

1 Metabolic differences between symbiont
2 subpopulations in the deep-sea tubeworm *Riftia*
3 *pachyptila*
4

5
6
7 Tjorven Hinzke^{1,2,3}, Manuel Kleiner⁴, Mareike Meister^{5,6}, Rabea Schlüter⁷, Christian
8 Hentschker⁸, Jan Pané-Farré⁹, Petra Hildebrandt⁸, Horst Felbeck¹⁰, Stefan M. Sievert¹¹, Florian
9 Bonn¹², Uwe Völker⁸, Dörte Becher⁵, Thomas Schweder^{1,2}, Stephanie Markert^{1,2*}

10 1- Institute of Pharmacy, University of Greifswald, Germany

11 2- Institute of Marine Biotechnology, Greifswald, Germany

12 3- Energy Bioengineering Group, University of Calgary, Calgary, Canada

13 4- Department of Plant and Microbial Biology, North Carolina State University, NC, USA

14 5- Institute of Microbiology, University of Greifswald, Germany

15 6- Leibniz Institute for Plasma Science and Technology, Greifswald, Germany

16 7- Imaging Center of the Department of Biology, University of Greifswald, Germany

17 8- Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Germany

18 9- Center for Synthetic Microbiology (SYNMIKRO), Philipps-University Marburg, Germany

19 10- Scripps Institution of Oceanography, University of California San Diego, CA, USA

20 11- Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, USA

21 12- Institute of Biochemistry, University Hospital, Goethe University School of Medicine Frankfurt, Germany

22
23 Key words: thiotrophic symbiosis, cell heterogeneity, metaproteomics, microbe-host
24 interactions, cell differentiation, chemosynthetic symbiosis
25
26

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

28 Abstract

29 The hydrothermal vent tube worm *Riftia pachytila* lives in intimate symbiosis with
30 intracellular sulfur-oxidizing gammaproteobacteria. Although the symbiont population
31 consists of a single 16S rRNA phylotype, bacteria in the same host animal exhibit a remarkable
32 degree of metabolic diversity: They simultaneously utilize two carbon fixation pathways and
33 various energy sources and electron acceptors. Whether these multiple metabolic routes are
34 employed in the same symbiont cells, or rather in distinct symbiont subpopulations, was
35 unclear. As *Riftia* symbionts vary considerably in cell size and shape, we enriched individual
36 symbiont cell sizes by density gradient centrifugation in order to test whether symbiont cells
37 of different sizes show different metabolic profiles. Metaproteomic analysis and statistical
38 evaluation using clustering and random forests, supported by microscopy and flow cytometry,
39 strongly suggest that *Riftia* symbiont cells of different sizes represent metabolically dissimilar
40 stages of a physiological differentiation process: Small symbionts actively divide and may
41 establish cellular symbiont-host interaction, as indicated by highest abundance of the cell
42 division key protein FtsZ and highly abundant chaperones and porins in this initial phase.
43 Large symbionts, on the other hand, apparently do not divide, but still replicate DNA, leading
44 to DNA endoreduplication. Highest abundance of enzymes for CO₂ fixation, carbon storage
45 and biosynthesis in large symbionts indicates that in this late differentiation stage the
46 symbiont's metabolism is efficiently geared towards the production of organic material. We
47 propose that this division of labor between smaller and larger symbionts benefits the
48 productivity of the symbiosis as a whole.

49

50

51 Introduction

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

58 Although the symbiont population consists of a single 16S rRNA phylotype (Polzin *et al.*, 2019),
59 it was previously shown to exhibit remarkable metabolic versatility: As demonstrated by
60 proteomic analyses, symbionts from the same host animal expressed enzymes of two CO₂
61 fixation pathways, the Calvin cycle and the reverse tricarboxylic acid (rTCA) cycle, as well as
62 enzymes for both, glycogen generation and glycogen degradation (Markert *et al.*, 2007,
63 Markert *et al.*, 2011, Gardebrecht *et al.*, 2012, Hinzke *et al.*, 2019). Moreover, proteins involved
64 in utilization of hydrogen sulfide and thiosulfate as energy sources were expressed
65 simultaneously by the same symbiont population; as were proteins for the use of nitrate and
66 oxygen as electron acceptors (Markert *et al.*, 2011). Based on these observations, we
67 hypothesized that individual, metabolically distinct symbiont subpopulations in the
68 trophosome may exist.

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

78 cycle (Bright *et al.*, 2000, Bright and Sorgo, 2003). Previous microscopy-based studies
79 indicated that small and large *Riftia* symbionts differ not only with regard to their frequency
80 of cell division, but also with regard to carbon incorporation rates, amount of stored glycogen,
81 and area of sulfur storage vesicles (Bright *et al.*, 2000, Sorgo *et al.*, 2002, Bright and Sorgo,
82 2003, Pflugfelder *et al.*, 2005). This suggests that individual cell sizes may indeed have
83 dissimilar metabolic properties.

84 In this study, we aimed to analyze and compare the metabolic profiles of individual *Riftia*
85 symbiont subpopulations. In contrast to previous molecular analyses that studied metabolic
86 capabilities of the *Riftia* symbiont population as a mixture of all cell sizes (e.g., Markert *et al.*,
87 2007, Markert *et al.*, 2011, Gardebrecht *et al.*, 2012), precluding comparisons between putative
88 subpopulations, we used a more sensitive approach: We enriched Endoriftia cells of different
89 sizes by gradient centrifugation of trophosome tissue homogenate and subjected these
90 enriched gradient fractions to separate metaproteomic analyses. Statistical evaluation using
91 clustering and random forests allowed us to deduce cell size-dependent differences in protein
92 abundance and metabolic functions. Catalyzed reporter deposition-fluorescence *in situ*
93 hybridization (CARD-FISH), transmission electron microscopy (TEM), hybridization chain
94 reaction (HCR)-FISH analyses, and flow cytometry complemented these experiments. Our
95 results suggest a division of labor between different developmental stages of the symbiont.

96

97

98 Material and Methods

99 Sample collection and enrichment of symbiont subpopulations

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

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

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

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

131 CARD-FISH

132 Enrichment of symbiont cell sizes in gradient fractions was analyzed employing fluorescence
133 microscopy with samples labelled by CARD-FISH. CARD-FISH labeling was performed as
134 previously described (Ponnudurai *et al.*, 2017), using the probe Rif445 (Nussbaumer *et al.*,
135 2006) and Alexa Fluor® 594-labeled tyramide. For counterstaining, 0.1% (w/v) 4,6-diamidino-
136 2-phenylindole (DAPI) was added to the embedding medium (4:1 Citifluor AF1 (Citifluor) and
137 Vectashield (Vector Laboratories)). CARD-FISH filters were analyzed using an Axio
138 Imager.M2 fluorescence microscope (Carl Zeiss Microscopy GmbH). For semi-automated cell
139 counting and to measure the longest cell dimension, we used a custom Fiji (Schindelin *et al.*,
140 2012) macro with the Fiji plugins Enhanced Local Contrast (CLAHE; Saalfeld, 2010) and Bi-
141 exponential edge preserving smoother (BEEPS; Thévenaz *et al.*, 2012). After image processing,
142 we excluded objects with a size of less than 2 μ m (as these were mainly artifacts) and set the
143 maximum object size to 20 μ m. To assign cell sizes to size classes (i.e., cell size ranges) we used
144 a quartile split: We calculated quartiles of cell sizes in non-enriched homogenate samples (i.e.,
145 25% of all cells in homogenate samples were assigned to each class). This resulted in the four
146 calculative size classes very small ($\geq 2 \mu\text{m} - < 3.912 \mu\text{m}$), small ($\geq 3.912 \mu\text{m} - < 5.314 \mu\text{m}$),
147 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
148 in all size classes were coccoid. Rod-shaped cells were almost exclusively present in the
149 smallest size class. Individual gradient fractions (subsamples) were screened for their
150 respective share of cells in each size class and the subsample with the highest percentage of
151 cells in the respective quartile was chosen for metaproteomic analysis. For example, if of all 24
152 subsamples of a sample, fraction 5 had the highest percentage of very small cells, i.e. most of

