bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Metabolic differences between symbiont subpopulations in the deep-sea tubeworm *Riftia pachyptila*

Tjorven Hinzke[1,2,3], Manuel Kleiner[4], Mareike Meister[5,6], Rabea Schlüter[7], Christian Hentschker[8], Jan Pané-Farré[9], Petra Hildebrandt[8], Horst Felbeck[10], Stefan M. Sievert[11], Florian Bonn[12], Uwe Völker[8], Dörte Becher[5], Thomas Schweder[1,2], Stephanie Markert[1,2*]

1- Institute of Pharmacy, University of Greifswald, Germany
2- Institute of Marine Biotechnology, Greifswald, Germany
3- Energy Bioengineering Group, University of Calgary, Calgary, Canada
4- Department of Plant and Microbial Biology, North Carolina State University, NC, USA
5- Institute of Microbiology, University of Greifswald, Germany
6- Leibniz Institute for Plasma Science and Technology, Greifswald, Germany
7- Imaging Center of the Department of Biology, University of Greifswald, Germany
8- Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Germany
9- Center for Synthetic Microbiology (SYNMIKRO), Philipps-University Marburg, Germany
10- Scripps Institution of Oceanography, University of California San Diego, CA, USA
11- Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, USA
12- Institute of Biochemistry, University Hospital, Goethe University School of Medicine Frankfurt, Germany

Key words: thiotrophic symbiosis, cell heterogeneity, metaproteomics, microbe-host interactions, cell differentiation, chemosynthetic symbiosis

*correspondence: stephanie.markert@uni-greifswald.de

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

The hydrothermal vent tube worm *Riftia pachyptila* lives in intimate symbiosis with intracellular sulfur-oxidizing gammaproteobacteria. Although the symbiont population consists of a single 16S rRNA phylotype, bacteria in the same host animal exhibit a remarkable degree of metabolic diversity: They simultaneously utilize two carbon fixation pathways and various energy sources and electron acceptors. Whether these multiple metabolic routes are employed in the same symbiont cells, or rather in distinct symbiont subpopulations, was unclear. As *Riftia* symbionts vary considerably in cell size and shape, we enriched individual symbiont cell sizes by density gradient centrifugation in order to test whether symbiont cells of different sizes show different metabolic profiles. Metaproteomic analysis and statistical evaluation using clustering and random forests, supported by microscopy and flow cytometry, strongly suggest that *Riftia* symbiont cells of different sizes represent metabolically dissimilar stages of a physiological differentiation process: Small symbionts actively divide and may establish cellular symbiont-host interaction, as indicated by highest abundance of the cell division key protein FtsZ and highly abundant chaperones and porins in this initial phase. Large symbionts, on the other hand, apparently do not divide, but still replicate DNA, leading to DNA endoreduplication. Highest abundance of enzymes for CO_2 fixation, carbon storage and biosynthesis in large symbionts indicates that in this late differentiation stage the symbiont's metabolism is efficiently geared towards the production of organic material. We propose that this division of labor between smaller and larger symbionts benefits the productivity of the symbiosis as a whole.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

The chemoautotrophic gammaproteobacterium *Candidatus* Endoriftia persephone, sulfur-oxidizing endosymbiont of the deep-sea tubeworm *Riftia pachyptila*, provides all nutrition for its gutless host (Cavanaugh *et al.*, 1981, Felbeck, 1981, Hand, 1987, Distel and Felbeck, 1988, Robidart *et al.*, 2008). *Ca.* E. persephone (here Endoriftia) densely populates *Riftia*'s trophosome, a specialized organ in the worm's trunk, where the bacteria are housed intracellularly in host bacteriocytes (Hand, 1987).

Although the symbiont population consists of a single 16S rRNA phylotype (Polzin *et al.*, 2019), it was previously shown to exhibit remarkable metabolic versatility: As demonstrated by proteomic analyses, symbionts from the same host animal expressed enzymes of two CO_2 fixation pathways, the Calvin cycle and the reverse tricarboxylic acid (rTCA) cycle, as well as enzymes for both, glycogen generation and glycogen degradation (Markert *et al.*, 2007, Markert *et al.*, 2011, Gardebrecht *et al.*, 2012, Hinzke *et al.*, 2019). Moreover, proteins involved in utilization of hydrogen sulfide and thiosulfate as energy sources were expressed simultaneously by the same symbiont population; as were proteins for the use of nitrate and oxygen as electron acceptors (Markert *et al.*, 2011). Based on these observations, we hypothesized that individual, metabolically distinct symbiont subpopulations in the trophosome may exist.

These presumptive subpopulations are likely congruent with symbionts of different cell sizes: Individual Endoriftia cells exhibit pronounced morphological diversity, ranging from small rods to small and large cocci in ultimate proximity to each other within the same host specimen (Hand, 1987, Bright *et al.*, 2000, Bright and Sorgo, 2003). In individual trophosome lobules, which measure approximately 200-500 µm in diameter, the smallest, rod-shaped symbiont cells are located close to the central blood vessel, while towards the lobule periphery, symbionts gradually increase in size and become coccoid, before they are degraded in the outermost lobule zone. Only small Endoriftia cells and the host bacteriocytes in which they reside appear to undergo cell division, indicating that small and large symbionts belong to a common cell

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cycle (Bright *et al.*, 2000, Bright and Sorgo, 2003). Previous microscopy-based studies indicated that small and large *Riftia* symbionts differ not only with regard to their frequency of cell division, but also with regard to carbon incorporation rates, amount of stored glycogen, and area of sulfur storage vesicles (Bright *et al.*, 2000, Sorgo *et al.*, 2002, Bright and Sorgo, 2003, Pflugfelder *et al.*, 2005). This suggests that individual cell sizes may indeed have dissimilar metabolic properties.

In this study, we aimed to analyze and compare the metabolic profiles of individual *Riftia* symbiont subpopulations. In contrast to previous molecular analyses that studied metabolic capabilities of the *Riftia* symbiont population as a mixture of all cell sizes (e.g., Markert *et al.*, 2007, Markert *et al.*, 2011, Gardebrecht *et al.*, 2012), precluding comparisons between putative subpopulations, we used a more sensitive approach: We enriched Endoriftia cells of different sizes by gradient centrifugation of trophosome tissue homogenate and subjected these enriched gradient fractions to separate metaproteomic analyses. Statistical evaluation using clustering and random forests allowed us to deduce cell size-dependent differences in protein abundance and metabolic functions. Catalyzed reporter deposition-fluorescence *in situ* hybridization (CARD-FISH), transmission electron microscopy (TEM), hybridization chain reaction (HCR)-FISH analyses, and flow cytometry complemented these experiments. Our results suggest a division of labor between different developmental stages of the symbiont.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Material and Methods

Sample collection and enrichment of symbiont subpopulations

Riftia samples for enrichment of symbiont subpopulations were collected at the East Pacific Rise hydrothermal vent field at 9°50' N, 104°17' W in a water depth of about 2,500 m during a research cruise with R/V Atlantis in November 2014 (AT26-23). Samples for electron microscopy were obtained during a second cruise (AT37-12) at the same site during March-April 2017 (Hinzke *et al.*, 2019). Sample details and numbers of biological replicates are summarized in Supplementary Table S1.

Trophosome sulfur content of the specimens was estimated based on the trophosome tissue's color: sulfur-rich (S-rich) specimens have a light yellowish trophosome, due to the sulfur stored in the symbionts, whereas trophosomes of sulfur-depleted (S-depleted) specimens appear dark green to black (Pflugfelder *et al.*, 2005). For proteomic analyses, we used four S-rich *Riftia* specimens and three S-depleted specimens.

To enrich symbiont cells of varying sizes (i.e., morphologically distinct symbiont subpopulations), *Riftia* specimens were dissected onboard the research vessel immediately after recovery of the worms and approximately 3 ml trophosome tissue were homogenized in a Dounce glass homogenizer in 6 ml imidazole-buffered saline (IBS, 0.49 M NaCl, 0.03 M MgSO4, 0.011 M CaCl2, 0.003 M KCl, 0.05 M imidazole). As described in Hinzke *et al.* (2018), the homogenate was subjected to rate-zonal density gradient centrifugation, which allows to separate particles based on their size (Graham, 2001). In brief, an 8-18% Histodenz™ density gradient was created using a dilution series of Histodenz™ in IBS (1% steps, 1 ml per step), which was stacked in a 15 ml tube so that Histodenz™ concentration was highest at the bottom and lowest at the top. 0.5 ml tissue homogenate was layered on top of this gradient and the gradient was centrifuged (1,000 x g, 5 min, 4 °C) in a swing-out rotor. Smaller symbiont cells were thus enriched in less dense gradient fractions (lower Histodenz™ concentrations) in the upper part of the gradient, while larger cells migrated to lower gradient fractions. After centrifugation, gradients were disassembled by carefully fractionizing the entire gradient

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

volume into 0.5 ml subsamples, giving a total of 24 fractions. Enrichment of distinct symbiont subpopulations in these subsamples was confirmed using catalyzed reporter deposition-fluorescence in situ hybridization (CARD-FISH, see below). For this purpose, 20 µl of each gradient fraction subsample and 15 µl of homogenate was fixed in 1% PFA in IBS, and symbiont cells were subsequently filtered onto GTTP polycarbonate filters (pore size 0.2 µm, Millipore) as described previously (Ponnudurai *et al.*, 2017).

CARD-FISH

Enrichment of symbiont cell sizes in gradient fractions was analyzed employing fluorescence microscopy with samples labelled by CARD-FISH. CARD-FISH labeling was performed as previously described (Ponnudurai *et al.*, 2017), using the probe Rif445 (Nussbaumer *et al.*, 2006) and Alexa Fluor® 594-labeled tyramide. For counterstaining, 0.1% (w/v) 4,6-diamidino-2-phenylindole (DAPI) was added to the embedding medium (4:1 Citifluor AF1 (Citifluor) and Vectashield (Vector Laboratories)). CARD-FISH filters were analyzed using an Axio Imager.M2 fluorescence microscope (Carl Zeiss Microscopy GmbH). For semi-automated cell counting and to measure the longest cell dimension, we used a custom Fiji (Schindelin *et al.*, 2012) macro with the Fiji plugins Enhanced Local Contrast (CLAHE; Saalfeld, 2010) and Bi-exponential edge preserving smoother (BEEPS; Thévenaz *et al.*, 2012). After image processing, we excluded objects with a size of less than 2 µm (as these were mainly artifacts) and set the maximum object size to 20 µm. To assign cell sizes to size classes (i.e., cell size ranges) we used a quartile split: We calculated quartiles of cell sizes in non-enriched homogenate samples (i.e., 25% of all cells in homogenate samples were assigned to each class). This resulted in the four calculative size classes very small (≥2 µm – <3.912 µm), small (≥3.912 µm – <5.314 µm), medium (≥5.314 µm – <6.83275 µm) and large (≥6.83275 µm – 20 µm). The majority of cells in all size classes were coccoid. Rod-shaped cells were almost exclusively present in the smallest size class. Individual gradient fractions (subsamples) were screened for their respective share of cells in each size class and the subsample with the highest percentage of cells in the respective quartile was chosen for metaproteomic analysis. For example, if of all 24 subsamples of a sample, fraction 5 had the highest percentage of very small cells, i.e. most of

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the cells in fraction 5 were between 2 µm and 3.912 µm in diameter (as measured by our Fiji macro), this fraction was chosen as representative of very small symbiont cells in the respective biological replicate (worm). The fraction containing the highest percentage of very small cells will be referred to as "fraction XS" in the following. The fractions containing the highest percentages of small, medium and large symbiont cells will be referred to as "S", "M", and "L", respectively. If the same subsample had the highest percentage of cells in two size classes, this subsample was chosen as representative for one of these size classes, and for the other size class, the subsample with the second highest percentage of cells in that class was used as representative. Cell size distributions in the four size class representatives are summarized in Figure 1.

Transmission electron microscopy (TEM)

Trophosome samples used for TEM in this study (see Supplementary Table S1 for details) were prepared and analyzed as described previously (Hinzke *et al.*, 2019). Tissue sections were recorded on sheet films (Kodak electron image film SO-163, Plano GmbH, Wetzlar) as described by Petersen *et al.* (2020). To create a composite high-resolution TEM image of a trophosome lobule (Figure 5A), we merged 50 individual micrographs of one section using Serif Affinity Photo (https://affinity.serif.com/en-us/photo/). All 50 partially overlapping images were loaded and the fully automated "Panorama Stitching" technique was applied, resulting in a panorama image still showing some vignette marks caused by inhomogeneous exposure at the former edges of individual images. The global smooth frequencies reflecting these exposure errors were removed using the frequency separation filter with a large radius. The gradation curve was manually corrected. For acquisition of the images in Figure 5B, a wide-angle dual speed CCD camera Sharpeye (Tröndle, Moorenweis, Germany) was used, operated by the ImageSP software. All micrographs were edited using Adobe Photoshop CS6.

HCR-FISH and confocal laser scanning microscopy (CLSM)

A gradient fraction enriched in large symbiont cells (see Supplementary Table S1 for details) that was fixed for CARD-FISH and immobilized on GTTP polycarbonate filters as described

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

above was used for hybridization chain reaction FISH (HCR-FISH) according to Choi *et al.* (2014). We used a HCR-FISH v2.0 Custom Kit (Molecular Technologies) according to the manufacturer's instructions. Probes targeted the *Riftia* symbiont's 16S rRNA (fluorescence marker: Alexa Fluor® 488), and the mRNAs of ATP-citrate lyase subunit AclB (Alexa Fluor® 647) and RubisCO (Alexa Fluor® 594; see Supplementary Table S2 for the probe sequences). In brief, filter sections were washed twice with 50% hybridization buffer (50% formamide, 5x sodium chloride sodium citrate (SCSC, 0.75 M NaCl, 75 mM $Na_3C_6H_5O_7$), 9 mM citric acid, pH 6.0, 0.1% Tween 20, 50 µg/ml heparin, 1x Denhardt's solution, 10% dextran sulfate) in 2x mPBS (89.8 mM Na_2HPO_4, 10.2 mM NaH_2PO_4, 0.9 M NaCl) at 45°C for 30 min for pre-hybridization, and incubated overnight (16 h, 45°C) with probe solution (1 pmol of each probe in 500 µl hybridization buffer). Excess probes were removed with several washing steps in 75-25% probe wash buffer (50% formamide, 5x SCSC, 9 mM citric acid, pH 6.0, 0.1% Tween 20, 50 µg/ml heparin in 5x SCSC) for 15 min at 45°C, 300 rpm, and subsequently in 5x SCSC for 30 min at 45°C and 300 rpm. Samples were pre-amplified with DNA amplification buffer (5x SCSC, 0.1% Tween 20, 10% dextran sulfate). Hairpins were activated by snap-cooling and added to the samples. After overnight incubation (16 h, room temperature) with the hairpin solution, samples were washed with 5x SCSC, containing 0.05% Tween 20 (room temperature, 300 rpm, four times 5 min, two times 30 min) and embedded in Mowiol 4-88 (Carl Roth GmbH) embedding medium prepared according to the manufacturer's instructions. Confocal microscopy was performed on a Zeiss LSM510 meta equipped with a 100x/1.3 oil immersion objective. Probes were excited with laser lines 633 (ATP-citrate lyase), 561 (RubisCO) and 488 (16S rRNA) and signals were detected with filters suitable for dye maximal emissions at 670 nm, 595 nm and 527.5 nm, respectively. Signal intensities and cell sizes (from 8 frames showing a total of 33 cells) were quantified using the Fiji software package (Schindelin *et al.*, 2012). Individual cells were defined as regions of interest (ROI), in which signal intensity per pixel was recorded. Mean pixel intensity of ROI was calculated and background was corrected. Global background values were calculated for every channel based on up to six ROIs randomly placed in each image frame. The following cell size parameters were calculated: (i) Feret's

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Diameter (the longest distance between any two points along the boundary of the ROI) and (ii) the area of the ROI.

