symbionts from a (potentially) diverse free-living population is expected to result in multiple symbiont subtypes coexisting within a single host individual (Stewart and Cavanaugh, 2009). In line with this expectation, intra-species subtype variations in a host have been reported in some horizontally transmitted bacterial endosymbionts (DeChaine *et al.*, 2006; Vrijenhoek *et al.*, 2007). However, until recently, little was known about the ecological significance of the genomic diversity within a symbiotic population.

The findings presented here advance our understanding of metabolic acclimation and genomic evolution in symbiotic bacteria.

Conclusions

Here, we have shown that a symbiont population with a single ribotype in an individual *B. septemdierum* host is composed of several heterogeneous subpopulations that differ in gene sets for key metabolic enzymes. Recently, genomic diversity in single nucleotide polymorphisms, short insertions/deletions and structural variants within microbial species has been also discovered in free-living populations and human gut microbes (Gonzaga *et al.*, 2012; Grote *et al.*, 2012; Schloissnig *et al.*, 2013; Kashtan *et al.*, 2014). Our findings shed light on the ecological significance of the genomic diversity not only within symbiotic bacterial species, but also natural free-living or other animal-associated microbial communities.

Conflict of Interest

The authors declare no conflict of interest.

References

Anantharaman K, Breier JA, Sheik CS, Dick GJ. (2013). Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. *Proc Natl Acad Sci USA* **110**: 330–335.

Beinart RA, Sanders JG, Faure B, Sylva SP, Lee RW, Becker EL *et al.* (2012). Evidence for the role of endosymbionts in regional-scale habitat partitioning by hydrothermal vent symbioses. *Proc Natl Acad Sci USA* **109**: E3241–E3250.

Blasco F, Iobbi C, Ratouchniak J, Bonnefoy V, Chippaux M. (1990). Nitrate reductases of Escherichia coli: sequence of the second nitrate reductase and comparison with that encoded by the narGHJI operon. *Mol Gen Genet* **222**: 104–111.

Brito B, Baginsky C, Palacios JM, Cabrera E, Ruiz-Argueso T, Imperial J. (2005). Biodiversity of uptake hydrogenase systems from legume endosymbiotic bacteria. *Biochem Soc Trans* **33**: 33–35.

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M *et al.* (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* **55**: 611–622.

DeChaine EG, Bates AE, Shank TM, Cavanaugh CM. (2006). Off-axis symbiosis found: characterization and biogeography of bacterial symbionts of *Bathymodiolus* mussels from Lost City hydrothermal vents. *Environ Microbiol* **8**: 1902–1912.

Distel DL, Lee HK, Cavanaugh CM. (1995). Intracellular coexistence of methano- and thioautotrophic bacteria in a hydrothermal vent mussel. *Proc Natl Acad Sci USA* **92**: 9598–9602.

Dubilier N, Bergin C, Lott C. (2008). Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat Rev Microbiol* **6**: 725–740.

Duperron S, Bergin C, Zielinski F, Blazejak A, Pernthaler A, McKiness ZP *et al.* (2006). A dual symbiosis shared by two mussel species, *Bathymodiolus azoricus* and *Bathymodiolus puteoserpentis* (Bivalvia: Mytilidae), from hydrothermal vents along the northern Mid-Atlantic Ridge. *Environ Microbiol* **8**: 1441–1447.

Fisher CR, Brooks JM, Vodenichar JS, Zande M, Childress JJ, Burke RA Jr. (1993). The co-occurrence of methanotrophic and chemoautotrophic sulfur-oxidizing bacterial symbionts in a deep-sea mussel. *Mar Ecol* **14**: 277–289.

Fujiwara Y, Takai K, Uematsu K, Tsuchida S, Hunt JC, Hashimoto J. (2000). Phylogenetic characterization of endosymbionts in three hydrothermal vent mussels: influence on host distributions. *Mar Ecol Prog Ser* **208**: 147–155.

Gonzaga A, Martin-Cuadrado AB, Lopez-Perez M, Megumi Mizuno C, Garcia-Heredia I, Kimes NE *et al.* (2012). Polyclonality of concurrent natural populations of Alteromonas macleodii. *Genome Biol Evol* **4**: 1360–1374.

Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ *et al.* (2012). Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *mBio* **3**: e00252–12.

Hansel A, Axelsson R, Lindberg P, Troshina OY, Wunschiers R, Lindblad P. (2001). Cloning and characterisation of a hyp gene cluster in the filamentous cyanobacterium *Nostoc* sp. strain PCC 73102. *FEMS Microbiol Lett* **201**: 59–64.

