

# Bacterial symbiont subpopulations have different roles in a deep-sea symbiosis

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

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

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

**Competing interests:** The authors declare that no competing interests exist.

**Funding:** See page 25

**Received:** 28 April 2020

**Accepted:** 05 December 2020

**Published:** 06 January 2021

**Reviewing editor:** Gisela Storz, National Institute of Child Health and Human Development, United States

© Copyright Hinzke et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

**Abstract** The hydrothermal vent tubeworm *Riftia pachyptila* hosts a single 16S rRNA phylotype of intracellular sulfur-oxidizing symbionts, which vary considerably in cell morphology and exhibit a remarkable degree of physiological diversity and redundancy, even in the same host. To elucidate whether multiple metabolic routes are employed in the same cells or rather in distinct symbiont subpopulations, we enriched symbionts according to cell size by density gradient centrifugation. Metaproteomic analysis, microscopy, and flow cytometry strongly suggest that *Riftia* symbiont cells of different sizes represent metabolically dissimilar stages of a physiological differentiation process: While small symbionts actively divide and may establish cellular symbiont-host interaction, large symbionts apparently do not divide, but still replicate DNA, leading to DNA endoreduplication. Moreover, in large symbionts, carbon fixation and biomass production seem to be metabolic priorities. We propose that this division of labor between smaller and larger symbionts benefits the productivity of the symbiosis as a whole.

## Introduction

The chemoautotrophic gammaproteobacterium *Candidatus* Endoriftia persephone, sulfur-oxidizing endosymbiont of the deep-sea tubeworm *Riftia pachyptila* (here *Riftia*), provides all nutrition for its gutless host (Cavanaugh et al., 1981; Felbeck, 1981; Hand, 1987; Robidart et al., 2008). *C. 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). Here, fueled by the oxidation of reduced sulfur compounds (mainly sulfide), the symbionts fix CO<sub>2</sub> and produce considerable amounts of biomass, which constitutes the host's source of nutrition (Cavanaugh et al., 1981; Felbeck et al., 1981; Childress et al., 1991; Goffredi et al., 1997a;

**Girguis and Childress, 2006**). *Riftia* acquires these nutrients mainly by directly digesting parts of the symbiont population (**Hand, 1987; Boetius and Felbeck, 1995; Hinzke et al., 2019**). In addition, the symbionts possibly also actively secrete carbon compounds into the bacteriocytes (**Felbeck and Jarchow, 1998**). Thus provided, *Riftia* grows rapidly and reaches nearly 2 m in body length (**Jones, 1981; Lutz et al., 1994; McClain et al., 2015**), despite its lack of a digestive system. This enormous productivity is underpinned by a multitude of symbiosis-specific adaptations of both host and symbiont. These include host proteins involved in substrate transport from the hydrothermal vent environment to the symbionts, such as ATPases and carbonic anhydrases for CO<sub>2</sub> uptake and conversion (**Goffredi et al., 1997b; De Cian et al., 2003**), and specialized extracellular hemoglobins which simultaneously and reversibly bind sulfide and oxygen (**Childress et al., 1984**). Moreover, symbiont cells in the highly vascularized trophosome contain globules of elemental sulfur, an intermediate of sulfide oxidation (**Bright and Sorgo, 2003; Pflugfelder et al., 2005**). As trophosome sulfur content is positively correlated with sulfide concentrations in the surrounding water and decreases in the absence of external sulfide (**Childress et al., 1991; Robidart et al., 2011; Scott et al., 2012**), sulfur in the symbionts likely serves as an energy reserve for times of sulfide deprivation.

Although the symbiont population consists of a single 16S rRNA phylotype (comprising one dominating and several minor genotypes; **Polzin et al., 2019**), it was previously shown to exhibit remarkable metabolic versatility and redundancy: As demonstrated by proteomic analyses, symbionts from the same host animal expressed enzymes of two CO<sub>2</sub> 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, whereas toward 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 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. Unlike previous molecular analyses that studied the uncultured *Riftia* symbiont's metabolic capabilities on the population level in 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. To assess whether these differences might be influenced by naturally occurring variations in energy supply, we conducted our analyses with symbionts from both sulfur-rich and sulfur-depleted trophosomes. 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, which leads to remarkable physiological heterogeneity within the symbiont population.

## Results

### Enrichment of individual symbiont cell sizes by gradient centrifugation

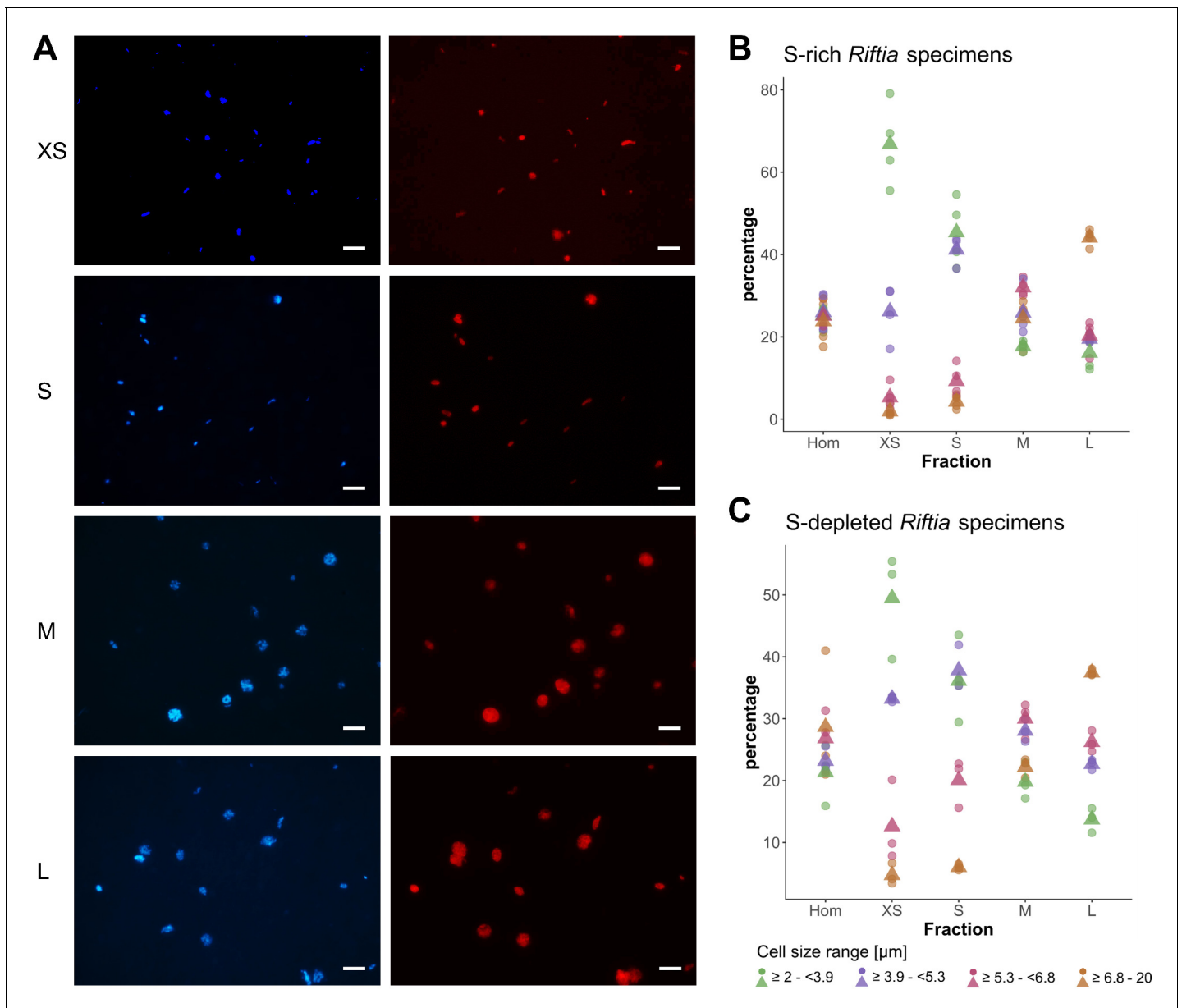
Our rate-zonal gradient centrifugation approach allowed us to enrich distinct symbiont cell sizes from *Riftia* trophosome tissue. Based on CARD-FISH microscopy, we defined four size ranges (**Figure 1**): very small symbiont cells ( $\geq 2.0$  –  $< 3.9$   $\mu\text{m}$  diameter), small ( $\geq 3.9$  –  $< 5.3$   $\mu\text{m}$ ), medium ( $\geq 5.3$  –  $< 6.8$   $\mu\text{m}$ ), and large symbiont cells ( $\geq 6.8$ – $20.0$   $\mu\text{m}$ ; see also Appendix section A). For subsequent comparative metaproteomic analyses, we chose those 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**).

### 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* 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 Appendix section 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, Figure 2—figure supplement 1**): 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 (**Figure 2—figure supplement 1**). For quantification of DNA in smaller and larger symbionts, we compared median fluorescence intensities (MFI) per particle between populations 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 1994 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 file 6**).

### Protein identifications and relative protein abundance

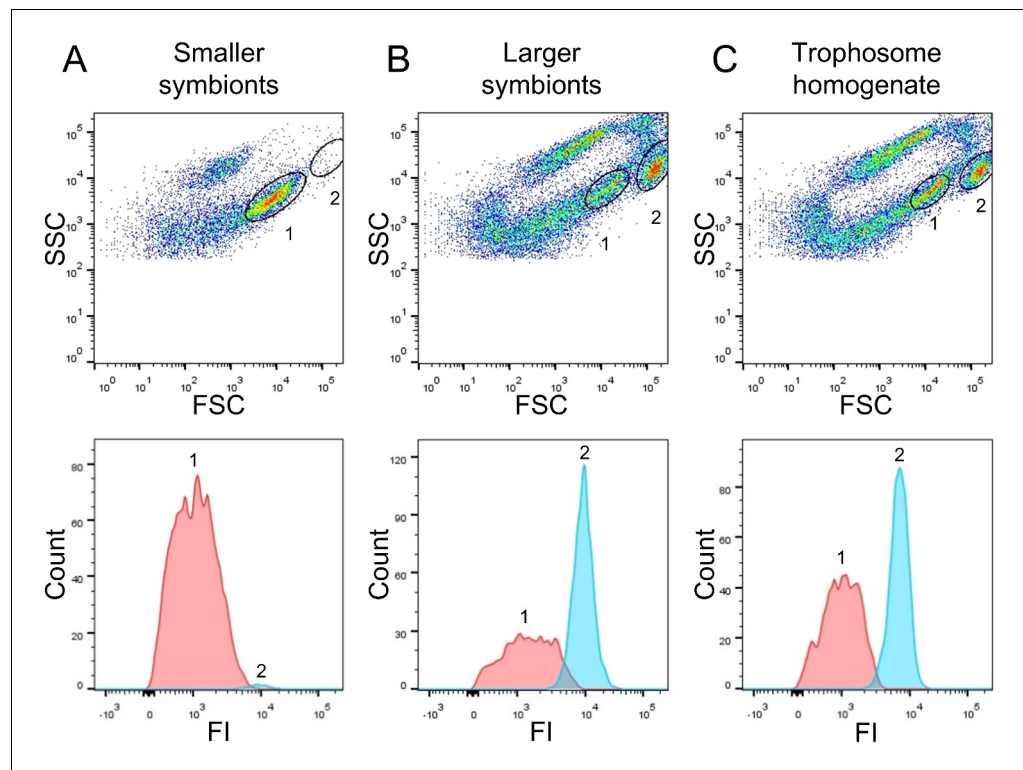
We identified a total of 1946 symbiont proteins across all sample types, including the four gradient fractions XS – L and non-enriched homogenate from both sulfur-rich and sulfur-depleted *Riftia* specimens (**Appendix 1—table 2**). Our sample fractionation by gradient centrifugation thus facilitated detection of around 60% of the symbiont's theoretical proteome, which encompasses 3182 proteins in PRJNA60889, and yielded substantially higher symbiont protein identification rates than non-enriched trophosome homogenate samples alone (1223 total symbiont protein identifications). After stringent filtering and normalization, a subset of 1212 symbiont proteins from gradient fractions XS – L was included in statistical analysis using abundance profile clustering and random forests (**Supplementary file 2**). A total of 465 proteins showed significant differences in relative abundance (**Appendix 1—figure 1**; note that the term 'significant' denominates trends that were consistent across all replicates in the context of our statistical approach). In **Figure 3** and **Supplementary file 2**, proteins that showed such significant changes in relative abundance are marked with asterisks. Of all proteins with significant abundance changes, 56% (261 proteins)



**Figure 1.** CARD-FISH images of *Riftia* symbiont cells after density gradient centrifugation. (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), whereas 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. Between 300 and 1300 individual cells were measured per filter (average: 590). (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 file 1**. Dots: individual % values, triangles: average % values. Please note the different scaling of the y axes in B and C.

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 file 2**).

In our comparison of energy-rich (S-rich) and energy-depleted (S-depleted) gradient fractions, very few proteins were detected exclusively in sulfur-rich samples (61 of 1212 proteins) or only in sulfur-depleted samples (77 proteins). The majority of symbiont proteins showed very similar



**Figure 2.** Flow-cytometry-based 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 (FI) 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 **Figure 2—figure supplement 1**, Methods, and **Supplementary file 6** for more details). This analysis was based on two *Riftia* specimens with medium sulfur content.

The online version of this article includes the following figure supplement(s) for figure 2:

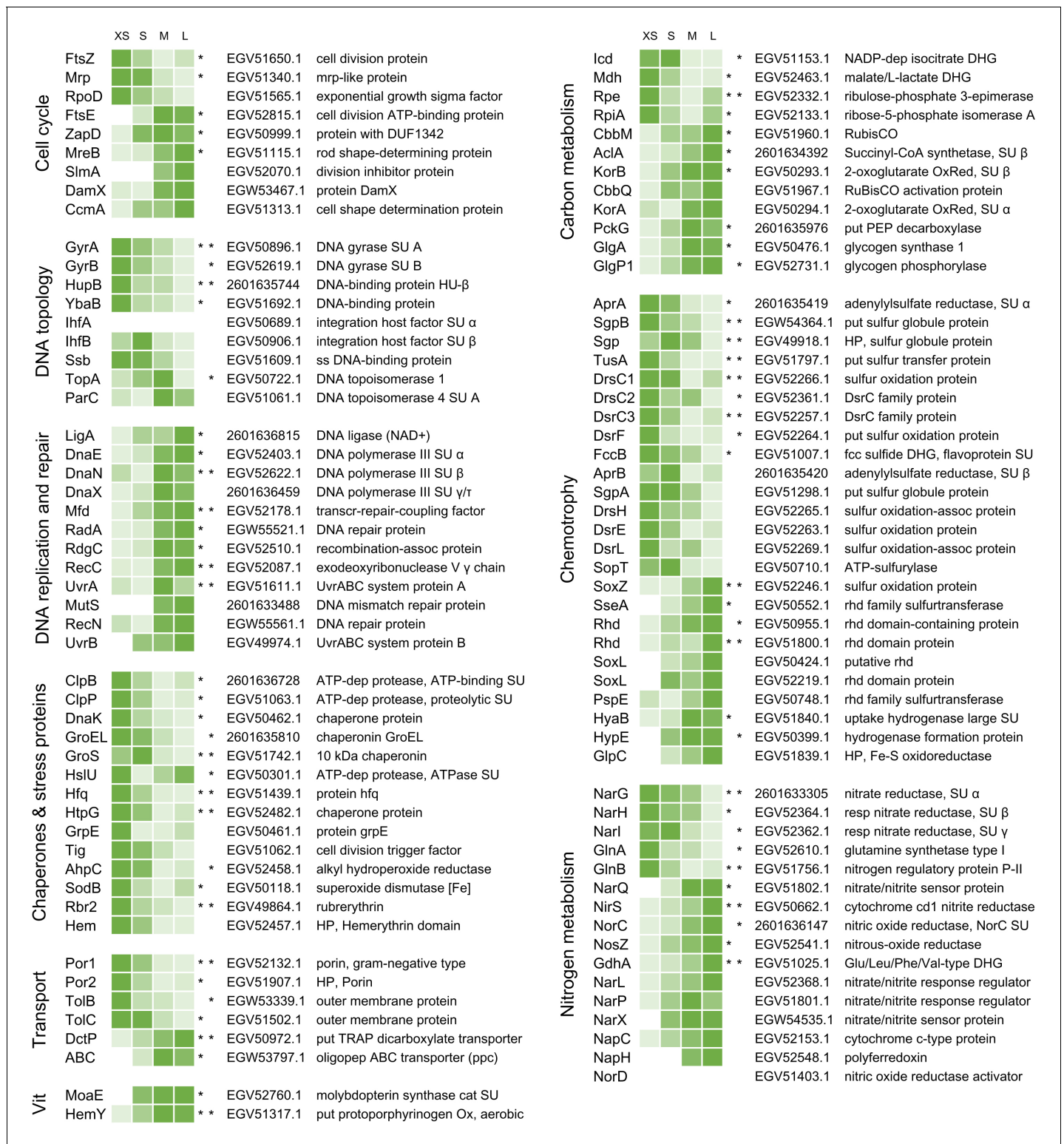
**Figure supplement 1.** Fluorescence microscopy and fluorescence-activated cell sorting (FACS) of *Riftia* symbiont cells.

abundance trends in sulfur-rich and sulfur-depleted samples. The following results will therefore not discriminate between both sample types and subsequent figures will focus on S-rich samples (unless otherwise stated). For a discussion of specific differences observed between symbionts from energy-rich and energy-starved trophosome tissue, see Appendix section 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 file 3a**). While the cell division protein FtsZ, as well as 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 Appendix section B).



**Figure 3.** Abundance trends of selected Endoriftia proteins of various functions in the four fractions XS to L. Trends are indicated by color shades from light green (lowest protein abundance across all four fractions) to dark green (highest abundance across all four fractions; note that colors do not allow comparison of protein abundance between proteins). Abundance values in the heat map are based on statistical evaluation of four biological replicates with sulfur-rich trophosomes (for abundance trends of sulfur-depleted samples refer to **Figure 3—figure supplement 1** and Appendix section B). Proteins marked with asterisks show statistically significant trends, that is, differences that are consistent across all replicates in S-rich (left asterisk) or S-depleted specimens (right asterisk), or both (two asterisks). White cells indicate that this protein was not detected in this sample or too low abundant. *Figure 3 continued on next page*

Figure 3 continued

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 files 2 and 3**, 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.

The online version of this article includes the following figure supplement(s) for figure 3:

**Figure supplement 1.** Abundance trends of selected Endoriffia proteins of various functions in the four fractions XS to L in sulfur-rich (S-rich) and sulfur-depleted (S-depl) *Riftia* specimens.

## 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 file 3b**).

## 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 file 3c**). On the other hand, all five detected tripartite ATP-independent periplasmic (TRAP) transporter subunits and 10 out of 13 ABC transporter components were relatively more abundant in fraction L (see also Appendix section 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 file 3d**). In contrast, the key enzymes of the two CO<sub>2</sub>-fixing pathways, Calvin cycle (ribulose-1.5-bisphosphate carboxylase/oxygenase, 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. Relative stable isotope fingerprint (SIF) values, that is, relative  $\delta^{13}\text{C}$  values, generally also increased from fraction XS to L (see Appendix sections B and D for details).

**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 file 3e, Appendix 1—figure 4**). 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, that is, 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 file 2**). 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.

**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 file 3f, Appendix section 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

	Hom	XS	S		
transmembrane protein 214-B				DN39679_c0_g1_i4::g.30017	Apoptosis
beta carbonic anhydrase 1				DN39690_c2_g7_i10::g.29161	Carbon metabolism
acid phosphatase type 7				DN41253_c0_g9_i2::g.125392	Degradation
glucoamylase 1				DN36692_c1_g2_i3::g.169923	Degradation (glycans)
lysosomal alpha-glucosidase				DN36692_c1_g2_i4::g.169924	Degradation (glycans)
cathepsin Z				DN38047_c1_g1_i1::g.177385	Degradation (proteins)
legumain				DN39514_c3_g1_i1::g.201492	Degradation (proteins)
hypoxia up-regulated protein 1				DN38980_c0_g1_i11::g.75054	Hypoxia
hypoxia up-regulated protein 1				DN38980_c0_g1_i5::g.75039	Hypoxia
PG-recognition protein SC1a/b				DN31073_c0_g1_i1::g.27245	Immune system
thiosulfate sulfurtransferase				DN37892_c0_g1_i3::g.78847	Sulfur metabolism

**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). Relative abundance trends are indicated by color shades from light blue (lowest protein abundance across the three sample types) to dark blue (highest abundance), based on mean values from four biological replicates with sulfur-rich trophosome. (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 Materials and methods). For a complete list of host proteins with significantly higher abundance in fractions XS and S (compared to Hom) see **Supplementary file 4**. 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. PG: peptidoglycan.

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

Fifty (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 file 3g**). 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 file 3h**).

### Potentially symbiosis-specific host proteins

Our density 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 file 4**). 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 (for example, acid phosphatase, digestive proteases, and glycan degradation enzymes), as well as hypoxia up-regulated proteins, a thiosulfate sulfurtransferase, and transmembrane protein 214-B.