Flow cytometry

Subsamples of fresh homogenate, and of three gradient fractions enriched in small symbionts, and three fractions enriched in large symbionts were fixed in 1% PFA as for CARD-FISH (see above) in two biological replicates (i.e. from two *Riftia* specimens). Right before flow cytometry analysis, fixed cells were carefully pelleted and incubated in 0.1 mg/ml RNAse A (from bovine pancreas, DNase-free, Carl Roth, Germany) for 30 min at 37°C to remove RNA, and stained with Syto9 (final concentration 0.5 µmol/l in PBS), a dye that selectively stains DNA and RNA (Stocks, 2004). The fluorescence signal was analyzed using a FACSAria high-speed cell sorter (Becton Dickinson Biosciences, San Jose, CA, USA) with 488 nm excitation from a blue Coherent Sapphire solid state laser at 18 mW. Optical filters were set up to detect the emitted Syto9 fluorescence signal at 530/30nm (FITC channel). All fluorescence data were recorded at logarithmic scale with the FACSDiva 8.02 software (Becton Dickinson). Prior to measurement of experimental samples, the proper function of the instrument was determined by using the cytometer setup and tracking software module (CS&T) together with the CS&T beads (Becton Dickinson Biosciences). First, in a SSC-area *versus* FSC-area dot plot the present populations were shown. The detection thresholds and photomultiplier (PMT) voltages were adjusted by using an unstained sample. The Syto9 signal from the scatter populations was monitored in a Syto9-area histogram. For each sample at least 10.000 events in the scatter gate were recorded. For further analysis, the Syto9-stained bacteria (population 1 and 2, see Figure 2) were sorted from the bivariate dot plot, SCC-A versus Syto9 (FITC-channel). Prior to sorting, the proper function of the cell sorter was determined using the AccuDrop routine. Data analysis was done with the software FlowJo™ V10. To evaluate the results of the sorting procedure, FACS-sorted cell populations as well as unsorted subsamples of homogenate and gradient fractions were examined using an Axio Imager.M2 fluorescence microscope (Carl Zeiss Microscopy GmbH).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Peptide sample preparation

Proteins were extracted as described in Hinzke and Markert (2017). Briefly, cells were mixed with lysis buffer (1% (w/v) sodium deoxycholate (SDC), 4% (w/v) sodium dodecyl sulfate (SDS) in 50 mM triethylammonium bicarbonate buffer (TEAB)), heated for 5 min at 95 °C and 600 rpm and cooled on ice. Samples were then placed in an ultrasonic bath for 5 min and subsequently cooled on ice. Cell debris was removed by centrifugation (14,000 x g, 10 min, room temperature). Protein concentration was determined using the Pierce BCA assay according to the manufacturer's instructions. Peptides were generated using a 1D gel-based approach as in Ponnudurai *et al.* (2017) with minor modifications. In brief, 20 µg of protein sample was mixed with Laemmli sample buffer containing DTT (final concentration 2% (w/v) SDS, 10% glycerol, 12.5 mM DTT, 0.001% (w/v) bromophenol blue in 0.06 M Tris-HCl; Laemmli, 1970) and separated using pre-cast 4-20% polyacrylamide gels (BioRad). After staining, protein lanes were cut into 10 pieces, destained (600 rpm, 37 °C, 200 mM NH_4HCO_3 in 30% acetonitrile) and digested with trypsin (sequencing grade, Promega) overnight at 37 °C, before peptides were eluted in an ultrasonic bath. Peptides were then directly used for LC-MS analysis.

LC-MS/MS analysis

MS/MS measurements were performed as described previously by Ponnudurai *et al.* (2017). In brief, samples were measured with an LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher, Waltham, MA, US) coupled to an EASY-nLC II (ThermoFisher) for peptide separation using a 100 min binary gradient. MS data were acquired in data-dependent MS/MS mode for the 20 most abundant precursor ions. After a full scan in the Orbitrap analyzer (R = 30,000), ions were fragmented via CID and recorded in the LTQ analyzer.

Protein identification and function prediction

Proteins were identified by searching the MS/MS spectra against the *Riftia* host and symbiont database (Hinzke *et al.*, 2019), which was constructed from the host transcriptome and three symbiont genome assemblies, i.e., NCBI project PRJNA60889 (endosymbiont of *Riftia*

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pachyptila (vent Ph05)), NCBI project PRJNA60887 (endosymbiont of *Tevnia jerichonana* (vent Tica)), and JGI IMG Gold Project Gp0016331 (endosymbiont of *Riftia pachyptila* (vent Mk28)). The cRAP database containing common laboratory contaminants (The Global Proteome Machine Organization) was added to complete the database. Database search was conducted using Proteome Discoverer v. 2.0.0.802 with the Sequest HT node as described in Kleiner *et al.* (2018) with a false discovery rate of 5% (FidoCT Protein Validator node, q-value <0.05).

To systematically screen the *Riftia* symbiont metagenome for dissimilatory sulfur metabolism-related proteins, candidates identified in different studies were searched against the *Ca*. E. persephone metaproteome database using bioedit (Hall, 1999; Supplementary Table S9). Host proteins were additionally annotated using the same tools as in Hinzke *et al.* (2019). Symbiont hydrogenase sequences were classified using HydDB (Søndergaard *et al.*, 2016).

Statistical evaluation of metaproteomics data and abundance quantification

Filtering and normalization

For samples from sulfur-rich specimens, four replicates for each of the four size classes were used (resulting in 16 samples); for analysis of symbionts from sulfur-depleted specimens, three replicates were available per size class (giving a total of 12 samples). For comparisons of protein abundance (i) across different samples, e.g., to determine a protein's abundance trend across gradient fractions XS to L, edgeR-RLE-normalized spectral counts were calculated (see below), while (ii) %orgNSAF values were used for abundance comparisons of different proteins within one sample, e.g., to determine the "most abundant" proteins in a sample.

(i) To allow for comparisons of protein abundance across different samples, spectral count data were first filtered so that they included only proteins that had at least five spectral counts in at least four out of 16 (S-rich specimens) or three out of 12 (S-depleted specimens) samples. The filtered dataset was then normalized using Relative Log Expression (RLE) normalization with the package edgeR v.3.24.3 (Robinson *et al.*, 2010) in R v. 3.5.1 (R Core Team, 2018; Supplementary Table S3a). The filtering and normalization step was included to avoid biasing

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the analysis towards symbiont proteins that were only detected in the high-density fractions M and L (enriched in larger symbiont cells), but which were absent in fractions of lower density (XS and S, containing primarily smaller cells). Fractions S and particularly XS contained relatively more host proteins, leading to a lower total number of detectable symbiont proteins. (Note that these values were not normalized to protein size, so that a protein's relative abundance changes can be followed across different samples, but abundances cannot be compared between proteins). We tested for significant differences in symbiont protein abundance between individual gradient factions (representing enrichments of different cell size classes) using two methods, i.e., hierarchical and profile clustering and random forests (see below).

(ii) To be able to compare relative symbiont protein abundances within samples and to identify particularly abundant proteins, %orgNSAF values were calculated from unfiltered spectral counts by normalization to protein size and to the sum of all proteins in a sample (Zybailov *et al.*, 2006, Mueller *et al.*, 2010). %orgNSAF values give an individual protein's percentage of all proteins in the same sample (Supplementary Table S3b). Note that %orgNSAF values cannot be compared across different samples, due to the unequal number of total host and symbiont proteins in different samples.

STEM analysis

For protein expression profile clustering, we employed the Short Time Series Expression Miner (STEM; Ernst and Bar-Joseph, 2006) v. 1.3.11., which fits gene expression profiles in ordered short series datasets (like the cell cycle stages of *Ca.* E. persephone), to model profiles representing different expression patterns. Filtered and RLE-normalized data were log-normalized, repeat data were defined to be from different time points and data were clustered using the STEM method with default options. For STEM filtering, the minimum correlation between repeats and the minimum absolute expression change were set to 0.5. All permutations were used. For correction, the false discovery rate (FDR) was set to 0.05. Profiles were clustered with a minimum correlation percentile of 0.5. Other parameters were left at default values. Proteins which were assigned to model profiles, i.e., all proteins which were not

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

316 removed by filtering and showed a consistent trend in all replicates, were used for further
317 analysis. This means that differences in protein abundance patterns were considered
318 significant if proteins were detected with a consistent abundance trend across all replicates
319 (increase, decrease or alternating increase and decrease of abundance from fraction XS to L).

320 *Random forests*

321 For random forest analysis, we used the ranger package v. 0.10.1 (Wright and Ziegler, 2015) in
322 R v. 3.5.1 (R Core Team, 2018). Random forests are a machine learning technique, which can
323 be used to find the variables – here proteins – that allow to predict which datasets or samples
324 are similar (and which ones are not; Degenhardt *et al.*, 2019). For variable importance
325 calculation, we employed the method from (Janitza *et al.*, 2018) as implemented in the ranger
326 package. This method uses a heuristic approach, where a null distribution for p-value
327 calculation is generated based on variables with importance scores of zero or negative
328 importance scores. For pairwise comparisons, the data set was subjected to an additional
329 filtering step, so that only proteins with a minimum of five spectral counts in at least six out of
330 eight (S-rich) or four out of six (S-depleted) samples were included. The comparison of all 16
331 S-rich samples included only such proteins which had a minimum of five spectral counts in at
332 least five samples, and the comparison of all 12 S-depleted samples included only proteins with
333 five or more spectral counts in a minimum of four samples. The filtered and RLE-normalized
334 data were used for random forest analysis as follows: 2,000 forests with 10,000 trees per forest
335 were grown for pairwise comparisons as well as for comparisons including the samples
336 representing all four size classes. Proteins which had a p-value below 0.05 in >90% of the
337 forests were included in further analyses.

338 *Significant differences in protein abundance*

339 Proteins that showed significant abundance differences as determined by STEM analysis or by
340 random forest analysis (see above) or by both methods were included in a common list. Please
341 note that this approach of determining significant protein abundance differences was not
342 based on individual p-values. For proteins with significant abundance differences, we clustered
343 the z-scored mean abundances using hierarchical clustering (Pearson correlation, complete

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

linkage) in R to visualize their abundance trends (Supplementary Figure S1). For this purpose, we employed the R base package stats (R Core Team, 2018) as well as the packages cluster (Maechler *et al.*, 2018) and ComplexHeatmap (Gu *et al.*, 2016). For comparison of S-rich and S-depleted symbionts of the same size class, we used the R package edgeR v. 3.24.3 (Robinson *et al.*, 2010), which uses a Bayes-moderated Poisson model for count data analysis, with an overdispersion-adapted analogon to Fisher's exact test for detecting differentially expressed genes (Robinson *et al.*, 2010).

Host proteins

Host proteins which were more abundant in symbiont-enriched fractions as compared to the non-enriched trophosome homogenate are candidates for direct host-symbiont interaction, as they might be secreted into symbiont compartments or even physically associated with symbiont cells. For evaluation of host protein enrichment, we used fractions XS and S enriched in the two smallest symbiont size classes (i.e. fractions collected from the upper part of the gradient). As the larger gradient fractions sometimes contained the gradient pellet, in which host proteins can also accumulate when host tissue fragments are pelleted, these fractions were not used for host protein analysis. Comparisons of relative host protein abundance between trophosome homogenate and fractions XS and S were performed using the R package edgeR v. 3.24.3. Spectral count data were filtered to include only proteins which had at least five spectral counts in at least four (for S-rich specimens) or three (in S-depleted specimens) samples and RLE-normalized abundance values were compared between samples. Proteins which were significant in the edgeR comparison and had a higher mean RLE-normalized abundance in fractions XS and S than in the homogenate sample were included in functional analysis.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (ProteomeXchange – ProteomeCentral) via the PRIDE partner repository (Vizcaíno *et al.*, 2016) with the dataset identifier PXD016986. The dataset will be released upon acceptance of the manuscript in a peer-reviewed journal.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

Enrichment of individual symbiont cell sizes by gradient centrifugation

Our rate-zonal gradient centrifugation approach allowed us to enrich distinct symbiont cell sizes from *Riftia* trophosome tissue. Based on CARD-FISH microscopy, we defined four size ranges (Figure 1): very small symbiont cells (≥2.0 – <3.9 µm diameter), small (≥3.9 – <5.3 µm), medium (≥5.3 – <6.8 µm) and large symbiont cells (≥6.8 – 20.0 µm; see also Supplementary Results and Discussion A). For subsequent comparative metaproteomic analyses, we chose the gradient fractions that were most enriched in one of these cell size ranges. In the following, these four gradient fractions are referred to as XS (containing the highest percentage of very small symbiont cells) to L (containing the highest percentage of large cells). The enrichment procedure was highly reproducible, particularly for symbionts isolated from sulfur-rich trophosome tissue (Figure 1).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

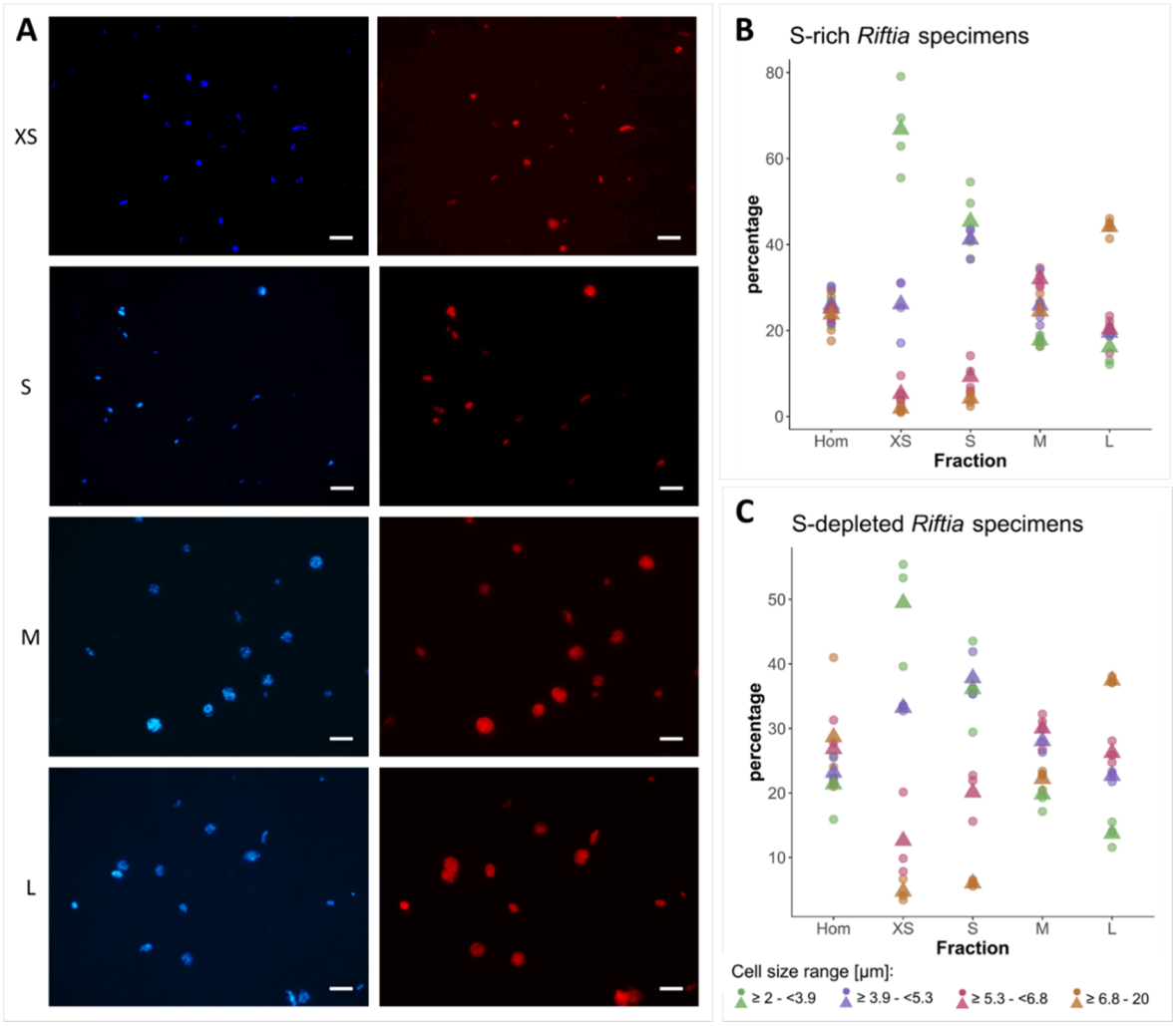


Figure 1: A) Catalyzed reporter deposition-fluorescence *in situ* hybridization (CARD-FISH) images of *Riftia* symbiont cells after density gradient centrifugation of trophosome homogenate. After the enrichment procedure, small bacterial cells had accumulated in the upper, less dense gradient fractions (top), while larger symbionts were enriched in the lower, denser fractions (bottom). Left: DAPI staining, right: 16S rRNA signal. For better visibility, brightness and contrast were adjusted in all images. **B and C)** Symbiont cell size distributions in individual gradient fractions. While all cell size groups were roughly equally abundant in non-enriched trophosome homogenate (Hom), fraction XS had the highest percentage of symbiont cells in the size range 2.0 µm - 3.9 µm, fraction S contained most symbiont cells of 3.9 µm – 5.3 µm, etc. Gradient centrifugation was performed using four biological replicates (n=4) of sulfur-rich trophosome tissue (B) and three biological replicates (n=3) of sulfur-depleted trophosome tissue (C). For an overview of which gradient fractions were chosen as fractions XS, S, M, and L in all samples see Supplementary Table S1. Dots: individual % values, triangles: average % values.

