Supplementary Online Material

Metabolic and physiological interdependencies in the *Bathymodiolus azoricus* symbiosis

Ruby Ponnudurai[1*], Manuel Kleiner[2*], Lizbeth Sayavedra[3], Jillian M. Petersen[3,4], Martin Moche[5], Andreas Otto[5], Dörte Becher[5], Takeshi Takeuchi[6], Noriyuki Satoh[6], Nicole Dubilier[3], Thomas Schweder[1,7] and Stephanie Markert[1,7]

[1]Ernst-Moritz-Arndt-University, Institute of Pharmacy, Greifswald, Germany

[2]University of Calgary, Department of Geoscience, Calgary, Canada

[3]Max-Planck-Institute for Marine Microbiology, Department of Symbiosis, Bremen, Germany

[4]University of Vienna, Division of Microbial Ecology, Vienna, Austria

[5]Ernst-Moritz-Arndt-University, Institute of Microbiology, Greifswald, Germany

[6]Okinawa Institute of Science and Technology, Marine Genomics Unit, Okinawa, Japan

[7]Institute of Marine Biotechnology, Greifswald, Germany

*These two authors contributed equally to this work.

Correspondence: stephanie.markert@uni-greifswald.de

Contents

Supplementary Methods……………………………………………………………….............2

Supplementary Results and Discussion…………………………………………………...13

References………………………………………………………………….........................27

The following tables and figures are provided as separate files:

Supplementary Table S1: A) Replicate numbers of all analyses performed in this study
B) Overview of MS measurements

Supplementary Table S2: A) List of all genomes in the *B. azoricus* protein database
B) Genomes of *Bathymodiolus* symbionts in the database
C) Number of protein identifications in MS analyses

Supplementary Table S3: A) Relative abundance of all identified proteins (average)
B) Relative abundance in NSAF % (all replicates)
C) Relative abundance in OrgNSAF % (all replicates)

Supplementary Table S4: A) Symbiosis-specific host proteins (expression ratios)
B) Symbiosis-specific host proteins (significance tests)

Supplementary Table S5: A) CRISPR-Cas and R-M genes in both symbionts
B) Genes encoding CRISPR repeats in the thiotroph

Supplementary Figure S1: Proportion of symbionts and host in enriched fractions
Supplementary Figure S2: Energy-generating pathways in *B. azoricus* symbionts
Supplementary Figure S3: Methane and formaldehyde oxidation in the methanotroph
Supplementary Figure S4: Nitrogen metabolism in the thiotrophic symbiont
Supplementary Figure S5: Genes and proteins involved in amino acid and cofactor synthesis

Supplementary Methods

Sampling sites, sampling times and replicate numbers of all analyses performed in this study are summarized in Supplementary Table S1A.

Sampling and sequencing of the methane-oxidizing symbionts from B. azoricus *and* Bathymodiolus *sp. mussels*

DNA extraction and sequencing of the methane-oxidizing (MOX) symbiont of *B. azoricus* was done as described previously by Sayavedra *et al.* and the MOX symbiont sequences were obtained by binning from the metagenome (Sayavedra *et al.*, 2015). Briefly, a *B. azoricus* mussel was collected at the Mid-Atlantic Ridge (MAR) Menez Gwen vent field at 37°45’35’’ N, 31°38’15.6’’ W during the MoMARETO cruise. DNA of the adductor muscle was extracted using a CTAB/PVP extraction procedure (2 % CTAB, 1 % PVP, 1.4 M NaCl, 0.2 % ß-mercaptoethanol, 100 mM Tris HCl pH 8, 0.1 mg ml^{-1} proteinase K). DNA was sequenced by OIST with 454-Titanium. The assembly was done with Newbler v.2.7 (454 Life Sciences Corporation) as described by Takeuchi and colleagues (Takeuchi *et al.*, 2012) resulting in 644,000 contigs of a total length of 510,449,434 bp. Sequences from the MOX symbiont were binned with Metawatt V. 1.7 (Strous et al., 2012). Only sequences longer than 800 bp were considered for further analyses.

Two *Bathymodiolus* sp. individuals were collected, one from the southern MAR Lilliput vent at 09°32'50.9” S, 13°12'33.3” W at the station ME782/335 and one at the Foggy Corner vent at 4°48'9.7" S, 12°22'16.79" W at the station ME782/274 during the RV Meteor cruise M78-2. Gill tissue of the *Bathymodiolus* sp. mussel from Lilliput was frozen at -20 °C and ground in the home laboratory with a glass tissue homogenizer. A gill sample of the *Bathymodiolus* sp. mussel from Foggy Corner was homogenized on board and stored in 70 % ethanol at -20 °C. Homogenates were centrifuged at low speed (18 min, 20 x g, 4 °C) using a swing-out rotor and the resultant supernatants were filtered through 8 µm GTTP filters (Millipore). 200 µl of the filtrate were centrifuged at high speed (5 min, 3500 x g, 4°C), and the resultant pellet resuspended in 0.1 M HCl in artificial seawater (ASW). ASW was prepared as described by

1 Widdel and Pfennig (1981), supplemented with a trace element mixture as described by Widdel
2 *et al.* (1983). The sample was incubated for 10 min at room temperature (RT), centrifuged (5
3 min, 3500 x g, 4 °C) and the resultant pellet was resuspended in 500 µl ASW followed by
4 another brief centrifugation (1 min, 3500 x g, 4 °C). To facilitate the isolation of the symbionts
5 by FACS, the 16S rRNA of the MOX symbiont was labeled with catalyzed reporter deposition-
6 fluorescence *in situ* hybridization (CARD-FISH) with the probe BMARm-845 as previously
7 described (Pernthaler *et al.*, 2002; Duperron *et al.*, 2006) with the following modifications:
8 Homogenates were hybridized for 2 h at 46 °C using hybridization buffer containing 30 %
9 formamide. Hybridized cells were washed using 1 ml pre-warmed washing-buffer (70 mM
10 NaCl, 5 mM EDTA pH 8, 20 mM Tris-HCl, 0.1 % SDS) and centrifuged (5 min, 3500 x g, RT).
11 After removing the supernatant, the samples were incubated in 1 ml pre-warmed washing-
12 buffer (15 minutes at 48 °C), followed by a centrifugation step (5 min, 3500 x g, RT). For signal
13 amplification, we incubated the resultant pellets in a solution containing 1x PBS, 0.0015 %
14 H_2O_2 and 0.33 µg ml^{-1} Alexa Fluor® 488-labeled tyramides (46 °C, 30 min). Samples were
15 washed twice using 1 ml of ASW followed by centrifugation (5 min, 3500 x g, RT). The washed
16 pellets were resuspended in 100 µl ASW. Symbiont cells were isolated by FACS (MoFlo;
17 Cytomation). DNA from the sorted methanotrophic symbiont cells was amplified by Multiple
18 Displacement Amplification (MDA) using the GenomiPhi V2 Illustra MDA Kit (GE Healthcare)
19 according to the manufacturer’s instructions. The products were screened with PCR using the
20 general bacterial primers GM3 and GM4 (Muyzer *et al.*, 1995). The 16S rRNA products of
21 MOX symbionts from both *Bathymodiolus* sp. individuals were identical, so the MDA products
22 were pooled for a second reaction of MDA amplification using the GenomiPhi HY Illustra MDA
23 Kit (GE Healthcare). DNA was sequenced at Genoscope (France) with 454-Titanium FLX.
24 1299554 reads were generated and assembled with Newbler v. 2.7 (454 Life Sciences
25 Corporation) resulting in 5123 contigs. The genome assemblies of the methanotrophic
26 symbiont of *B. azoricus* and of the methanotrophic symbiont of *Bathymodiolus* sp. were
27 submitted to NCBI and are available under the BioProject accession numbers PRJEB13769
28 and PRJEB13047, respectively. [Reviewers can access the data using the following link:

https://groupware.uni-greifswald.de/index.php?r=files/file/download&id=681780&random_code=b284JduEyMN&inline=false&security_token=jsydMmQcRkvtFLrIKfz8.]
All other symbiont and host genome sequences used in our study besides the MOX symbionts’ genomes were obtained from published genome projects (see Supplementary Table S2A and S2B and the database compilation section below).

Separation of symbiont and host fractions from gill tissue using differential pelleting followed by density-gradient rate-zonal centrifugation

Approximately 1 g of *B. azoricus* gill tissue was homogenized in 1x PBS (phosphate buffered saline: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4) with a Duall® homogenizer (tissue grind pestle and tube SZ22, Kontes Glass Company, Vineland, New Jersey). All steps were carried out at 4 °C. To remove host nuclei and large tissue fragments, the homogenate was then transferred to a 15 ml conical tube, filled up to 15 ml with 1x PBS and centrifuged (10 min, 700 x g) in a swing out rotor. To remove any remaining nuclei and tissue fragments from the resultant supernatant, the centrifugation step was repeated while the pellet (first pellet) was frozen at -80 °C. The supernatant was then transferred to a new tube and centrifuged in a fixed angle rotor (10 minutes at 15000 x g) to pellet the symbionts, mitochondria and small tissue fragments. The supernatant obtained from this step, putatively containing the cytosolic host proteins, was frozen at -80 °C. A discontinuous density gradient was assembled in 5 % steps from 5 % to 25 % (w/v) HistoDenzTM (Sigma® Saint Louis, Missouri, USA) dissolved in 1x PBS. The symbiont-containing pellet from the previous step was resuspended in 1x PBS, layered on top of this density gradient, and was centrifuged in a swing out rotor (7 min, 3000 x g). The density gradient was then disassembled into equal-sized fractions (gradient fractions). The enriched symbiont pellet in the gradient tube (gradient pellet) was frozen at -80 °C. The individual gradient fractions were washed twice with 1x PBS to remove the HistoDenz™ and frozen at -80 °C. Throughout the procedure, small subsamples were removed from the homogenates, pellets, supernatants and gradient fractions of three individual *B. azoricus* hosts

(i.e. in biological triplicates) and fixed over-night at 4 °C in 1 % formaldehyde (Fluka, Taufkirchen, Germany) for subsequent analysis of the sample composition using CARD-FISH (see Figure 2 in main text).