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

163 Transmission electron microscopy (TEM)

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

177 HCR-FISH and confocal laser scanning microscopy (CLSM)

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

180 above was used for hybridization chain reaction FISH (HCR-FISH) according to Choi *et al.*
181 (2014). We used a HCR-FISH v2.0 Custom Kit (Molecular Technologies) according to the
182 manufacturer's instructions. Probes targeted the *Riftia* symbiont's 16S rRNA (fluorescence
183 marker: Alexa Fluor[®] 488), and the mRNAs of ATP-citrate lyase subunit AclB (Alexa Fluor[®]
184 647) and RubisCO (Alexa Fluor[®] 594; see Supplementary Table S2 for the probe sequences).
185 In brief, filter sections were washed twice with 50% hybridization buffer (50% formamide, 5x
186 sodium chloride sodium citrate (SCSC, 0.75 M NaCl, 75 mM Na₃C₆H₅O₇), 9 mM citric acid, pH
187 6.0, 0.1% Tween 20, 50 µg/ml heparin, 1x Denhardt's solution, 10% dextran sulfate) in 2x
188 mPBS (89.8 mM Na₂HPO₄, 10.2 mM NaH₂PO₄, 0.9 M NaCl) at 45°C for 30 min for pre-
189 hybridization, and incubated overnight (16 h, 45°C) with probe solution (1 pmol of each probe
190 in 500 µl hybridization buffer). Excess probes were removed with several washing steps in 75-
191 25% probe wash buffer (50% formamide, 5x SCSC, 9 mM citric acid, pH 6.0, 0.1% Tween 20,
192 50 µg/ml heparin in 5x SCSC) for 15 min at 45°C, 300 rpm, and subsequently in 5x SCSC for
193 30 min at 45°C and 300 rpm. Samples were pre-amplified with DNA amplification buffer (5x
194 SCSC, 0.1% Tween 20, 10% dextran sulfate). Hairpins were activated by snap-cooling and
195 added to the samples. After overnight incubation (16 h, room temperature) with the hairpin
196 solution, samples were washed with 5x SCSC, containing 0.05% Tween 20 (room temperature,
197 300 rpm, four times 5 min, two times 30 min) and embedded in Mowiol 4-88 (Carl Roth
198 GmbH) embedding medium prepared according to the manufacturer's instructions. Confocal
199 microscopy was performed on a Zeiss LSM510 meta equipped with a 100x/1.3 oil immersion
200 objective. Probes were excited with laser lines 633 (ATP-citrate lyase), 561 (RubisCO) and 488
201 (16S rRNA) and signals were detected with filters suitable for dye maximal emissions at 670
202 nm, 595 nm and 527.5 nm, respectively. Signal intensities and cell sizes (from 8 frames
203 showing a total of 33 cells) were quantified using the Fiji software package (Schindelin *et al.*,
204 2012). Individual cells were defined as regions of interest (ROI), in which signal intensity per
205 pixel was recorded. Mean pixel intensity of ROI was calculated and background was corrected.
206 Global background values were calculated for every channel based on up to six ROIs randomly
207 placed in each image frame. The following cell size parameters were calculated: (i) Feret's

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

210 Flow cytometry

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

234 Peptide sample preparation

235 Proteins were extracted as described in Hinzke and Markert (2017). Briefly, cells were mixed
236 with lysis buffer (1% (w/v) sodium deoxycholate (SDC), 4% (w/v) sodium dodecyl sulfate (SDS)
237 in 50 mM triethylammonium bicarbonate buffer (TEAB)), heated for 5 min at 95 °C and 600
238 rpm and cooled on ice. Samples were then placed in an ultrasonic bath for 5 min and
239 subsequently cooled on ice. Cell debris was removed by centrifugation (14,000 x g, 10 min,
240 room temperature). Protein concentration was determined using the Pierce BCA assay
241 according to the manufacturer's instructions. Peptides were generated using a 1D gel-based
242 approach as in Ponnudurai *et al.* (2017) with minor modifications. In brief, 20 µg of protein
243 sample was mixed with Laemmli sample buffer containing DTT (final concentration 2% (w/v)
244 SDS, 10% glycerol, 12.5 mM DTT, 0.001% (w/v) bromophenol blue in 0.06 M Tris-HCl;
245 Laemmli, 1970) and separated using pre-cast 4-20% polyacrylamide gels (BioRad). After
246 staining, protein lanes were cut into 10 pieces, destained (600 rpm, 37 °C, 200 mM NH₄HCO₃
247 in 30% acetonitrile) and digested with trypsin (sequencing grade, Promega) overnight at 37 °C,
248 before peptides were eluted in an ultrasonic bath. Peptides were then directly used for LC-MS
249 analysis.

250 LC-MS/MS analysis

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

257 Protein identification and function prediction

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

261 *pachyptila* (vent Pho5)), NCBI project PRJNA60887 (endosymbiont of *Tevnia jerichonana*
262 (vent Tica)), and JGI IMG Gold Project Gp0016331 (endosymbiont of *Riftia pachyptila* (vent
263 Mk28)). The cRAP database containing common laboratory contaminants (The Global
264 Proteome Machine Organization) was added to complete the database. Database search was
265 conducted using Proteome Discoverer v. 2.0.0.802 with the Sequest HT node as described in
266 Kleiner *et al.* (2018) with a false discovery rate of 5% (FidoCT Protein Validator node, q-value
267 <0.05).

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

273 [Statistical evaluation of metaproteomics data and abundance quantification](#)

274 [Filtering and normalization](#)