Symbiont DNA quantification

Flow cytometry and fluorescence-activated cell sorting (FACS) indicated that DNA content in large *Riftia* symbionts is up to 10-fold higher compared to small symbionts. To identify distinguishable bacterial cell populations, we examined Syto9-stained cells in *Riftia*

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

trophosome homogenate and in gradient fractions from the upper and lower parts of the gradient (enriched in smaller and larger symbionts, respectively) with regard to their light scattering properties. Forward scatter (FSC) and side scatter (SSC) usually correlate with cell size and cell granularity, respectively (Bouvier *et al.*, 2001, Tracy *et al.*, 2010). Amongst a number of particle groups with different properties (see also Supplementary Results and Discussion C), we found two populations, 1 and 2, which were abundantly detected in non-enriched trophosome homogenate, but showed very dissimilar frequencies in fractions enriched in larger or smaller symbionts (Figure 2, Supplementary Figure S2): While population 1, which exhibited relatively lower FSC and SSC signals (indicative of smaller cell size and lower cell complexity), was highly abundant in fractions enriched in smaller symbionts, this population was notably less prominent in fractions enriched in larger symbionts. Simultaneously, population 2, which gave higher FSC and SSC signals (indicative of larger cell size and higher complexity), was highly abundant in fractions enriched in larger symbionts but nearly absent in gradient fractions enriched in smaller symbionts. This suggests that populations 1 and 2 consist of smaller and larger symbionts, respectively. This assumption was verified by FACS-separation of both populations from trophosome homogenate, and examination of the sorted cell suspensions by fluorescence microscopy along with unsorted enriched gradient fractions and homogenate samples for reference (Supplementary Figure S2). For quantification of DNA in smaller and larger symbionts, we compared median fluorescence intensities (MFI) per particle between population 1 and 2 in non-enriched homogenate and in enriched gradient fractions. In all sample types, MFI per particle was notably lower in population 1 (between 186 and 1,994 relative fluorescence units, rfu) than in population 2 (2,712 – 10,723 rfu). On average, MFI was 9.7-fold higher in population 2 than in population 1 (Supplementary Table S8).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

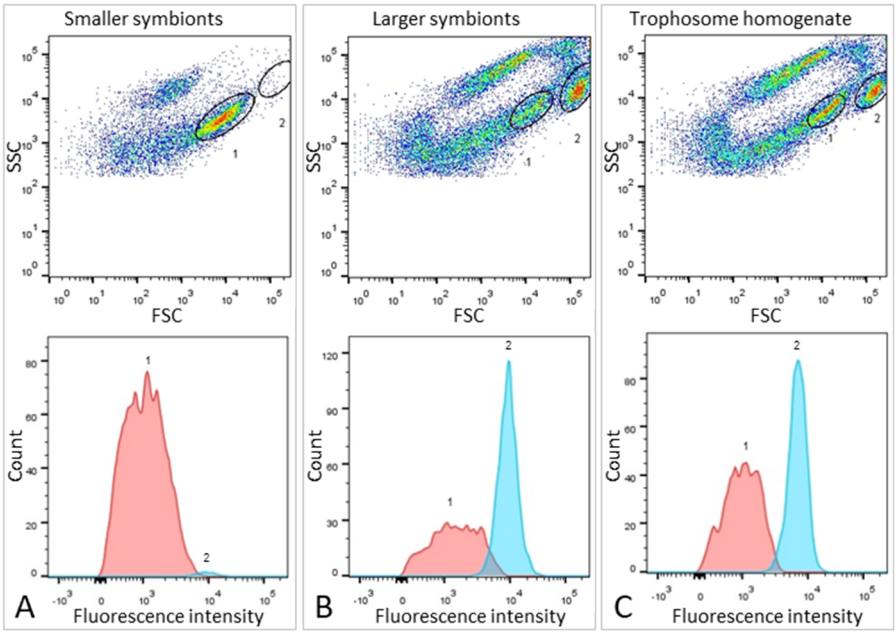


Figure 2: Flow cytometry for DNA quantification of *Riftia* symbionts. **A)** Dot plot of forward scatter (FSC) and side scatter (SSC), and histogram with fluorescence signal counts and fluorescence intensity per particle of a gradient fraction enriched in smaller symbionts. **B)** Gradient fraction enriched in larger symbionts. While cell population 1 was more prominent in A), population 2 was almost exclusively detected in B), and both populations were present in non-enriched trophosome homogenate (**C**), indicating that population 1 corresponds to smaller symbionts, whereas population 2 corresponds to larger symbiont cells. Cells were stained with Syto9 and median fluorescence intensity (MFI) per particle at wave length 530/30 nm was used as a measure of cellular DNA content (see Methods and Supplementary Table S8 for more details). This analysis was based on two *Riftia* specimens with medium sulfur content.

Protein identifications and relative protein abundance

We identified a total of 1,946 symbiont proteins across all sample types, including the four gradient fractions XS – L and non-enriched homogenate from both, sulfur-rich and sulfur-depleted *Riftia* specimens (Supplementary Table S6). Our sample fractionation by gradient centrifugation thus facilitated detection of around 60% of the symbiont's theoretical proteome, which encompasses 3,182 proteins in PRJNA60889, and yielded substantially higher symbiont protein identification rates than non-enriched trophosome homogenate samples alone (1,223 total symbiont protein identifications). After stringent filtering and normalization, a subset of 1,212 symbiont proteins from gradient fractions XS – L was included in statistical analysis using abundance profile clustering and random forests (Supplementary Table S3). A total of 465 proteins showed significant differences in relative abundance in S-rich and/or S-depleted samples (Supplementary Figure S1; note that the term "significant" denominates trends that were consistent across all replicates in the context of our statistical approach). In Figure 3 and

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Table S3, proteins that showed such significant changes in relative abundance are marked with asterisks.

Of all proteins with significant abundance changes, 56% (261 proteins) followed a clear, continuous abundance trend from fraction XS to L or *vice versa*, that is, protein abundance increased or decreased with increasing symbiont cell size (Supplementary Table S3). For the majority of symbiont proteins, abundance trends in samples obtained from sulfur-rich (energy-rich) and sulfur-depleted (energy-depleted) trophosome tissue were highly similar. Very few proteins were detected only in sulfur-rich samples (61 of 1,212 proteins) or exclusively in sulfur-depleted samples (77 proteins). For a discussion of specific differences observed between symbionts from energy-rich and energy-starved trophosome tissue, see Supplementary Results and Discussion B.

Symbiont protein functions

Cell cycle, DNA topology, replication and repair

Proteins involved in the bacterial cell cycle and in DNA topology, -replication and -repair were differentially expressed across fractions XS to L (Figure 3, Supplementary Table S4a). While the cell division protein FtsZ, DNA gyrase and DNA-binding proteins decreased significantly in abundance from fraction XS to L, abundance of other cell division-related proteins (e.g., FtsE, MreB, division inhibitor SlmA), and of proteins involved in DNA replication (e.g., DNA ligase, DNA polymerase) and repair (e.g., UvrAB) increased. Interestingly, FtsZ abundance was very low in S-depleted fractions, so that it was excluded from statistical analysis in these samples (see Supplementary Results and Discussion).

Chaperones and stress proteins

Many chaperones and other proteins involved in protein folding, as well as oxidative stress-related proteins were detected with significantly decreasing abundance from fraction XS to L, including (amongst others) the proteases ClpB, ClpP, GroEL, the abundant alkyl hydroperoxide reductase AhpC, superoxide dismutase SodB, and rubrerythrin (Figure 3, Supplementary Table S4b).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3: Abundance trends of selected Endoriftia proteins of various functions in the four fractions XS to L in sulfur-rich (S-rich) and sulfur-depleted (S-depl) *Riftia* specimens. Trends are indicated by color shades from light green/light grey (lowest protein abundance across all four fractions) to dark green/dark grey (highest abundance across all four fractions; note that colors do not allow comparison of protein abundance between proteins). Abundance values are based on statistical evaluation of four biological replicates (S-rich) and three biological replicates (S-depl). Proteins marked with asterisks show statistically significant trends, i.e., differences that are consistent across all replicates in S-rich or S-depl specimens (or both). White cells indicate that this protein was not detected in this sample or too low abundant to be included in statistical analyses. For an overview of all identified symbiont proteins and their relative abundances and for a summary of protein abundance trends sorted by metabolic category see Supplementary Tables S3 and S4, respectively. Accession numbers refer to NCBI/JGI entries. SU: subunit, DUF: domain of unknown function, ss: single-stranded, transcr: transcription, assoc: associated, dep: dependent, HP: hypothetical protein, put: putative, oligopep: oligopeptide, ppc: periplasmic component, DHG: dehydrogenase: RubisCO: ribulose-1.5-bisphosphate carboxylase/oxygenase, Ox: oxidase, OxRed: oxidoreductase, PEP: phosphoenolpyruvate, fcc: flavocytochrome c, rhd: rhodanese, resp: respiratory, cat: catalytic, Vit: vitamin and cofactor metabolism.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Transport

Outer membrane proteins such as two porins and TolBC showed significant abundance differences between the fractions, with highest relative abundance in fraction XS and lowest abundance in fractions L. Porin EGV52132.1 (Por1) was the most abundant symbiont protein throughout all sample types (Figure 3, Supplementary Table S4c). On the other hand, all five detected tripartite ATP-independent periplasmic (TRAP) transporter subunits and ten out of 13 ABC transporter components were relatively more abundant in fraction L (see also Supplementary Results and Discussion D).

Central metabolism

Carbon metabolism: Several tricarboxylic acid (TCA) cycle enzymes (e.g., Icd, Mdh), as well as enzymes of the pentose phosphate pathway (e.g., Rpe, RpiA) were detected with decreasing abundances from fraction XS to L (Figure 3, Supplementary Table S4d). In contrast, the key enzymes of the two CO_2-fixing pathways, Calvin cycle (RubisCO, CbbM) and rTCA cycle (ATP-citrate lyase, AclA; oxoglutarate oxidoreductase, KorAB), as well as most of the gluconeogenesis-related (e.g., PckG), and glycogen metabolism-related enzymes (e.g., GlgA, GlgP) increased in abundance from fraction XS to L.

Chemotrophy: Many sulfide oxidation-specific proteins, including both subunits of the abundant key enzyme adenylylsulfate reductase AprAB, as well as proteins involved in sulfur storage (sulfur globule proteins) had their highest abundance in fraction XS or S and their lowest abundance in fraction M or L (Figure 3, Supplementary Table S4e, Supplementary Figure S4). In contrast, thiosulfate oxidation-related proteins like SoxZ, SoxL and other rhodanese-like proteins were detected with significantly increasing abundance from fraction XS to fraction L. Four additional Sox proteins, i.e., SoxA, SoxB, SoxW and SoxY, which were detected at very low abundances across the sample types (and were therefore excluded from statistical analysis), were identified in fraction M and L, but were completely absent from fraction XS (Supplementary Table S3). Three proteins involved in energy generation by hydrogen oxidation, HyaB, HypE and GlpC, were also detected with increasing abundance from fraction XS to fraction L.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Nitrogen metabolism: Relative abundance of all three respiratory membrane-bound nitrate reductase subunits, NarGHI, decreased significantly from fraction XS to L, as did abundance of glutamine synthetase GlnA (Figure 3, Supplementary Table S4f, Supplementary Results and Discussion E). On the other hand, various other denitrification-related proteins (such as nitrite reductase NirS, nitrous oxide reductase NosZ, and nitrate/nitrite signal transduction systems) and glutamate dehydrogenase GdhA showed relatively higher abundances in fraction L (or M) than in fraction XS. The same trend was observed for the periplasmic nitrate reductase components NapC and NapH. Moreover, NapG, another NapH copy, the nitric oxide reductase subunit NorB, nitric oxide reductase activation protein NorQ, and the putative assimilatory nitrite reductase subunit NirB, whose overall abundances were too low to include them in statistical analysis, were only detected in fraction M and/or L.

Other categories

50 (75%) of the 67 proteins involved in cofactor- and vitamin synthesis in S-rich samples had their highest abundance in fraction M or L (Figure 3, Supplementary Table S4g). Also, of the 33 identified tRNA ligases and tRNA synthetases, 25 (75%) were most abundant in fraction M or L (in S-rich samples, Supplementary Table S4h).

Symbiosis-specific host proteins

Our density gradient fractionation procedure allowed not only for the identification of symbiont proteins with differential abundance across different Endoriftia size ranges, but also enabled us to single out host proteins that are potentially involved in direct interactions with the symbionts. As host proteins that are attached to the symbionts are pulled down with the symbiont cells during gradient centrifugation, these proteins should be significantly more abundant in symbiont-enriched fractions compared to the non-enriched trophosome homogenate (Figure 4, Supplementary Table S5). Besides many ribosomal and mitochondrial host proteins, which were also enriched, putatively symbiont-associated host proteins included the host's peptidoglycan-recognition protein SC1a/b, beta carbonic anhydrase 1, digestive proteins involved in protein- and carbohydrate degradation, e.g. acid phosphatase, digestive

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

proteases and glycan degradation enzymes, as well as hypoxia up-regulated proteins, a thiosulfate sulfurtransferase, and transmembrane protein 214-B.