Hartig E, Schiek U, Vollack KU, Zumft WG. (1999). Nitrate and nitrite control of respiratory nitrate reduction in denitrifying *Pseudomonas stutzeri* by a two-component regulatory system homologous to NarXL of Escherichia coli. *J Bacteriol* **181**: 3658–3665.

Hentschel U, Felbeck H. (1993). Nitrate respiration in the hydrothermal vent tubeworm *Riftia pachyptila*. *Nature* **366**: 338–340.

Hentschel U, Hand S, Felbeck H. (1996). The contribution of nitrate respiration to the energy budget of the symbiont-containing clam *Lucinoma aequizonata*: a calorimetric study. *J Exp Biol* **199**: 427–433.

Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A *et al.* (2014). Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* **344**: 416–420.

Kleiner M, Petersen JM, Dubilier N. (2012). Convergent and divergent evolution of metabolism in sulfur-oxidizing symbionts and the role of horizontal gene transfer. *Curr Opin Microbiol* **15**: 621–631.

Kuwahara H, Yoshida T, Takaki Y, Shimamura S, Nishi S, Harada M *et al.* (2007). Reduced genome of the

thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*. *Curr Biol* **17**: 881–886.

Lee MC, Marx CJ. (2012). Repeated, selection-driven genome reduction of accessory genes in experimental populations. *PLoS Genet* **8**: e1002651.

Liao L, Wankel SD, Wu M, Cavanaugh CM, Girguis PR. (2014). Characterizing the plasticity of nitrogen metabolism by the host and symbionts of the hydrothermal vent chemoautotrophic symbioses *Ridgeia piscesae*. *Mol Ecol* **23**: 1544–1557.

Mira A, Ochman H, Moran NA. (2001). Deletional bias and the evolution of bacterial genomes. *Trends Genet* **17**: 589–596.

Nakagawa S, Shimumura S, Takaki Y, Suzuki Y, Murakami S, Watanabe T *et al.* (2014). Allying with armored snails: the complete genome of gammaproteobacterial endosymbiont. *ISMEJ* **8**: 40–51.

Nakamura K, Takai K. (2014). Theoretical constraints of physical and chemical properties of hydrothermal fluids on variations in chemolithotrophic microbial communities in seafloor hydrothermal systems. *Prog Earth Planet Sci* **1**: 5.

Newton IL, Woyke T, Auchtung TA, Dilly GF, Dutton RJ, Fisher MC *et al.* (2007). The *Calyptogena magnifica* chemoautotrophic symbiont genome. *Science* **315**: 998–1000.

Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R *et al.* (2011). Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* **476**: 176–180.

Robidart JC, Bench SR, Feldman RA, Novoradovsky A, Podell SB, Gaasterland T *et al.* (2008). Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed through metagenomics. *Environ Microbiol* **10**: 727–737.

Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A *et al.* (2013). Genomic variation landscape of the human gut microbiome. *Nature* **493**: 45–50.

Stewart FJ, Cavanaugh CM. (2009). Pyrosequencing analysis of endosymbiont population structure: co-occurrence of divergent symbiont lineages in a single vesicomyid host clam. *Environ Microbiol* **11**: 2136–2147.

Treangen TJ, Sommer DD, Angly FE, Koren S, Pop M. (2011). Next generation sequence assembly with AMOS. *Curr Protoc Bioinformatics* **Chapter 11**: Unit 11.

Vignais PM, Billoud B. (2007). Occurrence, classification, and biological function of hydrogenases: an overview. *Chem Rev* **107**: 4206–4272.

Vrijenhoek RC, Duhaime M, Jones WJ. (2007). Subtype variation among bacterial endosymbionts of tubeworms (Annelida: Siboglinidae) from the Gulf of California. *Biol Bull* **212**: 180–184.

Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG *et al.* (2009). Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* **326**: 578–582.

Wentrup C, Wendeberg A, Huang JY, Borowski C, Dubilier N. (2013). Shift from widespread symbiont infection of host tissues to specific colonization of gills in juvenile deep-sea mussels. *ISME J* **7**: 1244–1247.

Won YJ, Hallam SJ, O'Mullan GD, Pan IL, Buck KR, Vrijenhoek RC. (2003). Environmental acquisition of thiotrophic endosymbionts by deep-sea mussels of the genus *Bathymodiolus*. *Appl Environ Microbiol* **69**: 6785–6792.

Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO *et al.* (2006). Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**: 950–955.

Zielinski FU, Gennerich HH, Borowski C, Wenzhofer F, Dubilier N. (2011). In situ measurements of hydrogen sulfide, oxygen, and temperature in diffuse fluids of an ultramafic-hosted hydrothermal vent field (Logatchev, 14 degrees 45 'N, Mid-Atlantic Ridge): implications for chemosymbiotic bathymodiolin mussels. *Geochem Geophys Geosys* **12**: Q0AE04.

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