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

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

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

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

305 *STEM analysis*

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

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

320 *Random forests*

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

338 *Significant differences in protein abundance*

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

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

351 *Host proteins*

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

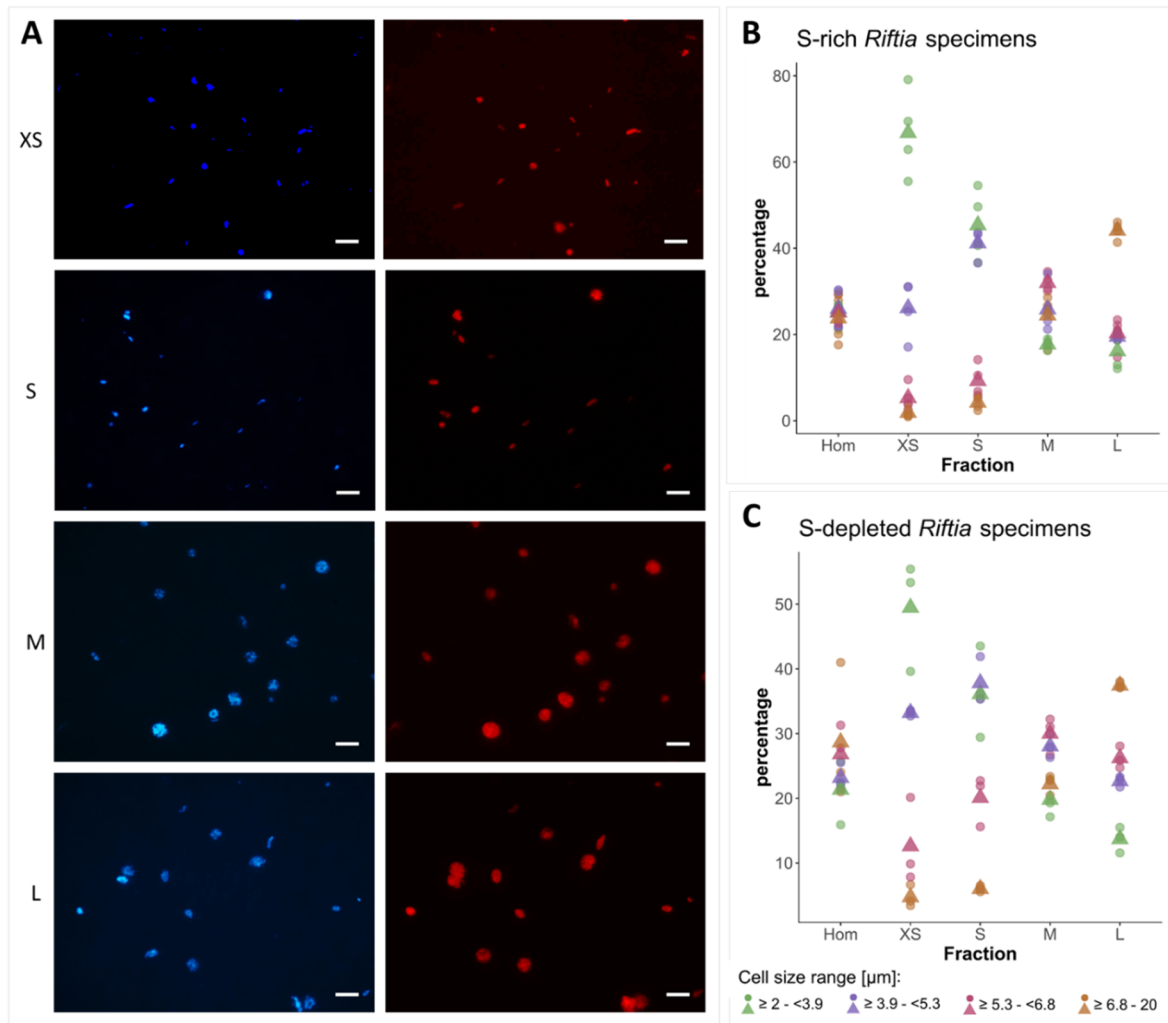
367 *Data availability*

368 The mass spectrometry proteomics data have been deposited to the ProteomeXchange
369 Consortium (ProteomeXchange – ProteomeCentral) via the PRIDE partner repository
370 (Vizcaíno *et al.*, 2016) with the dataset identifier PXDO16986. The dataset will be released upon
371 acceptance of the manuscript in a peer-reviewed journal.

372 Results

373 Enrichment of individual symbiont cell sizes by gradient centrifugation

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



384

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

397

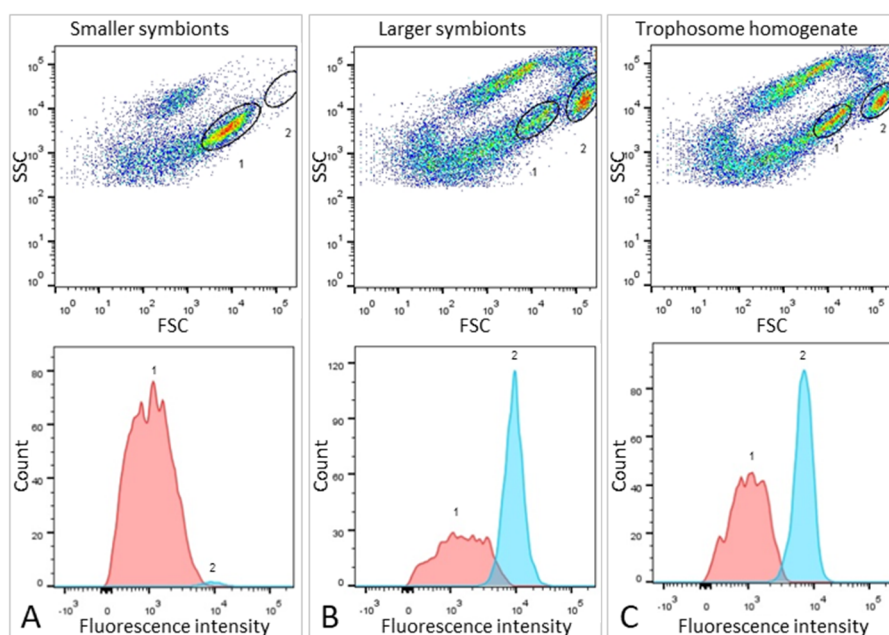
398 Symbiont DNA quantification

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

402 trophosome homogenate and in gradient fractions from the upper and lower parts of the
403 gradient (enriched in smaller and larger symbionts, respectively) with regard to their light
404 scattering properties. Forward scatter (FSC) and side scatter (SSC) usually correlate with cell
405 size and cell granularity, respectively (Bouvier *et al.*, 2001, Tracy *et al.*, 2010). Amongst a
406 number of particle groups with different properties (see also Supplementary Results and
407 Discussion C), we found two populations, 1 and 2, which were abundantly detected in non-
408 enriched trophosome homogenate, but showed very dissimilar frequencies in fractions
409 enriched in larger or smaller symbionts (Figure 2, Supplementary Figure S2): While
410 population 1, which exhibited relatively lower FSC and SSC signals (indicative of smaller cell
411 size and lower cell complexity), was highly abundant in fractions enriched in smaller
412 symbionts, this population was notably less prominent in fractions enriched in larger
413 symbionts. Simultaneously, population 2, which gave higher FSC and SSC signals (indicative
414 of larger cell size and higher complexity), was highly abundant in fractions enriched in larger
415 symbionts but nearly absent in gradient fractions enriched in smaller symbionts. This suggests
416 that populations 1 and 2 consist of smaller and larger symbionts, respectively. This assumption
417 was verified by FACS-separation of both populations from trophosome homogenate, and
418 examination of the sorted cell suspensions by fluorescence microscopy along with unsorted
419 enriched gradient fractions and homogenate samples for reference (Supplementary Figure S2).
420 For quantification of DNA in smaller and larger symbionts, we compared median fluorescence
421 intensities (MFI) per particle between population 1 and 2 in non-enriched homogenate and in
422 enriched gradient fractions. In all sample types, MFI per particle was notably lower in
423 population 1 (between 186 and 1,994 relative fluorescence units, rfu) than in population 2
424 (2,712 – 10,723 rfu). On average, MFI was 9.7-fold higher in population 2 than in population 1
425 (Supplementary Table S8).

426

427



428

429 **Figure 2:** Flow cytometry for DNA quantification of *Riftia* symbionts. **A)** Dot plot of forward scatter (FSC) and side
430 scatter (SSC), and histogram with fluorescence signal counts and fluorescence intensity per particle of a gradient
431 fraction enriched in smaller symbionts. **B)** Gradient fraction enriched in larger symbionts. While cell population
432 1 was more prominent in A), population 2 was almost exclusively detected in B), and both populations were
433 present in non-enriched trophosome homogenate (C), indicating that population 1 corresponds to smaller
434 symbionts, whereas population 2 corresponds to larger symbiont cells. Cells were stained with Syto9 and median
435 fluorescence intensity (MFI) per particle at wave length 530/30 nm was used as a measure of cellular DNA
436 content (see Methods and Supplementary Table S8 for more details). This analysis was based on two *Riftia*
437 specimens with medium sulfur content.