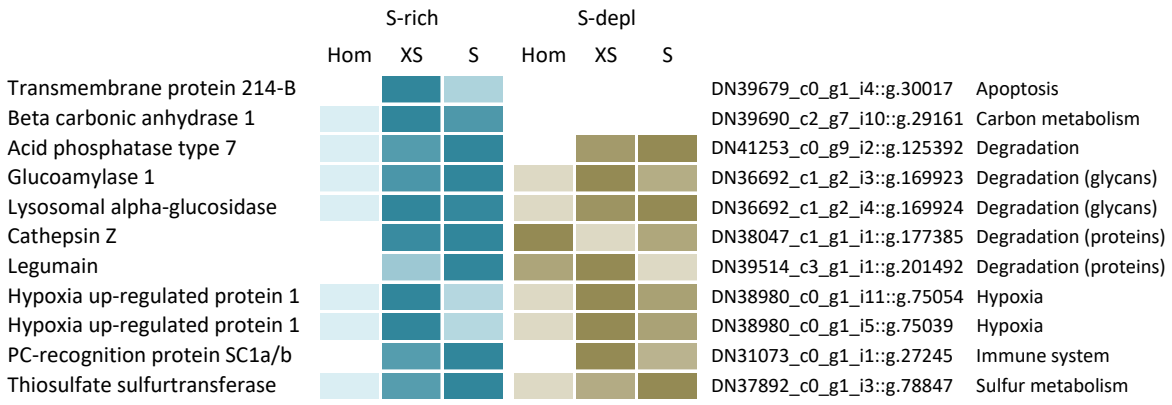


Figure 4: Selected *Riftia* host proteins with significantly higher relative abundance in the symbiont-enriched fractions XS and S compared to the non-enriched trophosome tissue homogenate (Hom) in sulfur-rich (S-rich) and sulfur-deleted (S-depl) *Riftia* specimens. Relative abundance trends are indicated by color shades from light blue/light brown (lowest protein abundance across the three sample types) to dark blue/dark brown (highest abundance), based on mean values from four biological replicates (S-rich) and three biological replicates (S-depl). (Note that colors do not allow comparison of protein abundance between proteins). Accession numbers refer to the combined host and symbiont database used for protein identification in this study (see Methods). For a complete list of host proteins with significantly higher abundance in fractions XS and S (compared to Hom) see Supplementary Table S5. This comparison includes only the symbiont-enriched fractions XS and S, but not fractions M and L, because these latter fractions were more likely to be contaminated by non-symbiosis-specific host proteins from host tissue fragments pelleted during centrifugation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

Symbiont growth and differentiation

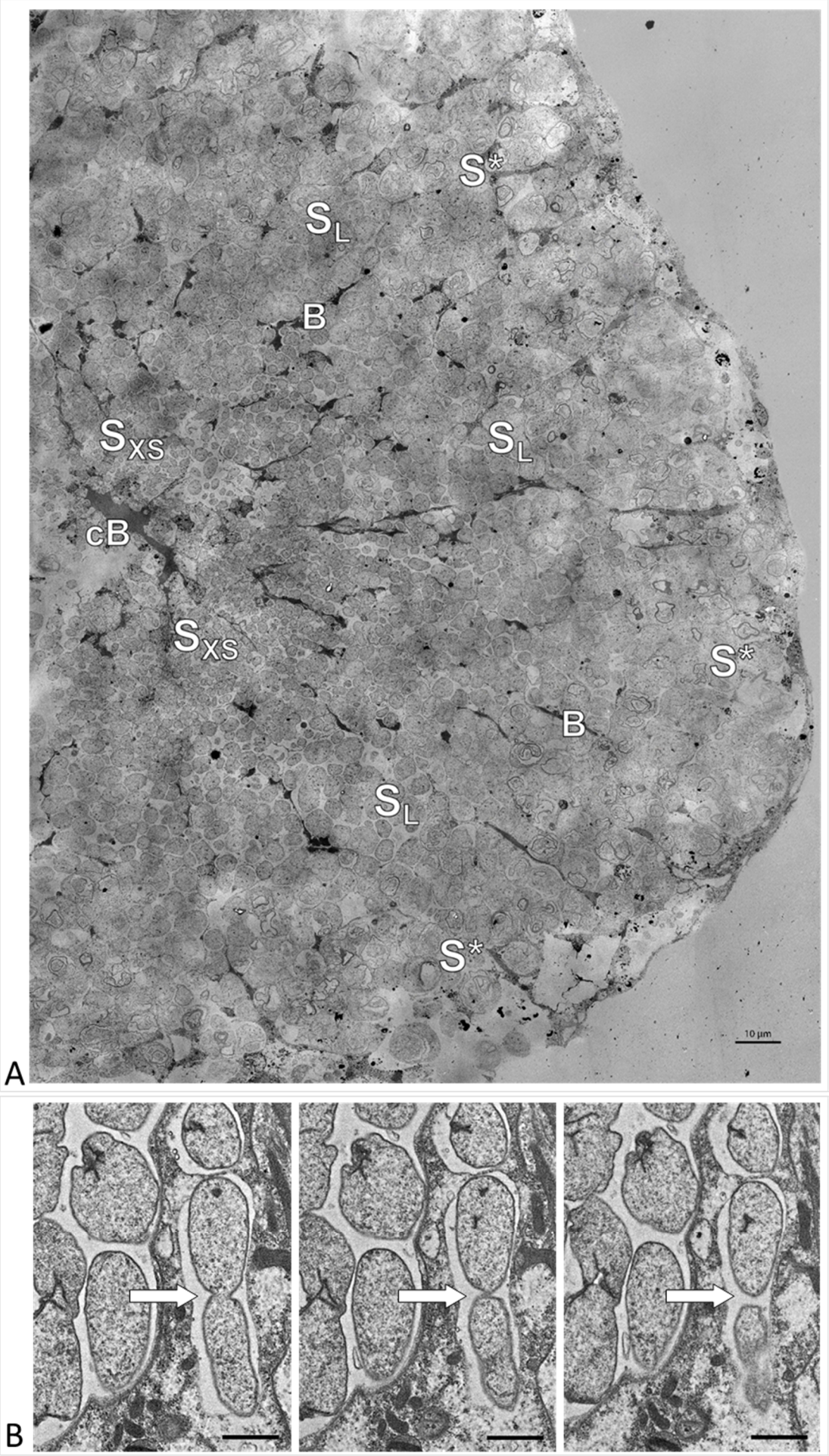
Cell division plays a more prominent role in small symbionts

As indicated by the significant decrease in abundance of the cell division key protein FtsZ from fraction XS to fraction L, small Endoriftia are more engaged in cell division than larger symbionts. In accordance with the microscopy-based hypothesis of Bright and Sorgo (2003), the smallest symbionts, which are *in situ* localized in the trophosome lobule center (Figure 5), thus apparently function as stem cells of the symbiont population. During cell division, FtsZ forms the Z ring, to which the other division-related proteins are successively recruited (reviewed in Weiss, 2004). Cell size and cell division therefore likely depend on the amount of FtsZ available (Chien *et al.*, 2012). This correlation is, for example, also reflected by a decrease of FtsZ concentration during differentiation of vegetative cells into non-dividing larger heterocysts in the cyanobacterium *Anabaena* (Klint *et al.*, 2007). Interestingly, while FtsZ abundance decreased across fractions, many other proteins which interact with FtsZ during cell division were detected with increasing abundance from fraction XS to L. This indicates that these proteins are also involved in processes other than cell division, e.g., in determining cell shape and stabilization. ZapD, for example, is involved in FtsZ filament organization, and its overexpression leads to cell filamentation (Durand-Heredia *et al.*, 2012). DamX overexpression, too, was observed to induce filamentation in *E. coli* (Lyngstadaas *et al.*, 1995), while overexpression of the cell shape determination protein CcmA in *E. coli* and *P. mirabilis* lead to enlarged, ellipsoidal cells (Hay *et al.*, 1999), and FtsEX is required for cell elongation rather than cell division in *B. subtilis* (Domínguez-Cuevas *et al.*, 2013). The actin homolog MreB is pivotal for rod-shape formation in bacteria and for cell stiffness in *E. coli*, could negatively regulate cell division, and participates in chromosome segregation (Wachi and Matsuhashi, 1989, Kruse *et al.*, 2006, Wang *et al.*, 2010, reviewed in Reimold *et al.*, 2013). In large Endoriftia, these proteins might therefore be involved in stabilizing growing symbiont cells. SlmA, which was only detected in fractions M and L in our study, was shown to disassemble FtsZ polymers, thus acting as a cell division inhibitor (Cho *et al.*, 2011), which

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

595 supports the idea of relatively less cell division in large *Riftia* symbionts. Although Endoriftia's
596 major cell division protein FtsZ was notably (1.75x) less abundant in fraction L (compared to
597 fraction XS), it was not completely absent. This may indicate that cell division is reduced with
598 increasing cell size, but not abandoned altogether, or it may point to additional FtsZ functions,
599 besides cell division (as also suggested for *Anabeana* (Klint *et al.*, 2007) and *E. coli* (Thanedar
600 and Margolin, 2004)).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



601

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 5: A) Electron micrograph of a cross section through a *Riftia* trophosome lobule. Surrounding an efferent central blood vessel (cB), small symbiont cells (S_{XS}) are visible in bacteriocytes in the central lobule zone. Symbiont cell size increases towards the periphery of the lobule (S_L: large symbiont cells). In the outermost bacteriocytes, symbiont cells are digested by host enzymes (S^*). Bacteriocytes are interspersed with smaller blood vessels (B), which facilitate blood flow from the lobule periphery to the lobule center (Felbeck and Turner, 1995). The image was assembled from 50 individual transmission electron micrographs of a trophosome section from a *Riftia* specimen with sulfur-depleted trophosome. The full resolution image is available as Supplementary Figure S5. Contrast and brightness were adapted. **B**) Cell division in small *Riftia* symbionts in the trophosome lobule center of a *Riftia* specimen with sulfur-rich trophosome. All micrographs show the same dividing Endoriftia cell in three subsequent tissue sections, revealing that both daughter cells are still connected, but are about to be separated (arrow). Scale bar: 1 µm.

Large symbionts have more genome copies and less compact chromosomes

Endoriftia's differentiation into large, non-dividing (but still replicating) cells coincides with endoreduplication cycles and an increase in genome copy number, as indicated by our flow cytometry analysis (Figure 2, Supplementary Figure S2, Supplementary Table S8). This observation is in agreement with earlier findings of Bright and Sorgo (2003), who noted more than one chromatin strand-containing area in large coccoid *Riftia* symbiont cells in electron microscopy images, whereas small rods and cocci featured only one chromatin strand area. The idea of endoreduplication in larger *Riftia* symbionts is also supported by the observation that large symbiont cells, which apparently divide less frequently than smaller cells (see above), still actively replicate DNA, as indicated by high abundances of DNA ligase and DNA polymerase III in fraction L. The observed decreasing abundance of DNA gyrase GyrAB with increasing cell size additionally corroborates this idea, as type II topoisomerases such as gyrase are not only involved in supercoiling and initiation of DNA replication (Levine *et al.*, 1998, Nöllmann *et al.*, 2007), but also essential for decatenation of newly replicated chromosomes in bacteria (Steck and Drlica, 1984, Guha *et al.*, 2018). Moreover, inhibition of topoisomerase II in eukaryotes leads to endoreduplication and polyploidy (Cortés and Pastor, 2003, Cortés *et al.*, 2003). Polyploidy in thiotrophic symbionts was also observed in the lucinid bivalve *Codakia orbicularis*, where larger symbiont cells contained more than four genome copies, while smaller cells had only one genome copy (Caro *et al.*, 2007), and in ectosymbionts of *Eubostrichus* nematodes, in which up to 16 nucleoids per large symbiont cell were reported (Polz *et al.*, 1992, Pende *et al.*, 2014). Moreover, also terminally differentiating *Rhizobia* undergo endoreduplication cycles (Mergaert *et al.*, 2006), and high genome copy numbers

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

have been reported for various bacterial insect symbionts, e.g., of aphids, cockroaches and sharpshooters (Komaki and Ishikawa, 2000, López-Sánchez *et al.*, 2008, Woyke *et al.*, 2010), suggesting that polyploidy is common in symbiotic bacteria. Possibly, enlarged polyploid cells might increase the metabolic activity and/or fitness of the Endoriftia cells: In *E. coli*, a *mreB* point mutation led to increased cell size, which gave the cells a measurable fitness advantage in presence of certain carbon sources (Monds *et al.*, 2014). Moreover, polyploidy was suggested to provide evolutionary advantages like a low mutation rate and resistance towards DNA-damaging conditions in haloarchaea (Zerulla and Soppa, 2014). In plants, endoreduplication is common and might increase transcription and metabolic activity of the cells (Kondorosi and Kondorosi, 2004), leading to enhanced productivity (Sattler *et al.*, 2016). More generally, in symbiotic associations, where the bacteria are stably and sufficiently provided with carbon and energy sources, the advantages of polyploidy might be greater than the associated costs (Angert, 2012).

Higher genome copy numbers in large symbionts seem to be accompanied by a lower degree of DNA condensation, compared to small Endoriftia, as indicated by notably lower abundances of the histone-like DNA-binding proteins HU (HupB) and integration host factor (IHF, IhfAB), and of DNA gyrase GyrAB in fraction L, compared to XS. Bacterial histone-like DNA-binding proteins like HU and IHF structure the chromosome and modulate the degree of supercoiling (reviewed in Dorman and Deighan, 2003). In *E. coli*, absence of HU leads to unfolding of the chromosome and cell filamentation (Dri *et al.*, 1991), and unspecific DNA-binding by IHF was shown to contribute to DNA compaction (Ali *et al.*, 2001). Moreover, bacterial DNA gyrase was also suggested to be involved in nucleoid compaction in *E. coli* (Stuger *et al.*, 2002). Co-occurrence of endoreduplication and decondensated DNA is also known in plant cells (Kondorosi and Kondorosi, 2004). As decondensation occurs in actively transcribed DNA regions (Wang *et al.*, 2014), it might facilitate protein synthesis and metabolic activity in large Endoriftia.

Since DNA condensation may function as a DNA protection mechanism (Ohniwa *et al.*, 2006, Mukherjee *et al.*, 2008, Yoshikawa *et al.*, 2008, Takata *et al.*, 2013), less condensed DNA might

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

be more prone to various kinds of damage and require the enhanced expression of DNA repair mechanisms. This would explain the observed higher abundance of several DNA repair proteins in fraction L, which was enriched in larger, older symbiont cells with (presumably) larger quantities of less condensed DNA, compared to the smaller symbiont cells. RadA, RdgC, RecCN, UvrAB, and Mfd, which are known to be involved in DNA recombination and repair in many bacteria (Kowalczykowski, 2000, Beam *et al.*, 2002, Tessmer *et al.*, 2005, Drees *et al.*, 2006, Truglio *et al.*, 2006, Deaconescu *et al.*, 2007), may compensate for this elevated vulnerability. In eukaryotes, chromatin decondensation was shown to facilitate access of the DNA damage response to double strand breaks, thus allowing for more efficient repair (Murga *et al.*, 2007).

Small symbionts may be exposed to elevated stress levels

Small symbionts might experience cell division-related or host-induced stress in the early phase of their cell cycle, as indicated by elevated levels of symbiont chaperones and stress response proteins, as well as of reactive oxygen species (ROS) scavengers in fraction XS. This is in line with observations in *Caulobacter crescendus*, where the DnaK-DnaJ and GroEL-GroES systems are crucial for cell division (Susin *et al.*, 2006), and in *E. coli*, where the protease ClpXP and the RNA chaperone Hfq are probably involved in cell division as well (Camberg *et al.*, 2009, Zambrano *et al.*, 2009). Interestingly, like the putative Endoriftia stem cells, eukaryotic embryonic stem cells also feature high levels of chaperone expression and stress tolerance (Prinsloo *et al.*, 2009). Although the reason for this congruence is yet unknown, possibly, cell division-related processes might require elevated levels of chaperones and stress proteins, e.g. to ensure correct assembly of all parts of the division machinery or to counteract some sort of yet to be determined host-induced stress.

Possibly, such host-induced stress may also involve the production of ROS in symbiont-containing bacteriocytes, similar to animal and plant hosts, which generate ROS to defend themselves against pathogenic bacteria (Heath, 2000, Lynch and Kuramitsu, 2000, D'Haeze and Holsters, 2004). Small symbionts, which are relatively loosely packed in their host cell vesicles (Figure 5A) and have a comparatively high surface-to-volume ratio, might be

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

692 particularly exposed to this presumptive ROS stress, while larger symbionts, which are more
693 tightly packed, may face lower ROS levels. This would explain the observed higher abundance
694 of the ROS scavengers rubrerythrin (Rbr2), superoxide dismutase (SodB), and
695 alkylhydroperoxide reductase (AhpC) in small symbionts. In line with this assumption, a
696 superoxide dismutase and also the chaperones ClpB, HtpG, and DnaK were suggested to be
697 involved in ROS protection in *Serratia symbiotica* (Renoz *et al.*, 2017), and ClpB protease
698 expression has been shown to increase during oxidative stress in the intracellular pathogen
699 *Francisella tularensis* (Twine *et al.*, 2006).