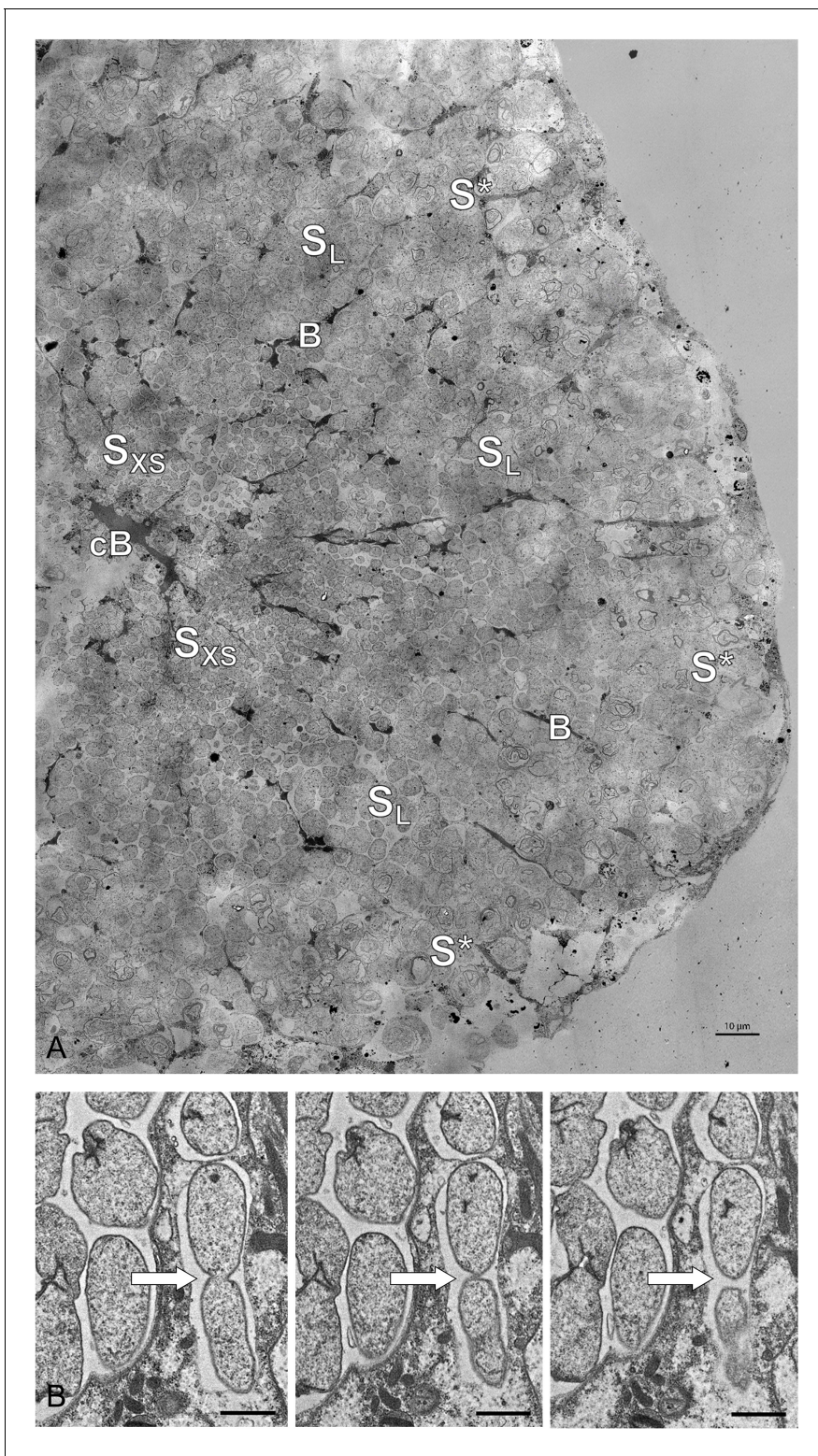
## Discussion

### 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 Sörgo, 2003**, the smallest





**Figure 5.** Electron microscopy of *Riftia* trophosome tissue. (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 (i.e. symbiont-containing host cells) in the central lobule zone. Symbiont cell size increases toward 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 **Figure 5—figure supplement 1**. Contrast and brightness Figure 5 continued on next page

Figure 5 continued

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  $\mu\text{m}$ . Despite thorough screening, we did not observe cell division in large Endoriftia cells in any of the TEM sections. This corroborates the idea that small and large symbiont subpopulations are developmental stages of the same Endoriftia strain. The online version of this article includes the following figure supplement(s) for figure 5:

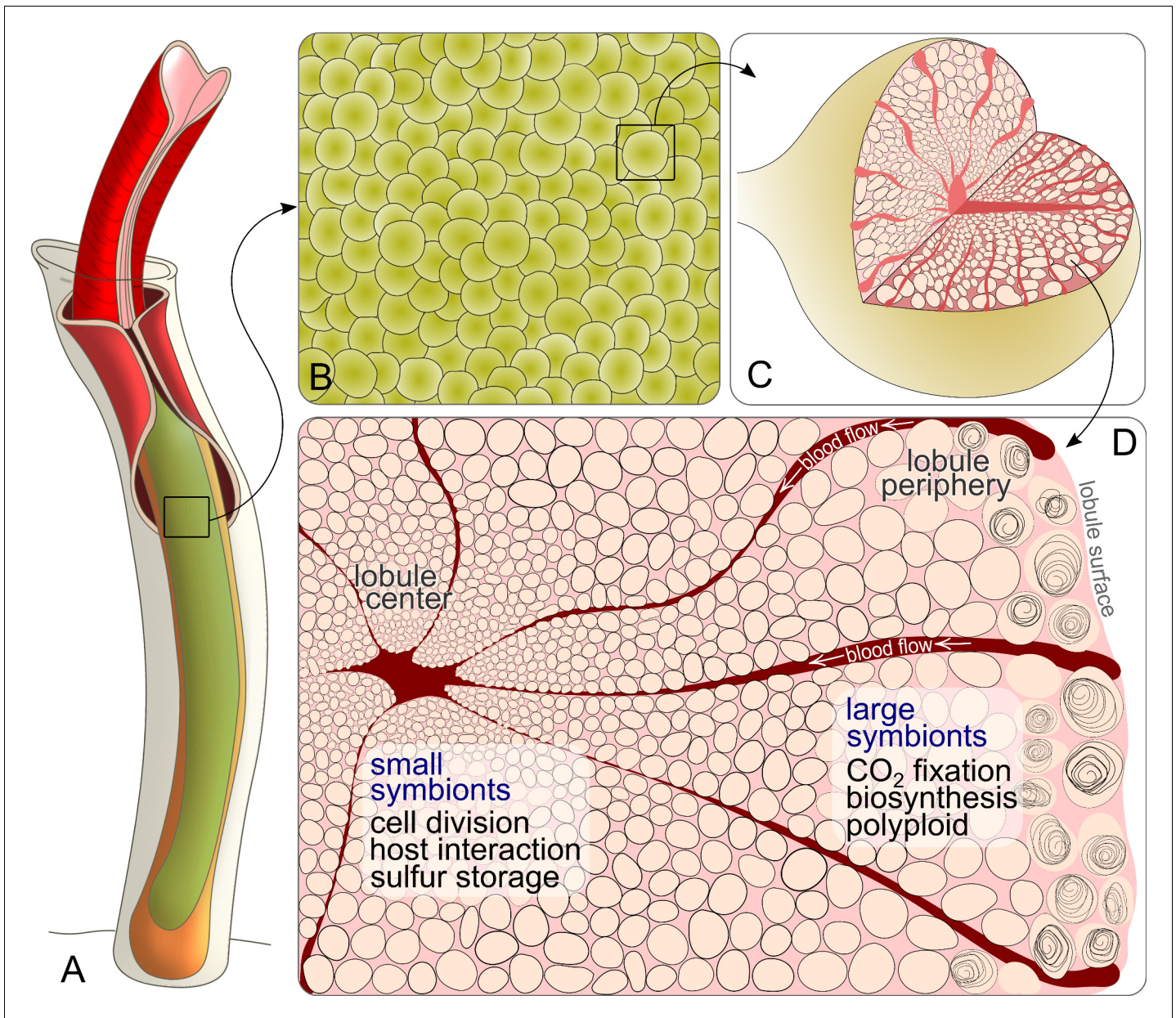
**Figure supplement 1.** High-resolution TEM image of the trophosome lobule section shown in **Figure 5**.

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**). (Although originally referring to eukaryotic cells, the term ‘differentiation’ is also applied to bacteria with strongly varying morphologies, such as *Rhizobia* in legume root nodules and heterocyst-forming cyanobacteria. We will therefore also use it in the following to describe the development from small to large Endoriftia in this study.)

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, for example, 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 supports the idea of relatively less cell division in large *Riftia* symbionts. Although Endoriftia’s major cell division protein FtsZ was notably (1.75x) less abundant in fraction L (compared to fraction XS), it was not completely absent. This may point to additional FtsZ functions, besides cell division (as also suggested for *Anabaena* [**Klint et al., 2007**] and *E. coli* [**Thanedar and Margolin, 2004**]).

### Large symbionts have more genome copies and less compact chromosomes

Endoriftia’s development into large, non-dividing (but still replicating) cells leads to endoreduplication cycles and an increase in genome copy number, as indicated by our flow cytometry analysis (**Figure 2, Figure 2—figure supplement 1, Supplementary file 6**). This observation is also 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 additionally 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**



**Figure 6.** Schematic drawing of *Riftia* symbiont cells inside the trophosome. (A) An adult tubeworm reaches 2 m in body length. Its symbiont-containing organ, the trophosome (green), fills most of the body cavity (coelomic cavity), and is immersed in coelomic (non-vascular) blood (not shown). (B) Close-up of the lobular trophosome tissue. (C) Single lobule (200–500  $\mu\text{m}$  in diameter) with interior blood vessels (blood-filled spaces) and symbiont cells visible. (D) Cross-section through a trophosome lobule (similar to that in **Figure 5A**) with small symbiont cells located in the center around an efferent blood vessel. Symbiont cell size increases toward the lobule periphery, where the largest symbionts are digested by the host (curls). Blood flow from lobule periphery to lobule center may cause gradients in nutrient availability. Based on the results of this study, the most striking characteristics of small and large symbionts, which determine their respective roles in the symbiosis, are indicated. Note that this is a simplified illustration, in which the various kinds of host cells (including their membranes, nuclei and organelles), as well as blood vessels or gonads that line the surface of the trophosome were omitted for clarity's sake. (Illustrations based on drawings, TEM images and descriptions by **van der Land and Nørrevang, 1977; Felbeck and Turner, 1995; Bright and Sorgo, 2003**).

**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, whereas 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 have been reported for various bacterial insect symbionts, for example, 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*, an *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 toward 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 *Riftia* 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 to 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 decondensed DNA is also known in plant cells (Kondorosi and Kondorosi, 2004). As decondensation occurs in actively transcribed DNA regions (Wang et al., 2015), 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 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 crescentus*, 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, one might speculate that cell-division-related processes require elevated levels of chaperones and stress proteins, for example, to ensure correct assembly of all parts of the division machinery or (in case of intracellular bacteria) 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'Haese 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 particularly exposed to this presumptive ROS

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

Interestingly, we found no indications of a strong bacterial stress response in fraction L, indicating that imminent digestion by the host poses no particular stress to the large symbionts. Possibly, bacterial degradation happens too fast to elicit a stress response, or a digestion-related stress response is suppressed during symbiosis, either by the symbionts themselves or by the host via a yet to be determined mechanism.

Host-microbe interactions may be particularly important in small Endoriftia. Abundant Endoriftia membrane proteins might play a key role in host interaction in small symbionts. Particularly, the high and differential abundance of porin Sym EGV52132.1, the most abundant symbiont protein in all fractions, which was nearly three times more abundant in fraction XS (11.7%orgNSAF) than in fraction L (4.0%orgNSAF), suggests that this protein may be of varying relative importance throughout the symbiont's differentiation process. Porins are water-filled channels in the outer membrane, through which small hydrophilic molecules can diffuse (Fernández and Hancock, 2012). In the oyster pathogen *Vibrio splendidus*, the porin OmpU serves as adhesin or invasin and is involved in recognition by the host cell (Duperthuy et al., 2011), while in *Neisseria gonorrhoeae*, a porin inhibits phagocytosis by human immune cells (Mosleh et al., 1998; Lorenzen et al., 2000). Interestingly, the phagocytosis-inhibiting action of *N. gonorrhoeae* porin apparently involves interference with the host's oxidative burst; that is, the porin allows the pathogen to evade killing by host-produced ROS (Lorenzen et al., 2000). Although the exact function of Endoriftia porin has not been elucidated yet, we suggest that it may have a similar function in resistance against host stress or ROS. This would 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 (Hinze 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), whereas *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 in 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-transcriptional regulator *hfq* often lead to reduced fitness and virulence in bacterial pathogens (reviewed in Chao and Vogel, 2010). Moreover, ClpB in *Listeria* is apparently specifically involved in virulence (Chastanet et al., 2004), as are

ClpX and ClpP in *Staphylococcus aureus* (Frees et al., 2003). In the insect symbiont *Wolbachia*, HU beta was suggested to directly interact with the host (Beckmann et al., 2013). Additional symbiont proteins that may protect small *Endoriftia* from host interference, and particularly so in S-depleted *Riftia* specimens, included an ankyrin protein and an FK506-binding protein (see Appendix section B).

### Interaction-specific host proteins

We detected a number of *Riftia* proteins which were co-enriched with symbiont cells in fractions XS and/or S and may therefore potentially be 'symbiosis-specific'. Although we cannot entirely exclude the possibility that non-symbiont-related host organelle-bound proteins may also have accumulated in this fraction (e.g. the ribosomal and mitochondrial proteins we identified), it appears very likely that the candidates discussed below may facilitate direct host-microbe interactions or enable the host to provide optimal conditions for the symbiont. (1) Peptidoglycan-recognition proteins, for example, are involved in innate immunity (Kang et al., 1998) and have previously been shown or suggested to participate in symbiotic interactions (Troll et al., 2009; Wang et al., 2009; Royet et al., 2011; Wippler et al., 2016). (2) Since oxygen concentrations in the trophosome might be comparatively low (benefitting the microaerophilic symbionts; Hinzke et al., 2019), the hypoxia up-regulated *Riftia* proteins we detected may present a protective adaptation of the host to these hypoxic conditions. In support of this idea, Hyou1 was shown to have a protective function during hypoxia in human cells (Ozawa et al., 1999). (3) Moreover, enrichment of beta carbonic anhydrase 1, which interconverts bicarbonate and CO<sub>2</sub> (Goffredi et al., 1999; De Cian et al., 2003), suggests that this host protein serves to optimally provide the symbionts with CO<sub>2</sub> for fixation. (4) The host transmembrane protein 214-B (TMP214-B), which was exclusively detected in symbiont-enriched fractions (but not in trophosome homogenate) may be involved in cell death of symbiont-containing bacteriocytes by an apoptosis-related mechanism. This would be in line with our previous suggestion that apoptosis-related proteins may play a role in symbiont and bacteriocyte cell death (Hinzke et al., 2019) and is further supported by the fact that TMP214-B was shown to be involved in apoptosis caused by endoplasmic reticulum stress (Li et al., 2013). (5) Finally, the detection of degradation proteins such as cathepsin Z, legumain, glucoamylase 1, and lysosomal alpha-glucosidase in fractions XS and S may imply 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 (although we did not see any evidence for this in the TEM images).

### 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 file 5) in fraction L suggest that large *Endoriftia* cells are relatively more engaged in the production of organic material than smaller symbiont cells. In support of this idea, we (1) observed notably higher RubisCO mRNA signal intensity in large symbiont cells than in smaller *Riftia* symbionts in our HCR-FISH analysis (Appendix section D, Appendix 1—figure 3). (2) Most cofactor- and vitamin metabolism-related proteins were more abundant in fractions M and/or L than in fractions XS or S. Moreover, (3) higher abundances of glycogen-producing enzymes in fraction L suggest that large symbionts invest relatively more of their biosynthetic capacities in storage of fixed carbon in the form of glycogen than smaller symbionts.

These observations concur with an autoradiographic study of Bright et al., 2000, who observed highest <sup>14</sup>C carbon incorporation in the *Riftia* trophosome lobule periphery and lowest short-term incorporation in the lobule center, and with results of Sorgo et al., 2002, who noted a glycogen gradient in the symbiont cells, with increasing glycogen density from the lobule center toward the periphery, that is, toward larger symbiont cells. Interestingly, in addition to our above observation that overall carbon fixation appears to play a more prominent role in larger symbionts, we also found clear indications that individual contributions of the two autotrophic pathways differ notably between symbionts of different sizes, as suggested by very dissimilar relative  $\delta^{13}\text{C}$  values in our SIF analysis (Appendix section D).

A possible explanation for all these findings would be a substrate gradient caused by the direction of the blood flow from lobule periphery to lobule center (*Felbeck and Turner, 1995; Figure 6, Appendix section G*), which may cause carbon concentrations to be higher around the largest symbionts. While such biochemical gradients were initially suggested as the immediate cause of the symbionts' diverse morphology (*Hand, 1987*), subsequent studies instead proposed that the symbionts' polymorphism as well as metabolic differences between them result from their cell cycle (*Sorgo et al., 2002; Bright and Sorgo, 2003*). However, both options are not mutually exclusive, but could actually complement each other, that is, different developmental symbiont stages could face dissimilar substrate concentrations.

The putative concentration gradient, which was suggested to trigger Endoriftia's differentiation (*Bright and Sorgo, 2003*), may thus 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<sub>2</sub> incorporation enzymes in large symbionts. Large *Riftia* symbionts thus presumably not only benefit from higher CO<sub>2</sub> levels, but also have more biosynthetic capacities at their disposal than small symbionts.

This facilitates a division of responsibilities between the symbiont subpopulations: Small Endoriftia invest a considerable part of their resources in cell division 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. Such a division of labor provides the combined advantage of multiple genome copies with resulting increased transcription efficiency and productivity in large cells, and a focus on cell division in small cells. At the same time, the respective disadvantages are avoided (large Endoriftia would need to invest much more resources in cell division and replication of their multiple genome copies than small cells, while smaller Endoriftia could not produce as much biomass as large symbionts). Increased metabolic efficiency as a consequence of specialized bacteria performing complementary tasks has also been reported for different bacterial species and strains working together in other symbioses (*Zheng et al., 2019; Ankras et al., 2020*). In *Riftia*, the largest symbionts that are digested at the trophosome lobule periphery are those with the highest nutritional value. They carry not only abundant carbon but also contain plenty of N-rich biomolecules in their multiple genome copies, and are likely easier to digest than smaller symbionts, due to their lower sulfur content (see below).

Small Endoriftia store more sulfur and are more involved in sulfide oxidation. Smaller symbionts produce relatively more sulfur globules for sulfur storage than larger symbiont cells, as indicated by relatively higher abundance of sulfur globule proteins in fraction XS (*Appendix 1—figure 4*). Enclosed in a proteinaceous envelope, Endoriftia's sulfur globules contain elemental sulfur that is formed during sulfide oxidation (*Wilmot and Vetter, 1990; Pflugfelder et al., 2005; Markert et al., 2011*). Higher numbers of these globules in small symbionts are in agreement with observations of *Hand, 1987*, who noted more sulfur deposits in central (small) than in peripheral (large) *Riftia* symbionts. Although this finding was not supported by a subsequent study (*Pflugfelder et al., 2005*), our results do point to different amounts of S storage in different Endoriftia subpopulations. Storing relatively more sulfur in small Endoriftia, but less in large symbionts might aid symbiont digestion by the host (S-poor symbionts being less toxic compared to S-rich cells), a strategy that was also suggested for the thiotrophic *C. orbicularis* symbiosis (*Caro et al., 2007*). As shown for the free-living thiotrophic model bacterium *Allochromatium vinosum*, activation of stored sulfur involves trafficking proteins such as TusA, which is involved in sulfur transfer to DsrEFH and DsrC (*Stockdreher et al., 2014*). In our study, the highly abundant TusA, several DsrC copies as well as DsrEFH were all detected with highest abundances in fraction XS, thus supporting the idea of relatively more re-mobilization of sulfur and subsequent utilization of reduced sulfur compounds in small Endoriftia. As the highly abundant adenylylsulfate reductase AprAB, the ATP-sulfurylase SopT, and sulfide dehydrogenase subunit FccB were also detected with higher abundances in fractions XS or S than in M or L, one might conclude that sulfide oxidation itself also plays a more prominent role in smaller symbionts than in larger symbionts. More sulfide oxidation would also be the most straightforward explanation for the observed higher amounts of stored sulfur in small

symbionts (see above). However, as we detected the dissimilatory sulfite reductase DsrAB, the third key enzyme of cytoplasmic sulfide oxidation, with a rather ambiguous abundance pattern (**Supplementary file 3e**), this idea remains speculative and requires further analysis.