438 Protein identifications and relative protein abundance

439 We identified a total of 1,946 symbiont proteins across all sample types, including the four
440 gradient fractions XS – L and non-enriched homogenate from both, sulfur-rich and sulfur-
441 depleted *Riftia* specimens (Supplementary Table S6). Our sample fractionation by gradient
442 centrifugation thus facilitated detection of around 60% of the symbiont's theoretical proteome,
443 which encompasses 3,182 proteins in PRJNA60889, and yielded substantially higher symbiont
444 protein identification rates than non-enriched trophosome homogenate samples alone (1,223
445 total symbiont protein identifications). After stringent filtering and normalization, a subset of
446 1,212 symbiont proteins from gradient fractions XS – L was included in statistical analysis
447 using abundance profile clustering and random forests (Supplementary Table S3). A total of
448 465 proteins showed significant differences in relative abundance in S-rich and/or S-depleted
449 samples (Supplementary Figure S1; note that the term “significant” denominates trends that
450 were consistent across all replicates in the context of our statistical approach). In Figure 3 and

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

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

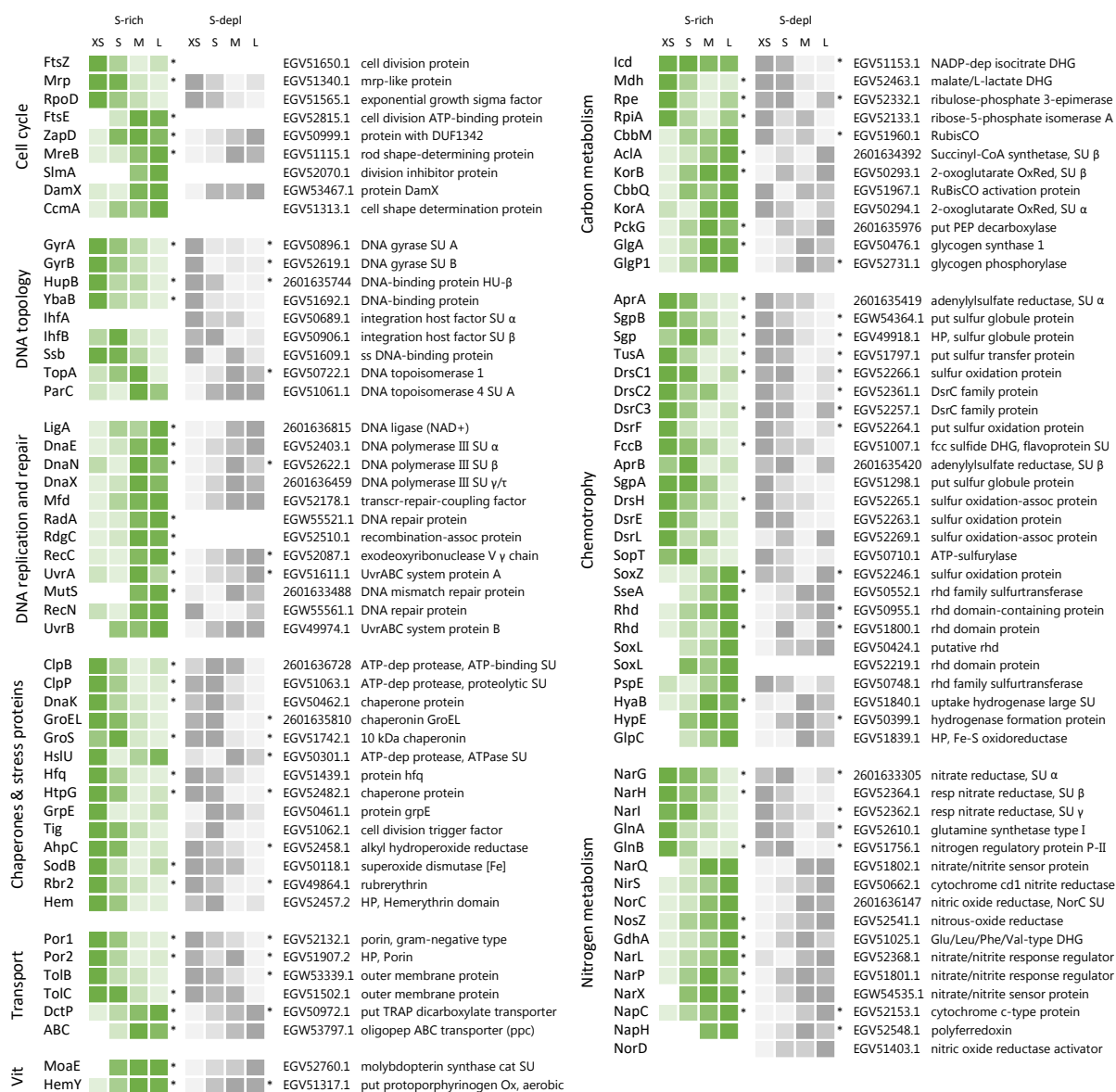
462 Symbiont protein functions

463 *Cell cycle, DNA topology, replication and repair*

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

472 *Chaperones and stress proteins*

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



478

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

494 *Transport*

495 Outer membrane proteins such as two porins and TolBC showed significant abundance
496 differences between the fractions, with highest relative abundance in fraction XS and lowest
497 abundance in fractions L. Porin EGV52132.1 (Por1) was the most abundant symbiont protein
498 throughout all sample types (Figure 3, Supplementary Table S4c). On the other hand, all five
499 detected tripartite ATP-independent periplasmic (TRAP) transporter subunits and ten out of
500 13 ABC transporter components were relatively more abundant in fraction L (see also
501 Supplementary Results and Discussion D).

502 *Central metabolism*

503 *Carbon metabolism:* Several tricarboxylic acid (TCA) cycle enzymes (e.g., Icd, Mdh),
504 as well as enzymes of the pentose phosphate pathway (e.g., Rpe, RpiA) were detected with
505 decreasing abundances from fraction XS to L (Figure 3, Supplementary Table S4d). In contrast,
506 the key enzymes of the two CO₂-fixing pathways, Calvin cycle (RubisCO, CbbM) and rTCA cycle
507 (ATP-citrate lyase, AclA; oxoglutarate oxidoreductase, KorAB), as well as most of the
508 gluconeogenesis-related (e.g., PckG), and glycogen metabolism-related enzymes (e.g., GlgA,
509 GlgP) increased in abundance from fraction XS to L.