700 Interestingly, neither S-depleted nor S-rich samples showed indications of a strong bacterial
701 stress response in fraction L, indicating that imminent digestion by the host poses no particular
702 stress to the large symbionts. Possibly, bacterial degradation happens too fast to elicit a stress
703 response, or a stress response is suppressed during symbiosis, either by the symbionts
704 themselves or by the host via a yet to be determined mechanism.

705 *Host-microbe interactions may be particularly important in small Endoriftia*

706 Abundant Endoriftia membrane proteins might play a key role in host interaction in small
707 symbionts. Particularly, the high and differential abundance of porin Sym EGV52132.1, the
708 most abundant symbiont protein in all fractions, which was nearly 3-times more abundant in
709 fraction XS (11.7 %orgNSAF) than in fraction L (4.0 %orgNSAF), suggests that this protein may
710 be of varying relative importance throughout the symbiont's differentiation process. Porins are
711 water-filled channels in the outer membrane, through which small hydrophilic molecules can
712 diffuse (Fernández and Hancock, 2012). In the oyster pathogen *Vibrio splendidus*, the porin
713 OmpU serves as adhesin or invasin and is involved in recognition by the host cell (Duperthuy
714 *et al.*, 2011), while in *Neisseria gonorrhoeae*, a porin inhibits phagocytosis by human immune
715 cells (Mosleh *et al.*, 1998, Lorenzen *et al.*, 2000). Interestingly, the phagocytosis-inhibiting
716 action of *N. gonorrhoeae* porin apparently involves interference with the host's oxidative
717 burst, i.e., the porin allows the pathogen to evade killing by host-produced ROS (Lorenzen *et*
718 *al.*, 2000). Although the exact function of Endoriftia porin has not been elucidated yet, we
719 suggest that it may have a similar function in resistance against host stress or ROS. This would

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

be in line with elevated levels of ROS scavengers in small *Riftia* symbionts (see above). Porins are furthermore not only known to be involved in recognition by the host (e.g., in the squid symbiont *Vibrio fischeri* (Nyholm *et al.*, 2009)), but were also shown to be involved in survival in and communication with the host in other intracellular and pathogenic bacteria, rendering *Vibrio cholerae* and *Xenorhabdus nematophila* more resistant against antimicrobial compounds (Mathur and Waldor, 2004, van der Hoeven and Forst, 2009). As *Riftia* trophosome tissue has antimicrobial effects (Klose *et al.*, 2016), and considering that *Riftia* might employ histone-derived antimicrobial peptides to modulate the symbiont's cell division (Hinzke *et al.*, 2019), Endoriftia porin may enable the symbionts to reject antimicrobial compounds produced by the host. This would be of particular importance for small symbionts, as it would ensure survival of the symbiont stem cell subpopulation and sustain their division capability.

Besides porin, the symbiont's outer membrane efflux pump TolC was also most abundant in fraction XS, suggesting that it may play a similar role in host interaction or persistence. TolC is a versatile export protein of Gram-negative bacteria, which interacts with different transporters of the cytoplasmic membrane to export proteins and drugs (reviewed in Koronakis *et al.*, 2004). In *Sinorhizobium meliloti*, TolC is apparently involved in establishing the symbiosis with legumes, possibly by conferring increased stress resistance and by secreting symbiosis factors (Cosme *et al.*, 2008), while *Erwinia chrysanthemi* TolC enables re-emission of the antimicrobial compound berberine and is thus essential for *Erwinia* growth in plant hosts (Barabote *et al.*, 2003).

Microbe-host interactions with particular relevance in smaller Endoriftia may furthermore also be mediated by chaperones and stress proteins, which were most abundant in fraction XS (see above). Chaperones have been shown to play a role in host interaction and intracellular survival in several pathogenic and symbiotic bacteria. For example, DnaK appears to be essential for growth of *Brucella suis* in phagocytes (Köhler *et al.*, 1996), while HtpG seems to be involved in virulence and intracellular survival of *Leptospira* (King *et al.*, 2014), *Salmonella* (Verbrugghe *et al.*, 2015) and *Edwardsiella tarda* (Dang *et al.*, 2011). Mutations in the post-

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

748 transcriptional regulator *hfq* often lead to reduced fitness and virulence in bacterial pathogens
749 (reviewed in Chao and Vogel, 2010). Moreover, ClpB in *Listeria* is apparently specifically
750 involved in virulence (Chastanet *et al.*, 2004), as are ClpX and ClpP in *Staphylococcus aureus*
751 (Frees *et al.*, 2003). In the insect symbiont *Wolbachia*, HU beta was suggested to directly
752 interact with the host (Beckmann *et al.*, 2013). Additional symbiont proteins that may protect
753 small *Endoriftia* from host interference, and particularly so in S-depleted *Riftia* specimens,
754 included an ankyrin protein and an FK506-binding protein (see Supplementary Results and
755 Discussion B).

756 *Interaction-specific host proteins*

757 We detected a number of 'symbiosis-specific' *Riftia* proteins, which were co-enriched with
758 symbiont cells in fractions XS and/or S and may thus facilitate direct host-microbe interactions
759 or enable the host to provide optimal conditions for the symbiont. PGRPs, for example, are
760 involved in innate immunity (Kang *et al.*, 1998) and have previously been shown or suggested
761 to participate in symbiotic interactions (Troll *et al.*, 2009, Wang *et al.*, 2009, Royet *et al.*, 2011,
762 Wippler *et al.*, 2016). Since oxygen concentrations in the trophosome might be comparatively
763 low (benefitting the microaerophilic symbionts; Hinzke *et al.*, 2019), the hypoxia up-regulated
764 *Riftia* proteins we detected may present a protective adaptation of the host to these hypoxic
765 conditions. In support of this idea, Hyou1 was shown to have a protective function during
766 hypoxia in human cells (Ozawa *et al.*, 1999). Moreover, enrichment of beta carbonic anhydrase
767 1, which interconverts bicarbonate and CO_2, suggests that this host protein serves to optimally
768 provide the symbionts with CO_2 for fixation. The host transmembrane protein 214-B (TMP214-
769 B), which was exclusively detected in symbiont-enriched fractions (but not in trophosome
770 homogenate) may be involved in cell death of symbiont-containing bacteriocytes by an
771 apoptosis-related mechanism. This would be in line with our previous suggestion that
772 apoptosis-related proteins may play a role in symbiont and bacteriocyte cell death (Hinzke *et*
773 *al.*, 2019), and is further supported by the fact that TMP214-B was shown to be involved in
774 apoptosis caused by endoplasmic reticulum stress (Li *et al.*, 2013). The detection of
775 degradation proteins such as cathepsin Z, legumain, glucoamylase 1 and lysosomal alpha-

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

glucosidase in fractions XS and S furthermore implies that the host digests not only large symbiont cells in the degradative trophosome lobule zone (see Figure 5A), but that small symbionts might also be exposed to host digestion.

Metabolic diversity among symbiont size classes

Large symbionts focus on carbon fixation and biosynthesis

Highest individual abundances of various carbon fixation and biosynthesis-related enzymes as well as highest overall abundances of all biosynthetic categories (including carbon-, amino acid-, lipid-, nitrogen- and cofactor metabolism; Supplementary Table S7) in fraction L suggests that large Endoriftia cells are relatively more engaged in the production of organic material than smaller symbiont cells. In support of this idea, we observed notably higher RubisCO mRNA signal intensity in large symbiont cells than in smaller *Riftia* symbionts in our HCR-FISH analysis (Supplementary Results and Discussion D, Supplementary Figure S3). This concurs with an autoradiographic study of Bright *et al.* (2000), who observed highest ^{14}C carbon incorporation in the *Riftia* trophosome lobule periphery and lowest short-term incorporation in the lobule center. As previously suggested (Hand, 1987), these and other observed differences might be due to a biochemical gradient, which could be caused by the direction of blood flow (from the lobule periphery to the lobule center; Felbeck and Turner, 1995). This presumptive concentration gradient may lead to differential availability of inorganic carbon (and other substrates, see below), which in turn likely results in differential regulation of bacterial gene expression, such as highest abundance of CO_2 incorporation enzymes in large symbionts. Large *Riftia* symbionts thus presumably not only benefit from higher CO_2 levels, but also have more biosynthetic capacities at their disposal than small symbionts: Small Endoriftia need to maintain cell division and, consequently, invest a considerable part of their resources in growth-related processes and the expression of putative host interaction-related proteins that ensure survival of the stem cell population (see above). In contrast, large symbionts apparently divide less frequently and may be less endangered of host interference (before they reach the degenerative lobule zone) and can thus allocate more energy to production of organic material. This would eventually benefit the host, which digests

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

804 the larger symbionts at the trophosome lobule periphery. Our observation that most cofactor-
805 and vitamin metabolism-related proteins were more abundant in fractions M and/or L than in
806 fractions XS or S supports the idea of relatively more biosynthesis in large Endoriftia.

807 Higher abundances of glycogen-producing enzymes in fraction L furthermore suggest that
808 large symbionts invest relatively more of their biosynthetic capacities in storage of fixed carbon
809 in the form of glycogen than smaller symbionts. This is in accordance with a previous study
810 (Sorgo *et al.*, 2002), which noted a glycogen gradient in the symbiont cells, with increasing
811 glycogen density from the lobule center towards the periphery, i.e., towards larger symbiont
812 cells.

813 *Small Endoriftia store more sulfur and are more involved in sulfide oxidation*

814 Smaller symbionts produce relatively more sulfur globules for sulfur storage than larger
815 symbiont cells, as indicated by relatively higher abundance of sulfur globule proteins in
816 fraction XS (Supplementary Figure S4). This is in agreement with observations of Hand (1987),
817 who noted more sulfur deposits in central (small) than in peripheral (large) *Riftia* symbionts.
818 Although this finding was not supported by a subsequent study (Pflugfelder *et al.*, 2005), our
819 results do point to different amounts of S storage in different Endoriftia subpopulations. As
820 shown for the free-living thiotrophic model bacterium *Allochromatium vinosum*, activation of
821 stored sulfur involves trafficking proteins such as TusA, which is involved in sulfur transfer to
822 DsrEFH and DsrC (Stockdreher *et al.*, 2014). In our study, the highly abundant TusA, several
823 DsrC copies as well as DsrEFH were all detected with highest abundances in fraction XS, thus
824 supporting the idea of relatively more re-mobilization of sulfur and subsequent utilization of
825 reduced sulfur compounds in small Endoriftia. As the highly abundant adenylylsulfate
826 reductase AprAB, the ATP-sulfurylase SopT and sulfide dehydrogenase subunit FccB were also
827 detected with higher abundances in fractions XS or S than in M or L, one might conclude that
828 sulfide oxidation itself also plays a more prominent role in smaller symbionts than in larger
829 symbionts. However, as we detected the dissimilatory sulfite reductase DsrAB, the third key
830 enzyme of cytoplasmic sulfide oxidation, with a rather ambiguous abundance pattern
831 (Supplementary Table S4e), this idea remains speculative and requires further analysis.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In large symbionts, thiosulfate oxidation plays a more prominent role

Larger symbionts may rely relatively more on thiosulfate oxidation – in addition to sulfide oxidation – than smaller Endoriftia, as suggested by highest abundance of SoxZ and detection of several other (low-abundant) Sox proteins in fraction L. Expression of the Sox (sulfur oxidation) complex was shown to be upregulated in the presence of thiosulfate in *A. vinosum* (Grimm *et al.*, 2011). We speculate that thiosulfate concentrations might be higher in the trophosome lobule periphery than in the lobule center, due to a concentration gradient (as proposed above for CO_2) and/or possibly also as a result of host thiosulfate production. The *Riftia* host appears to be able to oxidize toxic sulfide to the less toxic thiosulfate in its mitochondria (Hinzke *et al.*, 2019). Higher abundance of host thiosulfate sulfurtransferase in symbiont-enriched fractions compared to non-enriched trophosome homogenate in our present study suggests that this putative detoxification process could be particularly important in the symbiont-containing bacteriocytes. With sulfide supposedly reaching the trophosome lobule periphery first with the blood flow, free sulfide concentrations might be higher there and, consequently, host sulfide oxidation to thiosulfate might be more frequent in bacteriocytes at the lobule periphery than in the center. The idea of more thiosulfate oxidation in large Endoriftia is further substantiated by highest abundance of six rhodanese family proteins in fraction L, as rhodanese-like proteins can cleave thiosulfate into sulfite and sulfide and were proposed to be involved in thiosulfate oxidation (Hensen *et al.*, 2006, Welte *et al.*, 2009).

Interestingly, overall abundance of all proteins involved in the symbiont's energy-generating sulfur metabolism, the most abundant of all metabolic categories, remained relatively unchanged across the four fractions (Supplementary Table S7). This indicates that sulfur oxidation-based energy generation, a fundamental basis of all other metabolic processes, is equally important throughout the symbiont's differentiation process, even if individual contributions of reduced sulfur compounds may differ. (For a detailed overview of sulfur oxidation reactions in Endoriftia see Supplementary Results and Discussion F).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hydrogen oxidation is more relevant in large symbionts

In large symbionts, the use of hydrogen may furthermore play a more prominent role than in smaller symbiont cells, as suggested by increasing abundances of the Isp-type respiratory H_2-uptake [NiFe] hydrogenase large subunit HyaB, a Fe-S oxidoreductase (GlpC) encoded next to HyaB, and the hydrogenase expression/formation protein HypE from fraction XS to L. The small hydrogenase subunit HyaA (Sym_EGV51837.1) and an additional hydrogenase expression/formation protein (HoxM, Sym_EGV51835.1), both of which are encoded upstream of HyaB in the symbiont genome, were detected with increasing abundance towards fraction L as well (although at very low concentrations; Supplementary Table S3b), supporting the idea of relatively more hydrogen oxidation in large symbionts. Like for CO_2 and thiosulfate, this might be due to a concentration gradient with highest hydrogen concentrations at the lobule periphery and lowest concentrations towards the lobule center. Use of hydrogen as an energy source has been described or suggested for free-living sulfur oxidizing bacteria like *A. vinosum* (Weissgerber et al., 2011), and for a variety of thiotrophic symbionts of marine invertebrates (Petersen *et al.*, 2011). Taking advantage of hydrogen oxidation in addition to sulfide- and thiosulfate oxidation, i.e., using a broader repertoire of electron donors, would potentially enhance the metabolic flexibility, particularly of large Endoriftia. However, H_2 was recently suggested to be involved in maintaining intracellular redox homeostasis rather than working as electron donor in the *Riftia* symbiosis (Mitchell *et al.*, 2019), and hydrogenase may in fact also play a role in sulfur metabolism (as suggested for *A. vinosum* (Weissgerber *et al.*, 2014); Supplementary Results and Discussion F). Therefore, the exact role of hydrogen oxidation in Endoriftia and why it might be relatively more relevant in larger symbionts remains to be discussed.

Denitrification in Riftia symbionts appears to be modular

Our results suggest that small *Riftia* symbionts rely relatively more on the NarGHI-mediated first step of respiratory nitrate reduction to nitrite, while all subsequent steps of nitrite reduction via NO and N_2O to N_2 seem to be more prominent in larger symbionts. Since expression of nitrate reduction genes is usually inhibited by oxygen (Payne, 1973), high

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

887 NarGHI abundance in fraction XS suggests that O_2 levels might be particularly low in the
888 trophosome lobule center.