Assessing symbiont and host composition in density gradient fractions using CARD-FISH

To determine the relative abundance of symbionts in the homogenates, supernatants, pellets and gradient fractions, catalyzed reporter deposition-fluorescence in-situ hybridization (CARD-FISH) was used. Subsamples that were fixed over-night were washed three times in 1x PBS and then stored at -20 °C in 100 µl of a 50 % 1x PBS and 50 % ethanol mix. Prior to CARD-FISH, the cells were filtered onto GTTP polycarbonate filters with a pore size of 0.2 µm (Millipore, Billerica, MA). CARD-FISH was done as described previously (Pernthaler *et al.*, 2002) with the minor modification that the inactivation of endogenous peroxidases was done by incubating the filter discs in 0.01 M HCl for 10 min at room temperature. Double hybridization of rRNA was performed using the probes BMARt-193 for the thiotroph and BMARm-845 for the methanotroph (Duperron *et al.*, 2006). Tyramide signal amplification was done at 37 °C for 10 min followed by inactivation of HRP using methanol. Prior to microscopic evaluation, the cells were counterstained with 1 µg ml^{-1} 4’,6-diamidino-2-phenylindole (DAPI). Microscopic analysis and imaging of the CARD-FISH filters were performed using a Zeiss Axioplan epifluorescence microscope. Cell counting was done semi-automatically using a counting device developed at the Max Planck Institute for Marine Microbiolgy, Bremen. To calculate the relative abundance of thiotrophs, methanotrophs and host cells in each sample, cells with symbiont-specific probe signal were counted against at least 500 DAPI signals per filter section. DAPI signals, which did not overlap with the symbiont probe signals, were used as host cell count. Using this data, two enriched sample types were chosen for further proteomic analysis: the supernatant, containing soluble host proteins, and the gradient pellet, containing enriched thiotrophic and methanotrophic symbionts (Supplementary Figure S1).

1 *Protein extraction from enriched and whole tissue samples*

2 For the proteomic analyses, two enriched sample types were processed from three animals,
3 i.e. in three biological replicates: 1) the symbiont-enriched gradient pellet (containing both
4 symbionts, but particularly enriched in the thiotroph), and 2) the host supernatant, containing
5 soluble host proteins. Additionally, 3) whole gill tissue and 4) whole mussel foot tissue samples
6 were processed from the first two animals, i.e. in biological duplicates (see Supplementary
7 Table S1B for replicate numbers of individual MS measurements). For protein extraction,
8 gradient pellets were washed and resuspended in lysis buffer (10 mM Tris pH 7.5, 10 mM
9 EDTA pH 8.0, containing Roche-Complete Mini protease inhibitor cocktail). Tissue samples
10 were homogenized using a Duall® homogenizer and the resultant homogenate was transferred
11 to low-binding 1.7 ml reaction tubes (Sorenson BioScience Inc., Salt Lake City, UT, USA). All
12 samples, including the supernatant fractions, were then subjected to sonication using a
13 sonicator (Bandelin Sonopuls ultrasonic homogenizer, Berlin, Germany; 2 x 25 sec at 30 %
14 power and a cycle of 0.5 sec and with a 30 sec pause, 4 °C). The lysate was centrifuged to
15 remove crude cell debris (10 min, 4 °C, 15,300 x g) and the resultant protein raw extracts were
16 further subjected to ultracentrifugation (60 min, 4 °C, 100,000 x g) allowing for the enrichment
17 of membranes and membrane-associated proteins in the pellets and of soluble proteins in the
18 supernatants. Protein concentrations were determined using the method described by
19 Bradford *et al.* (1976). Aliquots of all four sample types were stored at -80 °C until MS analysis.
20 The enriched membrane fractions of gill samples and of gradient pellet samples obtained after
21 ultracentrifugation were further purified according to the protocol of Eymann *et al.* (2004). The
22 purified membrane protein extracts (solubilized in 30 µl of 50 mM triethylammonium
23 bicarbonate buffer, pH 7.8) were transferred to low binding 1.7 ml tubes and immediately used
24 for MS analysis. Due to sample scarcity, the biological replicates (n = 3 for the gradient pellets,
25 n = 2 for the gill samples) for these membrane protein samples were pooled together to obtain
26 enough protein for MS analysis.

27

1 *1D-PAGE LC-MS/MS*

2 Proteins from all samples (both membrane and soluble fractions) were separated using 1D-
3 PAGE. Approximately 20 µg of protein from all membrane- and soluble protein extracts were
4 dissolved in sample loading buffer (100 mM Tris-HCl pH 6.8, 10 % SDS, 20 % glycerol, 5 %
5 ß-mercaptoethanol, bromophenol blue) and separated on pre-cast 10 % polyacrylamide gels
6 (Bio-Rad). The gels were fixed and stained with Coomassie Brilliant Blue (G250, Sigma
7 Aldrich). Each sample lane was cut into 10 equal-sized pieces, thoroughly destained (37 °C,
8 200 rpm, 200 mM ammonium hydrogen carbonate, 30 % acetonitrile) and briefly vacuum-dried.
9 Proteins were in-gel digested overnight with trypsin (sequencing grade, Promega) at 37 °C.
10 Subsequently, peptides were eluted in water using an ultrasonic bath, and purified using
11 ZipTips (P10, U-C_{18}, Millipore). Peptides were separated by liquid chromatography (LC) on an
12 EASYnLC (Proxeon, Odense, Denmark) with self-packed columns (Luna 3µC18(2) 100A,
13 Phenomenex, Germany) in a one-column setup. Following loading at a flow rate of 700 nl/min
14 at a maximum of 220 bar of water in 0.1 % acetic acid, separation of peptides was achieved
15 by the application of a binary non-linear 70 min gradient from 5 % - 50 % acetonitrile in 0.1 %
16 acetic acid at a flow rate of 300 nl/min. The eluting peptides were simultaneously analyzed by
17 coupling the LC online to a mass spectrometer. The soluble protein extract of gradient pellet
18 and supernatant were each analyzed in two separate measurements a) on an LTQ-Orbitrap
19 Velos mass spectrometer (Thermo Fisher, Bremen, Germany) and also b) on an LTQ-Orbitrap
20 Classic mass spectrometer (Thermo Fisher, Bremen, Germany). Soluble protein extracts from
21 gill and foot tissue samples were analyzed only on the LTQ-Orbitrap Velos mass spectrometer.
22 All membrane protein extracts were run on the LTQ-Orbitrap Classic mass spectrometer.
23 During MS measurements, a spray voltage of 2.4 kV was applied in all cases. After a survey
24 scan (R = 30,000) MS/MS data were recorded for the twenty most intensive precursor ions in
25 the linear ion trap. Singly charged ions were not taken into account for MS/MS analysis. Lock
26 mass option was enabled throughout all analyses. After mass spectrometric measurements,
27 MS data were converted into mzXML format by msconvert (Chambers *et al.*, 2012) followed

by database searching. All proteomics raw data are available from the PRIDE ProteomeXchange repository (see accession details in the database compilation section below).

Protein identification, validation and quantitation

For identification of host and symbiont proteins, all MS/MS spectra were searched against a comprehensive protein database (see below for details on database compilation and Supplementary Table S2A for a list of all organisms whose protein sequences are included) on a SORCERER™ system (SageN, Milpitas, CA, USA) using the integrated SEQUEST algorithm (Thermo Fisher Scientific, San Jose, CA, USA; version 27, revision 11) without charge state deconvolution and de-isotoping. The search was done using the following parameters: parent mass tolerance 0.0065 Da; fragment ion tolerance 1 Da; up to 2 missed cleavages allowed (internal lysine and arginine residues), fully tryptic peptides only, and oxidation of methionine as variable modification (+ 15.99 Da). Scaffold (version 4.0.6.1, Proteome Software Inc., Portland, OR) was used for filtering and analysis of protein identifications. The protein identifications were filtered on the basis of their SEQUEST scores implemented in Scaffold (Keller *et al.*, 2002) using the following thresholds: XCorrs of at least 2.5 (charge +2) and 3.5 (charges > +2), DeltaCN > 0.08 and at least two unique peptides for each identified protein. The total protein level false discovery rate (FDR) across all samples, determined according to the method of Käll *et al.* (2008) using the number of identified decoy sequences that passed the filtering thresholds, was 0.66 %. For relative semi-quantitative analysis of proteins, normalized spectral abundance factor (NSAF) values were calculated for each sample (Florens *et al.*, 2006). NSAFs were subsequently also normalized for each organism (OrgNSAF) individually (Mueller *et al.*, 2010). The respective NSAF values are presented as percentages of all proteins in the same sample (NSAF %), or of all proteins of one specific organism in the same sample (OrgNSAF %), respectively. According to their NSAF % values, proteins of host, thiotrophic symbiont and methanotrophic symbiont were

classified as "high", "medium" and "low abundant" (see below for details on protein abundance threshold).