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

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

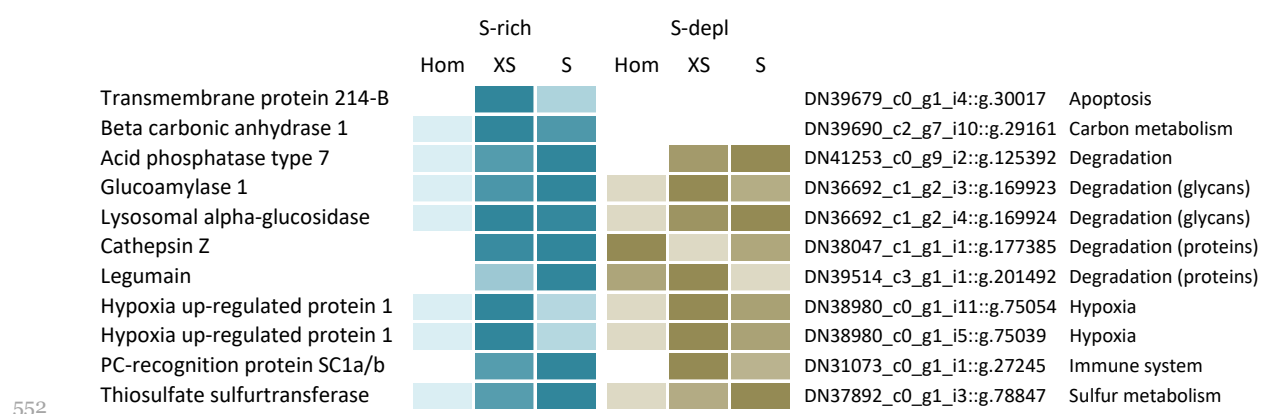
534 *Other categories*

535 50 (75%) of the 67 proteins involved in cofactor- and vitamin synthesis in S-rich samples had
536 their highest abundance in fraction M or L (Figure 3, Supplementary Table S4g). Also, of the
537 33 identified tRNA ligases and tRNA synthetases, 25 (75%) were most abundant in fraction M
538 or L (in S-rich samples, Supplementary Table S4h).

539 **Symbiosis-specific host proteins**

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

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



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

564

565

566

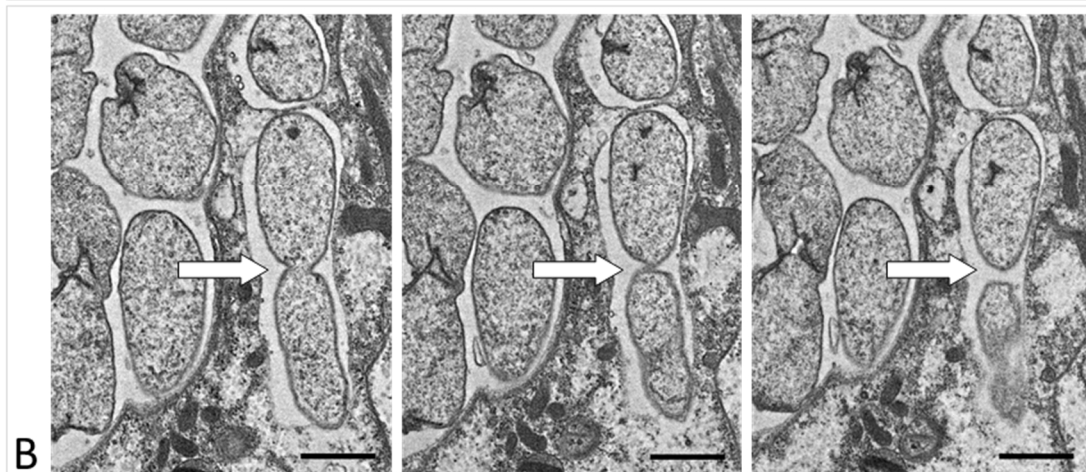
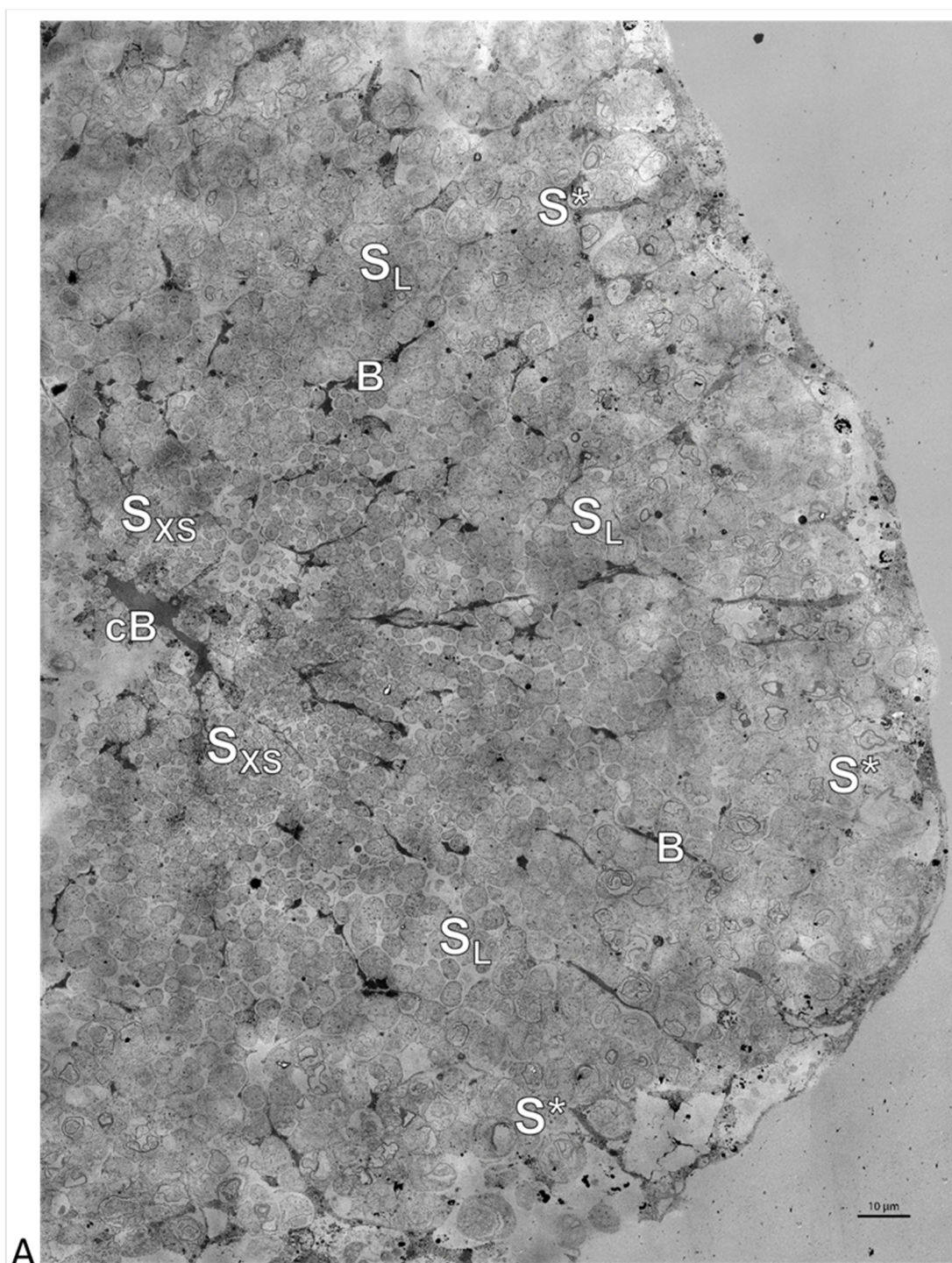
567 Discussion

568 Symbiont growth and differentiation

569 *Cell division plays a more prominent role in small symbionts*

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

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



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

613

614 *Large symbionts have more genome copies and less compact chromosomes*

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

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

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

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

664 be more prone to various kinds of damage and require the enhanced expression of DNA repair
665 mechanisms. This would explain the observed higher abundance of several DNA repair
666 proteins in fraction L, which was enriched in larger, older symbiont cells with (presumably)
667 larger quantities of less condensed DNA, compared to the smaller symbiont cells. RadA, RdgC,
668 RecCN, UvrAB, and Mfd, which are known to be involved in DNA recombination and repair in
669 many bacteria (Kowalczykowski, 2000, Beam *et al.*, 2002, Tessmer *et al.*, 2005, Drees *et al.*,
670 2006, Truglio *et al.*, 2006, Deaconescu *et al.*, 2007), may compensate for this elevated
671 vulnerability. In eukaryotes, chromatin decondensation was shown to facilitate access of the
672 DNA damage response to double strand breaks, thus allowing for more efficient repair (Murga
673 *et al.*, 2007).