889 Interestingly, although Endoriftia has the genomic potential for complete denitrification to N_2,
890 small and large symbionts seem to employ separate parts of the pathway. This is reminiscent
891 of free-living microbial communities, in which denitrification is modular, i.e., it is often not
892 carried out by individual organisms, but rather by the subsequent activity of several members
893 (Graf *et al.*, 2014), between which intermediates are passed on as 'metabolic handoffs'
894 (Anantharaman *et al.*, 2016). Moreover, nitrate reduction and subsequent denitrification steps
895 may occur as two temporally separated processes even in the same organism: During nitrate
896 reduction in *Staphylococcus carnosus*, nitrite reduction was inhibited and resumed only after
897 nitrate was depleted (Neubauer and Götz, 1996). A similar scenario might be assumed for
898 Endoriftia: Small symbionts apparently reduce nitrate to nitrite, which, potentially, yields
899 enough energy to cover their demand, while 'saving' nitrite as a handoff for future use. Once
900 the symbionts have become larger, expression of nitrite reductase, nitric oxide reductase and
901 nitrous oxide reductase in higher abundance enables them to further reduce the accumulated
902 intermediate nitrite. Whether these reactions could also be regulated in response to varying
903 oxygen concentrations is unclear. We speculate that an O_2 gradient may exist, which influences
904 the observed expression pattern.

905 *Regulation of gene expression may be less stringent in large symbionts*

906 Relative abundance of the RNA polymerase sigma factor RpoD decreased from fraction XS to
907 L (Figure 3, Supplementary Table S4h) in S-rich and S-depleted samples, pointing to relatively
908 more growth-related activities in small Endoriftia (see also Supplementary Results and
909 Discussion B). RpoD is the primary sigma factor for vegetative growth (σ70), which regulates
910 transcription of most genes involved in exponential growth in many bacteria (Helmann and
911 Chamberlin, 1988, Fujita *et al.*, 1994, Ishihama, 2000). This would be in agreement with the
912 idea of small *Riftia* symbionts being mainly occupied with cell division and proliferation in a
913 quasi-exponential growth phase, while large symbionts function as biosynthetic 'factories',
914 focusing on carbon fixation and biomass production. Interestingly, RpoS, the master

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transcriptional regulator of stationary phase gene expression and antagonist of RpoD, was not detected in any of our samples (although it is encoded in the symbiont genome). RpoS abundance increases upon stress and limitation during transition to the stationary phase in free-living model bacteria (Hengge-Aronis, 1993, Fujita *et al.*, 1994, Ishihama, 2000). Its absence in the *Riftia* symbiont's proteome suggests that, unlike free-living bacteria, the symbiont does not experience a stationary phase-like growth arrest even in later developmental stages, probably because it is ideally supplied with all necessary substrates by the host. This 'lack' of stress or limitation possibly results in less stringent regulation of symbiont gene expression, which could explain the metabolic diversity we observed particularly in large symbionts, such as multiple ways of energy generation (thiosulfate- and hydrogen oxidation in addition to sulfide oxidation) and two CO_2 fixation pathways. Under these premises, the previously observed simultaneous expression of seemingly redundant metabolic pathways in *Riftia* symbionts (Markert *et al.*, 2011) very likely reflects this presumptive "de-regulation" of gene expression in large parts of the symbiont population, which allows Endoriftia to fully exploit its versatile metabolic repertoire to the advantage of the symbiosis.

Conclusion

Our results show that Endoriftia cells of different differentiation stages likely employ distinct metabolic profiles, thus confirming our initial hypothesis. Whereas small Endoriftia ensure survival of the symbiont population, large Endoriftia are primarily engaged in biomass production. The driving force behind this differentiation remains to be elucidated. For *Rhizobium*, a steep O_2 concentration gradient inside legume nodules was proposed to be involved in signaling for symbiont differential gene expression (Soupene *et al.*, 1995). Similarly, some of the differences we observed in small and large Endoriftia might also be connected to the availability of electron donors or acceptors, and hence differentiation of Endoriftia cells might depend on substrate availability. Symbiont differentiation in *Riftia* might furthermore be induced by specific host effectors, e.g., histone-derived antimicrobial peptides, which were recently proposed to play a role in symbiont cell cycle regulation (Hinzke

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

et al., 2019), or other compounds that allow *Riftia* to modulate the symbiont's expression of certain metabolic pathways. Besides such direct interference, *Riftia* likely also exerts indirect influence on symbiont gene expression by providing copious amounts of all necessary substrates to the bacterial partner. We speculate that this constantly high nutrient availability inside the host causes Endoriftia's biosynthetic pathways to be regulated less stringently (compared to what we would expect in free-living bacteria). This would explain the previously observed metabolic versatility of symbionts in the same host: Large Endoriftia can afford to employ much of their metabolic repertoire at the same time. Such an 'advantageous deregulation', i.e., unhindered expression of multiple – even redundant – metabolic pathways, likely enables high symbiont productivity during symbiosis.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Ali, B. M. J., R. Amit, I. Braslavsky, A. B. Oppenheim, O. Gileadi and J. Stavans (2001). Compaction of single DNA molecules induced by binding of integration host factor (IHF). *Proceedings of the National Academy of Sciences of the United States of America* **98**: 10658-10663. DOI: 10.1073/pnas.181029198.

Anantharaman, K., C. T. Brown, L. A. Hug, I. Sharon, C. J. Castelle, A. J. Probst, B. C. Thomas, A. Singh, M. J. Wilkins, U. Karaoz, E. L. Brodie, K. H. Williams, S. S. Hubbard and J. F. Banfield (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature Communications* **7**(1): 13219. DOI: 10.1038/ncomms13219.

Angert, E. R. (2012). DNA Replication and Genomic Architecture of Very Large Bacteria. *Annual Review of Microbiology* **66**: 197-212. DOI: 10.1146/annurev-micro-090110-102827.

Barabote, R. D., O. L. Johnson, E. Zetina, S. K. San Francisco, J. A. Fralick and M. J. D. San Francisco (2003). *Erwinia chrysanthemi* TolC Is Involved in Resistance to Antimicrobial Plant Chemicals and Is Essential for Phytopathogenesis. *Journal of Bacteriology* **185**: 5772-5778. DOI: 10.1128/jb.185.19.5772-5778.2003.

Beam, C. E., C. J. Saveson and S. T. Lovett (2002). Role for *radA/sms* in recombination intermediate processing in *Escherichia coli*. *Journal of Bacteriology* **184**: 6836-6844. DOI: 10.1128/JB.184.24.6836-6844.2002.

Beckmann, J. F., T. W. Markowski, B. A. Witthuhn and A. M. Fallon (2013). Detection of the *Wolbachia*-encoded DNA binding protein, HU beta, in mosquito gonads. *Insect Biochemistry and Molecular Biology* **43**: 272-279. DOI: 10.1016/j.ibmb.2012.12.007.

Bouvier, T., M. Troussellier, A. Anzil, C. Courties and P. Servais (2001). Using light scatter signal to estimate bacterial biovolume by flow cytometry. *Cytometry: The Journal of the International Society for Analytical Cytology* **44**(3): 188-194. DOI: 10.1002/1097-0320(20010701)44:3<188::AID-CYTO1111>3.0.CO;2-C.

Bright, M., H. Keckeis and C. R. Fisher (2000). An autoradiographic examination of carbon fixation, transfer and utilization in the *Riftia pachyptila* symbiosis. *Marine Biology* **136**: 621-632. DOI: 10.1007/s002270050722.

Bright, M. and A. Sorgo (2003). Ultrastructural reinvestigation of the trophosome in adults of *Riftia pachyptila* (Annelida, Siboglinidae). *Invertebrate Biology* **122**(4): 347-368. DOI: 10.1111/j.1744-7410.2003.tb00099.x.

Camberg, J. L., J. R. Hoskins and S. Wickner (2009). ClpXP protease degrades the cytoskeletal protein, FtsZ, and modulates FtsZ polymer dynamics. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 10614-10619. DOI: 10.1073/pnas.0904886106.

Caro, A., O. Gros, P. Got, R. De Wit and M. Troussellier (2007). Characterization of the population of the sulfur-oxidizing symbiont of *Codakia orbicularis* (Bivalvia, Lucinidae) by single-cell analyses. *Applied and Environmental Microbiology* **73**: 2101-2109. DOI: 10.1128/AEM.01683-06.

Cavanaugh, C., S. L. Gardiner, M. L. Jones, H. W. Jannasch and J. B. Waterbury (1981). Procaryotic cells in the hydrothermal vent tubeworm *Riftia pachyptila*: possible chemoautotrophic symbionts. *Science* **213**: 340-342. DOI: 10.1126/science.213.4505.340.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

998 Chao, Y. and J. Vogel (2010). The role of Hfq in bacterial pathogens. *Current Opinion in*
999 *Microbiology* **13**: 24-33. DOI: 10.1016/j.mib.2010.01.001.

1000 Chastanet, A., I. Derre, S. Nair and T. Msadek (2004). *clpB*, a Novel Member of the *Listeria*
1001 *monocytogenes* CtsR Regulon, Is Involved in Virulence but Not in General Stress Tolerance.
1002 *Journal of Bacteriology* **186**(4): 1165-1174. DOI: 10.1128/jb.186.4.1165-1174.2004.

1003 Chien, A. C., N. S. Hill and P. A. Levin (2012). Cell size control in bacteria. *Current Biology* **22**:
1004 R340-R349. DOI: 10.1016/j.cub.2012.02.032.

1005 Cho, H., H. R. McManus, S. L. Dove and T. G. Bernhardt (2011). Nucleoid occlusion factor
1006 SlmA is a DNA-activated FtsZ polymerization antagonist. *Proceedings of the National*
1007 *Academy of Sciences of the United States of America* **108**: 3773-3778. DOI:
1008 10.1073/pnas.1018674108.

1009 Choi, H. M. T., V. A. Beck and N. A. Pierce (2014). Next-Generation in Situ Hybridization Chain
1010 Reaction: Higher Gain, Lower Cost, Greater Durability. *ACS Nano* **8**(5): 4284-4294. DOI:
1011 10.1021/nn405717p.

1012 Cortés, F. and N. Pastor (2003). Induction of endoreduplication by topoisomerase II catalytic
1013 inhibitors. *Mutagenesis* **18**: 105-112. DOI: 10.1093/mutage/18.2.105.

1014 Cortés, F., N. Pastor, S. Mateos and I. Domínguez (2003). Roles of DNA topoisomerases in
1015 chromosome segregation and mitosis. *Mutation Research - Reviews in Mutation Research*
1016 **543**: 59-66. DOI: 10.1016/S1383-5742(02)00070-4.

1017 Cosme, A. M., A. Becker, M. R. Santos, L. A. Sharypova, P. M. Santos and L. M. Moreira (2008).
1018 The Outer Membrane Protein TolC from *Sinorhizobium meliloti* Affects Protein Secretion,
1019 Polysaccharide Biosynthesis, Antimicrobial Resistance, and Symbiosis. *Molecular Plant-*
1020 *Microbe Interactions* **21**: 947-957. DOI: 10.1094/mpmi-21-7-0947.

1021 D'Haeze, W. and M. Holsters (2004). Surface polysaccharides enable bacteria to evade plant
1022 immunity. *Trends in Microbiology* **12**: 555-561. DOI: 10.1016/j.tim.2004.10.009.

1023 Dang, W., Y. h. Hu and L. Sun (2011). HtpG is involved in the pathogenesis of *Edwardsiella*
1024 *tarda*. *Veterinary Microbiology* **152**: 394-400. DOI: 10.1016/j.vetmic.2011.05.030.

1025 Deaconescu, A. M., N. Savery and S. A. Darst (2007). The bacterial transcription repair
1026 coupling factor. *Current Opinion in Structural Biology* **17**: 96-102. DOI:
1027 10.1016/j.sbi.2007.01.005.

1028 Degenhardt, F., S. Seifert and S. Szymczak (2019). Evaluation of variable selection methods for
1029 random forests and omics data sets. *Briefings in Bioinformatics* **20**(2): 492-503. DOI:
1030 10.1093/bib/bbx124.

1031 Distel, D. and H. Felbeck (1988). Pathways of inorganic carbon fixation in the endosymbiont-
1032 bearing lucinid clam *Lucinoma aequizonata*. Part 1. Purification and characterization of the
1033 endosymbiotic bacteria. *Journal of Experimental Zoology* **247**(1): 1-10. DOI:
1034 10.1002/jez.1402470102.

1035 Domínguez-Cuevas, P., I. Porcelli, R. A. Daniel and J. Errington (2013). Differentiated roles
1036 for MreB-actin isologues and autolytic enzymes in *Bacillus subtilis* morphogenesis. *Molecular*
1037 *Microbiology* **89**: 1084-1098. DOI: 10.1111/mmi.12335.

1038 Dorman, C. J. and P. Deighan (2003). Regulation of gene expression by histone-like proteins
1039 in bacteria. *Current Opinion in Genetics and Development* **13**: 179-184. DOI: 10.1016/S0959-
1040 437X(03)00025-X.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Drees, J. C., S. Chitteni-Pattu, D. R. McCaslin, R. B. Inman and M. M. Cox (2006). Inhibition of RecA protein function by the RdgC protein from *Escherichia coli*. *Journal of Biological Chemistry* **281**: 4708-4717. DOI: 10.1074/jbc.M513592200.

Dri, A. M., J. Rouviere-Yaniv and P. L. Moreau (1991). Inhibition of cell division in *hupA hupB* mutant bacteria lacking HU protein. *Journal of Bacteriology* **173**: 2852-2863. DOI: 10.1128/jb.173.9.2852-2863.1991.

Duperthuy, M., P. Schmitt, E. Garzón, A. Caro, R. D. Rosa, F. Le Roux, N. Lautrédou-Audouy, P. Got, B. Romestand, J. De Lorgeril, S. Kieffer-Jaquinod, E. Bachère and D. Destoumieux-Garzón (2011). Use of OmpU porins for attachment and invasion of *Crassostrea gigas* immune cells by the oyster pathogen *Vibrio splendidus*. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 2993-2998. DOI: 10.1073/pnas.1015326108.

Durand-Heredia, J., E. Rivkin, G. Fan, J. Morales and A. Janakiraman (2012). Identification of ZapD as a cell division factor that promotes the assembly of FtsZ in *Escherichia coli*. *Journal of Bacteriology* **194**: 3189-3198. DOI: 10.1128/JB.00176-12.

Ernst, J. and Z. Bar-Joseph (2006). STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* **7**(1): 191. DOI: 10.1186/1471-2105-7-191.

Felbeck, H. (1981). Chemoautotrophic Potential of the Hydrothermal Vent Tube Worm, *Riftia pachyptila* Jones (Vestimentifera). *Science* **213**(4505): 336-338. DOI: 10.1126/science.213.4505.336.

Felbeck, H. and P. J. Turner (1995). CO_2 transport in catheterized hydrothermal vent tubeworms, *Riftia pachyptila* (Vestimentifera). *Journal of Experimental Zoology* **272**(2): 95-102. DOI: 10.1002/jez.1402720203.

Fernández, L. and R. E. W. Hancock (2012). Adaptive and mutational resistance: Role of porins and efflux pumps in drug resistance. *Clinical Microbiology Reviews* **25**: 661-681. DOI: 10.1128/CMR.00043-12.

Frees, D., S. N. A. Qazi, P. J. Hill and H. Ingmer (2003). Alternative roles of ClpX and ClpP in *Staphylococcus aureus* stress tolerance and virulence. *Molecular Microbiology* **48**: 1565-1578. DOI: 10.1046/j.1365-2958.2003.03524.x.

Fujita, M., K. Tanaka, H. Takahashi and A. Amemural (1994). Transcription of the principal sigma-factor genes, *rpoD* and *rpoS*, in *Pseudomonas aeruginosa* is controlled according to the growth phase. *Molecular Microbiology* **13**: 1071-1077. DOI: 10.1111/j.1365-2958.1994.tb00498.x.

Gardebrecht, A., S. Markert, S. M. Sievert, H. Felbeck, A. Thürmer, D. Albrecht, A. Wollherr, J. Kabisch, N. Le Bris, R. Lehmann, R. Daniel, H. Liesegang, M. Hecker and T. Schweder (2012). Physiological homogeneity among the endosymbionts of *Riftia pachyptila* and *Tevnia jerichonana* revealed by proteogenomics. *The ISME Journal* **6**(4): 766-776. DOI: 10.1038/ismej.2011.137.