Compilation of the protein database

To create a comprehensive protein database, we included all available sequences pertaining to host and symbiont, a detailed list of which is given in Supplementary Table S2A. It included all ORFs from the recently published genome assemblies of the thiotrophic symbionts of *B. azoricus* from Menez Gwen and of *Bathymodiolus* sp. from South-MAR (Sayavedra *et al.*, 2015). Furthermore, all proteins from the newly sequenced genomes of the methanotrophic symbionts from the same hosts and sampling sites (see above) were included (Supplementary Table S2B). *B. azoricus* host protein sequences were procured from *B. azoricus* EST libraries (Bettencourt *et al.*, 2010) in the publicly accessible DeepSeaVent database (http://transcriptomics.biocant.pt/deepSeaVent/). Additionally, protein sequences of phylogenetically related free-living and symbiotic sulfur-oxidizing Gammaproteobacteria, free-living methane-oxidizing Gammaproteobacteria and of bivalves related to *B. azoricus* were retrieved from NCBI and JGI and added to the database. Redundant sequences were removed using the CD-hit-2D clustering program at 100 % clustering threshold (Li *et al.*, 2001). Amino acid sequences of common laboratory contaminants (ftp://ftp.thegpm.org/fasta/cRAP) were also added to the database. To determine false discovery rates (FDR) the sequences were reversed using the decoy.pl perl script (http://www.matrixscience.com/help/decoy_help.html) and appended to the database as decoy sequences. The protein sequence database and all proteome raw data were uploaded to the ProteomeXchange Consortium via the PRIDE (Vizcaíno *et al.*, 2016) partner repository with the dataset identifier PXD004061 (DOI: 10.6019/PXD004061). [Reviewer access: log in at http://www.ebi.ac.uk/pride/archive/ with Username: reviewer26938@ebi.ac.uk and Password: bRkDL3tp].

1 *Determination of NSAF protein abundance cut-offs*

2 Threshold values for high, medium and low abundance of proteins were calculated for host,
3 thiotroph and methanotroph individually from their respective OrgNSAF % (normalized spectral
4 abundance factor for individual organisms). For host abundance cut-offs, OrgNSAF % values
5 from biological replicates of the supernatant samples were averaged, whereas for the
6 symbionts, OrgNSAF % values from biological replicates of the gradient pellet samples were
7 averaged. Protein tables were then sorted in decreasing order based on OrgNSAF % averages
8 and the cumulative OrgNSAF % average was calculate by adding the average OrgNSAF % of
9 each line to the cumulative average of the previous line. The cumulative OrgNSAF % average
10 was plotted and cumulative average values corresponding to two elbow points in these curves
11 were determined. Proteins corresponding to cumulative mean values greater than the first
12 elbow were designated high abundant, proteins with values between the two elbows were
13 classified as medium abundant and such proteins with values lower than the second elbow
14 were designated low abundant.

15 *Statistical testing for significant expression differences of putative symbiosis-specific proteins*

16 To identify potentially symbiosis-relevant host and symbiont proteins we performed a statistical
17 comparison of protein abundances in different sample types (see main text Methods section
18 for details). Expression ratios were calculated from OrgNSAF % values using the average
19 across all replicates for the respective sample types (supernatant and gradient pellet: 3 Orbi
20 replicates and 2 Velos replicates = 5 replicates total, foot and gill samples: 2 Velos replicates,
21 see also Supplementary Table S1B; membrane fractions were excluded). For statistical
22 analyses of differences between samples, tables with NSAF or OrgNSAF values of all
23 replicates and MS analyses were loaded into the Perseus software (version 1.4.1.3,
24 http://www.perseus-framework.org/doku.php). Proteins that did not have at least two or four
25 (depending on comparison) expression values >0 were removed from the table. All remaining
26 values were multiplied by 10,000 and log_2-transformed. Missing values produced by $log_2(0)$

were replaced by sampling from a normal distribution assuming that the missing values are on the lower end of abundance (normal distribution parameters in Perseus: width 0.3, down shift 1.8, do separately for each column). A t-test with permutation-based false discovery rate (FDR) calculation was used to detect proteins that differed significantly in their expression level between two sample types. The statistical method implemented in Perseus that we used is based on the “significance analysis of microarrays” described by Tusher *et al.* (2001) which by using a permutation-based FDR accounts for the multiple-testing problem inherent in testing for significant expression differences for a large number of genes. The following parameters were used for the test: sample groupings were only preserved for technical replicates during randomizations, both sides, 250 randomizations, FDR of 5 % and a variable s0 parameter of 0 to 3.

Prediction of protein properties and assignment of metabolic functions

Each identified protein was checked for the presence of signal peptides and transmembrane helices using the SignalP (http://www.cbs.dtu.dk/services/SignalP/, Petersen *et al.*, 2011b) and TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) servers (Moller *et al.*, 2001). The PsortB online server (http://www.psort.org/psortb/) was used to predict the subcellular localization of each protein (Yu *et al.*, 2010). Proteins for which signal peptides were not predicted were also checked for non-classical secretion using the SecretomeP online server (http://www.cbs.dtu.dk/services/SecretomeP/, Bendtsen *et al.*, 2005). Assignment of metabolic functions for each protein was done by sequence-based comparison using blastp (http://www.ncbi.nlm.nih.gov/Blast/) against the NCBI (nr) database and against the UniProt knowledgebase (http://web.expasy.org/blast/, UniProt-Consortium, 2015). For hypothetical proteins and proteins with inadequate annotation, their likely functions were deduced using hidden markov model-based searches, such as PFAM (http://pfam.xfam.org/, Finn *et al.*, 2014) and HHpred (http://toolkit.tuebingen.mpg.de/hhpred, Biegert *et al.*, 2006) and on the basis of their genomic context using RAST SEED viewer 2.0 (http://rast.nmpdr.org/, Aziz *et al.*, 2008; Overbeek *et al.*, 2014).

Reconstruction of metabolic pathways

The annotated proteins were sorted into groups according to their metabolic functions. Proteins from each group were then mapped manually into pathways using classical metabolic maps from databases such as Metacyc (Caspi *et al.*, 2014) and KEGG (Kanehisa *et al.*, 2014) as templates. In cases where not all proteins of a pathway were identified in the proteome, we checked for the corresponding genes of missing proteins in the respective genomes using RAST SEED viewer. All metabolic maps were redrawn using Inkscape version 0.48. Graphical representation of data was performed using GraphPad Prism version 5.00 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com (http://www.graphpad.com/scientific-software/prism/).

1 ***Supplementary Results and Discussion***

2 We identified a total of 3128 proteins from *B. azoricus* and its symbionts (Supplementary Table
3 S2C). The draft genome assemblies of the *B. azoricus* thiotrophic symbiont yielded 717 protein
4 identifications (~25% of all predicted ORFs, Supplementary Table S2C). An additional 183
5 proteins were identified with genomes of thiotrophic symbionts of other *Bathymodiolus* species
6 and phylogenetically related sulfur-oxidizing Gammaproteobacteria. For the methanotroph we
7 identified 344 proteins (Supplementary Table S2C). The lower number of identified
8 methanotroph proteins is probably due to (i) the fragmented state of the current genome
9 assembly and (ii) the inherently lower numbers of methanotroph cells in *B. azoricus*. There is
10 currently no genome sequence available for the host. Nevertheless, we could identify 1884
11 proteins using sequences from a *B. azoricus* EST library and of other related bivalve molluscs
12 (Supplementary Table S2C). Based on these protein identifications and their relative
13 abundances (in NSAF %) we were able to draw a detailed picture of the metabolic network
14 within the *B. azoricus* consortium, which is described in the main text and in the following
15 sections.

16 *Energy metabolism in* B. azoricus *symbionts*

17 *Enzymes involved in the oxidation of sulfur substrates in the thiotrophic* B. azoricus *symbiont*:
18 Energy generation through the oxidation of thiosulfate, which appears to play a prominent role
19 in the thiotrophic *B. azoricus* symbiont (see main text), involves the periplasmic sulfur oxidation
20 (Sox) multienzyme complex. As also proposed for other gammaproteobacterial sulfur-oxidizing
21 symbionts (Harada *et al.*, 2009; Markert *et al.*, 2011; Kleiner *et al.*, 2012b), the *B. azoricus*
22 thiotroph's Sox complex, which oxidizes thiosulfate to form sulfate, lacks a SoxCD homolog.
23 Although a putative SoxD transcript of the *B. azoricus* thiotroph was identified in a gill
24 transcriptome (Egas *et al.*, 2012), no SoxCD homolog was detected in the genome or proteome
25 in this study. SoxCD is a sulfur dehydrogenase and in its absence, the sulfane sulfur bound to
26 SoxY cannot be oxidized further and may instead be transferred to growing periplasmic sulfur

intermediates (Sauve *et al.*, 2007). Apart from that, all components of the Sox multienzyme complex, SoxXYZAB, were identified in the proteome (Supplementary Figure S2A, Supplementary Table S3). The thiotroph’s SoxY (BazSymB_scaffold00004_21) and SoxB (BazSymA_Acontig01323_2) were amongst the most abundant proteins identified from membranes of gradient pellet and gill samples (Figure 3).

Unlike other sulfur oxidizers, the thiotrophic *B. azoricus* symbiont does not possess proteins for concentration and storage of sulfur (“sulfur globule proteins”), as confirmed by our study. However, orthorhombic precipitates of elemental sulfur have been found in the periplasm of the *B. azoricus* thiotroph (Pruski *et al.*, 2002), likely as intermediates of Sox-mediated thiosulfate oxidation and/or resulting from polysulfide formation by Sqr (see below and Supplementary Figure S2).