674 *Small symbionts may be exposed to elevated stress levels*

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

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

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

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

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

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

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

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

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

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

756 *Interaction-specific host proteins*

757 We detected a number of ‘symbiosis-specific’ *Riftia* proteins, which were co-enriched with
758 symbiont cells in fractions XS and/or S and may thus facilitate direct host-microbe interactions
759 or enable the host to provide optimal conditions for the symbiont. PGRPs, for example, are
760 involved in innate immunity (Kang *et al.*, 1998) and have previously been shown or suggested
761 to participate in symbiotic interactions (Troll *et al.*, 2009, Wang *et al.*, 2009, Royet *et al.*, 2011,
762 Wippler *et al.*, 2016). Since oxygen concentrations in the trophosome might be comparatively
763 low (benefitting the microaerophilic symbionts; Hinzke *et al.*, 2019), the hypoxia up-regulated
764 *Riftia* proteins we detected may present a protective adaptation of the host to these hypoxic
765 conditions. In support of this idea, Hyou1 was shown to have a protective function during
766 hypoxia in human cells (Ozawa *et al.*, 1999). Moreover, enrichment of beta carbonic anhydrase
767 1, which interconverts bicarbonate and CO₂, suggests that this host protein serves to optimally
768 provide the symbionts with CO₂ for fixation. The host transmembrane protein 214-B (TMP214-
769 B), which was exclusively detected in symbiont-enriched fractions (but not in trophosome
770 homogenate) may be involved in cell death of symbiont-containing bacteriocytes by an
771 apoptosis-related mechanism. This would be in line with our previous suggestion that
772 apoptosis-related proteins may play a role in symbiont and bacteriocyte cell death (Hinzke *et*
773 *al.*, 2019), and is further supported by the fact that TMP214-B was shown to be involved in
774 apoptosis caused by endoplasmic reticulum stress (Li *et al.*, 2013). The detection of
775 degradation proteins such as cathepsin Z, legumain, glucoamylase 1 and lysosomal alpha-

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

779 Metabolic diversity among symbiont size classes

780 *Large symbionts focus on carbon fixation and biosynthesis*

781 Highest individual abundances of various carbon fixation and biosynthesis-related enzymes as
782 well as highest overall abundances of all biosynthetic categories (including carbon-, amino
783 acid-, lipid-, nitrogen- and cofactor metabolism; Supplementary Table S7) in fraction L
784 suggests that large Endoriftia cells are relatively more engaged in the production of organic
785 material than smaller symbiont cells. In support of this idea, we observed notably higher
786 RubisCO mRNA signal intensity in large symbiont cells than in smaller *Riftia* symbionts in our
787 HCR-FISH analysis (Supplementary Results and Discussion D, Supplementary Figure S3).
788 This concurs with an autoradiographic study of Bright *et al.* (2000), who observed highest ¹⁴C
789 carbon incorporation in the *Riftia* trophosome lobule periphery and lowest short-term
790 incorporation in the lobule center. As previously suggested (Hand, 1987), these and other
791 observed differences might be due to a biochemical gradient, which could be caused by the
792 direction of blood flow (from the lobule periphery to the lobule center; Felbeck and Turner,
793 1995). This presumptive concentration gradient may lead to differential availability of
794 inorganic carbon (and other substrates, see below), which in turn likely results in differential
795 regulation of bacterial gene expression, such as highest abundance of CO₂ incorporation
796 enzymes in large symbionts. Large *Riftia* symbionts thus presumably not only benefit from
797 higher CO₂ levels, but also have more biosynthetic capacities at their disposal than small
798 symbionts: Small Endoriftia need to maintain cell division and, consequently, invest a
799 considerable part of their resources in growth-related processes and the expression of putative
800 host interaction-related proteins that ensure survival of the stem cell population (see above).
801 In contrast, large symbionts apparently divide less frequently and may be less endangered of
802 host interference (before they reach the degenerative lobule zone) and can thus allocate more
803 energy to production of organic material. This would eventually benefit the host, which digests

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

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

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

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

832 *In large symbionts, thiosulfate oxidation plays a more prominent role*

833 Larger symbionts may rely relatively more on thiosulfate oxidation – in addition to sulfide
834 oxidation – than smaller Endoriftia, as suggested by highest abundance of SoxZ and detection
835 of several other (low-abundant) Sox proteins in fraction L. Expression of the Sox (sulfur
836 oxidation) complex was shown to be upregulated in the presence of thiosulfate in *A. vinosum*
837 (Grimm *et al.*, 2011). We speculate that thiosulfate concentrations might be higher in the
838 trophosome lobule periphery than in the lobule center, due to a concentration gradient (as
839 proposed above for CO₂) and/or possibly also as a result of host thiosulfate production. The
840 *Riftia* host appears to be able to oxidize toxic sulfide to the less toxic thiosulfate in its
841 mitochondria (Hinzke *et al.*, 2019). Higher abundance of host thiosulfate sulfurtransferase in
842 symbiont-enriched fractions compared to non-enriched trophosome homogenate in our
843 present study suggests that this putative detoxification process could be particularly important
844 in the symbiont-containing bacteriocytes. With sulfide supposedly reaching the trophosome
845 lobule periphery first with the blood flow, free sulfide concentrations might be higher there
846 and, consequently, host sulfide oxidation to thiosulfate might be more frequent in
847 bacteriocytes at the lobule periphery than in the center. The idea of more thiosulfate oxidation
848 in large Endoriftia is further substantiated by highest abundance of six rhodanese family
849 proteins in fraction L, as rhodanese-like proteins can cleave thiosulfate into sulfite and sulfide
850 and were proposed to be involved in thiosulfate oxidation (Hensen *et al.*, 2006, Welte *et al.*,
851 2009).

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

859 *Hydrogen oxidation is more relevant in large symbionts*

860 In large symbionts, the use of hydrogen may furthermore play a more prominent role than in
861 smaller symbiont cells, as suggested by increasing abundances of the Isp-type respiratory H₂-
862 uptake [NiFe] hydrogenase large subunit HyaB, a Fe-S oxidoreductase (GlpC) encoded next to
863 HyaB, and the hydrogenase expression/formation protein HypE from fraction XS to L. The
864 small hydrogenase subunit HyaA (Sym_EGV51837.1) and an additional hydrogenase
865 expression/formation protein (HoxM, Sym_EGV51835.1), both of which are encoded
866 upstream of HyaB in the symbiont genome, were detected with increasing abundance towards
867 fraction L as well (although at very low concentrations; Supplementary Table S3b), supporting
868 the idea of relatively more hydrogen oxidation in large symbionts. Like for CO₂ and thiosulfate,
869 this might be due to a concentration gradient with highest hydrogen concentrations at the
870 lobule periphery and lowest concentrations towards the lobule center. Use of hydrogen as an
871 energy source has been described or suggested for free-living sulfur oxidizing bacteria like *A.*
872 *vinosum* (Weissgerber et al., 2011), and for a variety of thiotrophic symbionts of marine
873 invertebrates (Petersen et al., 2011). Taking advantage of hydrogen oxidation in addition to
874 sulfide- and thiosulfate oxidation, i.e., using a broader repertoire of electron donors, would
875 potentially enhance the metabolic flexibility, particularly of large Endoriftia. However, H₂ was
876 recently suggested to be involved in maintaining intracellular redox homeostasis rather than
877 working as electron donor in the *Riftia* symbiosis (Mitchell et al., 2019), and hydrogenase may
878 in fact also play a role in sulfur metabolism (as suggested for *A. vinosum* (Weissgerber et al.,
879 2014); Supplementary Results and Discussion F). Therefore, the exact role of hydrogen
880 oxidation in Endoriftia and why it might be relatively more relevant in larger symbionts
881 remains to be discussed.