### 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. Although *Endoriftia* prefers sulfide (**Wilmot and Vetter, 1990**), it was also shown to oxidize thiosulfate, and may supplement sulfide oxidation with thiosulfate oxidation for energy production (**Robidart et al., 2011**). Since expression of the thiosulfate-oxidizing Sox (sulfur oxidation) complex was shown to be upregulated in the presence of thiosulfate in *A. vinosum* (**Grimm et al., 2011**), we speculate that higher Sox abundance in large symbionts indicates higher thiosulfate concentrations in the trophosome lobule periphery than in the lobule center. This could be due to a concentration gradient (as proposed above for CO<sub>2</sub>) and/or possibly also 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**). The combined use of both sulfide and thiosulfate has also been reported for chemoautotrophic symbionts of snails and mussels (**Beinart et al., 2015**). In these symbioses, thiosulfate oxidation supported carbon fixation just as effectively as sulfide (or even more effectively). Using thiosulfate as their (additional) energy source may therefore not be an energetic disadvantage for *Riftia* symbionts. Moreover, thiosulfate oxidation in *Endoriftia* may have the positive side-effect of reducing competition for sulfide between larger and smaller symbionts.

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 file 5**). 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 Appendix section F).

### 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<sub>2</sub>-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 toward fraction L as well (although at very low concentrations; **Supplementary file 2b**), supporting the idea of relatively more hydrogen oxidation in large symbionts. Like for CO<sub>2</sub> and thiosulfate, this might be due to a concentration gradient with highest hydrogen concentrations at the lobule periphery and lowest concentrations toward 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, including *Riftia* (**Petersen et al., 2011**). However, above-seafloor hydrogen concentrations in diffuse-flow fluids at our particular sampling site were reported to be very low (<2 μM; **McNichol et al., 2018**), and it is presently unclear whether and how ambient hydrogen



would reach the intracellular symbionts in the trophosome. Nevertheless, taking advantage of hydrogen oxidation in addition to sulfide- and thiosulfate oxidation, that is, using a broader repertoire of electron donors, would potentially enhance the metabolic flexibility, particularly of large Endoriftia. However, H<sub>2</sub> 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); Appendix section 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.

### Small and large symbionts favor separate parts of the denitrification pathway

Our results suggest that small *Riftia* symbionts rely relatively more on the NarGHI-mediated first step of anaerobic respiratory nitrate reduction to nitrite. In contrast, Nap-mediated nitrate reduction and all subsequent denitrification steps via NO and N<sub>2</sub>O to N<sub>2</sub> seem to be more prominent in larger symbionts. While Endoriftia has the genomic potential for complete denitrification to N<sub>2</sub> and was experimentally shown to use nitrate for respiration (Hentschel and Felbeck, 1993), our results thus point to differential use of separate parts of this pathway in small and large symbionts. Since expression of membrane-bound nitrate reductase Nar is inhibited by oxygen (Moreno-Vivián et al., 1999), highest NarGHI abundance in fraction XS suggests that O<sub>2</sub> levels might be particularly low around smaller symbionts. This is presumably due to an oxygen gradient from lobule periphery to center (see also Appendix section G). Following this idea, increasing abundances of NapCHG, subunits of the periplasmic nitrate reductase Nap complex, from fraction XS to L indicate that Nap could take over the role of Nar in larger symbionts. As the inhibitory effect of oxygen on denitrification was proposed to be largely due to inhibition of nitrate transport across the cytoplasmic membrane to Nar (Denis et al., 1990; Moir and Wood, 2001), the periplasmic Nap nitrate reductase is not affected by elevated oxygen concentrations (Moreno-Vivián et al., 1999). Some denitrifying bacteria therefore couple Nap to nitrite reductase Nir, nitric oxide reductase Nor and nitrous oxide reductase Nos under oxic or microaerophilic conditions for aerobic denitrification (Ji et al., 2015). We propose that large Endoriftia adopt this strategy of aerobic denitrification when exposed to higher oxygen levels at the lobule periphery. Presumed higher nitrate concentrations around large Endoriftia may additionally facilitate increased expression of Nap, Nir, Nor, and Nos (compared to smaller symbionts). Moreover, while Nar is an energy-conserving enzyme and generates a proton motive force, which is potentially advantageous during cell division in small symbionts (see Appendix section B), the periplasmic Nap conserves less energy and does not contribute to the proton gradient (Bell et al., 1993). Only larger symbionts may therefore be able to afford Nap.

While the above scenario of anaerobic nitrate reduction in small Endoriftia and aerobic denitrification in large Endoriftia seems quite plausible in the context of our findings, some aspects of this hypothesis are still unclear. Nar, Nap, and Nir may not be exclusively involved in respiration, but could as well operate in nitrate assimilation (Liao et al., 2014; Hinzke et al., 2019). Highest NarGHI and Nap abundances in small and large Endoriftia, respectively, may therefore be due to either of the two functions or (more likely) a combination of both. Also, the fate of nitrite produced by NarGHI during dissimilatory nitrate reduction in small symbionts is dubious. Nitrite is toxic and needs to be removed from the cells either by conversion into ammonium or by export (Moir and Wood, 2001). Although assimilation appears to be the most likely option, Endoriftia's assimilatory nitrite reductase was not detected in fractions XS and S (but only in fractions M and L and at extremely low abundance; large subunit NirB) or not detected at all (small subunit NirD, Sym\_2601635198). Similarly, the nitrite extrusion protein NarK had its lowest abundance in fraction XS. Whether nitrite assimilation in Endoriftia relies on enzymes other than NirBD, or whether nitrite is removed from the small symbionts by other exporters or by diffusion, and possibly even passed on to large Endoriftia for further reduction, therefore needs to be elucidated in future studies.

### Regulation of gene expression may be less stringent in large symbionts

Relative abundance of the RNA polymerase sigma factor RpoD decreased from fraction XS to L (Figure 3, Supplementary file 3h) in S-rich and in S-depleted samples, pointing to relatively more growth-related activities in small Endoriftia (see also Appendix section B). RpoD is the primary sigma

factor for vegetative growth ( $\sigma^{70}$ ), which regulates transcription of most genes involved in exponential growth in many bacteria (Helmann and Chamberlin, 1988; Fujita et al., 1994; Ishihama, 2000). This would be in agreement with the idea of small *Riftia* symbionts being mainly occupied with cell division and proliferation in a quasi-exponential growth phase, while large symbionts function as bio-synthetic ‘factories’, focusing on carbon fixation and biomass production. Interestingly, RpoS, the master 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. The symbiont likely adapts to this ‘lack’ of stress or limitation by less stringent regulation of 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<sub>2</sub> fixation pathways. Calvin cycle and rTCA cycle are in fact in all likelihood expressed in the very same (large) symbiont cells, as suggested by our HCR-FISH experiments (Appendix 1—figure 3). Under these premises, the previously observed simultaneous expression of such 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 (i.e. in the large symbionts), 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. The symbionts divide responsibilities: Whereas small Endoriftia ensure survival of the symbiont population, large Endoriftia are primarily engaged in biomass production.

The factors that trigger Endoriftia’s development from smaller to larger cells remain to be elucidated. For *Rhizobium*, a steep O<sub>2</sub> concentration gradient inside legume nodules was proposed to be involved in signaling for symbiont differential gene expression (Soupène 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, for example, histone-derived antimicrobial peptides, which were recently proposed to play a role in symbiont cell cycle regulation (Hinze 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 bio-synthetic 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 multiple – even redundant – metabolic pathways at the same time. Division of labor between subpopulations, which enables such an ‘advantageous deregulation’ in specialized cells, thus likely enhances symbiont productivity during symbiosis. We suggest that such symbiont differentiation into functionally dissimilar subpopulations with complementary roles in the same host might be an important key to success of microbe-host associations in various environments.

## Materials 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 2500 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 (Hinze et al.,

2019). Sample details and numbers of biological replicates are summarized in **Supplementary file 1**.

Trophosome sulfur content of the specimens was estimated based on the trophosome tissue's color: sulfur-rich (S-rich) specimens have a light green or 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 seven healthy-looking *Riftia* specimens, of which four had S-rich trophosomes and three had S-depleted trophosomes. (Note that sulfur content does not provide any information on the symbiosis' state of health, that is, worms with S-depleted trophosomes are not 'unhealthier' than worms with S-rich trophosomes.) Since trophosome sulfur content is positively correlated with habitat sulfide concentrations (Childress et al., 1991; Robidart et al., 2011; Scott et al., 2012), it can serve as an indicator for the energetic situation of the symbiosis at the time of sampling.

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. 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 MgSO<sub>4</sub>, 0.011 M CaCl<sub>2</sub>, 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. On top of this gradient, 0.5 ml tissue homogenate was layered and the gradient was centrifuged (1000 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 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 labeled by CARD-FISH in five biological replicates (*Riftia* specimens) with S-rich trophosome and three biological replicates with S-depleted trophosomes. 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 ( $\geq 2 \mu\text{m} < 3.912 \mu\text{m}$ ), small ( $\geq 3.912 \mu\text{m} < 5.314 \mu\text{m}$ ), medium ( $\geq 5.314 \mu\text{m} < 6.83275 \mu\text{m}$ ), and large ( $\geq 6.83275 \mu\text{m} - 20 \mu\text{m}$ ). 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, the fifth fraction (counted from the top of the gradient) had the highest percentage of very small cells, that is, most of 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 is referred to as 'fraction XS' in the manuscript. The fractions containing the highest percentages of small, medium, and large symbiont cells are 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**. For each biological replicate, fractions XS, S, M, and L, as well as non-enriched trophosome homogenate were subjected to separate metaproteome analyses (see below).

### Transmission electron microscopy (TEM)

Trophosome samples used for TEM in this study (see **Supplementary file 1** 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 file 1** for details) that was fixed for CARD-FISH and immobilized on GTTP polycarbonate filters as described above was used for hybridization chain reaction FISH (HCR-FISH) according to **Choi et al., 2014**. We used an 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). Assisted by the probe manufacturer, we designed five individual probes per target RNA (for 16S and AclB) and four probes for RubisCO, each of which targeted a 50 nucleotides long sub-sequence of the respective RNA of interest. The sub-sequences were spread out across the length of each target RNA (see **Appendix 1—table 1** 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<sub>3</sub>C<sub>6</sub>H<sub>5</sub>O<sub>7</sub>), 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<sub>2</sub>HPO<sub>4</sub>, 10.2 mM NaH<sub>2</sub>PO<sub>4</sub>, 0.9 M NaCl) at 45°C for 30 min for pre-hybridization, and incubated overnight (16 hr, 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 hr, 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 eight frames showing a total of 33 cells on a filter from one biological replicate, n = 1) 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 diameter (the longest distance between any two points along the boundary of the ROI) and (ii) the area of the ROI. As described by *Nikolakakis et al., 2015*, two negative control preparations were performed to identify artifact signals: Two filters from the same biological replicate as above were incubated (a) without probes but with fluorophore-carrying hairpins ( $n = 1$ ; to test for non-specific amplification of hairpins) and (b) without probes or hairpins ( $n = 1$ ; to test for autofluorescence), and analyzed using an Axio Imager.M2 fluorescence microscope (Carl Zeiss Microscopy GmbH). Since *Endoriftia* is uncultivable as yet, a target-free control sample (i.e. an *Endoriftia* knockout mutant without the respective RNAs of interest) is not available at this point. Nonspecific binding of the probes to non-target RNAs was therefore excluded based on the following considerations: (i) In nucleic acid hybridizations, the hybridization rate decreases to zero when mismatches increase to more than 30% (*Wetmur, 1991*). (ii) There is only one symbiont phylogeny in *Riftia*, preventing false-positive binding to other bacterial RNAs, and (iii) false-positive matches to host RNAs could be excluded, since symbiont cells were co-labeled with the 16S rRNA-specific probe which confirmed the identity of the cells. Finally (iv), to ensure probe sequence specificity, we blast-searched all probe candidates against the *Riftia* symbiont genome (NCBI whole-genome shotgun sequences NZ\_AFOC00000000.1 and NZ\_AFZB00000000.1) at low stringency (BLASTn optimized for somewhat similar sequences, seed length: 7, expectation threshold: 1), and subjected all resulting matches to in silico hybridization behavior analyses using the mathFISH web tool (*Yilmaz et al., 2011*). These calculations confirmed that in *Riftia* symbiont cells, either of the two mRNA probes has only one target sequence to which it binds with a hybridization efficiency of 100%, whereas hybridization efficiency of off-target sequences was 0% (data not shown).

### Flow cytometry

Subsamples of fresh homogenate, of three gradient fractions enriched in small symbionts, and of 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  $\mu\text{mol/l}$  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 530/30 nm (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 and T) together with the CS and T beads (Becton Dickinson Biosciences). During sample measurements, the present populations were shown in a side scatter (SSC)-area versus forward scatter (FSC)-area dot plot. 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 (populations 1 and 2, see *Figure 2*) were sorted from the bivariate dot plot, SSC 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).

### Peptide sample preparation

For four biological replicates with S-rich trophosome and three biological replicates with S-depleted trophosome, non-enriched trophosome homogenate and the four density gradient fractions determined as XS, S, M, and L by CARD-FISH analyses (see above) were individually subjected to meta-proteomic analyses. 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<sub>4</sub>HCO<sub>3</sub> in 30% acetonitrile), and digested with trypsin (sequencing grade, Promega) overnight at 37°C, before peptides were eluted in an ultrasonic bath. Peptides were stored at -80°C until 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, USA), 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. Samples were measured in a randomized design.

### 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, that is, NCBI project PRJNA60889 (endosymbiont of *Riftia 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, 2017*) was added to complete the database. Database search was conducted using Proteome Discoverer v. 2.0.0.802 as described in *Kleiner et al., 2018*. Briefly, raw spectra were searched against the database using the Sequest HT node. False discovery rates (FDRs) for peptide spectrum matches were calculated and filtered using the Percolator Node (FDR < 0.05). FidoCT was used to infer proteins with a protein-level false discovery rate of 5% (q-value < 0.05, at least one unique peptide). The mass spectrometry proteomics data have been deposited to the PRIDE proteomics identification database (<https://www.ebi.ac.uk/pride/>; *Vizcaíno et al., 2016*) with the dataset identifier PXD016986.

To systematically screen the *Riftia* symbiont metagenome for dissimilatory sulfur metabolism-related proteins, candidates identified in different studies were searched against the Endoriftia metaproteome database using bioedit (*Hall, 1999; Supplementary file 7*). 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*).

Relative stable isotope fingerprints (relative SIFs, relative δ<sup>13</sup>C values) of all S-rich (n = 4) and S-depleted (n = 3) gradient fractions were calculated using Calis-p 2.0 (*Kleiner et al., 2018*), with fraction XS of S-rich trophosome as baseline.

### 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, for example, 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, for example, 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 Development Core Team, 2018; **Supplementary file 2a**). The filtering and normalization step was included to avoid biasing the analysis toward 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 fractions (representing enrichments of different cell size classes) using two methods, that is, profile clustering (STEM analysis) and random forests (see below).
- ii. To be able to compare relative symbiont protein abundances within samples and to identify particularly abundant proteins, normalized spectral abundance factor (NSAF) 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 of a given organism ('org') in the same sample (**Supplementary file 2b**). Note that %orgNSAF values in this analysis 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. persephe*), 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, that is, all proteins which were not removed by filtering and showed a consistent trend in all replicates, were used for further analysis. This means that differences in protein abundance patterns were considered significant if proteins were detected with a consistent abundance trend across all replicates (increase, decrease, or alternating increase and decrease of abundance from fraction XS to L).

### Random forests

For random forest analysis, we used the ranger package v. 0.10.1 (Wright and Ziegler, 2015) in R v. 3.5.1 (R Development Core Team, 2018). Random forests are a machine learning technique, which can be used to find the variables – here proteins – that allow to predict which datasets or samples are similar (and which ones are not; Degenhardt *et al.*, 2019). For variable importance calculation, we employed the method from Janitza *et al.*, 2018 as implemented in the ranger package. This method uses a heuristic approach, where a null distribution for p-value calculation is generated based on variables with importance scores of zero or negative importance scores. For pairwise comparisons, the data set was subjected to an additional filtering step, so that only proteins with a minimum of five spectral counts in at least six out of eight (S-rich) or four out of six (S-depleted) samples were included. The comparison of all 16 S-rich samples included only such proteins which had a minimum of five spectral counts in at least five samples, and the comparison of all twelve S-depleted samples included only proteins with five or more spectral counts in a minimum of four samples. The filtered and RLE-normalized data were used for random forest analysis as follows: 2000 forests with 10,000 trees per forest were grown for pairwise comparisons as well as for comparisons including

the samples representing all four size classes. Proteins which had a p-value below 0.05 in >90% of the forests were included in further analyses.

### Significant differences in protein abundance

Proteins that showed significant abundance differences as determined by STEM analysis or by random forest analysis (see above) or by both methods were included in a common list. Please note that this approach of determining significant protein abundance differences was not based on individual p-values. For proteins with significant abundance differences, we clustered the z-scored mean abundances using hierarchical clustering (Pearson correlation, complete linkage) in R to visualize their abundance trends (**Appendix 1—figure 1**). For this purpose, we employed the R base package stats (**R Development 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 lower 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 further analysis.

## 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 Wetzal 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 SM), fellowships of the Institute of Marine Biotechnology Greifswald (TH, MM), a German Academic Exchange Service (DAAD) grant (TH), the NC State Chancellor's Faculty Excellence Program Cluster on Microbiomes and Complex Microbial Communities (MK), the USDA National Institute of Food and Agriculture, Hatch project 1014212 (MK), the U.S. National Science Foundation (grants OCE-1131095 and OCE-1559198 to SMS), and The WHOI Investment in Science Fund (to SMS). We furthermore acknowledge support for article processing charges from the DFG (Grant 393148499) and the Open Access Publication Fund of the University of Greifswald.



## Additional information

### Funding

Funder	Grant reference number	Author
Institut für Marine Biotechnologie e.V.		Tjorven Hinzke Mareike Meister
German Academic Exchange Service		Tjorven Hinzke
NC State University	Chancellor's Faculty Excellence Program Cluster on Microbiomes and Complex Microbial Communities	Manuel Kleiner
National Institute of Food and Agriculture	Hatch project 1014212	Manuel Kleiner
National Science Foundation	OCE-1131095	Stefan M Sievert
Woods Hole Oceanographic Institution	The WHOI Investment in Science Fund	Stefan M Sievert
German Research Foundation	MA6346/2-1	Stephanie Markert
National Science Foundation	OCE-1559198	Stefan M Sievert
German Research Foundation	DFG Open Access Publication Fund: 393148499	Stephanie Markert

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

Tjorven Hinzke, Data curation, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing - original draft, review, editing; Manuel Kleiner, Supervision, Methodology, Writing - review and editing; Mareike Meister, Christian Hentschker, Petra Hildebrandt, Florian Bonn, Investigation; Rabea Schlüter, Resources, Investigation, Visualization; Jan Pané-Farré, Investigation, Visualization; Horst Felbeck, Supervision, Investigation; Stefan M Sievert, Resources, Funding acquisition, Writing - review and editing; Uwe Völker, Dörte Becher, Resources; Thomas Schweder, Resources, Supervision; Stephanie Markert, Conceptualization, Supervision, Funding acquisition, Investigation, Visualization, Methodology, Project administration, Writing - review and editing

### Author ORCIDs

Tjorven Hinzke  <https://orcid.org/0000-0003-1117-0235>  
 Manuel Kleiner  <http://orcid.org/0000-0001-6904-0287>  
 Dörte Becher  <http://orcid.org/0000-0002-9630-5735>  
 Thomas Schweder  <http://orcid.org/0000-0002-7213-3596>  
 Stephanie Markert  <https://orcid.org/0000-0003-0923-3305>

### Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.58371.sa1>

Author response <https://doi.org/10.7554/eLife.58371.sa2>

## Additional files

### Supplementary files

- Supplementary file 1. Sampling details for specimens and sample types used in this study. All animals were collected at the Crab Spa vent site in the East Pacific Rise (EPR) Tica area. For proteomic analyses, *Riftia* trophosome homogenate (Hom) was subjected to Histodenz-based density gradient centrifugation, separating symbiont cells according to their sizes. After centrifugation, the gradient

was carefully disassembled into 24 subsamples/fractions (numbered 1 to 24), all of which were analyzed by CARD-FISH to identify those fractions in which the percentage of very small, small, medium-sized and large symbiont cells was highest. These fractions were designated XS, S, M, and L for the respective worm and included in comparative proteomic analyses.