Oxidation of sulfide, although possibly less prominent in the thiotrophic *B. azoricus* symbiont, is initiated by the periplasmic oxidation of sulfide to polysulfides by the membrane-bound sulfur:quinone reductase (Sqr). We found three non-identical homologs of Sqr encoded in separate locations in the thiotroph’s genome. All three homologs were detected in the proteome (Supplementary Table S3, BazSymA_Acontig140454_0, BazSymA_V2Acontig295479_0, BazSymA_Acontig02297_1). Whether these Sqr proteins differ in their function is unclear; however, simultaneous expression of multiple Sqr proteins may aid in efficient removal of sulfide, protecting the host from its toxicity (Shibata *et al.*, 2001). After initial periplasmic thiosulfate or sulfide oxidation, the next steps of sulfide oxidation to sulfate proceed in the cytoplasm through the rDSR-APS-Sat pathway of gammaproteobacterial sulfur-oxidizing symbionts (Markert *et al.*, 2011; Kleiner *et al.*, 2012a; Kleiner *et al.*, 2012b). Most components of the reverse dissimilatory sulfite reductase (rDSR) complexes (DsrAB, DsrC, DsrEH, DsrL, DsrKMO), the adenylylsulfate reductase (APS) complex (AprAB) and sulfate adenylyltransferase (Sat) were identified in the thiotrophic symbiont’s proteome (Supplementary Figure S2A, Supplementary Table S3). Two sulfate transporters (SulP, BAZSYMA_Acontig02997_2, BAZSYMA_Acontig01342_0) are encoded in the genome, but

1 neither was detected in the proteome, likely due to their extensive membrane integration,
2 which makes detection by proteomics difficult.

3 *Hydrogen oxidation appears to be less relevant in the* B. azoricus *thiotroph from Menez Gwen*:
4 In our study, the thiotrophic *B. azoricus* symbiont expressed the membrane-bound uptake
5 hydrogenase HupL (BazSymB_scaffold00037_2), the key enzyme involved in hydrogen
6 oxidation. An earlier study by Petersen and colleagues (2011a) reported hydrogen oxidation
7 in the sulfide-oxidizing symbiont of *B. puteoserpentis* mussels from Logatchev vent (MAR),
8 where energy from H_2 oxidation promoted carbon fixation at rates similar to sulfide oxidation.
9 HupL abundance as determined in our study was lower than that of the sulfur oxidation-related
10 enzymes, accounting for only a small fraction of the thiotroph’s proteome (0.05 % OrgNSAF of
11 the gradient pellet proteome and 0.1 % OrgNSAF of the gradient pellet membrane proteome).
12 This relatively low HupL expression likely reflects the relatively low *in situ* concentration of H_2
13 in the vent fluids surrounding *B. azoricus* at the sampling site (Menez Gwen). Previous studies
14 showed that (i) H_2 concentration in end-member fluids of Menez Gwen is considerably lower
15 (0.024/0.048 mmol/kg) compared to that from Logatchev (12 mmol/kg, Charlou *et al.*, 2002)
16 and (ii) basalt-hosted vents, such as Menez Gwen, produce fluids comparatively rich in H_2S
17 but low in H_2 (Petersen *et al.*, 2011a; Zielinski *et al.*, 2011). Furthermore, the uptake rate of H_2
18 in *Bathymodiolus* mussels from Logatchev and southern MAR vent sites was shown to be
19 influenced by the *in situ* concentrations of available H_2, i.e. the lower the availability, the lower
20 the uptake (Petersen et al 2011). Thus, we propose that although the *B. azoricus* symbiosis
21 uses hydrogen as an energy source, H_2 oxidation may not be as important as sulfur oxidation,
22 under the conditions reflected in this study.

23 *Energy-generating methane and formaldehyde oxidation in the methanotroph - pathway*
24 *details:* The methanotrophic symbiont of *B. azoricus* generates electrons from oxidation of
25 methane to formaldehyde in a two-step oxidation process: i) Methane is oxidized to methanol
26 through the copper-containing particulate methane monooxygenase (pMMO) complex

1 (BAZMOX_contig00686_3, BAZMOX_contig00686_4, BAGiLS_001099) located on its
2 intracytoplasmic membranes (see main text and Supplementary Figures S2B and S3). ii)
3 Methanol is further oxidized to formaldehyde through the periplasmic XoxF-type methanol
4 dehydrogenase (BAZMOX_contig18015_2, Fiala-Médioni *et al.*, 2002). The resulting
5 formaldehyde enters the cytoplasm, where it is channelled into carbon assimilation reactions
6 (see below) or oxidized to CO_2 for energy generation with the aid of either one of two cofactors,
7 tetrahydrofolate (H_4F) or dephospho-tetrahydromethanopterin (H_4MPT, Thauer, 1998; Maden,
8 2000). While the initial step of formaldehyde condensation with H_4F is spontaneous
9 (Supplementary Figure S3), formaldehyde binding to H_4MPT is mediated by the formaldehyde-
10 activating enzyme Fae (Vorholt, 2002). Fae was abundantly expressed (2.5 % gill OrgNSAF,
11 BAZMOX_contig158934_1) in the methanotroph’s proteome (Supplementary Table S3). In the
12 methylotrophic Alphaproteobacterium *Methylobacterium extorquens* high Fae levels allow for
13 rapid detoxification of accumulating formaldehyde, and H_4MPT-based formaldehyde oxidation
14 was therefore suggested to be the preferred energy generation pathway, while the H_4F
15 pathway might be less relevant (Vorholt *et al.*, 2000; Vorholt, 2002). Considering the high Fae
16 abundance detected in our study, the majority of the formaldehyde is likely to enter the H_4MPT-
17 linked energy-generating pathway in the *B. azoricus* methanotroph (Supplementary Figure
18 S3). H_4F-based formaldehyde oxidation might occur less frequently, i.e. in case of excess
19 formaldehyde (as also observed by Crowther *et al.*, 2008). In several subsequent enzymatic
20 conversions both the H_4F-dependent pathway and the H_4MPT pathway produce formate,
21 which is finally oxidized to CO_2 by the enzyme formate dehydrogenase (Fdh) in an energy-
22 generating reaction. 17-49 % of the consumed methane was observed to be transformed to
23 CO_2 in *B. azoricus* and *B. puteoserpentis* (reviewed in Petersen and Dubilier, 2009). It is
24 unclear from our data, however, whether formate is oxidized completely to CO_2. The Fdh
25 protein was not detected in the proteome in this study, and the respective gene was also
26 missing from the *B. azoricus* methanotroph’s genome. This may be attributed to the
27 incompleteness of our genome assembly, as an Fdh-encoding gene is present in the
28 *Bathymodiolus* sp. methanotroph genome assembly from South MAR (METH93_3). It may be

1 speculated that if the methanotroph indeed generates CO_2, as suggested by Petersen and
2 Dubilier (2009), it could be redirected to the thiotroph for carbon fixation by the methanotrophic
3 CA (BAZMOX_contig02303_3), forming an internal carbon cycling mechanism between the
4 two endosymbionts.

5 *Carbon metabolism in* B. azoricus

6 *The thiotroph uses a modified Calvin cycle for CO_2 fixation:* The thiotrophic symbiont of *B.*
7 *azoricus* possesses a variant of the classical Calvin-Benson-Bassham (CBB) cycle (Figure
8 4A), where a reversible pyrophosphate-dependent phosphofructokinase, PP_i-PFK (PfkA,
9 BazSymB_scaffold00001_50) replaces the function of two missing enzymes: fructose-1,6-
10 bisphosphatase and sedoheptulose-1,7-bisphosphatase (Kleiner *et al.*, 2012a). In some sulfur-
11 oxidizing symbionts, a pyrophosphate-energized proton pump (H^+PPA) is co-transcribed with
12 PP_i-PFK. PP_i produced by the PP_i-PFK working in reverse in the Calvin cycle can be used by
13 H^+PPA to generate a proton motive force, thus making the CBB cycle more energy-efficient
14 (Markert *et al.*, 2011; Kleiner *et al.*, 2012b). H^+PPA, however, was not found in the genome or
15 the proteome of the *B. azoricus* thiotroph, indicating that the modified Calvin cycle may not be
16 particularly energy-efficient in this symbiont. Whether the PPi-PFK offers other yet unknown
17 selective advantages to the symbiont still remains an open question.

18 *The methanotroph's RuMP pathway may not be complete:* From our data it is unclear whether
19 the methanotroph's RuMP pathway is capable of completely oxidizing formaldehyde to
20 pyruvate and acetyl-CoA, as the enzyme phosphoglycerate mutase was not identified in the
21 methanotroph's genome. However, theoretically, the *B. azoricus* methanotroph could also
22 adopt an alternative route for this using the Entner-Doudoroff pathway (Figure 4B). This
23 pathway requires two enzymes, phosphogluconate dehydratase (Edd) and phospho-2-keto-3-
24 deoxygluconate (KDPG) aldolase of which the former was identified in the proteome
25 (METH418_1, 0.007 % gill OrgNSAF).