882 *Denitrification in Riftia symbionts appears to be modular*

883 Our results suggest that small *Riftia* symbionts rely relatively more on the NarGHI-mediated
884 first step of respiratory nitrate reduction to nitrite, while all subsequent steps of nitrite
885 reduction via NO and N₂O to N₂ seem to be more prominent in larger symbionts. Since
886 expression of nitrate reduction genes is usually inhibited by oxygen (Payne, 1973), high

887 NarGHI abundance in fraction XS suggests that O₂ levels might be particularly low in the
888 trophosome lobule center.

889 Interestingly, although Endoriftia has the genomic potential for complete denitrification to N₂,
890 small and large symbionts seem to employ separate parts of the pathway. This is reminiscent
891 of free-living microbial communities, in which denitrification is modular, i.e., it is often not
892 carried out by individual organisms, but rather by the subsequent activity of several members
893 (Graf *et al.*, 2014), between which intermediates are passed on as ‘metabolic handoffs’
894 (Anantharaman *et al.*, 2016). Moreover, nitrate reduction and subsequent denitrification steps
895 may occur as two temporally separated processes even in the same organism: During nitrate
896 reduction in *Staphylococcus carnosus*, nitrite reduction was inhibited and resumed only after
897 nitrate was depleted (Neubauer and Götz, 1996). A similar scenario might be assumed for
898 Endoriftia: Small symbionts apparently reduce nitrate to nitrite, which, potentially, yields
899 enough energy to cover their demand, while ‘saving’ nitrite as a handoff for future use. Once
900 the symbionts have become larger, expression of nitrite reductase, nitric oxide reductase and
901 nitrous oxide reductase in higher abundance enables them to further reduce the accumulated
902 intermediate nitrite. Whether these reactions could also be regulated in response to varying
903 oxygen concentrations is unclear. We speculate that an O₂ gradient may exist, which influences
904 the observed expression pattern.

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

906 Relative abundance of the RNA polymerase sigma factor RpoD decreased from fraction XS to
907 L (Figure 3, Supplementary Table S4h) in S-rich and S-depleted samples, pointing to relatively
908 more growth-related activities in small Endoriftia (see also Supplementary Results and
909 Discussion B). RpoD is the primary sigma factor for vegetative growth (σ_{70}), which regulates
910 transcription of most genes involved in exponential growth in many bacteria (Helmann and
911 Chamberlin, 1988, Fujita *et al.*, 1994, Ishihama, 2000). This would be in agreement with the
912 idea of small *Riftia* symbionts being mainly occupied with cell division and proliferation in a
913 quasi-exponential growth phase, while large symbionts function as biosynthetic ‘factories’,
914 focusing on carbon fixation and biomass production. Interestingly, RpoS, the master

915 transcriptional regulator of stationary phase gene expression and antagonist of RpoD, was not
916 detected in any of our samples (although it is encoded in the symbiont genome). RpoS
917 abundance increases upon stress and limitation during transition to the stationary phase in
918 free-living model bacteria (Hengge-Aronis, 1993, Fujita *et al.*, 1994, Ishihama, 2000). Its
919 absence in the *Riftia* symbiont's proteome suggests that, unlike free-living bacteria, the
920 symbiont does not experience a stationary phase-like growth arrest even in later
921 developmental stages, probably because it is ideally supplied with all necessary substrates by
922 the host. This 'lack' of stress or limitation possibly results in less stringent regulation of
923 symbiont gene expression, which could explain the metabolic diversity we observed
924 particularly in large symbionts, such as multiple ways of energy generation (thiosulfate- and
925 hydrogen oxidation in addition to sulfide oxidation) and two CO₂ fixation pathways. Under
926 these premises, the previously observed simultaneous expression of seemingly redundant
927 metabolic pathways in *Riftia* symbionts (Markert *et al.*, 2011) very likely reflects this
928 presumptive "de-regulation" of gene expression in large parts of the symbiont population,
929 which allows Endoriftia to fully exploit its versatile metabolic repertoire to the advantage of
930 the symbiosis.

931 Conclusion

932 Our results show that Endoriftia cells of different differentiation stages likely employ distinct
933 metabolic profiles, thus confirming our initial hypothesis. Whereas small Endoriftia ensure
934 survival of the symbiont population, large Endoriftia are primarily engaged in biomass
935 production. The driving force behind this differentiation remains to be elucidated. For
936 *Rhizobium*, a steep O₂ concentration gradient inside legume nodules was proposed to be
937 involved in signaling for symbiont differential gene expression (Soupene *et al.*, 1995).
938 Similarly, some of the differences we observed in small and large Endoriftia might also be
939 connected to the availability of electron donors or acceptors, and hence differentiation of
940 Endoriftia cells might depend on substrate availability. Symbiont differentiation in *Riftia*
941 might furthermore be induced by specific host effectors, e.g., histone-derived antimicrobial
942 peptides, which were recently proposed to play a role in symbiont cell cycle regulation (Hinze