- Supplementary file 2. Symbiont proteins identified in this study, which were included in statistical analyses. Relative abundance of symbiont proteins in fractions XS, S, M, and L in sulfur-rich (S-rich) and sulfur-depleted (S-depl) trophosomes is displayed as edgeR-RLE-corrected spectral count values, which represent average values of four biological replicates (S-rich) and three biological replicates (S-depl). Abundance trends, that is increase or decrease of relative protein abundance across the four gradient fractions is indicated by spark lines (columns „Trend“) and by color shades from light green/light gray (lowest protein abundance across all four fractions) to dark green/dark gray (highest abundance). Significant changes in S-rich or S-depl specimens (or both) are indicated by \* (these trends are consistent in all replicates according to STEM trend analysis and pairwise comparison between fractions by random forests; for a detailed definition of significance as applied in this study see Materials and methods). Protein accession numbers refer to NCBI/Uniprot entries (EGV- and EGW- accessions) and JGI entries (all other accessions). The prefix ‘Sym\_’ indicates that this accession number refers to a symbiont protein in our combined host-and-symbiont database, while host proteins have the prefix ‘Host\_’ (note that the prefixes were omitted in **Figure 3** and **Figure 4** in the main text for readability’s sake). Please note that this table includes only such identified symbiont proteins, which were detected with at least five spectral counts in a minimum of four (of 16) individual replicate S-rich samples or a minimum of three (of 12) individual S-depl samples in fractions XS - L (see Materials and methods for details). Note also, that edgeR-RLE-corrected spectral count values as displayed here can be used to compare a given protein’s abundance between the individual fractions, but do not allow for comparisons between proteins of the same sample. A complete list of all symbiont protein identifications, including low abundant proteins and proteins detected in unenriched homogenate samples, which allows for abundance comparison between proteins is presented in b. (b) Unfiltered list of all symbiont proteins identified in this study in density gradient fractions XS, S, M, and L and in unenriched trophosome tissue homogenate (Hom) from sulfur-rich (S-rich) and sulfur-depleted (S-depl) *Riftia* specimens. Relative protein abundance is displayed as %orgNSAF (normalized spectral counts, see Materials and methods), which give a protein’s abundance as percentage of all symbiont proteins in the same sample, allowing for comparison between individual proteins within a given sample. %orgNSAF values are average values of four biological replicates (S-rich) and three biological replicates (S-depl). A protein’s ‘abundance rank’ indicates overall abundance across all samples (rank one being the most abundant protein). The 100 most abundant proteins according to this ranking are highlighted in yellow in column B. Protein accession numbers refer to NCBI/Uniprot entries (EGV- and EGW- accessions) and JGI entries (all other accessions). The prefix ‘Sym\_’ indicates that this accession number refers to a symbiont protein in our combined host-and-symbiont database, while host proteins have the prefix ‘Host\_’ (note that the prefixes were omitted in **Figure 3** and **Figure 4** in the main text for readability’s sake). Proteins in gray font are low-abundant proteins, which were not included in statistical analyses. %orgNSAF values of these low-abundant proteins are less reliable and should be interpreted with care. Proteins that were exclusively identified in homogenate samples were also excluded from statistical analyses and are therefore also set in gray font. Please note that %orgNSAF values cannot be compared across sample types, due to the unequal total identification numbers in the individual sample types. For cross-sample comparisons, please refer to the edgeR-corrected values in a, which contains all proteins in black font.

- Supplementary file 3. Abundance trends of Endoriftia proteins in various metabolic categories across 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 gray (lowest protein abundance across all four fractions) to dark green/dark gray (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, that is, 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 was too low abundant to be included in statistical analyses. For

an overview of all identified symbiont proteins and their relative abundances see **Supplementary file 2**. Protein accession numbers refer to NCBI/Uniprot entries (EGV- and EGW-accessions) and JGI entries (all other accessions). The prefix 'Sym\_' indicates that this accession number refers to a symbiont protein in our combined host-and-symbiont database (note that the prefix was omitted in **Figure 3** and **Figure 4** in the main text for readability's sake).

- **Supplementary file 4.** *Riftia* host proteins with significantly higher abundance (FDR 0.05) in fractions XS and S (containing predominantly very small and small symbiont cells, respectively), compared to the non-enriched trophosome homogenate (Hom). Spectral count data from sulfur-rich (S-rich) and S-depleted specimens were normalized separately in edgeR using RLE normalization (RLE-SC). For details regarding statistical analysis see Materials and methods. In the significance columns, '1' refers to significantly higher abundance in fractions XS and S, '0' to non-significant differences. Blast-KOALA functional categories: FE: family eukaryote, GE: genus eukaryote. TH: transmembrane helices. SP: signal peptide. WoLF PSORT subcellular location prediction: cyto: cytosol, cysk: cytoskeleton, E.R.: endoplasmic reticulum, extr: extracellular, golg: golgi apparatus, lyso: lysosome, mito: mitochondrial, nucl: nucleus, pero: peroxisome, plas: plasma membrane. TargetP secretory pathway prediction: M: mitochondrion, S: secretory pathway, '\_': other location. Phobius/SignalP signal peptide prediction: Y: signal peptide predicted, N: no signal peptide predicted. Accession numbers refer to the combined host-and-symbiont database used for protein identification in this study (see Materials and methods, **Hinzke et al., 2019**). The prefix 'Host\_' indicates that this accession number refers to a host protein (note that the prefix was omitted in **Figure 4** in the main text for readability's sake).

- **Supplementary file 5.** Total (summed up) relative abundance of Endoriftia proteins involved in specific metabolic categories in fractions XS, S, M, and L in sulfur-rich (S-rich) *Riftia* specimens (average values, n = 4) and sulfur-depleted (S-depl) *Riftia* specimens (average values, n = 3). Only those 1212 symbiont proteins presented in **Supplementary file 2a**, which are included in the edgeR statistical evaluation, are included (proteins with low abundance and/or only one or two replicate values were excluded). To allow comparison and summing of protein abundances across proteins within one sample, edgeR-RLE-corrected spectral count values were normalized (a) to protein size and (b) to the sum of all proteins before summing up the proteins within categories (100% = all proteins in **Supplementary file 2a**). These results indicate that morphological differences between individual symbiont differentiation stages are accompanied by a gradual change in metabolic function. During differentiation from small to large cells, *Riftia* symbionts rearrange their metabolic priorities, allocating resources to those processes that are most important in their respective life phase and role in the symbiosis.

- **Supplementary file 6.** *Riftia* trophosome homogenate and gradient fractions enriched in small and large symbionts, respectively, were stained with Syto9 and subjected to flow cytometry analysis in a FACSAria high-speed cell sorter with 488 nm excitation (see Materials and methods for details). Two cell populations were identified, Pop1 and Pop2, which correspond to smaller and larger symbiont cells, respectively (see main text **Figure 2** and **Figure 2—figure supplement 1**). Median fluorescence intensity (FI) per particle, a measure of DNA content per cell, was compared between the two populations 1 and 2 to quantify differences in genome copy number between smaller and larger symbionts (column 'ratio'). Note that FI ratios were not calculated for samples consisting of sorted populations (bottom rows), because these samples contained high cell numbers of either of the two populations, but very low cell numbers of the respective other population, preventing meaningful comparison. Analyses were performed with samples from two *Riftia* specimens (two biological replicates, BR).

- **Supplementary file 7.** Proteins identified as likely involved in dissimilatory sulfur metabolism in *Ca. E. persephone* after Blast-comparison against proteins identified in the literature: **Weissgerber et al., 2014**; **Rodriguez et al., 2011**; **Gregersen et al., 2011**. Significant - protein abundance significantly different between fractions containing symbionts of different size (see Materials and methods for details on statistical analysis). Y - yes, N - no, M - maybe.

- Transparent reporting form

### Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (ProteomeXchange - ProteomeCentral) via the PRIDE partner repository (Vizcaino et al., 2016) with the dataset identifier PXD016986.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Hinze T, Kleiner M, Meister M, Schlüter R, Hentschke C, Pané-Farré J, Hildebrandt P, Felbeck H, Sievert SM, Bonn F, Völker U, Becher D, Schweder T, Markert S	2020	Metabolic differences between morphologically distinct symbiont populations in the tubeworm <i>Riftia pachyptila</i>	<a href="https://www.ebi.ac.uk/pride/archive/projects/PXD016986">https://www.ebi.ac.uk/pride/archive/projects/PXD016986</a>	PRIDE, PXD016986

### References

- Ali BM, Amit R, Braslavsky I, Oppenheim AB, Gileadi O, Stavans J. 2001. Compaction of single DNA molecules induced by binding of integration host factor (IHF). *PNAS* **98**:10658–10663. DOI: <https://doi.org/10.1073/pnas.181029198>, PMID: 11535804
- Angert ER. 2012. DNA replication and genomic architecture of very large bacteria. *Annual Review of Microbiology* **66**:197–212. DOI: <https://doi.org/10.1146/annurev-micro-090110-102827>, PMID: 22994492
- Ankrah NYD, Wilkes RA, Zhang FQ, Zhu D, Kaweesi T, Aristilde L, Douglas AE. 2020. Syntrophic splitting of central carbon metabolism in host cells bearing functionally different symbiotic bacteria. *The ISME Journal* **14**:1982–1993. DOI: <https://doi.org/10.1038/s41396-020-0661-z>, PMID: 32350409
- Barabote RD, Johnson OL, Zetina E, San Francisco SK, Fralick JA, San Francisco MJ. 2003. *Erwinia chrysanthemi* *tolC* is involved in resistance to antimicrobial plant chemicals and is essential for phytopathogenesis. *Journal of Bacteriology* **185**:5772–5778. DOI: <https://doi.org/10.1128/JB.185.19.5772-5778.2003>, PMID: 13129948
- Beam CE, Saveson CJ, Lovett ST. 2002. Role for *radA/sms* in recombination intermediate processing in *Escherichia coli*. *Journal of Bacteriology* **184**:6836–6844. DOI: <https://doi.org/10.1128/JB.184.24.6836-6844.2002>, PMID: 12446634
- Beckmann JF, Markowski TW, Witthuhn BA, Fallon AM. 2013. Detection of the *Wolbachia*-encoded DNA binding protein, HU beta, in mosquito gonads. *Insect Biochemistry and Molecular Biology* **43**:272–279. DOI: <https://doi.org/10.1016/j.ibmb.2012.12.007>, PMID: 23287400
- Beinart RA, Gartman A, Sanders JG, Luther GW, Girguis PR. 2015. The uptake and excretion of partially oxidized sulfur expands the repertoire of energy resources metabolized by hydrothermal vent symbioses. *Proceedings of the Royal Society B: Biological Sciences* **282**:20142811. DOI: <https://doi.org/10.1098/rspb.2014.2811>
- Bell LC, Page MD, Berks BC, Richardson DJ, Ferguson SJ. 1993. Insertion of transposon Tn5 into a structural gene of the membrane-bound nitrate reductase of *Thiosphaera pantotropha* results in anaerobic overexpression of periplasmic nitrate reductase activity. *Microbiology* **139**:3205–3214. DOI: <https://doi.org/10.1099/00221287-139-12-3205>
- Bertero MG, Rothery RA, Palak M, Hou C, Lim D, Blasco F, Weiner JH, Strynadka NCJ. 2003. Insights into the respiratory electron transfer pathway from the structure of nitrate reductase A. *Nature Structural & Molecular Biology* **10**:681–687. DOI: <https://doi.org/10.1038/nsb969>
- Boetius A, Felbeck H. 1995. Digestive enzymes in marine invertebrates from hydrothermal vents and other reducing environments. *Marine Biology* **122**:105–113. DOI: <https://doi.org/10.1007/BF00349283>
- Bouvier T, Troussellier M, Anzil A, Courties C, Servais P. 2001. Using light scatter signal to estimate bacterial biovolume by flow cytometry. *Cytometry* **44**:188–194. DOI: [https://doi.org/10.1002/1097-0320\(20010701\)44:3<188::AID-CYTO1111>3.0.CO;2-C](https://doi.org/10.1002/1097-0320(20010701)44:3<188::AID-CYTO1111>3.0.CO;2-C), PMID: 11429769
- Bresler V, Fishelson L. 2003. Polyploidy and polyteny in the gigantic eubacterium *Epulopiscium fishelsoni*. *Marine Biology* **143**:17–21. DOI: <https://doi.org/10.1007/s00227-003-1055-2>
- Bright M, Keckeis H, Fisher CR. 2000. An autoradiographic examination of carbon fixation, transfer and utilization in the *Riftia pachyptila* symbiosis. *Marine Biology* **136**:621–632. DOI: <https://doi.org/10.1007/s002270050722>
- Bright M, Sorgo A. 2003. Ultrastructural reinvestigation of the trophosome in adults of *Riftia pachyptila* (Annelida, Siboglinidae). *Invertebrate Biology* **122**:347–368. DOI: <https://doi.org/10.1111/j.1744-7410.2003.tb00099.x>
- Camberg JL, Hoskins JR, Wickner S. 2009. ClpXP protease degrades the cytoskeletal protein, FtsZ, and modulates FtsZ polymer dynamics. *PNAS* **106**:10614–10619. DOI: <https://doi.org/10.1073/pnas.0904886106>, PMID: 19541655

- Caro A**, Gros O, Got P, De Wit R, Troussellier M. 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: <https://doi.org/10.1128/AEM.01683-06>, PMID: 17259363
- Cavanaugh CM**, Gardiner SL, Jones ML, Jannasch HW, Waterbury JB. 1981. Prokaryotic cells in the hydrothermal vent tube worm *Riftia pachyptila* Jones: Possible Chemoautotrophic Symbionts. *Science* **213**:340–342. DOI: <https://doi.org/10.1126/science.213.4505.340>, PMID: 17819907
- Chao Y**, Vogel J. 2010. The role of hfq in bacterial pathogens. *Current Opinion in Microbiology* **13**:24–33. DOI: <https://doi.org/10.1016/j.mib.2010.01.001>, PMID: 20080057
- Chastanet A**, Derre I, Nair S, Msadek T. 2004. *clpB*, a novel member of the *Listeria monocytogenes* CtsR regulon, is involved in virulence but not in general stress tolerance. *Journal of Bacteriology* **186**:1165–1174. DOI: <https://doi.org/10.1128/JB.186.4.1165-1174.2004>, PMID: 14762012
- Chen J**, Brevet A, Fromant M, Lévêque F, Schmitter JM, Blanquet S, Plateau P. 1990. Pyrophosphatase is essential for growth of *Escherichia coli*. *Journal of Bacteriology* **172**:5686–5689. DOI: <https://doi.org/10.1128/JB.172.10.5686-5689.1990>, PMID: 2170325
- Chien AC**, Hill NS, Levin PA. 2012. Cell size control in bacteria. *Current Biology* **22**:R340–R349. DOI: <https://doi.org/10.1016/j.cub.2012.02.032>, PMID: 22575476
- Childress JJ**, Arp AJ, Fisher CR. 1984. Metabolic and blood characteristics of the hydrothermal vent tube-worm *Riftia pachyptila*. *Marine Biology* **83**:109–124. DOI: <https://doi.org/10.1007/BF00394718>
- Childress JJ**, Fisher CR, Favuzzi JA, Kochevar RE, Sanders NK, Alayse AM. 1991. Sulfide-driven autotrophic balance in the bacterial symbiont-containing hydrothermal vent tubeworm, *Riftia pachyptila* Jones. *The Biological Bulletin* **180**:135–153. DOI: <https://doi.org/10.2307/1542437>, PMID: 29303639
- Cho H**, McManus HR, Dove SL, Bernhardt TG. 2011. Nucleoid occlusion factor SlmA is a DNA-activated FtsZ polymerization antagonist. *PNAS* **108**:3773–3778. DOI: <https://doi.org/10.1073/pnas.1018674108>, PMID: 21321206
- Choi HM**, Beck VA, Pierce NA. 2014. Next-generation in situ hybridization chain reaction: higher gain, lower cost, greater durability. *ACS Nano* **8**:4284–4294. DOI: <https://doi.org/10.1021/nn405717p>, PMID: 24712299
- Cortés F**, Pastor N, Mateos S, Domínguez I. 2003. Roles of DNA topoisomerases in chromosome segregation and mitosis. *Mutation Research* **543**:59–66. DOI: [https://doi.org/10.1016/S1383-5742\(02\)00070-4](https://doi.org/10.1016/S1383-5742(02)00070-4), PMID: 12510017
- Cortés F**, Pastor N. 2003. Induction of endoreduplication by topoisomerase II catalytic inhibitors. *Mutagenesis* **18**:105–112. DOI: <https://doi.org/10.1093/mutage/18.2.105>, PMID: 12621064
- Cosme AM**, Becker A, Santos MR, Sharypova LA, Santos PM, Moreira LM. 2008. The outer membrane protein TolC from *Sinorhizobium meliloti* affects protein secretion, polysaccharide biosynthesis, antimicrobial resistance, and symbiosis. *Molecular Plant-Microbe Interactions* **21**:947–957. DOI: <https://doi.org/10.1094/MPMI-21-7-0947>, PMID: 18533835
- D'Haese W**, Holsters M. 2004. Surface polysaccharides enable bacteria to evade plant immunity. *Trends in Microbiology* **12**:555–561. DOI: <https://doi.org/10.1016/j.tim.2004.10.009>, PMID: 15539115
- Dahl C**, Franz B, Hensen D, Kesselheim A, Zigann R. 2013. Sulfite oxidation in the purple sulfur bacterium *Allochrochromatium vinosum*: identification of SoeABC as a major player and relevance of SoxYZ in the process. *Microbiology* **159**:2626–2638. DOI: <https://doi.org/10.1099/mic.0.071019-0>, PMID: 24030319
- Dang W**, Hu YH, Sun L. 2011. HtpG is involved in the pathogenesis of *Edwardsiella tarda*. *Veterinary Microbiology* **152**:394–400. DOI: <https://doi.org/10.1016/j.vetmic.2011.05.030>, PMID: 21664076
- Davidson AL**, Dassa E, Orelle C, Chen J. 2008. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiology and Molecular Biology Reviews* **72**:317–364. DOI: <https://doi.org/10.1128/MMBR.00031-07>, PMID: 18535149
- De Cian MC**, Andersen AC, Bailly X, Lallier FH. 2003. Expression and localization of carbonic anhydrase and ATPases in the symbiotic tubeworm *Riftia pachyptila*. *Journal of Experimental Biology* **206**:399–409. DOI: <https://doi.org/10.1242/jeb.00074>, PMID: 12477910
- Deaconescu AM**, Savery N, Darst SA. 2007. The bacterial transcription repair coupling factor. *Current Opinion in Structural Biology* **17**:96–102. DOI: <https://doi.org/10.1016/j.sbi.2007.01.005>, PMID: 17239578
- Degenhardt F**, Seifert S, Szymczak S. 2019. Evaluation of variable selection methods for random forests and omics data sets. *Briefings in Bioinformatics* **20**:492–503. DOI: <https://doi.org/10.1093/bib/bbx124>, PMID: 29045534
- Denis KS**, Dias FM, Rowe JJ. 1990. Oxygen regulation of nitrate transport by diversion of electron flow in *Escherichia coli*. *J Biol Chem* **265**:18095–18097.
- Domínguez-Cuevas P**, Porcelli I, Daniel RA, Errington J. 2013. Differentiated roles for MreB-actin isologues and autolytic enzymes in *Bacillus subtilis* morphogenesis. *Molecular Microbiology* **89**:1084–1098. DOI: <https://doi.org/10.1111/mmi.12335>, PMID: 23869552
- Dorman CJ**, Deighan P. 2003. Regulation of gene expression by histone-like proteins in bacteria. *Current Opinion in Genetics & Development* **13**:179–184. DOI: [https://doi.org/10.1016/S0959-437X\(03\)00025-X](https://doi.org/10.1016/S0959-437X(03)00025-X), PMID: 12672495
- Drees JC**, Chitteni-Pattu S, McCaslin DR, Inman RB, Cox MM. 2006. Inhibition of RecA protein function by the RdgC protein from *Escherichia coli*. *Journal of Biological Chemistry* **281**:4708–4717. DOI: <https://doi.org/10.1074/jbc.M513592200>
- Dri AM**, Rouviere-Yaniv J, Moreau PL. 1991. Inhibition of cell division in hupA hupB mutant bacteria lacking HU protein. *Journal of Bacteriology* **173**:2852–2863. DOI: <https://doi.org/10.1128/JB.173.9.2852-2863.1991>, PMID: 2019558