The methanotroph may also use the serine pathway for carbon assimilation: Generally, methanotrophs assimilate carbon via the RuMP pathway (classified as type I methanotrophs), or via an alternate route, the serine pathway (type II methanotrophs), while a few others use the RuMP as major route while additionally possessing elements of the serine pathway (type X, Hanson and Hanson, 1996; Ward *et al.*, 2004; Chistoserdova and Lidstrom, 2013). In this study, we identified a few key enzymes of the serine pathway, such as a serine-glyoxylate aminotransferase (SgaA, METH759_0; identified from the sequence of the methanotrophic symbiont of *Bathymodiolus sp.* from a southern MAR vent), and a malyl-CoA lyase (MclA, BAZMOX_contig18662_0, Supplementary Figure S3). A homolog of malate thiokinase (MtkA, generating malyl-CoA and phosphoenol carboxylase) was not detected in our study but is encoded in the genome of the methanotrophic *B. azoricus* symbiont from Lucky Strike vent (L. Sayavedra, unpublished data) and may thus be present in the *B. azoricus* methanotroph after all. However, for the pathway to close, consumed glyoxylate needs to be regenerated (Anthony, 1982). Candidate enzymes such as isocitrate lyase and those of the glyoxylate-regeneration cycle, which can putatively perform this function (Korotkova *et al.*, 2002), are absent in the *B. azoricus* methanotroph. Hence, the serine pathway seems to be only partially active, as also described for *Methylococcus capsulatus* (Ward *et al.*, 2004), and cannot operate as a major pathway for formaldehyde assimilation in the methanotroph. The incompleteness of the methanotrophic symbiont’s genome assembly and the possibility that these enzymes are expressed at *in situ* levels too low to be identified through our approach hinders our understanding of the exact function of this pathway in *B. azoricus*. However, it has been suggested that the expression of some basic enzymes of the serine pathway may serve two functions: Firstly, when formaldehyde is in excess, tetrahydrofolate and tetrahydromethanopterin (see methane oxidation section) can channel carbon into the serine pathway, thus preventing accumulation of the toxic formaldehyde. Secondly, the intermediates of the serine cycle could function as precursors for cellular biosynthesis (Chistoserdova *et al.*, 2005).

1 *The methanotrophic symbiont may be able to live on multicarbon substrates*: The genome of
2 the methanotrophic *B. azoricus* symbiont encodes TRAP transporters for dicarboxylates
3 (BAZMOX_contig03539_1, BAZMOX_contig03539_2) and also a phosphoenolpyruvate-
4 dihydroxyacetone phosphotransferase system for transport of dihydroxyacetone and fructose
5 (BAZMOX_contig03539_0, METH02_0, METH531561754). The methanotroph could thus in
6 principle take up organic substrates in the absence of methane and catabolize them, given
7 that a complete TCA cycle is expressed under these conditions (as proposed in the main text).
8 While methane is usually abundantly available to the methanotrophic symbiont of *B. azoricus*,
9 rendering a complete TCA cycle unnecessary, hydrothermal vent resources fluctuate and a
10 temporary scarcity of methane is conceivable. During these periods, the methanotrophic
11 symbiont may potentially import multicarbon substrates or break down glycogen (Figure 4B;
12 see below). Moreover, a complete TCA cycle and thus the ability to survive on alternative
13 organic substrates may also be advantageous for a potential free-living stage of the
14 methanotroph’s life cycle, during which steady access to methane may not be possible.

15 B. azoricus *and its methanotrophic symbiont synthesize storage compounds*: The
16 methanotrophic symbiont of *B. azoricus* expressed enzymes of the glycogen synthesis
17 pathway such as glycogen synthase (GlgA, BAZMOX_contig07384_1) and glucose-1-
18 phosphate adenylyltransferase (GlgC, YP_004918751.1, Figure 4B). The alpha-glucan
19 branching enzyme, involved in inserting branch points in the glycogen chain, however, was not
20 detected, although it is encoded in the methanotroph’s genome (GlgB,
21 BAZMOX_contig07384_2). The methanotroph is known to synthesize polysaccharides such
22 as glycogen when there is an excess of formaldehyde (Linton and Cripps, 1978; Pieja *et al.*,
23 2011a; Pieja *et al.*, 2011b), thereby storing carbon for periods of low nutrient availability. In
24 contrast, the thiotrophic symbiont does not seem to possess the genetic machinery for the
25 biosynthesis of storage compounds, such as glycogen. A eukaryotic glycogen synthase (GlgA)
26 was identified in the proteome with a sequence of a bivalve relative of *B. azoricus*
27 (AAS93900.1). Interestingly, this host protein was only expressed in the gill (Supplementary

Table S3) and not in the foot, indicating enhanced glycogen storage activity in the symbiont-containing tissue, possibly based on symbiont digestion-related release of carbon compounds. Glycogen-like storage compounds have also been detected in the adipogranular cells of mantle connective tissue in *B. azoricus* (Lobo-da-Cunha *et al.*, 2006), suggesting that glycogen can also be a regular carbon reserve in non-symbiont-containing tissues.

Nitrogen metabolism in the B. azoricus *symbiosis*

Ammonium and nitrate are nitrogen sources for the symbiosis: Previous reports detected ammonia and nitrate in the range of 8-10 μmol l^{-1} and 0-2 μmol l^{-1}, respectively, in end-member fluids of the MAR Menez Gwen vent field (Sarradin *et al.*, 1999). Our metaproteome analysis suggests that both the *B. azoricus* host and its symbionts assimilate nitrogen from ammonia into their amino acids using the ammonia-assimilating enzymes glutamine synthetase, GlnA, (BAGiLS_000948, BazSymA_Acontig00191_2, METH59_2) and glutamate synthase, GltBD (BazSymA_Acontig00571_2, BazSymB_scaffold00012_17, BAZMOX_contig02646_0, Supplementary Table S3, Supplementary Figures S4 and S5).

The thiotrophic symbiont, already shown to possess genes for the assimilatory reduction of nitrate (Kleiner *et al.*, 2012a), also expressed an assimilatory nitrate reductase (NasA, BazSymA_Acontig00081_3) and a nitrite reductase (NirB, BazSymA_Acontig00081_0) in its proteome. So far, nothing was known about the nitrate assimilation capabilities of the methanotrophic *B. azoricus* symbiont. A study reported the absence of nitrate utilization in an undescribed *Bathymodiolus* sp. found at hydrocarbon seeps of the Gulf of Mexico, which exclusively contains a methanotrophic symbiont (Lee and Childress, 1994). In our study, we did identify a methanotrophic NirB (BAZMOX_contig06966_3) but a homolog for NasA could not be found in the methanotroph's genome or proteome (Supplementary Table S3). Since the *nasA* gene is present in other related free-living methane oxidizers (Murrell and Dalton, 1983), we speculate that the *B. azoricus* methanotroph may possess a complete nitrate assimilation pathway, which was simply not identified in our study due to limited genome information.

1 *The thiotroph can use nitrate as an electron acceptor*: Several enzymes involved in respiratory
2 nitrate reduction (i.e. denitrification), such as the membrane-bound respiratory nitrate
3 reductase subunit NarG (BazSymA_Acontig00119_5), a copper-containing periplasmic nitrite
4 reductase (NirK, BazSymB_scaffold00114_1) and a quinol nitric oxide reductase subunit
5 (NorB, BazSymA_Acontig03118_2) were identified in the thiotroph’s proteome
6 (Supplementary Figure S5, Supplementary Table S3), indicating use of nitrate as a terminal
7 electron acceptor. Whether the symbiont can perform complete reduction of nitrate remains
8 unclear, since a gene for nitrous oxide reductase, NosZ, catalyzing the final step of nitrate
9 reduction to dinitrogen gas, was not detected in the genome or the proteome (Supplementary
10 Figure S4). A bacterial transcript of *nosZ* was reported in the gill transcriptome of *B. azoricus*
11 (Egas *et al.*, 2012). Nevertheless, even if NosZ were absent, our results show that the thiotroph
12 can respire nitrate to nitrous oxide. Although genes for dissimilatory nitrate reduction were not
13 detected in the methanotrophic *B. azoricus* symbiont’s genome in our analysis, a NarG
14 homolog (ZP_08782554.1) was identified in the gill fraction using the sequence of the related
15 free-living methanotroph *Methylobacter tundripaludum* SV96, indicating that the genetic
16 machinery for nitrate respiration may be present in the methanotroph (Supplementary Table
17 S3). Nitrate can act as an electron acceptor during anoxic conditions in free-living and
18 symbiotic bacteria from the deep-sea (Hentschel and Felbeck, 1993; Girguis *et al.*, 2000;
19 Vetriani *et al.*, 2014). However, *in situ* levels of oxygen in the seawater were around 107 μM
20 during mussel collection, which is high enough to support aerobic respiration in marine
21 bacteria. We identified homologs of terminal cytochrome oxidases for oxygen respiration in our
22 thiotroph proteome (Supplementary Table S3). This indicates that under oxygenated
23 conditions nitrogen respiration enzymes might be constitutively expressed in the thiotrophic
24 symbiont. This constitutive expression was also reported for the *Lucinoma aequizonata*
25 symbiont (Hentschel *et al.*, 1996) and the γ3-symbiont of *O. algarvensis* (Kleiner *et al.*, 2012b).
26 In the *B. azoricus* thiotroph, the use of nitrate as an oxidant would not only reduce competition
27 between host and symbiont for oxygen but may also enable it to overcome anoxic periods
28 occurring in its natural environment.