Graf, D. R. H., C. M. Jones and S. Hallin (2014). Intergenomic Comparisons Highlight Modularity of the Denitrification Pathway and Underpin the Importance of Community Structure for N_2O Emissions. *PLOS ONE* **9**(12): e114118. DOI: 10.1371/journal.pone.0114118.

Graham, J. M. (2001). Biological centrifugation. *The Basics*, BIOS Scientific Publishers Ltd., Oxford, UK.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1083 Grimm, F., B. Franz and C. Dahl (2011). Regulation of dissimilatory sulfur oxidation in the
1084 purple sulfur bacterium *Allochromatium vinosum*. *Frontiers in Microbiology* **2**: 51. DOI:
1085 10.3389/fmicb.2011.00051.

1086 Gu, Z., R. Eils and M. Schlesner (2016). Complex heatmaps reveal patterns and correlations in
1087 multidimensional genomic data. *Bioinformatics* **32**(18): 2847-2849. DOI:
1088 10.1093/bioinformatics/btw313.

1089 Guha, S., S. Udupa, W. Ahmed and V. Nagaraja (2018). Rewired Downregulation of DNA
1090 Gyrase Impacts Cell Division, Expression of Topology Modulators, and Transcription in
1091 *Mycobacterium smegmatis*. *Journal of Molecular Biology* **430**: 4986-5001. DOI:
1092 10.1016/j.jmb.2018.10.001.

1093 Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis
1094 program for Windows 95/98/NT. *Nucleic acids symposium series [London]: Information*
1095 *Retrieval Ltd.* **41**(41): c1979-c2000.

1096 Hand, S. C. (1987). Trophosome ultrastructure and the characterization of isolated
1097 bacteriocytes from invertebrate-sulfur bacteria symbioses. *Biological Bulletin* **173**: 260-276.
1098 DOI: 10.2307/1541878.

1099 Hay, N. A., D. J. Tipper, D. Gygi and C. Hughes (1999). A novel membrane protein influencing
1100 cell shape and multicellular swarming of *Proteus mirabilis*. *Journal of Bacteriology* **181**:
1101 2008-2016.

1102 Heath, M. C. (2000). Nonhost resistance and nonspecific plant defenses. *Current Opinion in*
1103 *Plant Biology* **3**: 315-319. DOI: 10.1016/S1369-5266(00)00087-X.

1104 Helmann, J. D. and M. J. Chamberlin (1988). Structure and function of bacterial sigma factors.
1105 *Annual Review of Biochemistry* **57**: 839-872. DOI: 10.1146/annurev.biochem.57.1.839.

1106 Hengge-Aronis, R. (1993). Survival of hunger and stress: the role of *rpoS* in early stationary
1107 phase gene regulation in *E. coli*. *Cell* **72**(2): 165-168. DOI: 10.1016/0092-8674(93)90655-A.

1108 Hensen, D., D. Sperling, H. G. Trüper, D. C. Brune and C. Dahl (2006). Thiosulphate oxidation
1109 in the phototrophic sulphur bacterium *Allochromatium vinosum*. *Molecular Microbiology*
1110 **62**: 794-810. DOI: 10.1111/j.1365-2958.2006.05408.x.

1111 Hinzke, T. and S. Markert (2017). Efficient protein extraction for proteomics and
1112 metaproteomics (also suitable for low biomass samples). protocols.io. DOI:
1113 10.17504/protocols.io.kg6ctze

1114 Hinzke, T., M. Kleiner and S. Markert (2018). Centrifugation-Based Enrichment of Bacterial
1115 Cell Populations for Metaproteomic Studies on Bacteria–Invertebrate Symbioses. *Microbial*
1116 *Proteomics: Methods and Protocols*. D. Becher. New York, USA, Springer New York**:** 319-334
1117 DOI: 10.1007/978-1-4939-8695-8_22.

1118 Hinzke, T., M. Kleiner, C. Breusing, H. Felbeck, R. Häsler, S. M. Sievert, R. Schlüter, P.
1119 Rosenstiel, T. B. H. Reusch, T. Schweder and S. Markert (2019). Host-Microbe Interactions in
1120 the Chemosynthetic *Riftia pachyptila* Symbiosis. *mBio* **10**(6): e02243-02219. DOI:
1121 10.1128/mBio.02243-19.

1122 Ishihama, A. (2000). Functional modulation of *Escherichia coli* RNA polymerase. *Annual*
1123 *Reviews in Microbiology* **54**(1): 499-518. DOI: 10.1146/annurev.micro.54.1.499.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Janitza, S., E. Celik and A. L. Boulesteix (2018). A computationally fast variable importance test for random forests for high-dimensional data. *Advances in Data Analysis and Classification* **12**: 885-915. DOI: 10.1007/s11634-016-0276-4.

Kang, D., G. Liu, A. Lundström, E. Gelius and H. Steiner (1998). A peptidoglycan recognition protein in innate immunity conserved. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 10078-10082. DOI: 10.1073/pnas.95.17.10078.

King, A. M., G. Pretre, T. Bartpho, R. W. Sermswan, C. Toma, T. Suzuki, A. Eshghi, M. Picardeau, B. Adler and G. L. Murray (2014). High-Temperature Protein G Is an Essential Virulence Factor of *Leptospira interrogans*. *Infection and Immunity* **82**: 1123-1131. DOI: 10.1128/iai.01546-13.

Kleiner, M., X. Dong, T. Hinzke, J. Wippler, E. Thorson, B. Mayer and M. Strous (2018). Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. *Proceedings of the National Academy of Sciences of the United States of America* **115**: E5576-E5584. DOI: 10.1073/pnas.1722325115.

Klint, J., U. Rasmussen and B. Bergman (2007). FtsZ may have dual roles in the filamentous cyanobacterium *Nostoc/Anabaena* sp. strain PCC 7120. *Journal of Plant Physiology* **164**: 11-18. DOI: 10.1016/j.jplph.2005.08.021.

Klose, J., K. Aistleitner, M. Horn, L. Krenn, V. Dirsch, M. Zehl and M. Bright (2016). Trophosome of the deep-sea tubeworm *Riftia pachyptila* inhibits bacterial growth. *PLoS ONE* **11**: e0146446. DOI: 10.1371/journal.pone.0146446.

Köhler, S., J. Teyssier, A. Cloeckaert, B. Rouot and J. P. Liautard (1996). Participation of the molecular chaperone DnaK in intracellular growth of *Brucella suis* within U937-derived phagocytes. *Molecular Microbiology* **20**: 701-712. DOI: 10.1111/j.1365-2958.1996.tb02510.x.

Komaki, K. and H. Ishikawa (2000). Genomic copy number of intracellular bacterial symbionts of aphids varies in response to developmental stage and morph of their host. *Insect Biochemistry and Molecular Biology* **30**: 253-258. DOI: 10.1016/S0965-1748(99)00125-3.

Kondorosi, E. and A. Kondorosi (2004). Endoreduplication and activation of the anaphase-promoting complex during symbiotic cell development. *FEBS Letters* **567**: 152-157. DOI: 10.1016/j.febslet.2004.04.075.

Koronakis, V., J. Eswaran and C. Hughes (2004). Structure and Function of TolC: The Bacterial Exit Duct for Proteins and Drugs. *Annual Review of Biochemistry* **73**: 467-489. DOI: 10.1146/annurev.biochem.73.011303.074104.

Kowalczykowski, S. C. (2000). Initiation of genetic recombination and recombination-dependent replication. *Trends in Biochemical Sciences* **25**: 156-165. DOI: 10.1016/S0968-0004(00)01569-3.

Kruse, T., B. Blagoev, A. Løbner-Olesen, M. Wachi, K. Sasaki, N. Iwai, M. Mann and K. Gerdes (2006). Actin homolog MreB and RNA polymerase interact and are both required for chromosome segregation in *Escherichia coli*. *Genes and Development* **20**: 113-124. DOI: 10.1101/gad.366606.

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**: 680-685. DOI: 10.1038/227680a0.

Levine, C., H. Hiasa and K. J. Marians (1998). DNA gyrase and topoisomerase IV: Biochemical activities, physiological roles during chromosome replication, and drug sensitivities.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Biochimica et Biophysica Acta - Gene Structure and Expression **1400**: 29-43. DOI: 10.1016/S0167-4781(98)00126-2.

Li, C., J. Wei, Y. Li, X. He, Q. Zhou, J. Yan, J. Zhang, Y. Liu, Y. Liu and H.-b. Shu (2013). Transmembrane Protein 214 (TMEM214) Mediates Endoplasmic Reticulum Stress-induced Caspase 4 Enzyme Activation and Apoptosis. *Journal of Biological Chemistry* **288**: 17908-17917. DOI: 10.1074/jbc.M113.458836.

López-Sánchez, M. J., A. Neef, R. Patiño-Navarrete, L. Navarro, R. Jiménez, A. Latorre and A. Moya (2008). Blattabacteria, the endosymbionts of cockroaches, have small genome sizes and high genome copy numbers. *Environmental Microbiology* **10**: 3417-3422. DOI: 10.1111/j.1462-2920.2008.01776.x.

Lorenzen, D. R., D. Gunther, J. Pandit, T. Rudel, E. Brandt and T. F. Meyer (2000). *Neisseria gonorrhoeae* porin modifies the oxidative burst of human professional phagocytes. *Infection and Immunity* **68**: 6215-6222. DOI: 10.1128/IAI.68.11.6215-6222.2000.

Lynch, M. and H. Kuramitsu (2000). Expression and role of superoxide dismutases (SOD) in pathogenic bacteria. *Microbes and Infection* **2**: 1245-1255. DOI: 10.1016/S1286-4579(00)01278-8.

Lyngstadaas, A., A. Løbner-Olesen and E. Boye (1995). Characterization of three genes in the *dam*-containing operon of *Escherichia coli*. *MGG Molecular & General Genetics* **247**: 546-554. DOI: 10.1007/BF00290345.

Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert and K. Hornik. (2018). cluster: Cluster Analysis Basics and Extensions. R package version 2.0.7-1.

Markert, S., C. Arndt, H. Felbeck, D. Becher, S. M. Sievert, M. Hugler, D. Albrecht, J. Robidart, S. Bench, R. A. Feldman, M. Hecker and T. Schweder (2007). Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila*. *Science* **315**(5809): 247-250. DOI: 10.1126/science.1132913.

Markert, S., A. Gardebrecht, H. Felbeck, S. M. Sievert, J. Klose, D. Becher, D. Albrecht, A. Thurmer, R. Daniel, M. Kleiner, M. Hecker and T. Schweder (2011). Status quo in physiological proteomics of the uncultured *Riftia pachyptila* endosymbiont. *Proteomics* **11**(15): 3106-3117. DOI: 10.1002/pmic.201100059.

Mathur, J. and M. K. Waldor (2004). The *Vibrio cholerae* ToxR-regulated porin OmpU confers resistance to antimicrobial peptides. *Infection and Immunity* **72**: 3577-3583. DOI: 10.1128/IAI.72.6.3577-3583.2004.

Mergaert, P., T. Uchiumi, B. Alunni, G. Evanno, A. Cheron, O. Catrice, A. E. Mausset, F. Barloy-Hubler, F. Galibert, A. Kondorosi and E. Kondorosi (2006). Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 5230-5235. DOI: 10.1073/pnas.0600912103.

Mitchell, J. H., J. M. Leonard, J. Delaney, P. R. Girguis and K. M. Scott (2019). Hydrogen Does Not Appear To Be a Major Electron Donor for Symbiosis with the Deep-Sea Hydrothermal Vent Tubeworm *Riftia pachyptila*. *Applied and Environmental Microbiology* **86**(1): e01522-01519. DOI: 10.1128/aem.01522-19.

Monds, R. D., T. K. Lee, A. Colavin, T. Ursell, S. Quan, T. F. Cooper and K. C. Huang (2014). Systematic Perturbation of Cytoskeletal Function Reveals a Linear Scaling Relationship between Cell Geometry and Fitness. *Cell Reports* **9**: 1528-1537. DOI: 10.1016/j.celrep.2014.10.040.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mosleh, I. M., L. A. Huber, P. Steinlein, C. Pasquali, D. Günther and T. F. Meyer (1998). *Neisseria gonorrhoeae* porin modulates phagosome maturation. *Journal of Biological Chemistry* **273**: 35332-35338. DOI: 10.1074/jbc.273.52.35332.

Mueller, R. S., V. J. Denef, L. H. Kalnejais, K. B. Suttle, B. C. Thomas, P. Wilmes, R. L. Smith, D. K. Nordstrom, R. B. McCleskey, M. B. Shah, N. C. VerBerkmoes, R. L. Hettich and J. F. Banfield (2010). Ecological distribution and population physiology defined by proteomics in a natural microbial community. *Molecular Systems Biology* **6**: 374. DOI: 10.1038/Msb.2010.30.

Mukherjee, A., A. O. Sokunbi and A. Grove (2008). DNA protection by histone-like protein HU from the hyperthermophilic eubacterium *Thermotoga maritima*. *Nucleic Acids Research* **36**: 3956-3968. DOI: 10.1093/nar/gkn348.

Murga, M., I. Jaco, Y. Fan, R. Soria, B. Martinez-Pastor, M. Cuadrado, S. M. Yang, M. A. Blasco, A. I. Skoultchi and O. Fernandez-Capetillo (2007). Global chromatin compaction limits the strength of the DNA damage response. *Journal of Cell Biology* **178**: 1101-1108. DOI: 10.1083/jcb.200704140.

Neubauer, H. and F. Götz (1996). Physiology and interaction of nitrate and nitrite reduction in *Staphylococcus carnosus*. *Journal of Bacteriology* **178**: 2005-2009. DOI: 10.1128/jb.178.7.2005-2009.1996.

Nöllmann, M., N. J. Crisona and P. B. Arimondo (2007). Thirty years of *Escherichia coli* DNA gyrase: From *in vivo* function to single-molecule mechanism. *Biochimie* **89**: 490-499. DOI: 10.1016/j.biochi.2007.02.012.

Nussbaumer, A. D., C. R. Fisher and M. Bright (2006). Horizontal endosymbiont transmission in hydrothermal vent tubeworms. *Nature* **441**: 345-348. DOI: 10.1038/nature04793.

Nyholm, S. V., J. J. Stewart, E. G. Ruby and M. J. McFall-Ngai (2009). Recognition between symbiotic *Vibrio fischeri* and the haemocytes of *Euprymna scolopes*. *Environmental Microbiology* **11**(2): 483-493. DOI: 10.1111/j.1462-2920.2008.01788.x.

Ohniwa, R. L., K. Morikawa, J. Kim, T. Ohta, A. Ishihama, C. Wada and K. Takeyasu (2006). Dynamic state of DNA topology is essential for genome condensation in bacteria. *EMBO Journal* **25**: 5591-5602. DOI: 10.1038/sj.emboj.7601414.

Ozawa, K., K. Kuwabara, M. Tamatani, K. Takatsuji, Y. Tsukamoto, S. Kaneda, H. Yanagi, D. M. Stern, Y. Eguchi, Y. Tsujimoto, S. Ogawa and M. Tohyama (1999). 150-kDa Oxygen-regulated Protein (ORP150) Suppresses Hypoxia-induced Apoptotic Cell Death. *Journal of Biological Chemistry* **274**: 6397-6404. DOI: 10.1074/jbc.274.10.6397.

Payne, W. (1973). Reduction of nitrogenous oxides by microorganisms. *Bacteriological Reviews* **37**(4): 409.

Pende, N., N. Leisch, H. R. Gruber-Vodicka, N. R. Heindl, J. Ott, T. Den Blaauwen and S. Bulgheresi (2014). Size-independent symmetric division in extraordinarily long cells. *Nature Communications* **5**: Article 4803. DOI: 10.1038/ncomms5803.

Petersen, I., R. Schlüter, K. J. Hoff, V. Liebscher, G. Bange, K. Riedel and J. Pané-Farré (2020). Non-invasive and label-free 3D-visualization shows *in vivo* oligomerization of the staphylococcal alkaline shock protein 23 (Asp23). *Scientific Reports* **10**(1): 125. DOI: 10.1038/s41598-019-56907-9.