- Duperthuy M**, Schmitt P, Garzón E, Caro A, Rosa RD, Le Roux F, Lautredu-Audouy N, Got P, Romestand B, de Lorgeril J, Kieffer-Jaquinod S, Bachère E, Destoumieux-Garzón D. 2011. Use of OmpU porins for attachment and invasion of *Crassostrea gigas* immune cells by the oyster pathogen *Vibrio splendidus*. *PNAS* **108**:2993–2998. DOI: <https://doi.org/10.1073/pnas.1015326108>, PMID: 21282662
- Durand-Heredia J**, Rivkin E, Fan G, Morales J, Janakiraman A. 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: <https://doi.org/10.1128/JB.00176-12>, PMID: 22505682
- Ernst J**, Bar-Joseph Z. 2006. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* **7**:191. DOI: <https://doi.org/10.1186/1471-2105-7-191>, PMID: 16597342
- Eshwar AK**, Tasara T, Stephan R, Lehner A. 2015. Influence of FkpA variants on survival and replication of *Cronobacter* spp in human macrophages. *Research in Microbiology* **166**:186–195. DOI: <https://doi.org/10.1016/j.resmic.2015.02.005>, PMID: 25724920
- Felbeck H**. 1981. Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila* Jones (Vestimentifera). *Science* **213**:336–338. DOI: <https://doi.org/10.1126/science.213.4505.336>, PMID: 17819905
- Felbeck H**, Childress JJ, Somero GN. 1981. Calvin-Benson cycle and sulphide oxidation enzymes in animals from sulphide-rich habitats. *Nature* **293**:291–293. DOI: <https://doi.org/10.1038/293291a0>
- Felbeck H**, Jarchow J. 1998. Carbon release from purified chemoautotrophic bacterial symbionts of the hydrothermal vent tubeworm *Riftia pachyptila*. *Physiological Zoology* **71**:294–302. DOI: <https://doi.org/10.1086/515931>, PMID: 9634176
- Felbeck H**, Turner PJ. 1995. CO<sub>2</sub> transport in catheterized hydrothermal vent tubeworms, *Riftia pachyptila* (vestimentifera). *Journal of Experimental Zoology* **272**:95–102. DOI: <https://doi.org/10.1002/jez.1402720203>
- Fernández L**, Hancock RE. 2012. Adaptive and mutational resistance: role of porins and efflux pumps in drug resistance. *Clinical Microbiology Reviews* **25**:661–681. DOI: <https://doi.org/10.1128/CMR.00043-12>, PMID: 23034325
- Frees D**, Qazi SN, Hill PJ, Ingmer H. 2003. Alternative roles of ClpX and ClpP in *Staphylococcus aureus* stress tolerance and virulence. *Molecular Microbiology* **48**:1565–1578. DOI: <https://doi.org/10.1046/j.1365-2958.2003.03524.x>, PMID: 12791139
- Fujita M**, Tanaka K, Takahashi H, Amemura A. 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: <https://doi.org/10.1111/j.1365-2958.1994.tb00498.x>, PMID: 7531806
- Gardebrecht A**, Markert S, Sievert SM, Felbeck H, Thürmer A, Albrecht D, Wollherr A, Kabisch J, Le Bris N, Lehmann R, Daniel R, Liesegang H, Hecker M, Schweder T. 2012. Physiological homogeneity among the endosymbionts of *Riftia pachyptila* and *Tevnia jerichonana* revealed by proteogenomics. *The ISME Journal* **6**:766–776. DOI: <https://doi.org/10.1038/ismej.2011.137>, PMID: 22011719
- Geier B**, Sogin EM, Michellod D, Janda M, Kompauer M, Spengler B, Dubilier N, Liebeke M. 2020. Spatial metabolomics of in situ host-microbe interactions at the micrometre scale. *Nature Microbiology* **5**:498–510. DOI: <https://doi.org/10.1038/s41564-019-0664-6>, PMID: 32015496
- Girguis PR**, Childress JJ. 2006. Metabolite uptake, stoichiometry and chemoautotrophic function of the hydrothermal vent tubeworm *Riftia pachyptila*: responses to environmental variations in substrate concentrations and temperature. *Journal of Experimental Biology* **209**:3516–3528. DOI: <https://doi.org/10.1242/jeb.02404>, PMID: 16943492
- Goehring NW**, Beckwith J. 2005. Diverse paths to midcell: assembly of the bacterial cell division machinery. *Current Biology* **15**:R514–R526. DOI: <https://doi.org/10.1016/j.cub.2005.06.038>, PMID: 16005287
- Goffredi S**, Childress J, Desaulniers N, Lee R, Lallier F, Hammond D. 1997a. Inorganic carbon acquisition by the hydrothermal vent tubeworm *Riftia pachyptila* depends upon high external PCO<sub>2</sub> and upon proton-equivalent ion transport by the worm. *The Journal of Experimental Biology* **200**:883–896. PMID: 9318669
- Goffredi SK**, Childress JJ, Desaulniers NT, Lallier FJ. 1997b. Sulfide acquisition by the vent worm *Riftia pachyptila* appears to be via uptake of HS<sup>-</sup>, rather than H<sub>2</sub>S. *The Journal of Experimental Biology* **200**:2609–2616. PMID: 9359367
- Goffredi SK**, Girguis PR, Childress JJ, Desaulniers NT. 1999. Physiological functioning of carbonic anhydrase in the hydrothermal vent tubeworm *Riftia pachyptila*. *The Biological Bulletin* **196**:257–264. DOI: <https://doi.org/10.2307/1542950>, PMID: 28296490
- Graham JM**. 2001. *Biological Centrifugation*. Oxford, United Kingdom: BIOS Scientific Publishers Ltd.
- Gregersen LH**, Bryant DA, Frigaard NU. 2011. Mechanisms and evolution of oxidative sulfur metabolism in green sulfur bacteria. *Frontiers in Microbiology* **2**:116. DOI: <https://doi.org/10.3389/fmicb.2011.00116>, PMID: 21833341
- Grein F**, Pereira IA, Dahl C. 2010. Biochemical characterization of individual components of the *Allochrocatium vinosum* DsrMKJOP transmembrane complex aids understanding of complex function in vivo. *Journal of Bacteriology* **192**:6369–6377. DOI: <https://doi.org/10.1128/JB.00849-10>, PMID: 20952577
- Grimm F**, Franz B, Dahl C. 2011. Regulation of dissimilatory sulfur oxidation in the purple sulfur bacterium *Allochrocatium vinosum*. *Frontiers in Microbiology* **2**:51. DOI: <https://doi.org/10.3389/fmicb.2011.00051>, PMID: 21927612
- Gu Z**, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**:2847–2849. DOI: <https://doi.org/10.1093/bioinformatics/btw313>, PMID: 27207943
- Guha S**, Udupa S, Ahmed W, Nagaraja V. 2018. Rewired downregulation of DNA gyrase impacts cell division, expression of topology modulators, and transcription in *Mycobacterium smegmatis*. *Journal of Molecular Biology* **430**:4986–5001. DOI: <https://doi.org/10.1016/j.jmb.2018.10.001>, PMID: 30316784

- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series* **41**:95–98. DOI: [https://doi.org/10.14601/Phytopathol\\_Mediterr-14998u1.29](https://doi.org/10.14601/Phytopathol_Mediterr-14998u1.29)
- Hand SC. 1987. Trophosome ultrastructure and the characterization of isolated bacteriocytes from invertebrate-sulfur bacteria symbioses. *The Biological Bulletin* **173**:260–276. DOI: <https://doi.org/10.2307/1541878>, PMID: 29314988
- Harada M, Yoshida T, Kuwahara H, Shimamura S, Takaki Y, Kato C, Miwa T, Miyake H, Maruyama T. 2009. Expression of genes for sulfur oxidation in the intracellular chemoautotrophic symbiont of the deep-sea bivalve *Calyptogena okutanii*. *Extremophiles* **13**:895–903. DOI: <https://doi.org/10.1007/s00792-009-0277-8>, PMID: 19730970
- Hay NA, Tipper DJ, Gygi D, Hughes C. 1999. A novel membrane protein influencing cell shape and multicellular swarming of *Proteus mirabilis*. *Journal of Bacteriology* **181**:2008–2016. DOI: <https://doi.org/10.1128/JB.181.7.2008-2016.1999>, PMID: 10094676
- Heath MC. 2000. Nonhost resistance and nonspecific plant defenses. *Current Opinion in Plant Biology* **3**:315–319. DOI: [https://doi.org/10.1016/S1369-5266\(00\)00087-X](https://doi.org/10.1016/S1369-5266(00)00087-X), PMID: 10873843
- Helmann JD, Chamberlin MJ. 1988. Structure and function of bacterial sigma factors. *Annual Review of Biochemistry* **57**:839–872. DOI: <https://doi.org/10.1146/annurev.bi.57.070188.004203>, PMID: 3052291
- Hengge-Aronis R. 1993. Survival of hunger and stress: the role of *rpoS* in early stationary phase gene regulation in *E. coli*. *Cell* **72**:165–168. DOI: [https://doi.org/10.1016/0092-8674\(93\)90655-A](https://doi.org/10.1016/0092-8674(93)90655-A), PMID: 8425216
- Hensen D, Sperling D, Trüper HG, Brune DC, Dahl C. 2006. Thiosulphate oxidation in the phototrophic sulphur bacterium *Allochrochromatium vinosum*. *Molecular Microbiology* **62**:794–810. DOI: <https://doi.org/10.1111/j.1365-2958.2006.05408.x>, PMID: 16995898
- Hentschel U, Felbeck H. 1993. Nitrate respiration in the hydrothermal vent tubeworm *Riftia pachyptila*. *Nature* **366**:338–340. DOI: <https://doi.org/10.1038/366338a0>
- Hinzke T, Kleiner M, Markert S. 2018. Centrifugation-Based enrichment of bacterial cell populations for metaproteomic studies on bacteria-invertebrate symbioses. *Methods in Molecular Biology* **1841**:319–334. DOI: [https://doi.org/10.1007/978-1-4939-8695-8\\_22](https://doi.org/10.1007/978-1-4939-8695-8_22), PMID: 30259496
- Hinzke T, Kleiner M, Breusing C, Felbeck H, Häsler R, Sievert SM, Schlüter R, Rosenstiel P, Reusch TBH, Schweder T, Markert S. 2019. Host-Microbe interactions in the chemosynthetic *Riftia pachyptila* symbiosis. *mBio* **10**:e02243-19. DOI: <https://doi.org/10.1128/mBio.02243-19>, PMID: 31848270
- Hinzke T, Markert S. 2017. Efficient protein extraction for proteomics and metaproteomics (also suitable for low biomass samples). *protocols.io* **10**:kg6ctze. DOI: <https://doi.org/10.17504/protocols.io.kg6ctze>
- Horne SM, Kottom TJ, Nolan LK, Young KD. 1997. Decreased intracellular survival of an *fkpA* mutant of *Salmonella typhimurium* copenhagen. *Infection and Immunity* **65**:806–810. DOI: <https://doi.org/10.1128/IAI.65.2.806-810.1997>, PMID: 9009347
- Ishihama A. 2000. Functional modulation of *Escherichia coli* RNA polymerase. *Annual Review of Microbiology* **54**:499–518. DOI: <https://doi.org/10.1146/annurev.micro.54.1.499>, PMID: 11018136
- Janitzka S, Celik E, Boulesteix A-L. 2018. A computationally fast variable importance test for random forests for high-dimensional data. *Advances in Data Analysis and Classification* **12**:885–915. DOI: <https://doi.org/10.1007/s11634-016-0276-4>
- Ji B, Yang K, Zhu L, Jiang Y, Wang H, Zhou J, Zhang H. 2015. Aerobic denitrification: a review of important advances of the last 30 years. *Biotechnology and Bioengineering* **20**:643–651. DOI: <https://doi.org/10.1007/s12257-015-0009-0>
- Jones ML. 1981. *Riftia pachyptila* Jones: Observations on the vestimentiferan worm from the Galapagos Rift. *Science* **213**:333–336. DOI: <https://doi.org/10.1126/science.213.4505.333>, PMID: 17819904
- Kang D, Liu G, Lundström A, Gelius E, Steiner H. 1998. A peptidoglycan recognition protein in innate immunity conserved from insects to humans. *PNAS* **95**:10078–10082. DOI: <https://doi.org/10.1073/pnas.95.17.10078>
- King AM, Pretre G, Bartpho T, Sermswan RW, Toma C, Suzuki T, Eshghi A, Picardeau M, Adler B, Murray GL. 2014. High-temperature protein G is an essential virulence factor of *Leptospira interrogans*. *Infection and Immunity* **82**:1123–1131. DOI: <https://doi.org/10.1128/IAI.01546-13>, PMID: 24366253
- Kleiner M, Dong X, Hinzke T, Wippler J, Thorson E, Mayer B, Strous M. 2018. Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. *PNAS* **115**:E5576–E5584. DOI: <https://doi.org/10.1073/pnas.1722325115>, PMID: 29844191
- Klemme J-H. 1976. Regulation of intracellular pyrophosphatase-activity and conservation of the phosphoanhydride-energy of inorganic pyrophosphate in microbial metabolism. *Zeitschrift für Naturforschung C* **31**:544–550. DOI: <https://doi.org/10.1515/znc-1976-9-1011>
- Klint J, Rasmussen U, Bergman B. 2007. FtsZ may have dual roles in the filamentous cyanobacterium *Nostoc/Anabaena* sp. strain PCC 7120. *Journal of Plant Physiology* **164**:11–18. DOI: <https://doi.org/10.1016/j.jplph.2005.08.021>, PMID: 16603273
- Klose J, Aistleitner K, Horn M, Krenn L, Dirsch V, Zehl M, Bright M. 2016. Trophosome of the deep-sea tubeworm *Riftia pachyptila* inhibits bacterial growth. *PLOS ONE* **11**:e0146446. DOI: <https://doi.org/10.1371/journal.pone.0146446>, PMID: 26730960
- Köhler S, Teyssier J, Cloeckeaert A, Rouot B, Liautard JP. 1996. Participation of the molecular chaperone DnaK in intracellular growth of *Brucella suis* within U937-derived phagocytes. *Molecular Microbiology* **20**:701–712. DOI: <https://doi.org/10.1111/j.1365-2958.1996.tb02510.x>, PMID: 8793869

- Komaki K**, Ishikawa H. 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: [https://doi.org/10.1016/S0965-1748\(99\)00125-3](https://doi.org/10.1016/S0965-1748(99)00125-3), PMID: 10732993
- Kondorosi E**, Kondorosi A. 2004. Endoreduplication and activation of the anaphase-promoting complex during symbiotic cell development. *FEBS Letters* **567**:152–157. DOI: <https://doi.org/10.1016/j.febslet.2004.04.075>, PMID: 15165909
- Koronakis V**, Eswaran J, Hughes C. 2004. Structure and function of TolC: the bacterial exit duct for proteins and drugs. *Annual Review of Biochemistry* **73**:467–489. DOI: <https://doi.org/10.1146/annurev.biochem.73.011303.074104>, PMID: 15189150
- Kowalczykowski SC**. 2000. Initiation of genetic recombination and recombination-dependent replication. *Trends in Biochemical Sciences* **25**:156–165. DOI: [https://doi.org/10.1016/S0968-0004\(00\)01569-3](https://doi.org/10.1016/S0968-0004(00)01569-3), PMID: 10754547
- Kruse T**, Blagoev B, Løbner-Olesen A, Wachi M, Sasaki K, Iwai N, Mann M, Gerdes K. 2006. Actin homolog MreB and RNA polymerase interact and are both required for chromosome segregation in *Escherichia coli*. *Genes & Development* **20**:113–124. DOI: <https://doi.org/10.1101/gad.366606>, PMID: 16391237
- Laemmli UK**. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680–685. DOI: <https://doi.org/10.1038/227680a0>, PMID: 5432063
- Lahti R**. 1983. Microbial inorganic pyrophosphatases. *Microbiological Reviews* **47**:169–178. DOI: <https://doi.org/10.1128/MR.47.2.169-178.1983>, PMID: 6135978
- Levine C**, Hiasa H, Mariani KJ. 1998. DNA gyrase and topoisomerase IV: biochemical activities, physiological roles during chromosome replication, and drug sensitivities. *Biochimica Et Biophysica Acta (BBA) - Gene Structure and Expression* **1400**:29–43. DOI: [https://doi.org/10.1016/S0167-4781\(98\)00126-2](https://doi.org/10.1016/S0167-4781(98)00126-2)
- Li C**, Wei J, Li Y, He X, Zhou Q, Yan J, Zhang J, Liu Y, Liu Y, Shu H-B. 2013. Transmembrane protein 214 (TMEM214) mediates endoplasmic reticulum stress-induced caspase 4 enzyme activation and apoptosis. *Journal of Biological Chemistry* **288**:17908–17917. DOI: <https://doi.org/10.1074/jbc.M113.458836>
- Liao L**, Wankel SD, Wu M, Cavanaugh CM, Girguis PR. 2014. Characterizing the plasticity of nitrogen metabolism by the host and symbionts of the hydrothermal vent chemoautotrophic symbioses *Ridgeia piscesae*. *Molecular Ecology* **23**:1544–1557. DOI: <https://doi.org/10.1111/mec.12460>, PMID: 24237389
- López-Sánchez MJ**, Neef A, Patiño-Navarrete R, Navarro L, Jiménez R, Latorre A, Moya A. 2008. *Blattabacteria*, the endosymbionts of cockroaches, have small genome sizes and high genome copy numbers. *Environmental Microbiology* **10**:3417–3422. DOI: <https://doi.org/10.1111/j.1462-2920.2008.01776.x>, PMID: 19025557
- Lorenzen DR**, Günther D, Pandit J, Rudel T, Brandt E, Meyer TF. 2000. *Neisseria gonorrhoeae* porin modifies the oxidative burst of human professional phagocytes. *Infection and Immunity* **68**:6215–6222. DOI: <https://doi.org/10.1128/IAI.68.11.6215-6222.2000>, PMID: 11035728
- Lutz RA**, Shank TM, Fornari DJ, Haymon RM, Lilley MD, Von Damm KL, Desbruyeres D. 1994. Rapid growth at deep-sea vents. *Nature* **371**:663–664. DOI: <https://doi.org/10.1038/371663a0>
- Lynch M**, Kuramitsu H. 2000. Expression and role of superoxide dismutases (SOD) in pathogenic bacteria. *Microbes and Infection* **2**:1245–1255. DOI: [https://doi.org/10.1016/S1286-4579\(00\)01278-8](https://doi.org/10.1016/S1286-4579(00)01278-8), PMID: 11008114
- Lyngstadaas A**, Løbner-Olesen A, Boye E. 1995. Characterization of three genes in the dam-containing operon of *Escherichia coli*. *Molecular and General Genetics MGG* **247**:546–554. DOI: <https://doi.org/10.1007/BF00290345>, PMID: 7603433
- Maechler M**, Rousseeuw P, Struyf A, Hubert M, Hornik K. 2018. cluster: Cluster Analysis Basics and Extensions. R Package. 2.0.7-1.
- Maeshima M**. 2000. Vacuolar H<sup>+</sup>-pyrophosphatase. *Biochimica Et Biophysica Acta (BBA) - Biomembranes* **1465**:37–51. DOI: [https://doi.org/10.1016/S0005-2736\(00\)00130-9](https://doi.org/10.1016/S0005-2736(00)00130-9)
- Mall A**, Sobotta J, Huber C, Tschirner C, Kowarschik S, Bačnik K, Mergelsberg M, Boll M, Hügler M, Eisenreich W, Berg IA. 2018. Reversibility of citrate synthase allows autotrophic growth of a thermophilic bacterium. *Science* **359**:563–567. DOI: <https://doi.org/10.1126/science.aao2410>, PMID: 29420287
- Markert S**, Arndt C, Felbeck H, Becher D, Sievert SM, Hügler M, Albrecht D, Robidart J, Bench S, Feldman RA, Hecker M, Schweder T. 2007. Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila*. *Science* **315**:247–250. DOI: <https://doi.org/10.1126/science.1132913>, PMID: 17218528
- Markert S**, Gardebrecht A, Felbeck H, Sievert SM, Klose J, Becher D, Albrecht D, Thürmer A, Daniel R, Kleiner M, Hecker M, Schweder T. 2011. Status quo in physiological proteomics of the uncultured *Riftia pachyptila* endosymbiont. *Proteomics* **11**:3106–3117. DOI: <https://doi.org/10.1002/pmic.201100059>, PMID: 21710568
- Mathur J**, Waldor MK. 2004. The *Vibrio cholerae* ToxR-regulated porin OmpU confers resistance to antimicrobial peptides. *Infection and Immunity* **72**:3577–3583. DOI: <https://doi.org/10.1128/IAI.72.6.3577-3583.2004>, PMID: 15155667
- McClain CR**, Balk MA, Benfield MC, Branch TA, Chen C, Cosgrove J, Dove AD, Gaskins LC, Helm RR, Hochberg FG, Lee FB, Marshall A, McMurray SE, Schanche C, Stone SN, Thaler AD. 2015. Sizing ocean giants: patterns of intraspecific size variation in marine megafauna. *PeerJ* **3**:e715. DOI: <https://doi.org/10.7717/peerj.715>, PMID: 25649000
- McNichol J**, Stryhanyuk H, Sylva SP, Thomas F, Musat N, Seewald JS, Sievert SM. 2018. Primary productivity below the seafloor at deep-sea hot springs. *PNAS* **115**:6756–6761 DOI: <https://doi.org/10.1073/pnas.1804351115>, PMID: 29891698
- Mergaert P**, Uchiumi T, Alunni B, Evanno G, Cheron A, Catrice O, Mausset AE, Barloy-Hubler F, Galibert F, Kondorosi A, Kondorosi E. 2006. Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. *PNAS* **103**:5230–5235. DOI: <https://doi.org/10.1073/pnas.0600912103>, PMID: 16547129