1 *Alternative routes to replenish oxaloacetate and succinate are missing in the thiotroph*

2 The thiotrophic symbiont’s genome does not encode the enzymes Mdh and Sdh, which are
3 required for the replenishment of the essential TCA intermediates oxaloacetate and succinate
4 (see main text, Figures 4A and 5A). A functional substitute for the malate dehydrogenase Mdh,
5 for instance the irreversible membrane-bound malate:quinone oxidoreductase (Mqo) found in
6 the obligate deep-sea autotroph *Thiomicrospira crunogena* XCL-2 (Scott *et al.*, 2006), could in
7 principle allow regeneration of oxaloacetate, but no such alternates were found in the
8 thiotrophic *B. azoricus* symbiont’s genome (Figure 5B). Theoretically, oxaloacetate could also
9 be derived from aspartate through the aspartate transaminase AspC expressed in the thiotroph
10 (BazSymA_Acontig06653_0), provided that cellular aspartate levels are sufficient to drive this
11 enzyme in reverse (Figure 5A and B, Supplementary Table S3). However, as indicated by our
12 data, the thiotroph consumes aspartate for the biosynthesis of lysine, threonine and methionine
13 at relatively high rates (see main text), which makes aspartate an unlikely source for
14 oxaloacetate regeneration. Enyzmes that could functionally replace Sdh or alternate routes for
15 the replenishment of succinate and succinyl-CoA were also missing in the thiotrophic
16 symbiont’s genome (Figure 5B). The need for succinyl-CoA for porphyrin biosynthesis might
17 theoretically be circumvented as porphyrin can also be synthesized from glutamate (Scott *et*
18 *al.*, 2006) and the respective enzymes were expressed by the thiotroph (see Supplementary
19 Figure S5, proto- and siroheme cofactors). However, succinyl-CoA is also involved in lysine
20 and methionine biosynthesis (as is oxaloacetate) and its replenishment therefore mandatory.
21 The thiotrophic symbiont may thus not be able to restore its oxaloacetate and succinate stocks
22 autonomously. Instead, it might rely on an external supply of oxaloacetate and succinate or
23 precursors of these compounds from either of its symbiotic partners. The methanotrophic
24 symbiont encodes both Mdh and PckA and may thus theoretically be capable of providing
25 oxaloacetate to the thiotroph (PckA was identified in the proteome: BAZMOX_contig10981_0;
26 see Figure 4B). However, owing to the incomplete genome information we do not know if the
27 methanotroph also has the potential to export oxaloacetate or C4 intermediates from its

cytoplasm into the bacteriocyte cytosol. It is therefore questionable whether the methanotroph might be involved in replenishing oxalaoactetate or succinate to the thiotroph. Our results do, however, indicate that the host likely supplies oxaloacetate and succinate to the thiotrophic symbiont (see main text).

Potential causes of high expression of chaperones and bacterial nucleoid proteins in B. azoricus *symbionts*

The chaperones GroEL (BazSymB_scaffold00015_18, BAZMOX_contig109183_0, BAZMOX_contig53815_0, METH500_1), GroES (BazSymB_scaffold00015_17, BAZMOX_contig53815_1), DnaJ (BazSymA_Acontig00623_1) and DnaK (BazSymB_scaffold00038_12, BAZMOX_contig04778_1, BAZMOX_contig243452_0) were extraordinarily abundant in both *B. azoricus* symbionts, as were the thiotroph's DNA-binding protein Hns (BAT01474) and HU beta (HupB, BazSymB_scaffold00007_16, see main text and Supplementary Table S3). On a molecular level, both chaperones and DNA-binding proteins (also called histone chaperones) share parallel roles: Chaperones facilitate folding and transport of newly synthesized proteins and the disaggregation or repair of damaged proteins. DNA-binding proteins act at the level of DNA, mediating the degree of DNA compaction (Spurio *et al.*, 1992; Köhler and Marahiel, 1997). Both protein classes are generally abundant in most bacteria, e.g. in *E. coli* and *B. subtilis*, where GroEL and HupB/Hns are among the top 5 % of most abundant proteins (Chi *et al.*, 2011; Mancuso *et al.*, 2012). Their expression is upregulated in response to a variety of stress conditions that endanger DNA and protein integrity, such as thermal stress, radioactivity, oxidative stress and osmotic stress (Farr and Kogoma, 1991; Atlung and Ingmer, 1997; Esser *et al.*, 1999; Prasad *et al.*, 2003; Ghosh and Grove, 2006; Lieber *et al.*, 2009; Nguyen *et al.*, 2009; Wang and Maier, 2015).

In our study, thermal stress is not a likely cause of protein or DNA damage in *B. azoricus* symbionts, as the measured temperature during time of collection was 9 °C, which corresponds to normal habitat conditions (Pruski and Dixon, 2003). Although other common

1 stress factors in hydrothermal vent environments, such as radioactivity (Cherry *et al.*, 1992),
2 heavy metals (Rousse *et al.*, 1998) and free radical generation by spontaneous reactions of
3 H_2S with oxygen (Tapley, 1999) cannot be excluded, it is also conceivable that high levels of
4 chaperones and DNA-binding proteins in the *B. azoricus* symbionts might not necessarily be
5 related to stress response, but rather have a symbiosis-specific role. This would be in
6 agreement with the observation that symbionts of invertebrates inhabiting shallow water, such
7 as *O. algarvensis* and *Codakia orbicularis*, where the above mentioned factors are absent,
8 also show particularly high expression levels of chaperones and DNA-binding proteins (Kleiner
9 *et al.*, 2012b; T. Schweder, personal communication).
10 Chaperones and DNA-binding proteins may be of particular relevance in some symbioses,
11 because they were hypothesized to buffer the symbiont genomes against deleterious
12 mutations that accumulate during vertical transmission. Due to population bottlenecks,
13 vertically transmitted symbionts, such as the aphid symbiont *Buchnera aphidicola*, experience
14 an increased rate of evolution as compared to their free-living counterparts, and therefore carry
15 a high mutational load in their greatly reduced genomes (Itoh *et al.*, 2002). Overexpression of
16 chaperones in these symbionts salvages misfolded proteins arising from DNA mutations
17 (Fares *et al.*, 2002; Maisnier-Patin *et al.*, 2005). *B. aphidicola*, for example, constitutively
18 overexpresses GroEL, which makes up 10% of the total cellular protein under optimum
19 conditions (Baumann *et al.*, 1996). There are indications that DNA-binding proteins may
20 protect against deleterious mutations at the DNA level (Warnecke *et al.*, 2012), although these
21 effects have not yet been investigated in symbionts. If, like in the insect symbioses, the
22 overexpression of chaperones and DNA-binding proteins in *B. azoricus* symbionts is a
23 response to the accumulation of deleterious mutations, this might mean that the *B. azoricus*
24 symbionts, too, experience an increased rate of evolution. Although the specific conditions that
25 lead to high mutational rates in *Buchnera*, such as sexual isolation and population bottle necks
26 during transmission, do likely not apply to the *B. azoricus* thiotroph, the process of adapting to
27 its intracellular niche might still be accompanied by an increased number of mutation events.

Another possible explanation for elevated levels of chaperones and DNA-binding proteins would be that the symbionts face symbiosis-specific intracellular challenges, such as host-mediated immune responses or digestion. Pathogenic bacteria induce an elaborate heat shock response to override unfavorable conditions caused by the host immune reaction (Neckers and Tatu, 2008). While the thiotrophic symbiont of *B. azoricus* is not related to any known pathogen (Sayavedra *et al.*, 2015) it might use a similar strategy to survive in the intracellular environment. The exact nature of this hypothetical ‘host stress’ remains to be elucidated.

Symbiont proteins involved in immunity to phages

B. azoricus symbionts encode two well-known forms of bacterial defense against bacteriophages (Labrie *et al.*, 2010). The thiotroph’s genome encodes four CRISPR-associated (Cas) proteins, multiple CRISPR repeats and spacers, and 40 restriction-modification (R-M) associated proteins, while the methanotroph’s genome encodes 24 R-M proteins (Supplementary Table S5). This is surprising, because in their intracellular niches endosymbionts are expected to rarely experience exposure to phages (Vasu and Nagaraja, 2013). We detected some of the CRISPR-Cas and R-M proteins in whole gill samples and symbiont enrichments, suggesting that the symbionts actively express phage defense mechanisms while residing within bacteriocytes, although at relatively low abundances (Supplementary Tables S3 and S5A). Cas proteins were observed in very high concentrations, i.e. constituting 90 % of total spectral counts in phage-infected *Streptococcus thermophilus*, while their expression is constitutively low in the absence of phages (Agari *et al.*, 2010; Young *et al.*, 2012). Low abundance of Cas and R-M proteins in the *B. azoricus* symbionts therefore likely reflects constitutively low expression rather than a response to recent phage invasion. Moreover, the thiotroph’s Cas genes are spread out over the genome rather than being encoded in a common locus adjacent to CRISPR repeats, indicating that functionality against viral encounters might be restricted (van der Oost *et al.*, 2009). The free-living relatives of the *B. azoricus* thiotroph, SUP05 and Arctic96BD-19, do not encode CRISPR loci in their

genomes, and only a single R-M-associated gene was found in the SUP05 genome (Walsh *et al.*, 2009; Swan *et al.*, 2011). The Arctic96BD-19 genomes did not encode any R-M genes. Finally, the best BLAST hits to the thiotroph’s Cas and R-M genes were from methylotrophic Gammaproteobacteria, pathogenic *Acinetobacter* spp. and other distantly related bacterial taxa and archaea, raising the possibility that these genes were acquired by horizontal gene transfer. The unexpected presence of phage defense mechanisms in *B. azoricus* symbionts despite their intracellular niches may thus point to other symbiosis-related roles for CRISPR-Cas and R-M proteins, besides cellular defense. Cas proteins have, for instance, been suggested to help pathogenic bacteria to evade the host’s innate immunity during infection (Sampson and Weiss, 2013). A similar function could be hypothesized in the *B. azoricus* symbionts.