Petersen, J. M., F. U. Zielinski, T. Pape, R. Seifert, C. Moraru, R. Amann, S. Hourdez, P. R. Girguis, S. D. Wankel, V. Barbe, E. Pelletier, D. Fink, C. Borowski, W. Bach and N. Dubilier

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(2011). Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* **476**: 176-180. DOI: 10.1038/nature10325.

Pflugfelder, B., C. R. Fisher and M. Bright (2005). The color of the trophosome: elemental sulfur distribution in the endosymbionts of *Riftia pachyptila* (Vestimentifera; Siboglinidae). *Marine Biology* **146**: 895-901. DOI: 10.1007/s00227-004-1500-x.

Polz, M. F., H. Felbeck, R. Novak, M. Nebelsick and J. A. Ott (1992). Chemoautotrophic, Sulfur-Oxidizing Symbiotic Bacteria on Marine Nematodes: Morphological and Biochemical Characterization. *Microbial Ecology*: 313-329. DOI: 10.1007/BF00167789.

Polzin, J., P. Arevalo, T. Nussbaumer, M. F. Polz and M. Bright (2019). Polyclonal symbiont populations in hydrothermal vent tubeworms and the environment. *Proceedings of the Royal Society B: Biological Sciences* **286**: 20181281. DOI: 10.1098/rspb.2018.1281.

Ponnudurai, R., M. Kleiner, L. Sayavedra, J. M. Petersen, M. Moche, A. Otto, D. Becher, T. Takeuchi, N. Satoh and N. Dubilier (2017). Metabolic and physiological interdependencies in the *Bathymodiolus azoricus* symbiosis. *The ISME journal* **11**(2): 463. DOI: 10.1038/ismej.2016.124.

Prinsloo, E., M. M. Setati, V. M. Longshaw and G. L. Blatch (2009). Chaperoning stem cells: A role for heat shock proteins in the modulation of stem cell self-renewal and differentiation? *BioEssays* **31**: 370-377. DOI: 10.1002/bies.200800158.

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reimold, C., H. J. Defeu Soufo, F. Dempwolff and P. L. Graumann (2013). Motion of variable-length MreB filaments at the bacterial cell membrane influences cell morphology. *Molecular Biology of the Cell* **24**: 2340-2349. DOI: 10.1091/mbc.e12-10-0728.

Renoz, F., A. Champagne, H. Degand, A.-M. Faber, P. Morsomme, V. Foray and T. Hance (2017). Toward a better understanding of the mechanisms of symbiosis: a comprehensive proteome map of a nascent insect symbiont. *PeerJ* **5**: e3291. DOI: 10.7717/peerj.3291.

Robidart, J. C., S. R. Bench, R. A. Feldman, A. Novoradovsky, S. B. Podell, T. Gaasterland, E. E. Allen and H. Felbeck (2008). Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed through metagenomics. *Environmental Microbiology* **10**(3): 727-737. DOI: 10.1111/j.1462-2920.2007.01496.x.

Robinson, M. D., D. J. McCarthy and G. K. Smyth (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140. DOI: 10.1093/bioinformatics/btp616.

Royet, J., D. Gupta and R. Dziarski (2011). Peptidoglycan recognition proteins : modulators of the microbiome and inflammation. *Nature Reviews Immunology* **11**: 837-851. DOI: 10.1038/nri3089.

Saalfeld, S. (2010). Enhance Local Contrast (CLAHE) - a Fiji plugin; available at: https://imagej.net/Enhance_Local_Contrast_(CLAHE).

Sattler, M. C., C. R. Carvalho and W. R. Clarindo (2016). The polyploidy and its key role in plant breeding. *Planta* **243**: 281-296. DOI: 10.1007/s00425-015-2450-x.

Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tomancak and A. Cardona (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**(7): 676-682. DOI: 10.1038/nmeth.2019.

Søndergaard, D., C. N. S. Pedersen and C. Greening (2016). HydDB: A web tool for hydrogenase classification and analysis. *Scientific Reports* **6**: 34212. DOI: 10.1038/srep34212.

Sorgo, A., F. Gaill, J.-P. Lechaire, C. Arndt and M. Bright (2002). Glycogen storage in the *Riftia pachyptila* trophosome: contribution of host and symbionts. *Marine Ecology Progress Series* **231**: 115-120. DOI: 10.3354/meps231115.

Soupene, E., M. Foussard, P. Boistard, G. Truchet and J. Batut (1995). Oxygen as a key developmental regulator of *Rhizobium meliloti* N_2-fixation gene expression within the alfalfa root nodule. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 3759-3763. DOI: 10.1073/pnas.92.9.3759.

Steck, T. R. and K. Drlica (1984). Bacterial chromosome segregation: Evidence for DNA gyrase involvement in decatenation. *Cell* **36**: 1081-1087. DOI: 10.1016/0092-8674(84)90058-8.

Stockdreher, Y., M. Sturm, M. Josten, H. G. Sahl, N. Dobler, R. Zigann and C. Dahl (2014). New proteins involved in sulfur trafficking in the cytoplasm of *Allochromatium vinosum*. *Journal of Biological Chemistry* **289**: 12390-12403. DOI: 10.1074/jbc.M113.536425.

Stocks, S. M. (2004). Mechanism and use of the commercially available viability stain, BacLight. *Cytometry Part A* **61**: 189-195. DOI: 10.1002/cyto.a.20069.

Stuger, R., C. L. Woldringh, C. C. V. D. Weijden, N. O. E. Vischer, M. Bakker, R. J. M. V. Spanning, J. L. Snoep and H. V. Westerhoff (2002). DNA supercoiling by gyrase is linked to nucleoid compaction. *Molecular Biology Reports* **29**: 79–82. DOI: 10.1023/A:1020318705894.

Susin, M. F., R. L. Baldini, F. Gueiros-Filho and S. L. Gomes (2006). GroES/GroEL and DnaK/DnaJ have distinct roles in stress responses and during cell cycle progression in *Caulobacter crescentus*. *Journal of Bacteriology* **188**: 8044-8053. DOI: 10.1128/JB.00824-06.

Takata, H., T. Hanafusa, T. Mori, M. Shimura, Y. Iida, K. Ishikawa, K. Yoshikawa, Y. Yoshikawa and K. Maeshima (2013). Chromatin Compaction Protects Genomic DNA from Radiation Damage. *PLoS ONE* **8**: e75622. DOI: 10.1371/journal.pone.0075622.

Tessmer, I., T. Moore, R. G. Lloyd, A. Wilson, D. A. Erie, S. Allen and S. J. B. Tendler (2005). AFM studies on the role of the protein RdgC in bacterial DNA recombination. *Journal of Molecular Biology* **350**: 254-262. DOI: 10.1016/j.jmb.2005.04.043.

Thanedar, S. and W. Margolin (2004). FtsZ Exhibits Rapid Movement and Oscillation Waves in Helix-like Patterns in *Escherichia coli*. *Current Biology* **14**: 1167-1173. DOI: 10.1016/j.cub.2004.06.048.

The Global Proteome Machine Organization. The Global Proteome Machine: cRAP protein sequences. Available at: http://thegpm.org/crap/ [Accessed November 28, 2017].

Thévenaz, P., D. Sage and M. Unser (2012). Bi-Exponential Edge-Preserving Smoother. *IEEE Transactions on Image Processing* **21**(9): 3924--3936. DOI: 10.1109/TIP.2012.2200903.

Tracy, B. P., S. M. Gaida and E. T. Papoutsakis (2010). Flow cytometry for bacteria: enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes. *Current Opinion in Biotechnology* **21**(1): 85-99. DOI: https://doi.org/10.1016/j.copbio.2010.02.006.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Troll, J. V., D. M. Adin, A. M. Wier, N. Paquette, N. Silverman, W. E. Goldman, F. J. Stadermann, E. V. Stabb and M. J. Mcfall-ngai (2009). Peptidoglycan induces loss of a nuclear peptidoglycan recognition protein during host tissue development in a beneficial animal-bacterial symbiosis. *Cellular Microbiology* **11**: 1114-1127. DOI: 10.1111/j.1462-5822.2009.01315.x.

Truglio, J. J., D. L. Croteau, B. Van Houten and C. Kisker (2006). Prokaryotic Nucleotide Excision Repair: The UvrABC System. *Chemical Reviews* **106**: 233-252. DOI: 10.1021/cr040471u.

Twine, S. M., N. C. S. Mykytczuk, M. D. Petit, H. Shen, J. W. Conlan and J. F. Kelly (2006). *In vivo* proteomic analysis of the intracellular bacterial pathogen, *Francisella tularensis*, isolated from mouse spleen. *Biochemical and Biophysical Research Communications* **345**: 1621-1633. DOI: 10.1016/j.bbrc.2006.05.070.

van der Hoeven, R. and S. Forst (2009). OpnS, an outer membrane porin of *Xenorhabdus nematophila*, confers a competitive advantage for growth in the insect host. *Journal of Bacteriology* **191**: 5471-5479. DOI: 10.1128/JB.00148-09.

Verbrugghe, E., A. Van Parys, B. Leyman, F. Boyen, F. Haesebrouck and F. Pasmans (2015). HtpG contributes to *Salmonella typhimurium* intestinal persistence in pigs. *Veterinary Research* **46**: 118. DOI: 10.1186/s13567-015-0261-5.

Vizcaíno, J. A., A. Csordas, N. del-Toro, J. A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-Riverol, F. Reisinger, T. Ternent, Q.-W. Xu, R. Wang and H. Hermjakob (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Research* **44**(D1): D447-D456. DOI: 10.1093/nar/gkv1145.

Wachi, M. and M. Matsuhashi (1989). Negative control of cell division by *mreB*, a gene that functions in determining the rod shape of *Escherichia coli* cells. *Journal of Bacteriology* **171**: 3123-3127. DOI: 10.1128/jb.171.6.3123-3127.1989.

Wang, J., Y. Wu, G. Yang and S. Aksoy (2009). Interactions between mutualist *Wigglesworthia* and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypanosome transmission. *Proceedings of the National Academy of Sciences of the United States of America* **106**(29): 12133-12138. DOI: 10.1073/pnas.0901226106.

Wang, S., H. Arellano-Santoyo, P. A. Combs and J. W. Shaevitz (2010). Actin-like cytoskeleton filaments contribute to cell mechanics in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 9182-9185. DOI: 10.1073/pnas.0911517107.

Wang, Y., S. Maharana, M. D. Wang and G. V. Shivashankar (2014). Super-resolution microscopy reveals decondensed chromatin structure at transcription sites. *Scientific Reports* **4**(1): 4477. DOI: 10.1038/srep04477.

Weiss, D. S. (2004). Bacterial cell division and the septal ring. *Molecular Microbiology* **54**(3): 588-597. DOI: 10.1111/j.1365-2958.2004.04283.x.

Weissgerber, T., R. Zigann, D. Bruce, Y.-j. Chang, J. C. Detter, C. Han, L. Hauser, C. D. Jeffries, M. Land and A. C. Munk (2011). Complete genome sequence of *Allochromatium vinosum* DSM 180 T. *Standards in Genomic Sciences* **5**(3): 311-330. DOI: 10.4056/sigs.2334270.

Weissgerber, T., M. Sylvester, L. Kröninger and C. Dahl (2014). A comparative quantitative proteomic study identifies new proteins relevant for sulfur oxidation in the purple sulfur bacterium *Allochromatium vinosum*. *Applied and Environmental Microbiology* **80**: 2279-2292. DOI: 10.1128/aem.04182-13.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Welte, C., S. Hafner, C. Krätzer, A. Quentmeier, C. G. Friedrich and C. Dahl (2009). Interaction between Sox proteins of two physiologically distinct bacteria and a new protein involved in thiosulfate oxidation. *FEBS Letters* **583**: 1281-1286. DOI: 10.1016/j.febslet.2009.03.020.

Wippler, J., M. Kleiner, C. Lott, A. Gruhl, P. E. Abraham, R. J. Giannone, J. C. Young, R. L. Hettich and N. Dubilier (2016). Transcriptomic and proteomic insights into innate immunity and adaptations to a symbiotic lifestyle in the gutless marine worm *Olavius algarvensis*. *BMC Genomics* **17**: 942. DOI: 10.1186/s12864-016-3293-y.

Woyke, T., D. Tighe, K. Mavromatis, A. Clum, A. Copeland, W. Schackwitz, A. Lapidus, D. Wu, J. P. Mccutcheon, B. R. Mcdonald, N. A. Moran, J. Bristow and J. F. Cheng (2010). One bacterial cell, one complete genome. *PLoS ONE* **5**(4): e10314. DOI: 10.1371/journal.pone.0010314.

Wright, M. N. and A. Ziegler (2015). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software* **77**: 1-17. DOI: 10.18637/jss.v077.i01.

Yoshikawa, Y., T. Mori, N. Magome, K. Hibino and K. Yoshikawa (2008). DNA compaction plays a key role in radioprotection against double-strand breaks as revealed by single-molecule observation. *Chemical Physics Letters* **456**: 80-83. DOI: 10.1016/j.cplett.2008.03.009.

Zambrano, N., P. P. Guichard, Y. Bi, B. Cayrol, S. Marco and V. Arluison (2009). Involvement of HFq protein in the post-transcriptional regulation of *E. coli* bacterial cytoskeleton and cell division proteins. *Cell Cycle* **8**: 2470-2472. DOI: 10.4161/cc.8.15.9090.

Zerulla, K. and J. Soppa (2014). Polyploidy in haloarchaea: Advantages for growth and survival. *Frontiers in Microbiology* **5**: 274. DOI: 10.3389/fmicb.2014.00274.

Zybailov, B., A. L. Mosley, M. E. Sardiu, M. K. Coleman, L. Florens and M. P. Washburn (2006). Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *Journal of Proteome Research* **5**: 2339-2347. DOI: 10.1021/pr060161n.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

We thank captains and crews of R/V Atlantis and DSV Alvin who supported sampling during cruises AT26-23 and AT37-12. We are grateful to Jana Matulla and Annette Meuche for excellent technical assistance in sample preparation for proteomics and electron microscopy, respectively. Thanks to Ruby Ponnudurai and Frank Unfried for help with CARD-FISH, and to Alexander Graf, Mathis Appelbaum, Judith Zimmermann and Silke Wetzel for advice on epifluorescence microscopy and staining. We greatly appreciate Elisa Kasbohm's help with random forest analyses. We are very grateful to Jörg Bernhardt for stitching the transmission electron micrographs to produce a panorama image with high resolution. This work was supported by the German Research Foundation DFG (grant MA 6346/2-1 to S.M.), a fellowship of the Institute of Marine Biotechnology Greifswald (T.H., M.M.), a German Academic Exchange Service (DAAD) grant (T.H.), the NC State Chancellor's Faculty Excellence Program Cluster on Microbiomes and Complex Microbial Communities (M.K.), the USDA National Institute of Food and Agriculture, Hatch project 1014212 (M.K.), the U.S. National Science Foundation (grants OCE-1131095 and OCE-1559198 to S.M.S), and The WHOI Investment in Science Fund (to S.M.S).

Author contributions

S.M. conceived the study, S.M., T.H., M.K. designed the experiments, T.H. and H.F. took samples. T.H. prepared samples for CARD-FISH and metaproteomics analysis, analyzed data, conducted statistical analyses and prepared figures. T.H. and S.M. wrote the manuscript with input from all coauthors. T.S. was involved in project coordination, S.M.S. obtained funding for the research cruises and coordinated sampling as chief scientist. C.H. and F.B. performed MS measurements of metaproteomics samples, D.B. coordinated MS measurements. M.M. and J.P.-F. contributed to fluorescence microscopy and R.S. conducted electron microscopy analyses. P.H. performed flow cytometry analyses, and U.V. coordinated flow cytometry measurements.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.032177; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Conflicts of interest

The authors declare no conflicts of interest.