- Mitchell JH, Leonard JM, Delaney J, Girguis PR, Scott KM. 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**:e01522-19. DOI: <https://doi.org/10.1128/AEM.01522-19>, PMID: 31628148
- Moir JW, Wood NJ. 2001. Nitrate and nitrite transport in bacteria. *Cellular and Molecular Life Sciences* **58**:215–224. DOI: <https://doi.org/10.1007/PL00000849>, PMID: 11289303
- Monds RD, Lee TK, Colavin A, Ursell T, Quan S, Cooper TF, Huang KC. 2014. Systematic perturbation of cytoskeletal function reveals a linear scaling relationship between cell geometry and fitness. *Cell Reports* **9**:1528–1537. DOI: <https://doi.org/10.1016/j.celrep.2014.10.040>, PMID: 25456141
- Moreno-Vivián C, Cabello P, Martínez-Luque M, Blasco R, Castillo F. 1999. Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. *Journal of Bacteriology* **181**:6573–6584. DOI: <https://doi.org/10.1128/JB.181.21.6573-6584.1999>, PMID: 10542156
- Mosleh IM, Huber LA, Steinlein P, Pasquali C, Günther D, Meyer TF. 1998. *Neisseria gonorrhoeae* porin modulates phagosome maturation. *Journal of Biological Chemistry* **273**:35332–35338. DOI: <https://doi.org/10.1074/jbc.273.52.35332>
- Mueller RS, Deneff VJ, Kalnejais LH, Suttle KB, Thomas BC, Wilmes P, Smith RL, Nordstrom DK, McCleskey RB, Shah MB, Verberkmoes NC, Hettich RL, Banfield JF. 2010. Ecological distribution and population physiology defined by proteomics in a natural microbial community. *Molecular Systems Biology* **6**:374. DOI: <https://doi.org/10.1038/msb.2010.30>, PMID: 20531404
- Mukherjee A, Sokunbi AO, Grove A. 2008. DNA protection by histone-like protein HU from the hyperthermophilic Eubacterium *Thermotoga maritima*. *Nucleic Acids Research* **36**:3956–3968. DOI: <https://doi.org/10.1093/nar/gkn348>, PMID: 18515342
- Mulligan C, Kelly DJ, Thomas GH. 2007. Tripartite ATP-independent periplasmic transporters: application of a relational database for genome-wide analysis of transporter gene frequency and organization. *Journal of Molecular Microbiology and Biotechnology* **12**:218–226. DOI: <https://doi.org/10.1159/000099643>, PMID: 17587870
- Mulligan C, Fischer M, Thomas GH. 2011. Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. *FEMS Microbiology Reviews* **35**:68–86. DOI: <https://doi.org/10.1111/j.1574-6976.2010.00236.x>, PMID: 20584082
- Murga M, Jaco I, Fan Y, Soria R, Martinez-Pastor B, Cuadrado M, Yang SM, Blasco MA, Skoultchi AI, Fernandez-Capetillo O. 2007. Global chromatin compaction limits the strength of the DNA damage response. *Journal of Cell Biology* **178**:1101–1108. DOI: <https://doi.org/10.1083/jcb.200704140>, PMID: 17893239
- Nikolakakis K, Lehnert E, McFall-Ngai MJ, Ruby EG. 2015. Use of hybridization chain reaction-fluorescent *in situ* hybridization to track gene expression by both partners during initiation of symbiosis. *Applied and Environmental Microbiology* **81**:4728–4735. DOI: <https://doi.org/10.1128/AEM.00890-15>, PMID: 25956763
- Nöllmann M, Crisona NJ, Arimondo PB. 2007. Thirty years of *Escherichia coli* DNA gyrase: from *in vivo* function to single-molecule mechanism. *Biochimie* **89**:490–499. DOI: <https://doi.org/10.1016/j.biochi.2007.02.012>, PMID: 17397985
- Nussbaumer AD, Fisher CR, Bright M. 2006. Horizontal endosymbiont transmission in hydrothermal vent tubeworms. *Nature* **441**:345–348. DOI: <https://doi.org/10.1038/nature04793>, PMID: 16710420
- Nyholm SV, Stewart JJ, Ruby EG, McFall-Ngai MJ. 2009. Recognition between symbiotic *Vibrio fischeri* and the haemocytes of *Euprymna scolopes*. *Environmental Microbiology* **11**:483–493. DOI: <https://doi.org/10.1111/j.1462-2920.2008.01788.x>, PMID: 19196278
- Ohniwa RL, Morikawa K, Kim J, Ohta T, Ishihama A, Wada C, Takeyasu K. 2006. Dynamic state of DNA topology is essential for genome condensation in bacteria. *The EMBO Journal* **25**:5591–5602. DOI: <https://doi.org/10.1038/sj.emboj.7601414>, PMID: 17093499
- Oke V, Long SR. 1999. Bacteroid formation in the *Rhizobium*-legume symbiosis. *Current Opinion in Microbiology* **2**:641–646. DOI: [https://doi.org/10.1016/S1369-5274\(99\)00035-1](https://doi.org/10.1016/S1369-5274(99)00035-1), PMID: 10607628
- Ozawa K, Kuwabara K, Tamatani M, Takatsuji K, Tsukamoto Y, Kaneda S, Yanagi H, Stern DM, Eguchi Y, Tsujimoto Y, Ogawa S, Tohyama M. 1999. 150-kDa Oxygen-regulated protein (ORP150) suppresses hypoxia-induced apoptotic cell death. *Journal of Biological Chemistry* **274**:6397–6404. DOI: <https://doi.org/10.1074/jbc.274.10.6397>
- Pearson A. 2010. Pathways of carbon assimilation and their impact on organic matter values  $\delta^{13}\text{C}$ . In: Timmis K. N (Ed). *Handbook of Hydrocarbon and Lipid Microbiology*. Springer. p. 143–156. DOI: [https://doi.org/10.1007/978-3-540-77587-4\\_9](https://doi.org/10.1007/978-3-540-77587-4_9)
- Pende N, Leisch N, Gruber-Vodicka HR, Heindl NR, Ott J, den Blaauwen T, Bulgheresi S. 2014. Size-independent symmetric division in extraordinarily long cells. *Nature Communications* **5**:4803. DOI: <https://doi.org/10.1038/ncomms5803>, PMID: 25221974
- Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R, Hourdez S, Girguis PR, Wankel SD, Barbe V, Pelletier E, Fink D, Borowski C, Bach W, Dubilier N. 2011. Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* **476**:176–180. DOI: <https://doi.org/10.1038/nature10325>, PMID: 21833083
- Petersen I, Schlüter R, Hoff KJ, Liebscher V, Bange G, Riedel K, Pané-Farré J. 2020. Non-invasive and label-free 3D-visualization shows *in vivo* oligomerization of the staphylococcal alkaline shock protein 23 (Asp23). *Scientific Reports* **10**:125. DOI: <https://doi.org/10.1038/s41598-019-56907-9>, PMID: 31924851
- Pflugfelder B, Fisher CR, Bright M. 2005. The color of the trophosome: elemental sulfur distribution in the endosymbionts of *Riftia pachyptila* (Vestimentifera; Siboglinidae). *Marine Biology* **146**:895–901. DOI: <https://doi.org/10.1007/s00227-004-1500-x>

- Polz MF**, Felbeck H, Novak R, Nebelsick M, Ott JA. 1992. Chemoautotrophic, sulfur-oxidizing symbiotic bacteria on marine nematodes: morphological and biochemical characterization. *Microbial Ecology* **24**:313–329. DOI: <https://doi.org/10.1007/BF00167789>, PMID: 24193210
- Polzin J**, Arevalo P, Nussbaumer T, Polz MF, Bright M. 2019. Polyclonal symbiont populations in hydrothermal vent tubeworms and the environment. *Proceedings of the Royal Society B: Biological Sciences* **286**:20181281. DOI: <https://doi.org/10.1098/rspb.2018.1281>
- Ponnudurai R**, Kleiner M, Sayavedra L, Petersen JM, Moche M, Otto A, Becher D, Takeuchi T, Satoh N, Dubilier N, Schweder T, Markert S. 2017. Metabolic and physiological interdependencies in the *Bathymodiolus azoricus* symbiosis. *The ISME Journal* **11**:463–477. DOI: <https://doi.org/10.1038/ismej.2016.124>, PMID: 27801908
- Ponsard J**, Cambon-Bonavita MA, Zbinden M, Lepoint G, Joassin A, Corbari L, Shillito B, Durand L, Cuffe-Gauchard V, Compère P. 2013. Inorganic carbon fixation by chemosynthetic ectosymbionts and nutritional transfers to the hydrothermal vent host-shrimp *Rimicaris exoculata*. *The ISME Journal* **7**:96–109. DOI: <https://doi.org/10.1038/ismej.2012.87>, PMID: 22914596
- Prinsloo E**, Setati MM, Longshaw VM, Blatch GL. 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: <https://doi.org/10.1002/bies.200800158>, PMID: 19274656
- R Development Core Team**. 2018. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reimold C**, Defeu Soufo HJ, Dempwolff F, Graumann PL. 2013. Motion of variable-length MreB filaments at the bacterial cell membrane influences cell morphology. *Molecular Biology of the Cell* **24**:2340–2349. DOI: <https://doi.org/10.1091/mbc.e12-10-0728>, PMID: 23783036
- Reitzer L**. 2003. Nitrogen assimilation and global regulation in *Escherichia coli*. *Annual Review of Microbiology* **57**:155–176. DOI: <https://doi.org/10.1146/annurev.micro.57.030502.090820>, PMID: 12730324
- Renoz F**, Champagne A, Degand H, Faber AM, Morsomme P, Foray V, Hance T. 2017. Toward a better understanding of the mechanisms of symbiosis: a comprehensive proteome map of a nascent insect symbiont. *PeerJ* **5**:e3291. DOI: <https://doi.org/10.7717/peerj.3291>, PMID: 28503376
- Robidart JC**, Bench SR, Feldman RA, Novoradovsky A, Podell SB, Gaasterland T, Allen EE, Felbeck H. 2008. Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed through metagenomics. *Environmental Microbiology* **10**:727–737. DOI: <https://doi.org/10.1111/j.1462-2920.2007.01496.x>, PMID: 18237306
- Robidart JC**, Roque A, Song P, Girguis PR. 2011. Linking hydrothermal geochemistry to organismal physiology: physiological versatility in *Riftia pachyptila* from sedimented and basalt-hosted vents. *PLOS ONE* **6**:e21692. DOI: <https://doi.org/10.1371/journal.pone.0021692>, PMID: 21779334
- Robinson MD**, McCarthy DJ, Smyth GK. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**:139–140. DOI: <https://doi.org/10.1093/bioinformatics/btp616>, PMID: 19910308
- Rodriguez J**, Hiras J, Hanson TE. 2011. Sulfite oxidation in *Chlorobaculum tepidum*. *Frontiers in Microbiology* **2**:112. DOI: <https://doi.org/10.3389/fmicb.2011.00112>, PMID: 21747809
- Royet J**, Gupta D, Dziarski R. 2011. Peptidoglycan recognition proteins: modulators of the microbiome and inflammation. *Nature Reviews Immunology* **11**:837–851. DOI: <https://doi.org/10.1038/nri3089>, PMID: 22076558
- Saalfeld S**. 2010. Enhance Local Contrast (CLAHE). [https://imagej.net/Enhance\\_Local\\_Contrast\\_\(CLAHE\)](https://imagej.net/Enhance_Local_Contrast_(CLAHE))
- Sattler MC**, Carvalho CR, Clarindo WR. 2016. The polyploidity and its key role in plant breeding. *Planta* **243**:281–296. DOI: <https://doi.org/10.1007/s00425-015-2450-x>, PMID: 26715561
- Schindelin J**, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**:676–682. DOI: <https://doi.org/10.1038/nmeth.2019>, PMID: 22743772
- Scott KM**, Boller AJ, Dobrinski KP, Le Bris N. 2012. Response of hydrothermal vent vestimentiferan *Riftia pachyptila* to differences in habitat chemistry. *Marine Biology* **159**:435–442. DOI: <https://doi.org/10.1007/s00227-011-1821-5>
- Seah BKB**, Antony CP, Huettel B, Zarzycki J, Schada von Borzyskowski L, Erb TJ, Kouris A, Kleiner M, Liebeke M, Dubilier N, Gruber-Vodicka HR. 2019. Sulfur-oxidizing symbionts without canonical genes for autotrophic CO<sub>2</sub> fixation. *mBio* **10**:e01112-19. DOI: <https://doi.org/10.1128/mBio.01112-19>, PMID: 31239380
- Søndergaard D**, Pedersen CN, Greening C. 2016. HydDB: a web tool for hydrogenase classification and analysis. *Scientific Reports* **6**:34212. DOI: <https://doi.org/10.1038/srep34212>, PMID: 27670643
- Sorgo A**, Gaill F, Lechère JP, Arndt C, Bright M. 2002. Glycogen storage in the *Riftia pachyptila* trophosome: contribution of host and symbionts. *Marine Ecology Progress Series* **231**:115–120. DOI: <https://doi.org/10.3354/meps231115>
- Souppèe E**, Foussard M, Boistard P, Truchet G, Batut J. 1995. Oxygen as a key developmental regulator of *Rhizobium meliloti* N<sub>2</sub>-fixation gene expression within the alfalfa root nodule. *PNAS* **92**:3759–3763. DOI: <https://doi.org/10.1073/pnas.92.9.3759>, PMID: 7731979
- Steck TR**, Drlica K. 1984. Bacterial chromosome segregation: evidence for DNA gyrase involvement in decatenation. *Cell* **36**:1081–1087. DOI: [https://doi.org/10.1016/0092-8674\(84\)90058-8](https://doi.org/10.1016/0092-8674(84)90058-8), PMID: 6323018
- Stewart FJ**, Dmytrenko O, Delong EF, Cavanaugh CM. 2011. Metatranscriptomic analysis of sulfur oxidation genes in the endosymbiont of *Solemya velum*. *Frontiers in Microbiology* **2**:134. DOI: <https://doi.org/10.3389/fmicb.2011.00134>, PMID: 21738524

- Stockdreher Y**, Sturm M, Josten M, Sahl H-G, Dobler N, Zigann R, Dahl C. 2014. New proteins involved in sulfur trafficking in the cytoplasm of *Allochromatium vinosum*. *Journal of Biological Chemistry* **289**:12390–12403. DOI: <https://doi.org/10.1074/jbc.M113.536425>
- Stocks SM**. 2004. Mechanism and use of the commercially available viability stain, BacLight. *Cytometry* **61**:189–195. DOI: <https://doi.org/10.1002/cyto.a.20069>, PMID: 15382024
- Stuger R**, Woldringh CL, van der Weijden CC, Vischer NO, Bakker BM, van Spanning RJ, Snoep JL, Westerhoff HV. 2002. DNA supercoiling by gyrase is linked to nucleoid compaction. *Molecular Biology Reports* **29**:79–82. DOI: <https://doi.org/10.1023/a:1020318705894>, PMID: 12241080
- Susin MF**, Baldini RL, Gueiros-Filho F, Gomes SL. 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: <https://doi.org/10.1128/JB.00824-06>, PMID: 16980445
- Takata H**, Hanafusa T, Mori T, Shimura M, Iida Y, Ishikawa K, Yoshikawa K, Yoshikawa Y, Maeshima K. 2013. Chromatin compaction protects genomic DNA from radiation damage. *PLOS ONE* **8**:e75622. DOI: <https://doi.org/10.1371/journal.pone.0075622>, PMID: 24130727
- Tessmer I**, Moore T, Lloyd RG, Wilson A, Erie DA, Allen S, Tendler SJB. 2005. AFM studies on the role of the protein RdgC in bacterial DNA recombination. *Journal of Molecular Biology* **350**:254–262. DOI: <https://doi.org/10.1016/j.jmb.2005.04.043>
- Thanedar S**, Margolin W. 2004. FtsZ exhibits rapid movement and oscillation waves in helix-like patterns in *Escherichia coli*. *Current Biology* **14**:1167–1173. DOI: <https://doi.org/10.1016/j.cub.2004.06.048>, PMID: 15242613
- The Global Proteome Machine Organization**. 2017. The global proteome machine: crap protein sequences. <http://thegpm.org/crap/> [Accessed November 28, 2017].
- Thévenaz P**, Sage D, Unser M. 2012. Bi-exponential edge-preserving smoother. *IEEE Transactions on Image Processing* **21**:3924–3936. DOI: <https://doi.org/10.1109/TIP.2012.2200903>, PMID: 22645271
- Tracy BP**, Gaida SM, Papoutsakis ET. 2010. Flow cytometry for bacteria: enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes. *Current Opinion in Biotechnology* **21**:85–99. DOI: <https://doi.org/10.1016/j.copbio.2010.02.006>, PMID: 20206495
- Troll JV**, Adin DM, Wier AM, Paquette N, Silverman N, Goldman WE, Stadermann FJ, Stabb EV, McFall-Ngai MJ. 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: <https://doi.org/10.1111/j.1462-5822.2009.01315.x>, PMID: 19416268
- Truglio JJ**, Croteau DL, Van Houten B, Kisker C. 2006. Prokaryotic nucleotide excision repair: the UvrABC system. *Chemical Reviews* **106**:233–252. DOI: <https://doi.org/10.1021/cr040471u>, PMID: 16464004
- Twine SM**, Mykytczuk NC, Petit MD, Shen H, Sjöstedt A, Wayne Conlan J, Kelly JF. 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: <https://doi.org/10.1016/j.bbrc.2006.05.070>, PMID: 16730660
- van der Hoeven R**, Forst S. 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: <https://doi.org/10.1128/JB.00148-09>, PMID: 19465651
- van der Land J**, Nørrevang A. 1977. *Structure and Relationships of Lamellibrachia (Annelida, Vestimentifera)*. Det Kongelige Danske Videnskabernes Selskab, Biologiske Skrifter.
- Venceslau SS**, Stockdreher Y, Dahl C, Pereira IAC. 2014. The “bacterial heterodisulfide” DsrC is a key protein in dissimilatory sulfur metabolism. *Biochimica Et Biophysica Acta (BBA) - Bioenergetics* **1837**:1148–1164. DOI: <https://doi.org/10.1016/j.bbabi.2014.03.007>
- Verbrugghe E**, Van Parys A, Leyman B, Boyen F, Haesebrouck F, Pasmans F. 2015. HtpG contributes to *Salmonella typhimurium* intestinal persistence in pigs. *Veterinary Research* **46**:118. DOI: <https://doi.org/10.1186/s13567-015-0261-5>, PMID: 26466674
- Vizcaíno JA**, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, Xu Q-W, Wang R, Hermjakob H. 2016. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Research* **44**:D447–D456. DOI: <https://doi.org/10.1093/nar/gkv1145>
- Wachi M**, Matsushashi M. 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: <https://doi.org/10.1128/JB.171.6.3123-3127.1989>, PMID: 2656641
- Wang J**, Wu Y, Yang G, Aksoy S. 2009. Interactions between mutualist *Wigglesworthia* and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypanosome transmission. *PNAS* **106**:12133–12138. DOI: <https://doi.org/10.1073/pnas.0901226106>, PMID: 19587241
- Wang S**, Arellano-Santoyo H, Combs PA, Shaevitz JW. 2010. Actin-like cytoskeleton filaments contribute to cell mechanics in bacteria. *PNAS* **107**:9182–9185. DOI: <https://doi.org/10.1073/pnas.0911517107>, PMID: 20439764
- Wang Y**, Maharana S, Wang MD, Shivashankar GV. 2015. Super-resolution microscopy reveals decondensed chromatin structure at transcription sites. *Scientific Reports* **4**:4477. DOI: <https://doi.org/10.1038/srep04477>
- Weiss DS**. 2004. Bacterial cell division and the septal ring. *Molecular Microbiology* **54**:588–597. DOI: <https://doi.org/10.1111/j.1365-2958.2004.04283.x>
- Weissgerber T**, Zigann R, Bruce D, Chang YJ, Detter JC, Han C, Hauser L, Jeffries CD, Land M, Munk AC, Tapia R, Dahl C. 2011. Complete genome sequence of *Allochromatium vinosum* DSM 180(T). *Standards in Genomic Sciences* **5**:311–330. DOI: <https://doi.org/10.4056/sigs.2335270>, PMID: 22675582

- Weissgerber T, Dobler N, Polen T, Latus J, Stockdreher Y, Dahl C. 2013. Genome-wide transcriptional profiling of the purple sulfur bacterium *Allochromatium vinosum* DSM 180T during growth on different reduced sulfur compounds. *Journal of Bacteriology* **195**:4231–4245. DOI: <https://doi.org/10.1128/JB.00154-13>, PMID: 23873913
- Weissgerber T, Sylvester M, Kröninger L, Dahl C. 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: <https://doi.org/10.1128/AEM.04182-13>, PMID: 24487535
- Welte C, Hafner S, Krätzer C, Quentmeier A, Friedrich CG, Dahl C. 2009. Interaction between Sox proteins of two physiologically distinct bacteria and a new protein involved in thiosulfate oxidation. *FEBS Letters* **583**: 1281–1286. DOI: <https://doi.org/10.1016/j.febslet.2009.03.020>, PMID: 19303410
- Wetmur JG. 1991. DNA probes: applications of the principles of nucleic acid hybridization. *Critical Reviews in Biochemistry and Molecular Biology* **26**:227–259. DOI: <https://doi.org/10.3109/10409239109114069>, PMID: 1718662
- Wilmot DB, Vetter RD. 1990. The bacterial symbiont from the hydrothermal vent tubeworm *Riftia pachyptila* is a sulfide specialist. *Marine Biology* **106**:273–283. DOI: <https://doi.org/10.1007/BF01314811>
- Wippler J, Kleiner M, Lott C, Gruhl A, Abraham PE, Giannone RJ, Young JC, Hettich RL, Dubilier N. 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: <https://doi.org/10.1186/s12864-016-3293-y>
- Woyke T, Tighe D, Mavromatis K, Clum A, Copeland A, Schackwitz W, Lapidus A, Wu D, McCutcheon JP, McDonald BR, Moran NA, Bristow J, Cheng JF. 2010. One bacterial cell, one complete genome. *PLOS ONE* **5**: e10314. DOI: <https://doi.org/10.1371/journal.pone.0010314>, PMID: 20428247
- Wright MN, Ziegler A. 2015. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software* **77**:1–17. DOI: <https://doi.org/10.18637/jss.v077.i01>
- Yilmaz LS, Parnerkar S, Noguera DR. 2011. mathFISH, a web tool that uses thermodynamics-based mathematical models for *in silico* evaluation of oligonucleotide probes for fluorescence *in situ* hybridization. *Applied and Environmental Microbiology* **77**:1118–1122. DOI: <https://doi.org/10.1128/AEM.01733-10>, PMID: 21148691
- Yoshikawa Y, Mori T, Magome N, Hibino K, Yoshikawa K. 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: <https://doi.org/10.1016/j.cplett.2008.03.009>
- Zambrano N, Guichard PP, Bi Y, Cayrol B, Marco S, Arluisson V. 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: <https://doi.org/10.4161/cc.8.15.9090>, PMID: 19571680
- Zerulla K, Soppa J. 2014. Polyploidy in Haloarchaea: advantages for growth and survival. *Frontiers in Microbiology* **5**:274. DOI: <https://doi.org/10.3389/fmicb.2014.00274>, PMID: 24982654
- Zheng H, Perreau J, Powell JE, Han B, Zhang Z, Kwong WK, Tringe SG, Moran NA. 2019. Division of labor in honey bee gut microbiota for plant polysaccharide digestion. *PNAS* **116**:25909–25916. DOI: <https://doi.org/10.1073/pnas.1916224116>, PMID: 31776248
- Zybailov B, Mosley AL, Sardi ME, Coleman MK, Florens L, Washburn MP. 2006. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *Journal of Proteome Research* **5**:2339–2347. DOI: <https://doi.org/10.1021/pr060161n>, PMID: 16944946

## Appendix 1

### Supplementary results and discussion

#### A. Symbiotic Endoriftia cells exist in a remarkable size range

Symbiont cell sizes in *Riftia* trophosome tissue range from 1 to 2  $\mu\text{m}$  to more than 15  $\mu\text{m}$  (main text: **Figure 1, Figure 5**). This is in line with previous microscopy-based observations, which suggested that the symbiont cells differentiate from small rod-shaped cells in the trophosome lobule center to larger coccoid cells towards the lobule periphery (**Bright and Sorgo, 2003**). With an about 10-fold increase in diameter, Endoriftia cells enlarge their volume by a factor of  $\sim 1000$  during their development from smallest to largest coccoid symbiont cells. Considerable enlargement of bacterial cells in the course of symbiotic differentiation has also been observed in the intracellular thiotrophic symbiont of the shallow water clam *Codakia orbicularis* (increases 10-fold in length; **Caro et al., 2007**), in *Sinorhizobium melliloti* in alfalfa nodules (increases four- to seven-fold in length; **Oke and Long, 1999**), in symbionts of the nematode *Eubostrichus* (increase up to 13-fold in length; **Pende et al., 2014**), and in the giant bacterium *Epulopiscium fishelsoni*, intestinal symbiont of surgeon fish (increases up to 3,000-fold in volume; **Bresler and Fishelson, 2003**). Such enormous size gradients are rather the exception than the rule in bacteria, however. Cell sizes, that is, length or diameter, of free-living model bacteria like *Bacillus subtilis* or *Escherichia coli* usually vary only by factor 2 (during cell division), that is, mathematically, these bacteria may increase their volume twofold (assuming a cylindrical shape) to eightfold (assuming a spherical shape) at most (**Chien et al., 2012**). This suggests that the remarkably large size range observed for Endoriftia presents a consequence of its symbiotic life style.