Host proteins with immune- or defense-related functions

A set of 23 immune-related host proteins was detected in our study. Seven of them showed significantly elevated expression levels in gills and supernatant as compared to symbiont-free foot tissue samples, or were not detected in foot tissue at all, indicating a symbiosis-specific function (Table 1). These host proteins included a lectin, a coagulation factor domain protein, and a cytolysin. The cytolysin was 5.57 times more abundant in the symbiont-enriched gradient pellet than in whole gill tissue, indicating that it might be involved in direct interaction with the bacterial cells (causing it to be pulled down in the gradient pellet). The presence of an active innate immune system targeting pathogens in the hemolymph of *B. azoricus* is known (Bettencourt *et al.*, 2007; Bettencourt *et al.*, 2009; Martins *et al.*, 2014), but its potential role in interactions with the beneficial symbionts has not yet been addressed. Moreover, the exact functions of innate immunity proteins in symbiotic invertebrates are poorly understood. However, considering their physical association with the symbiont cells, these host proteins might play a role in mediating beneficial host-microbe interactions such as recognition and attachment (Bosch, 2013), rather than in defense against pathogens.

References

Agari Y, Sakamoto K, Tamakoshi M, Oshima T, Kuramitsu S, Shinkai A. (2010). Transcription profile of *Thermus thermophilus* CRISPR systems after phage infection. *J Mol Biol* **395:** 270-281.

Anthony C. (1982). *The biochemistry of methylotrophs*. Academic Press: London; New York.

Atlung T, Ingmer H. (1997). H-NS: a modulator of environmentally regulated gene expression. *Molecular microbiology* **24:** 7-17.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA *et al*. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9:** 75.

Baumann P, Baumann L, Clark MA. (1996). Levels of *Buchnera aphidicola* Chaperonin GroEL During Growth of the Aphid *Schizaphis graminum*. *Curr Microbiol* **32:** 279-285.

Bendtsen JD, Kiemer L, Fausboll A, Brunak S. (2005). Non-classical protein secretion in bacteria. *BMC Microbiol* **5:** 58.

Bettencourt R, Roch P, Stefanni S, Rosa D, Colaco A, Santos RS. (2007). Deep sea immunity: unveiling immune constituents from the hydrothermal vent mussel *Bathymodiolus azoricus*. *Mar Environ Res* **64:** 108-127.

Bettencourt R, Dando P, Collins P, Costa V, Allam B, Serrao Santos R. (2009). Innate immunity in the deep sea hydrothermal vent mussel *Bathymodiolus azoricus*. *Comp Biochem Physiol A Mol Integr Physiol* **152:** 278-289.

Bettencourt R, Pinheiro M, Egas C, Gomes P, Afonso M, Shank T *et al*. (2010). High-throughput sequencing and analysis of the gill tissue transcriptome from the deep-sea hydrothermal vent mussel *Bathymodiolus azoricus*. *BMC Genomics* **11:** 559.

Biegert A, Mayer C, Remmert M, Soding J, Lupas AN. (2006). The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acids Res* **34:** W335-339.

Bosch TC. (2013). Cnidarian-microbe interactions and the origin of innate immunity in metazoans. *Annu Rev Microbiol* **67:** 499-518.

Bradford MM. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72:** 248-254.

Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA *et al*. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* **42:** D459-471.

Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S *et al*. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **30:** 918-920.

Charlou JL, Donval JP, 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 (36 degrees 14 ' N, MAR). *Chem Geol* **191:** 345-359.

Cherry R, Desbruyeres D, Heyraud M, Nolan C. (1992). High levels of natural radioactivity in hydrothermal vent polychaetes. *C R Acad Sci III* **315:** 21-26.

Chi BK, Gronau K, Mäder U, Hessling B, Becher D, Antelmann H. (2011). S-Bacillithiolation Protects Against Hypochlorite Stress in *Bacillus subtilis* as Revealed by Transcriptomics and Redox Proteomics. *Mol Cell Proteomics* **10:** M111.009506.

Chistoserdova L, Vorholt JA, Lidstrom ME. (2005). A genomic view of methane oxidation by aerobic bacteria and anaerobic archaea. *Genome Biol* **6:** 208.

Chistoserdova L, Lidstrom M (2013). Aerobic Methylotrophic Prokaryotes. In: Rosenberg E, DeLong E, Lory S, Stackebrandt E, Thompson F (eds). *The Prokaryotes*. Springer Berlin Heidelberg. pp 267-285.

Crowther GJ, Kosaly G, Lidstrom ME. (2008). Formate as the main branch point for methylotrophic metabolism in *Methylobacterium extorquens* AM1. *J Bacteriol* **190:** 5057-5062.

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.

Egas C, Pinheiro M, Gomes P, Barroso C, Bettencourt R. (2012). The transcriptome of *Bathymodiolus azoricus* gill reveals expression of genes from endosymbionts and free-living deep-sea bacteria. *Mar Drugs* **10:** 1765-1783.

Esser D, Rudolph R, Jaenicke R, Böhm G. (1999). The HU protein from *Thermotoga maritima*: recombinant expression, purification and physicochemical characterization of an extremely hyperthermophilic DNA-binding protein. *J Mol Biol* **291:** 1135-1146.

Eymann C, Dreisbach A, Albrecht D, Bernhardt J, Becher D, Gentner S *et al*. (2004). A comprehensive proteome map of growing *Bacillus subtilis* cells. *Proteomics* **4:** 2849-2876.

Fares MA, Ruiz-Gonzalez MX, Moya A, Elena SF, Barrio E. (2002). Endosymbiotic bacteria: GroEL buffers against deleterious mutations. *Nature* **417:** 398.

Farr SB, Kogoma T. (1991). Oxidative stress responses in *Escherichia coli* and *Salmonella typhimurium*. *Microbiol Rev* **55:** 561-585.

Fiala-Médioni A, McKiness ZP, Dando P, Boulegue J, Mariotti A, Alayse-Danet AM *et al*. (2002). Ultrastructural, biochemical, and immunological characterization of two populations of the mytilid mussel *Bathymodiolus azoricus* from the Mid-Atlantic Ridge: evidence for a dual symbiosis. *Mar Biol* **141:** 1035–1043.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR *et al*. (2014). Pfam: the protein families database. *Nucleic Acids Res* **42:** D222-230.

Florens L, Carozza MJ, Swanson SK, Fournier M, Coleman MK, Workman JL *et al*. (2006). Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* **40:** 303-311.

Ghosh S, Grove A. (2006). The Deinococcus radiodurans-encoded HU protein has two DNA-binding domains. *Biochemistry* **45:** 1723-1733.

Girguis PR, Lee RW, Desaulniers N, Childress JJ, Pospesel M, Felbeck H *et al*. (2000). Fate of nitrate acquired by the tubeworm *Riftia pachyptila*. *Appl Environ Microbiol* **66:** 2783-2790.

Hanson RS, Hanson TE. (1996). Methanotrophic bacteria. *Microbiol Rev* **60:** 439-471.

Harada M, Yoshida T, Kuwahara H, Shimamura S, Takaki Y, Kato C *et al*. (2009). Expression of genes for sulfur oxidation in the intracellular chemoautotrophic symbiont of the deep-sea bivalve *Calyptogena okutanii*. *Extremophiles* **13:** 895-903.

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.

Itoh T, Martin W, Nei M. (2002). Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. *Proc Natl Acad Sci USA* **99:** 12944-12948.

Käll L, Storey JD, MacCoss MJ, Noble WS. (2008). Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J Proteome Res* **7:** 29-34.

Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* **42:** D199-205.

Keller A, Nesvizhskii AI, Kolker E, Aebersold R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **74:** 5383-5392.

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

Kleiner M, Wentrup C, Lott C, Teeling H, Wetzel S, Young J *et al*. (2012b). Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc Natl Acad Sci USA* **109:** E1173-1182.

Köhler P, Marahiel MA. (1997). Association of the histone-like protein HBsu with the nucleoid of *Bacillus subtilis*. *J Bacteriol* **179:** 2060-2064.

Korotkova N, Chistoserdova L, Kuksa V, Lidstrom ME. (2002). Glyoxylate regeneration pathway in the methylotroph *Methylobacterium extorquens* AM1. *J Bacteriol* **184:** 1750-1758.

Labrie SJ, Samson JE, Moineau S. (2010). Bacteriophage resistance mechanisms. *Nat Rev Microbiol* **8:** 317-327.

Lee RW, Childress JJ. (1994). Assimilation of inorganic nitrogen by marine invertebrates and their chemoautotrophic and methanotrophic symbionts. *Appl Environ Microbiol* **60:** 1852-1858.

Li W, Jaroszewski L, Godzik A. (2001). Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* **17:** 282-283.

Lieber A, Leis A, Kushmaro A, Minsky A, Medalia O. (2009). Chromatin organization and radio resistance in the bacterium *Gemmata obscuriglobus*. *J Bacteriol* **191:** 1439-1445.

Linton JD, Cripps RE. (1978). The occurrence and identification of intracellular polyglucose storage granules in *Methylococcus* NCIB 11083 grown in chemostat culture on methane. *Arch Microbiol* **117:** 41-48.

Lobo-da-Cunha A, Kádár E, Serrão Santos R. (2006). Histochemical and ultrastructural characterisation of mantle storage cells in the hydrothermal-vent bivalve *Bathymodiolus azoricus*. *Mar Biol* **150:** 253-260.

Maden BE. (2000). Tetrahydrofolate and tetrahydromethanopterin compared: functionally distinct carriers in C-1 metabolism. *Biochem J* **350:** 609-629.