#### B. Comparative analysis of enriched symbiont fractions from S-rich vs. S-depleted *Riftia* specimens

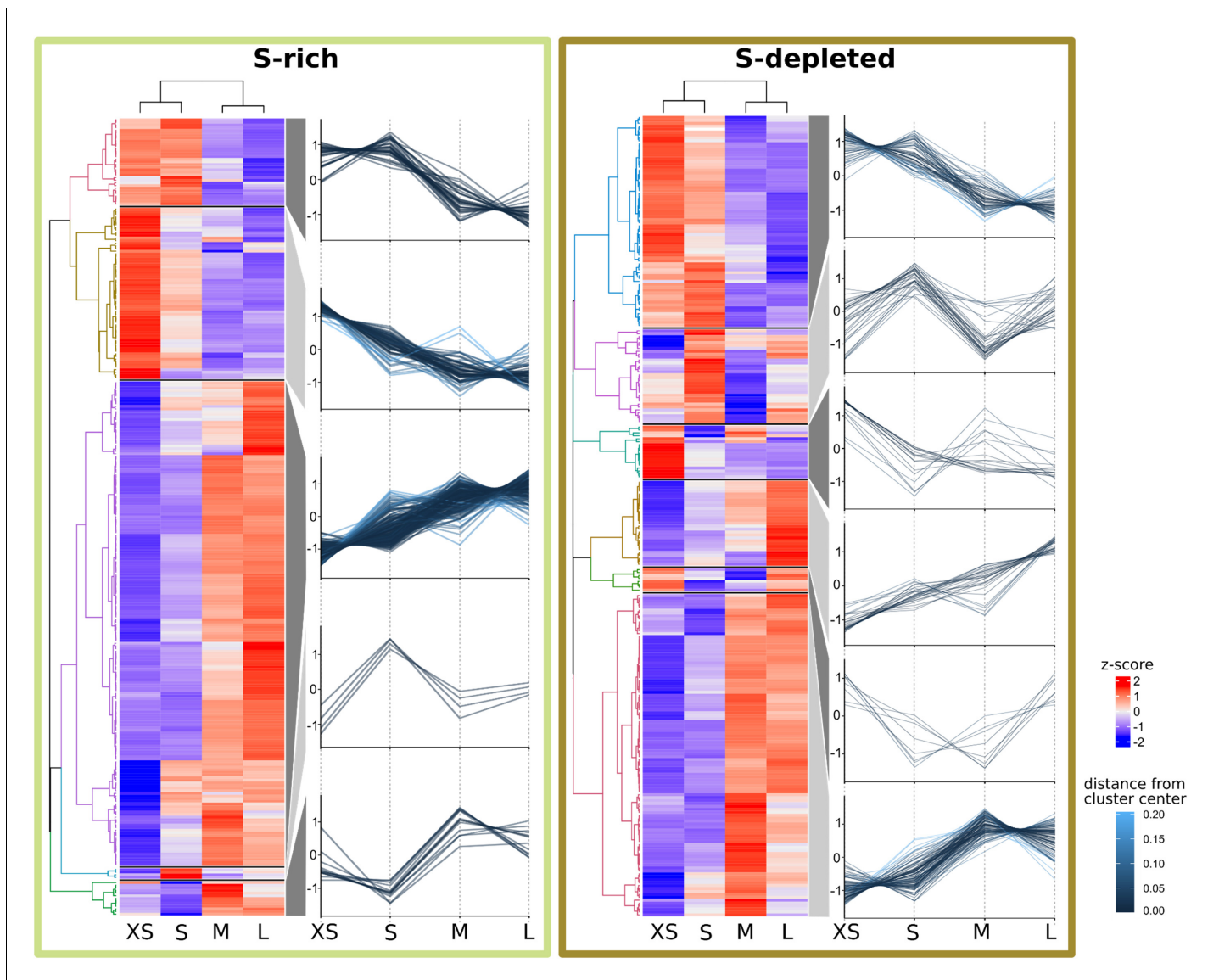
##### Overview

Our comparative analyses of symbiont-enriched fractions XS to L revealed that in both S-rich and S-depleted samples, protein profiles differed with increasing symbiont cell size (**Appendix 1—figure 1**). Many groups of proteins (e.g. carbohydrate metabolism-related proteins) showed similar trends across size classes in S-rich and S-depleted specimens, even if individual protein abundances differed. Statistical testing for significant differences in protein abundance between S-rich and S-depleted fractions of the same size class returned only very few (edgeR) or no (random forest) hits. This may in part be due to the less effective enrichment of symbionts from S-depleted trophosome tissue homogenate. However, very similar abundance patterns in symbionts from sulfur-rich and sulfur-depleted hosts might also reflect the fact that symbionts are very well buffered against environmental changes (as previously suggested, **Hinzke et al., 2019**) and, therefore, functional differences between symbiont morphotypes in S-rich vs. S-depleted symbionts might largely be negligible. Some of these differences, however, seemed to be specific for the respective energy situation and are outlined below.

##### Cell division

In sulfur-depleted hosts, *Riftia* symbionts appear to divide less frequently than in sulfur-rich specimens, as indicated by lower abundance of the major cell division protein FtsZ in all S-depleted fractions compared to their S-rich counterparts (**Supplementary file 2**; please note that, due to its low abundance, FtsZ was not included in statistical analysis in S-depleted samples). In S-depleted fraction XS, FtsZ abundance was about 3.5 times lower than in S-rich fraction XS.

Less symbiont cell division in S-depleted *Riftia* accords with the idea of severe energy limitation in sulfur-depleted symbionts and is in agreement with our previous finding that symbiont proteinaceous biomass is lower in trophosomes of S-depleted specimens (**Hinzke et al., 2019**). In this previous study, we suggested that S-depleted hosts digest a larger part of their symbiont population as compared to S-rich tubeworms. As the host mainly digests large symbionts at the trophosome lobule periphery (**Figure 5** main text; **Bright and Sorgo, 2003**), one might expect that more digestion leads to relatively more smaller symbionts in S-depleted trophosomes as compared to S-rich hosts.



**Appendix 1—figure 1.** Abundance trends of 465 *Riftia* symbiont proteins with significant abundance differences between the four analyzed gradient fractions XS (enriched in very small symbiont cells) to L (containing the highest percentage of large symbiont cells) in S-rich and S-depleted *Riftia* trophosomes. Heat maps show relative protein abundances (z-scores of edgeR-RLE-corrected spectral count values; see Methods for details) and line graphs indicate trends in the observed differences.

This was, however, not the case, as symbiont size distribution was quite comparable in trophosome homogenates of S-rich and S-depleted trophosomes (see **Figure 1**, main text). Such similar distributions would be in line with the hypothesis that the host may digest small symbionts, too, besides large *Endoriftia* (see main text). However, as we could not see evidence for small symbionts in the process of digestion in any of our TEM images, we rather propose that both more symbiont digestion and less symbiont cell division co-occur in S-depleted worm specimens, leading to the previously observed loss in total symbiont biomass.

### Carbon fixation

Enzymes of the two autotrophic pathways, Calvin cycle and rTCA cycle, did not show significant differences in protein abundance patterns when comparing S-rich and S-depleted gradient fractions. However, based on our SIF analysis (**Appendix 1—figure 2**, see below), both pathways could be differentially used between the two sample sets: Throughout all gradient fractions, S-depleted samples

had a notably more positive  $\delta^{13}\text{C}$  value than the respective S-rich fractions. This indicates that in S-depleted Endoriftia, the relative contribution of the rTCA cycle to overall carbon fixation is higher than in S-rich symbionts. This is in good agreement with earlier findings of comparative proteomic analyses (Markert et al., 2007). Since the rTCA cycle is more energy-efficient than the Calvin cycle, this shift towards a higher contribution of the less costly autotrophic pathway might be beneficial during energy limitation.

### Growth-related processes

Highest abundance of RNA polymerase subunits, transcription elongation factors, transcription anti-termination protein and various translation-related proteins in fraction XS of S-rich and S-depleted specimens indicates that small symbionts devote relatively more energy and resources to protein synthesis than large symbionts (Supplementary file 3h). This is in agreement with the idea that small Endoriftia function as actively dividing and growing 'stem cells' of the Endoriftia population, whereas large symbionts have the role of highly efficient biomass producers (see main text).

This proposed greater importance of growth-related processes in small symbionts may result in higher intracellular pyrophosphate levels, as suggested by high abundance of pyrophosphatases in fraction XS. The highly abundant pyrophosphate-energized proton pump HppA (Sym\_EGV49909.1) and the inorganic pyrophosphatase Ppa (Sym\_EGV49908.1) had their highest abundances in fraction XS in S-rich samples (Supplementary file 2). Pyrophosphatases play an important role in energy metabolism by catalyzing the hydrolysis of inorganic pyrophosphates ( $\text{PP}_i$ ), which are produced at particularly high rates by biosynthetic reactions in growing cells (Klemme, 1976; Chen et al., 1990). By removing  $\text{PP}_i$ , pyrophosphatases shift the thermodynamic equilibrium to favor reactions like DNA, RNA, and protein synthesis (Lahti, 1983).

HppA may furthermore have an additional growth-related function: During  $\text{PP}_i$  hydrolysis, HppA pumps protons into the periplasm, thus establishing a proton motive force (Maeshima, 2000). As cell division is an energy-expensive process, which requires not only ATP but also proton motive force (Goehring and Beckwith, 2005), HppA may be upregulated to accommodate this increased demand in small, dividing Endoriftia. At the same time, HppA presumably increases energy efficiency of the Calvin cycle (Markert et al., 2011). Interestingly, HppA abundance was notably lower in S-depleted XS fractions, supporting the idea of reduced cell division in energy-depleted symbionts (see above).

Besides HppA, highly abundant nitrate reductase NarGHI showed a very similar abundance pattern with (i) decreasing abundance from fraction XS to L and (ii) notably lower abundances in S-depleted fractions compared to their respective S-rich counterparts (Supplementary file 2). NarGHI, which catalyzes the first step of anaerobic denitrification (see main text), also produces a proton motive force (Bertero et al., 2003). We therefore speculate that it may also support cell division in small symbionts, and that its lower abundance in S-depleted samples correlates with less cell division in these symbionts.

### Host interactions

Proteins which may protect the symbiont from digestion by the host could be most important in small symbiont cells and particularly so in S-depleted *Riftia*, as suggested by highest abundances of an ankyrin protein and of the FK506-binding protein FkpA in S-depleted fraction S (Supplementary file 2).

In S-rich and S-depleted samples, the ankyrin-like symbiont protein (Sym\_EGV51005.1) decreased in abundance from fraction S to L. Endoriftia ankyrin repeat-containing proteins were previously suggested to be involved in microbe-host interactions, possibly to counteract digestion by the host (Hinze et al., 2019). As small Endoriftia are the main dividing symbiont subpopulation and thus ensure survival of the symbiont population as a whole, digesting those cells would harm not only the symbiont, but also the host itself. The ankyrin protein could fulfill a protective role especially for these smaller symbionts.

The *Riftia* symbiont's FK506-binding protein (Sym\_EGV50540.1), which showed a comparable abundance trend, might have a similar role. In *Salmonella typhimurium* and *Cronobacter*, FkpA is

involved in survival inside host cells (*Horne et al., 1997; Eshwar et al., 2015*), suggesting that the Endoriftia FkpA, too, provides protection for the intracellular symbiont.

### C. Flow cytometry of *Riftia* symbionts

According to our flow cytometry data, *Riftia* trophosome homogenate and enriched gradient fractions were quite heterogeneous (**Figure 2—figure supplement 1**), with a number of other populations present besides populations 1 (small symbionts) and 2 (large symbionts). This heterogeneity may, to a minor degree, result from contaminating host organelles, but is presumably mostly due to the fact that (a) symbionts exist not only as small or large cells, but also adopt any intermediate size, and (b) intracellularly stored sulfur influences the cells' light-scattering properties (especially side scatter, SSC), considerably (as shown for thiotrophic lucinid symbionts; *Caro et al., 2007*).

We sorted one of the additional populations, with SSC between  $10^4$  and  $10^5$  and FSC between  $10^3$  and  $10^4$  (i.e. with higher SSC but lower FSC than populations 1 and 2) to examine it separately. Fluorescence microscopy revealed that this population consisted mostly of medium-sized symbionts, which – unlike populations 1 and 2 – contained numerous sulfur globules (images not shown). It could be assumed that other symbiont cell populations, for example, small S-rich and large S-rich cells, might also be present. This hypothesis awaits confirmation in future studies. To estimate symbiont DNA content in the present study, we only included populations 1 and 2, which were readily comparable due to their similar sulfur content (i.e. there were hardly any sulfur globules visible).

As also described for a thiotrophic lucinid symbiont (*Caro et al., 2007*), cell populations were not entirely congruent across the two bioreplicates in our Endoriftia flow cytometry analyses. Consequently, individual fluorescence intensity (FI) values varied by a factor of 2 (on average; **Supplementary file 6**). Nevertheless, both replicates clearly showed the same trend, that is, higher FI per particle in population 2 compared to population 1 across all samples, strongly indicating multiple genome copy numbers in large symbionts.

### D. CO<sub>2</sub> metabolism is differentially regulated across Endoriftia cell sizes Small and large Endoriftia differ with regard to their autotrophic pathway use

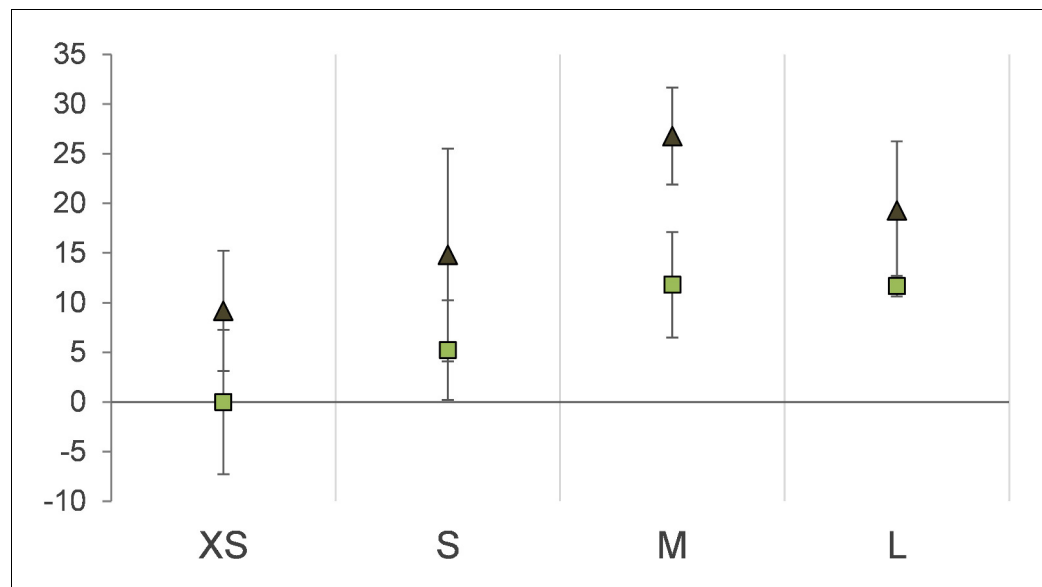
Relative contributions of Calvin cycle and rTCA cycle to autotrophic net carbon fixation appear to differ distinctly between small and large Endoriftia, as revealed by our analysis of stable carbon isotope fingerprints (SIFs, **Appendix 1—figure 2**). The Calvin cycle, with its type II RubisCO, fixes preferentially <sup>12</sup>CO<sub>2</sub> rather than <sup>13</sup>CO<sub>2</sub>, that is, it discriminates against <sup>13</sup>C with a shift of –5 to –25‰. In contrast, the rTCA cycle discriminates only with –2 to –13‰ against <sup>13</sup>C, which makes it possible to distinguish between carbon fixed by either of the two pathways or by varying contributions of both (*Pearson, 2010*). The differences in SIF values we observed point to differential use of Calvin cycle and rTCA cycle in small vs. large symbionts: Fraction XS had the most negative of all δ<sup>13</sup>C values, indicating that relatively more of the carbon in these samples was fixed by the Calvin cycle key enzyme RubisCO than in the other fractions, that is, the Calvin cycle's relative contribution to autotrophy is highest in small symbionts. With growing symbiont cell size, this contribution appears to decrease, as indicated by more positive δ<sup>13</sup>C values. Highest relative δ<sup>13</sup>C values in fractions M and L suggest that the relative contribution of the rTCA cycle to fixed carbon was highest in large symbionts.

In addition to this, we observed that δ<sup>13</sup>C values were notably more positive in all S-depleted samples, compared to their S-rich counterparts. This strongly indicates that the rTCA cycle, which is presumably more energy-efficient than the Calvin cycle, is relatively more used in S-depleted (i.e. energy-depleted) symbionts of all sizes (see Appendix section B above).