Maisnier-Patin S, Roth JR, Fredriksson A, Nystrom T, Berg OG, Andersson DI. (2005). Genomic buffering mitigates the effects of deleterious mutations in bacteria. *Nature Genet* **37:** 1376-1379.

Mancuso F, Bunkenborg J, Wierer M, Molina H. (2012). Data extraction from proteomics raw data: An evaluation of nine tandem MS tools using a large Orbitrap data set. *J Proteomics* **75:** 5293-5303.

Markert S, Gardebrecht A, Felbeck H, Sievert SM, Klose J, Becher D *et al*. (2011). Status quo in physiological proteomics of the uncultured *Riftia pachyptila* endosymbiont. *Proteomics* **11:** 3106-3117.

Martins E, Figueras A, Novoa B, Santos RS, Moreira R, Bettencourt R. (2014). Comparative study of immune responses in the deep-sea hydrothermal vent mussel *Bathymodiolus azoricus* and the shallow-water mussel *Mytilus galloprovincialis* challenged with *Vibrio* bacteria. *Fish Shellfish Immun* **40:** 485-499.

Moller S, Croning MDR, Apweiler R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17:** 646-653.

Mueller RS, Denef VJ, Kalnejais LH, Suttle KB, Thomas BC, Wilmes P *et al*. (2010). Ecological distribution and population physiology defined by proteomics in a natural microbial community. *Mol Syst Biol* **6:** 374.

Murrell JC, Dalton H. (1983). Ammonia Assimilation in *Methylococcus capsulatus* (Bath) and Other Obligate Methanotrophs. *J Gen Microbiol* **129:** 1197-1206.

Muyzer G, Teske A, Wirsen CO, Jannasch HW. (1995). Phylogenetic Relationships of *Thiomicrospira* Species and Their Identification in Deep-Sea Hydrothermal Vent Samples by Denaturing Gradient Gel-Electrophoresis of 16S rDNA Fragments. *Arch Microbiol* **164:** 165-172.

Neckers L, Tatu U. (2008). Molecular Chaperones in Pathogen Virulence: Emerging New Targets for Therapy. *Cell Host Microbe* **4:** 519-527.

Nguyen HH, de la Tour CB, Toueille M, Vannier F, Sommer S, Servant P. (2009). The essential histone-like protein HU plays a major role in *Deinococcus radiodurans* nucleoid compaction. *Mol Microbiol* **73:** 240-252.

Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T *et al*. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* **42:** D206-214.

Pernthaler A, Pernthaler J, Amann R. (2002). Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl Environ Microbiol* **68:** 3094-3101.

Petersen JM, Dubilier N. (2009). Methanotrophic symbioses in marine invertebrates. *Environ Microbiol Rep* **1:** 319-335.

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

Petersen TN, Brunak S, von Heijne G, Nielsen H. (2011b). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8:** 785-786.

Pieja AJ, Rostkowski KH, Criddle CS. (2011a). Distribution and selection of poly-3-hydroxybutyrate production capacity in methanotrophic proteobacteria. *Microb Ecol* **62:** 564-573.

Pieja AJ, Sundstrom ER, Criddle CS. (2011b). Poly-3-hydroxybutyrate metabolism in the type II methanotroph *Methylocystis parvus* OBBP. *Appl Environ Microbiol* **77:** 6012-6019.

Prasad J, McJarrow P, Gopal P. (2003). Heat and osmotic stress responses of probiotic *Lactobacillus rhamnosus* HN001 (DR20) in relation to viability after drying. *Appl Environ Microbiol* **69:** 917-925.

Pruski AM, Rousse N, Fiala-Médioni A, Boulègue J. (2002). Sulphur signature in the hydrothermal vent mussel *Bathymodiolus azoricus* from the Mid-Atlantic Ridge. *J Mar Biol Assoc UK* **82:** 463-468.

Pruski AM, Dixon DR. (2003). Toxic vents and DNA damage: first evidence from a naturally contaminated deep-sea environment. *Aquat Toxicol* **64:** 1-13.

Rousse N, Boulegue J, Cosson RP, Fiala-Medioni A. (1998). Bioaccumulation des métaux chez le mytilidae hydrothermal Bathymodiolus sp. de la ride médio-atlantique. *Oceanol Acta* **21:** 597-607.

Sampson TR, Weiss DS. (2013). Alternative Roles for CRISPR/Cas Systems in Bacterial Pathogenesis. *PLoS Pathog* **9:** e1003621.

Sarradin PM, Caprais JC, Riso R, Kerouel R, Aminot A. (1999). Chemical environment of the hydrothermal mussel communities in the Lucky Strike and Menez Gwen vent fields, Mid Atlantic ridge. *Cah Biol Mar* **40:** 93-104.

Sauve V, Bruno S, Berks BC, Hemmings AM. (2007). The SoxYZ complex carries sulfur cycle intermediates on a peptide swinging arm. *J Biol Chem* **282:** 23194-23204.

Sayavedra L, Kleiner M, Ponnudurai RP, Wetzel S, Pelletier E, Barbe V *et al.* (2015). Abundant toxin-related genes in the genomes of beneficial symbionts from deep-sea hydrothermal vent mussels. *eLife* **4:** e07966.

Scott KM, Sievert SM, Abril FN, Ball LA, Barrett CJ, Blake RA *et al.* (2006). The genome of deep-sea vent chemolithoautotroph *Thiomicrospira crunogena* XCL-2. *PLoS Biol* **4:** e383.

Shibata H, Takahashi M, Yamaguchi I, Kobayashi S. (2001). Efficient removal of sulfide following integration of multiple copies of the sulfide-quinone oxidoreductase gene (sqr) into the *Escherichia coli* chromosome. *J Biosci Bioeng* **91:** 493-499.

Spurio R, Durrenberger M, Falconi M, La Teana A, Pon CL, Gualerzi CO. (1992). Lethal overproduction of the *Escherichia coli* nucleoid protein H-NS: ultramicroscopic and molecular autopsy. *Mol Gen Genet* **231:** 201-211.

Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D *et al*. (2011). Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean. *Science* **333:** 1296-1300.

Takeuchi T, Kawashima T, Koyanagi R, Gyoja F, Tanaka M, Ikuta T *et al*. (2012). Draft genome of the pearl oyster *Pinctada fucata*: a platform for understanding bivalve biology. *DNA Res* **19:** 117-130.

Tapley DW, Buettner, G.R., Shick, J.M.,. (1999). Free radicals and chemoluminescence as products of the spontaneous oxidation of sulfide in seawater, and their biological implications. *Biol Bull* **196:** 52-56.

Thauer RK. (1998). Biochemistry of methanogenesis: a tribute to Marjory Stephenson. 1998 Marjory Stephenson Prize Lecture. *Microbiology* **144 (Pt 9):** 2377-2406.

Tusher VG, Tibshirani R, Chu G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* **98:** 5116-5121.

UniProt-Consortium. (2015). UniProt: a hub for protein information. *Nucleic Acids Res* **43:** D204-212.

van der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJJ. (2009). CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem Sci* **34:** 401-407.

Vasu K, Nagaraja V. (2013). Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol Mol Biol Rev* **77:** 53-72.

Vetriani C, Voordeckers JW, Crespo-Medina M, O'Brien CE, Giovannelli D, Lutz RA. (2014). Deep-sea hydrothermal vent Epsilonproteobacteria encode a conserved and widespread nitrate reduction pathway (Nap). *ISME J* **8:** 1510-1521.

Vizcaíno JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I *et al*. (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* **44:** D447-D456.

Vorholt JA, Marx CJ, Lidstrom ME, Thauer RK. (2000). Novel formaldehyde-activating enzyme in *Methylobacterium extorquens* AM1 required for growth on methanol. *J Bacteriol* **182:** 6645-6650.

Vorholt JA. (2002). Cofactor-dependent pathways of formaldehyde oxidation in methylotrophic bacteria. *Arch Microbiol* **178:** 239-249.

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.

Wang G, Maier RJ. (2015). Bacterial histone-like proteins: roles in stress resistance. *Curr Genet* **61:** 489-492.

Ward N, Larsen O, Sakwa J, Bruseth L, Khouri H, Durkin AS *et al*. (2004). Genomic insights into methanotrophy: The complete genome sequence of *Methylococcus capsulatus* (Bath). *PLoS Biol* **2:** 1616-1628.

Warnecke T, Supek F, Lehner B. (2012). Nucleoid-associated proteins affect mutation dynamics in *E. coli* in a growth phase-specific manner. *PLoS Comput Biol* **8:** e1002846.

Widdel F, Pfennig N. (1981). Studies on Dissimilatory Sulfate-Reducing Bacteria That Decompose Fatty-Acids.1. Isolation of New Sulfate-Reducing Bacteria Enriched with Acetate from Saline Environments - Description of *Desulfobacter postgatei* gen. nov. sp. nov. *Arch Microbiol* **129:** 395-400.

Widdel F, Kohring G-W, Mayer F. (1983). Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids III. Characterization of the filamentous gliding *Desulfonema limicola* gen. nov. sp. nov., and *Desulfonema magnum* sp. nov. *Arch Microbiol* **134:** 286-294.

Young JC, Dill BD, Pan CL, Hettich RL, Banfield JF, Shah M *et al*. (2012). Phage-Induced Expression of CRISPR-Associated Proteins Is Revealed by Shotgun Proteomics in *Streptococcus thermophilus*. *PloS ONE* **7:** e38077.

Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R *et al*. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26:** 1608-1615.

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 Geosy* **12:** Q0AE04.