### The carbon fixation key enzyme RubisCO is more abundant in large Endoriftia

The Calvin cycle key enzyme RubisCO was detected with notably higher mRNA-based fluorescence intensities in large Endoriftia cells, compared to smaller symbionts. This is in agreement with our proteomic results (see main text), and supports the conclusion that large symbionts are more involved in carbon fixation and, generally, in biomass production, than small symbionts



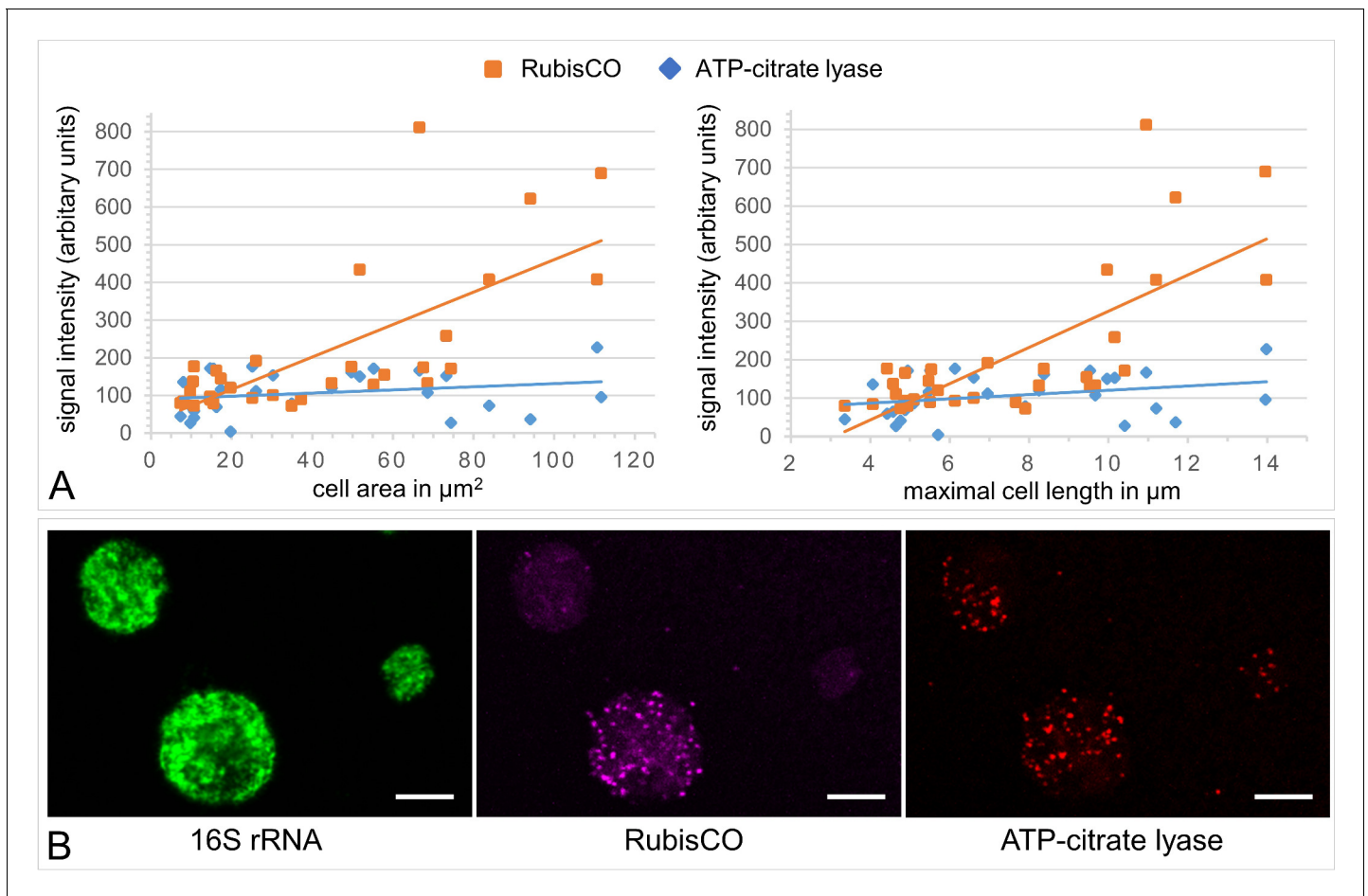


**Appendix 1—figure 2.** Protein stable carbon isotope values ( $\delta^{13}\text{C}$  values) of *Riftia* gradient fractions enriched in symbionts of different cell size (XS – L) relative to fraction XS of S-rich trophosome as baseline. Light squares: S-rich symbionts (average values,  $n = 4$ ), dark triangles: S-depleted symbionts ( $n = 3$ ). Error bars indicate standard errors of the mean.

(Appendix 1—figure 3). Potential differences in mRNA fluorescence intensity of the rTCA cycle key enzyme ATP-citrate lyase subunit B (*AclB*) were poorly distinguishable between smaller and larger symbiont cells. However, an *AclB* signal was quite clearly visible in the very same large cells in which RubisCO was detected. This strongly suggests that Calvin cycle and rTCA cycle are expressed simultaneously by the same symbiont cells.

### Expression patterns of TCA cycle enzymes are ambiguous

Like RubisCO, the rTCA cycle key enzyme ATP-citrate lyase small subunit (*AclA*, Sym\_2601634392) was detected with significantly increasing abundance from fraction XS to L in our proteomic analyses, suggesting that carbon fixation plays a relatively more important role in large *Riftia* symbionts than in small symbionts (see main text). However, expression of other (r)TCA cycle enzymes was surprisingly inconsistent, that is, while abundance of some enzymes increased towards fraction L (including the key enzymes *AclA* and *KorAB*), other enzymes showed the opposite, albeit non-significant, trend and became less abundant (e.g. *AclB*, isocitrate dehydrogenase *Icd*; **Supplementary file 3d**, **Supplementary file 2**). Further contributing to this ambiguous pattern, the *AclB* mRNA signal was detected with very similar (and very low) abundances in small and large *Endoriftia* cells (see **Appendix 1—figure 3**). A possible explanation for these observations might be that *Endoriftia*'s (r)TCA cycle enzymes can run in either direction, depending on cellular requirements. While certain key reactions of TCA and rTCA cycle have long been considered as irreversible, this seems not always to be the case, as, for instance, reported for citrate synthase, key enzyme of the oxidative TCA cycle, which can also operate in the reverse direction, cleaving citrate (**Mall et al., 2018**). *Endoriftia*'s citrate synthase (although encoded in the genome) was not detected at all on the protein level in this study, allowing for the speculation that *AclAB* might functionally replace citrate synthase in the oxidative version of the TCA cycle by running in reverse, possibly even producing ATP in the process. Assuming that the observed discrepancies in *Endoriftia* (r)TCA cycle enzyme abundance trends are thus indeed caused by flexible changes in the enzymes' operating directions, *Icd* could, for example, produce oxaloacetate (e.g. for glutamate synthesis) and NADH in small symbionts, while in large symbionts, *Icd* might fix  $\text{CO}_2$  by running in the reverse direction. Further studies are required to solve the regulation of the symbiont (r)TCA cycle. The recently described combination of matrix-assisted laser desorption/ionization mass spectrometry and FISH (metaFISH),

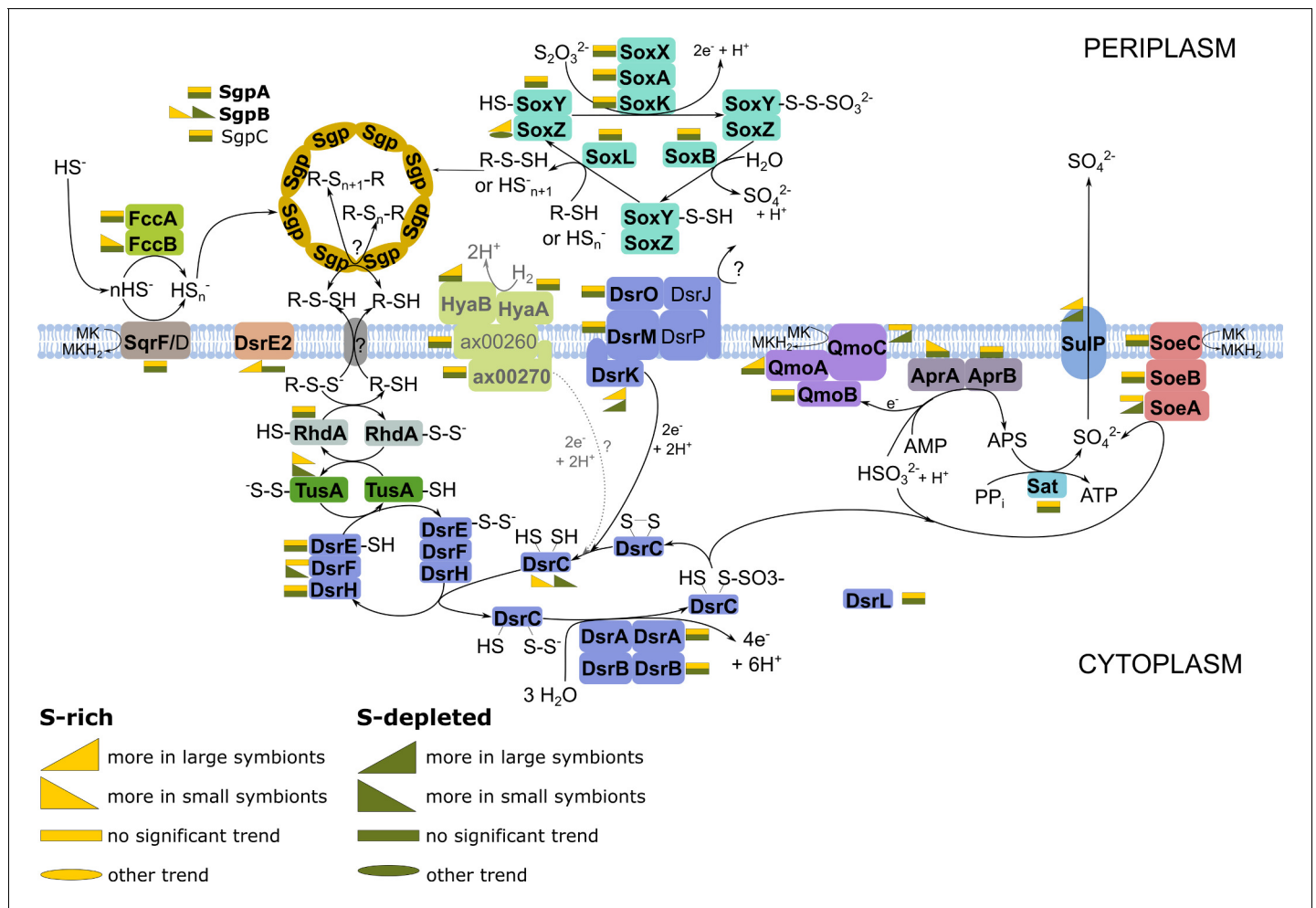


**Appendix 1—figure 3.** A gradient fraction enriched in large symbionts (but also containing small symbiont cells) was fixed as for CARD-FISH analysis and incubated with fluorescently labeled RNA probes against the Endoriftia 16S rRNA and the mRNAs of Calvin cycle key enzyme RubisCO and rTCA cycle key enzyme ATP-citrate lyase (subunit AclB) before examination by confocal laser scanning microscopy (CLSM, see Materials and methods). (A) Background-corrected mean signal intensities per pixel calculated from a total of 33 cells (in eight images on a filter from one biological replicate,  $n = 1$ ) plotted against cell area (left) and Feret's diameter of the cell (right). Straight lines indicate the linear between mean pixel intensities and cell size. Average RubisCO mRNA signal intensity increased notably with cell size (orange lines), while AclB signal intensity increased only very slightly (blue lines). (B) CLSM image of Endoriftia cells (the same cells are visible with different fluorophores in the three panels). Supporting the quantitation in (A) and in line with our proteomic results, the RubisCO signal is markedly more intense in large symbiont cells than in small cells, while the AclB signal is very weak and signal intensity differences between large and small cells seem to be minor. Scale bar = 5  $\mu\text{m}$ . Image brightness and contrast were manually adjusted. Hybridizations of several filters from one biological replicate ( $n = 1$ ) without probes but with fluorophore-carrying hairpins, and without probes or hairpins were used as negative controls and produced no fluorescence signals (images not shown).

which allows for discrimination of symbiont subpopulations based on the metabolites they produce (Geier et al., 2020), might be a promising tool for this purpose.

### Large symbionts may take up organic compounds in addition to $\text{CO}_2$

Our detection of five *Riftia* symbiont TRAP transporter subunits and four ABC transporter components putatively involved in uptake of organic material with increasing relative abundance from fraction XS to L indicates that Endoriftia imports small organic compounds, particularly in the late stage of differentiation, that is, in large cells. ABC transporters can mediate uptake of small molecules (such as sugars, amino acids, or vitamins), and metal ions (Davidson et al., 2008), while TRAP transporters facilitate import of C4-dicarboxylates like fumarate, succinate, and malate (Dct type; Mulligan et al., 2011) or of amino acids like glutamate and glutamine (TAXI type; Mulligan et al., 2007). All these compounds may be relevant heterotrophic substrates in large Endoriftia, which could channel amino acids and peptides into protein biosynthesis, while sugars could be stored as



**Appendix 1—figure 4.** Energy-generating oxidation of reduced sulfur compounds in Endoriftia. Proteins in bold were detected in this study. (Figure adapted from Grein et al., 2010; Markert et al., 2011; Rodriguez et al., 2011; Stewart et al., 2011; Dahl et al., 2013; Stockdreher et al., 2014; Weissgerber et al., 2014). As the role of hydrogen as electron donor in the Riftia symbioses was recently questioned (Mitchell et al., 2019), the associated reactions are labeled in gray and with a question mark.

glycogen. Heterotrophy in thiotrophic symbionts was previously shown for a ciliate symbiont (Seah et al., 2019) and for ectosymbionts of shrimp (Ponsard et al., 2013). Although the Riftia symbiont’s potential for mixotrophy, that is, for both autotrophy and heterotrophy, had been predicted from the symbiont’s genome, it was previously assumed that heterotrophy might be particularly relevant in free-living Endoriftia, but not during symbiosis (Robidart et al., 2008). Our results challenge this assumption and suggest that Endoriftia relies on mixotrophy even when in symbiosis, which would allow re-cycling of carbon from host to symbiont.

### E. Small Endoriftia might be nitrogen-limited

Small Endoriftia may rely relatively more on the glutamine synthetase-glutamate synthase (GS-GOGAT) pathway for ammonia assimilation, while large symbionts cells seem to preferably use glutamate dehydrogenase (GDH) for this purpose. In both S-rich and S-depleted samples, a glutamine synthetase copy (GlnA), glutamate synthase subunit GltB and nitrogen regulatory protein P-II (GlnB) were detected with decreasing abundance from fraction XS to L (Figure 3 main text, Supplementary file 3f). In contrast, glutamate dehydrogenase (GdhA) showed the opposite trend with lowest abundance in XS and highest abundance in L (S-rich) or M (S-depleted). The GS-GOGAT pathway, which is energetically more expensive than the GDH-pathway, was shown to be used under energy-rich conditions or during nitrogen limitation in E. coli (reviewed in Reitzer, 2003). GS-

GOGAT was furthermore shown to have a higher affinity towards ammonium than GDH (**Reitzer, 2003**). This suggests that small symbionts could be nitrogen-limited, either due to a concentration gradient (with highest nitrogen levels in the peripheral lobule zones), and/or due to their own high demand for nitrogen compounds for growth. Further investigations are required to evaluate this speculation.

### F. Sulfur metabolism

While many of the energy-generating reactions of the uncultured *Riftia* symbiont's sulfur metabolism have been elucidated previously (**Markert et al., 2011**), several details remained vague. Our new proteome data enabled us to propose a more detailed model of the *Endoriftia* sulfur metabolism (**Appendix 1—figure 4, Supplementary file 7**).

**Appendix 1—table 1.** Nucleotide sequences used for Hybridization chain reaction (HCR) FISH analyses in this study.

See Materials and methods for details.

>Endoriftia_RubisCO-1-3	CAACGGGGTAGGCGATCTTCATCAGCTCTTTGGCTTCATCGATCTCATAG
>Endoriftia_RubisCO-1-6	CACATATCCTGGATGTTGACCGCAGGACCGTCGTACAGGCGCAGGTATTT
>Endoriftia_RubisCO-1-11	ATGCACCCAGCAGACGGGTCATCTTGATGTGTACGAAAGCGGTGTAACCA
>Endoriftia_RubisCO-1-14	AAAGGACTCGAAGGCGCGAGCGAACTCTTTGTGCTCTTCGCGTACTCGA
>Endoriftia_AclB-1-1	AGACGGCGGTAGATAGACCACCGCCACATTGAATTCACAGCCGTCGTCAA
>Endoriftia_AclB-1-6	CGAACTTCTCCATGAACCACTCTTCTTGGCGACGGCATTGTACCCGGAA
>Endoriftia_AclB-1-8	CTGGTGTGCGTGGATCTTCGATGCCTGCTTTCTTGAACAGCTCCATCAT
>Endoriftia_AclB-1-12	GTGGGCGAAACCGGTATGGGTGAGGAAACCGATATAGCCTTTGTTGACCT
>Endoriftia_AclB-1-18	AAACAGGAACGTGGTGAAGGCGGCAGATTCCATGGTCGCGTCGCTGATCT
>Endoriftia_16srRNA-1	TATTAGCTCGGATTTCTCCGAGTTGTCCCCACTACTGGGCAGATTCCTA
>Endoriftia_16srRNA-5	ACGGAGTTAGCCGGTGCTTCTTCTAAAGGTAACGTCAAGACCCAAGGGTA
>Endoriftia_16srRNA-9	TTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGTACCCACGCTT
>Endoriftia_16srRNA-13	TCGGCTCCCGAAGGCACCAATCTATCTCTAGAAAGTTCCGAGGATGTCAA
>Endoriftia_16srRNA-14	GTTCCCCTAGGGCTACCTGTTACGACTTCACCCAGTCATGAATCACAA

**Appendix 1—table 2.** Overview of symbiont protein identification numbers in all sample types in this study, that is in gradient fractions XS - L and in non-enriched trophosome homogenate (Hom). ID count: number of identified proteins. Numbers are based on four biological replicates for sulfur-rich samples and three biological replicates for sulfur-depleted samples. Note that not all proteins were included in statistical analyses (StAn; see Materials and methods for details). GF: gradient fractions.

	sulfur-rich trophosome					sulfur-depleted trophosome					total
	Hom	XS	S	M	L	Hom	XS	S	M	L	
ID count	1151	1022	1296	1603	1722	1017	1099	1260	1605	1572	1946
ID count (Hom only)	1151					1017					1223
ID count (total all GF)			1821					1727			1898
ID count (total all sample types)			1867					1773			1946
Proteins in StAn		940	1081	1135	1134		1008	1091	1150	1143	1212
Proteins in StAn (total all GF)			1135					1151			1212

DsrC: The Endoriftia genome encodes several copies of DsrC family proteins, four of which were detected as proteins in this study (**Supplementary file 2**). One of them, Sym\_EGV52266.1, was one of the most abundant symbiont proteins, pointing to considerable physiological importance of this protein. Similar to the situation in Endoriftia, three putative DsrC copies were found in the *Calyptogenia okutanii* symbiont (**Harada et al., 2009**), and DsrC was also the single most abundant sulfur metabolism mRNA in the *Solemya velum* symbiont (**Stewart et al., 2011**). DsrC has been described to fulfill a key role in dissimilatory sulfur metabolism, including a putative function in transcription regulation and a function as a sulfur trap to allow for maximum DsrAB efficiency (**Venceslau et al., 2014**). Considering this role of DsrC as enhancer of sulfide oxidation efficiency, highest abundance of all Endoriftia DsrC copies in fraction XS (and lowest DsrC abundance in fraction M or L), corroborates our hypothesis of relatively more H<sub>2</sub>S oxidation for energy generation in small *Riftia* symbionts (see main text).

SoeABC: In addition to AprAB and SopT, two of the key enzymes of cytoplasmic sulfide oxidation, we also found SoeABC to be expressed in Endoriftia. In *Allochromatium vinosum*, SoeABC catalyzes direct oxidation of sulfite to sulfur, independently of AMP (**Dahl et al., 2013**).

SreABC: We found the putatively sulfur oxidation-related proteins SreABC in the metagenome and detected SreA on the protein level in Endoriftia. While the exact function of SreABC in the oxidation of reduced sulfur compounds is unclear, for *A. vinosum* it was speculated that the Sre proteins could oxidize polysulfides, which are intermediates generated during sulfide oxidation to sulfur (**Weissgerber et al., 2013**).

HyaAB: Endoriftia's uptake hydrogenase HyaAB might be involved in sulfur oxidation. In *A. vinosum*, concentration of the Lsp-type hydrogenase HydLS was shown to increase substantially in the presence of sulfide (**Weissgerber et al., 2014**), leading to the proposition that hydrogen-derived electrons may be fed into sulfide oxidation via hydrogenase as illustrated in **Appendix 1—figure 4**. *A. vinosum*'s HydL (Alvin\_2036) and Endoriftia's HyaB (EGV51840.1) protein sequences are 75.69% identical (NCBI BlastP), indicating that both may have similar functions in sulfur oxidation.

## G. Considerations on the blood flow direction in *Riftia* trophosome lobules

Blood flow across *Riftia* trophosome lobules is commonly accepted to proceed from the lobule periphery inwards to the lobule center, based on observations of **van der Land and Nørrevang, 1977**, **Felbeck and Turner, 1995** and **Bright and Sorgo, 2003** (see also **Figure 6** in the main text). This direction would facilitate a potential substrate gradient with highest concentrations of CO<sub>2</sub>, N<sub>2</sub>, O<sub>2</sub>, reduced sulfur compounds and other substances at the lobule periphery and lowest concentrations in the lobule center. Such a periphery-to-center gradient is in agreement with and corroborates many of our findings. For example, (1) lower oxygen concentrations in the lobule center accord with increasing abundances of all four cytochrome c oxidase subunits (Cco: Sym\_EGV51283.1, Sym\_EGV51282.1, Sym\_2601635240, Sym\_EGV51284.1; see **Supplementary file 2**) from fraction XS to M or L. Lowest oxygen levels around the smallest symbionts would furthermore explain highest levels of nitrate-reducing NarGHI in fraction XS. (2) Highest CO<sub>2</sub> levels at the lobule periphery support higher abundances of CO<sub>2</sub> fixation enzymes in large symbionts, and (3) highest abundances of thiosulfate oxidation proteins in large symbionts may be related to more host sulfide oxidation at the lobule periphery, where sulfide concentrations are highest. Providing highest raw substrate concentrations to the large symbionts would support them in their role as major biomass producers in the symbiosis. Moreover, (4) nutrients released during symbiont digestion in the lobule periphery would likely enter the blood stream and subsequently be transported inwards. Decreasing relative abundances of TRAP-type transporters from fraction L to XS (see section D above) suggest that these compounds are imported by the symbionts along the way, with largest Endoriftia benefiting from highest nutrient concentrations.

*Riftia*'s blood flow and ensuing putative substrate gradients thus go hand in hand with symbiont differentiation. Quite possibly, they might even be involved in triggering this differentiation.