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

953

954

955 References

- 956 Ali, B. M. J., R. Amit, I. Braslavsky, A. B. Oppenheim, O. Gileadi and J. Stavans (2001).
957 Compaction of single DNA molecules induced by binding of integration host factor (IHF).
958 *Proceedings of the National Academy of Sciences of the United States of America* **98**: 10658-
959 10663. DOI: 10.1073/pnas.181029198.
- 960 Anantharaman, K., C. T. Brown, L. A. Hug, I. Sharon, C. J. Castelle, A. J. Probst, B. C. Thomas,
961 A. Singh, M. J. Wilkins, U. Karaoz, E. L. Brodie, K. H. Williams, S. S. Hubbard and J. F. Banfield
962 (2016). Thousands of microbial genomes shed light on interconnected biogeochemical
963 processes in an aquifer system. *Nature Communications* **7**(1): 13219. DOI:
964 10.1038/ncomms13219.
- 965 Angert, E. R. (2012). DNA Replication and Genomic Architecture of Very Large Bacteria.
966 *Annual Review of Microbiology* **66**: 197-212. DOI: 10.1146/annurev-micro-090110-102827.
- 967 Barabote, R. D., O. L. Johnson, E. Zetina, S. K. San Francisco, J. A. Fralick and M. J. D. San
968 Francisco (2003). *Erwinia chrysanthemi* TolC Is Involved in Resistance to Antimicrobial Plant
969 Chemicals and Is Essential for Phytopathogenesis. *Journal of Bacteriology* **185**: 5772-5778.
970 DOI: 10.1128/jb.185.19.5772-5778.2003.
- 971 Beam, C. E., C. J. Saveson and S. T. Lovett (2002). Role for *radA/sms* in recombination
972 intermediate processing in *Escherichia coli*. *Journal of Bacteriology* **184**: 6836-6844. DOI:
973 10.1128/JB.184.24.6836-6844.2002.
- 974 Beckmann, J. F., T. W. Markowski, B. A. Witthuhn and A. M. Fallon (2013). Detection of the
975 *Wolbachia*-encoded DNA binding protein, HU beta, in mosquito gonads. *Insect Biochemistry*
976 *and Molecular Biology* **43**: 272-279. DOI: 10.1016/j.ibmb.2012.12.007.
- 977 Bouvier, T., M. Troussellier, A. Anzil, C. Courties and P. Servais (2001). Using light scatter
978 signal to estimate bacterial biovolume by flow cytometry. *Cytometry: The Journal of the*
979 *International Society for Analytical Cytology* **44**(3): 188-194. DOI: 10.1002/1097-
980 0320(20010701)44:3<188::AID-CYTO1111>3.O.CO;2-C.
- 981 Bright, M., H. Keckeis and C. R. Fisher (2000). An autoradiographic examination of carbon
982 fixation, transfer and utilization in the *Riftia pachyptila* symbiosis. *Marine Biology* **136**: 621-
983 632. DOI: 10.1007/s002270050722.
- 984 Bright, M. and A. Sörgo (2003). Ultrastructural reinvestigation of the trophosome in adults of
985 *Riftia pachyptila* (Annelida, Siboglinidae). *Invertebrate Biology* **122**(4): 347-368. DOI:
986 10.1111/j.1744-7410.2003.tb00099.x.
- 987 Camberg, J. L., J. R. Hoskins and S. Wickner (2009). ClpXP protease degrades the cytoskeletal
988 protein, FtsZ, and modulates FtsZ polymer dynamics. *Proceedings of the National Academy*
989 *of Sciences of the United States of America* **106**: 10614-10619. DOI:
990 10.1073/pnas.0904886106.
- 991 Caro, A., O. Gros, P. Got, R. De Wit and M. Troussellier (2007). Characterization of the
992 population of the sulfur-oxidizing symbiont of *Codakia orbicularis* (Bivalvia, Lucinidae) by
993 single-cell analyses. *Applied and Environmental Microbiology* **73**: 2101-2109. DOI:
994 10.1128/AEM.01683-06.
- 995 Cavanaugh, C., S. L. Gardiner, M. L. Jones, H. W. Jannasch and J. B. Waterbury (1981).
996 Prokaryotic cells in the hydrothermal vent tubeworm *Riftia pachyptila*: possible
997 chemoautotrophic symbionts. *Science* **213**: 340-342. DOI: 10.1126/science.213.4505.340.

- 998 Chao, Y. and J. Vogel (2010). The role of Hfq in bacterial pathogens. *Current Opinion in*
999 *Microbiology* **13**: 24-33. DOI: 10.1016/j.mib.2010.01.001.
- 1000 Chastanet, A., I. Derre, S. Nair and T. Msadek (2004). *clpB*, a Novel Member of the *Listeria*
1001 *monocytogenes* CtsR Regulon, Is Involved in Virulence but Not in General Stress Tolerance.
1002 *Journal of Bacteriology* **186**(4): 1165-1174. DOI: 10.1128/jb.186.4.1165-1174.2004.
- 1003 Chien, A. C., N. S. Hill and P. A. Levin (2012). Cell size control in bacteria. *Current Biology* **22**:
1004 R340-R349. DOI: 10.1016/j.cub.2012.02.032.
- 1005 Cho, H., H. R. McManus, S. L. Dove and T. G. Bernhardt (2011). Nucleoid occlusion factor
1006 SlmA is a DNA-activated FtsZ polymerization antagonist. *Proceedings of the National*
1007 *Academy of Sciences of the United States of America* **108**: 3773-3778. DOI:
1008 10.1073/pnas.1018674108.
- 1009 Choi, H. M. T., V. A. Beck and N. A. Pierce (2014). Next-Generation in Situ Hybridization Chain
1010 Reaction: Higher Gain, Lower Cost, Greater Durability. *ACS Nano* **8**(5): 4284-4294. DOI:
1011 10.1021/nn405717p.
- 1012 Cortés, F. and N. Pastor (2003). Induction of endoreduplication by topoisomerase II catalytic
1013 inhibitors. *Mutagenesis* **18**: 105-112. DOI: 10.1093/mutage/18.2.105.
- 1014 Cortés, F., N. Pastor, S. Mateos and I. Domínguez (2003). Roles of DNA topoisomerases in
1015 chromosome segregation and mitosis. *Mutation Research - Reviews in Mutation Research*
1016 **543**: 59-66. DOI: 10.1016/S1383-5742(02)00070-4.
- 1017 Cosme, A. M., A. Becker, M. R. Santos, L. A. Sharypova, P. M. Santos and L. M. Moreira (2008).
1018 The Outer Membrane Protein TolC from *Sinorhizobium meliloti* Affects Protein Secretion,
1019 Polysaccharide Biosynthesis, Antimicrobial Resistance, and Symbiosis. *Molecular Plant-*
1020 *Microbe Interactions* **21**: 947-957. DOI: 10.1094/mpmi-21-7-0947.
- 1021 D'Haese, W. and M. Holsters (2004). Surface polysaccharides enable bacteria to evade plant
1022 immunity. *Trends in Microbiology* **12**: 555-561. DOI: 10.1016/j.tim.2004.10.009.
- 1023 Dang, W., Y. h. Hu and L. Sun (2011). HtpG is involved in the pathogenesis of *Edwardsiella*
1024 *tarda*. *Veterinary Microbiology* **152**: 394-400. DOI: 10.1016/j.vetmic.2011.05.030.
- 1025 Deaconescu, A. M., N. Savery and S. A. Darst (2007). The bacterial transcription repair
1026 coupling factor. *Current Opinion in Structural Biology* **17**: 96-102. DOI:
1027 10.1016/j.sbi.2007.01.005.
- 1028 Degenhardt, F., S. Seifert and S. Szymczak (2019). Evaluation of variable selection methods for
1029 random forests and omics data sets. *Briefings in Bioinformatics* **20**(2): 492-503. DOI:
1030 10.1093/bib/bbx124.
- 1031 Distel, D. and H. Felbeck (1988). Pathways of inorganic carbon fixation in the endosymbiont-
1032 bearing lucinid clam *Lucinoma aequizonata*. Part 1. Purification and characterization of the
1033 endosymbiotic bacteria. *Journal of Experimental Zoology* **247**(1): 1-10. DOI:
1034 10.1002/jez.1402470102.
- 1035 Domínguez-Cuevas, P., I. Porcelli, R. A. Daniel and J. Errington (2013). Differentiated roles
1036 for MreB-actin isologues and autolytic enzymes in *Bacillus subtilis* morphogenesis. *Molecular*
1037 *Microbiology* **89**: 1084-1098. DOI: 10.1111/mmi.12335.
- 1038 Dorman, C. J. and P. Deighan (2003). Regulation of gene expression by histone-like proteins
1039 in bacteria. *Current Opinion in Genetics and Development* **13**: 179-184. DOI: 10.1016/S0959-
1040 437X(03)00025-X.

- 1041 Drees, J. C., S. Chitteni-Pattu, D. R. McCaslin, R. B. Inman and M. M. Cox (2006). Inhibition
1042 of RecA protein function by the RdgC protein from *Escherichia coli*. *Journal of Biological*
1043 *Chemistry* **281**: 4708-4717. DOI: 10.1074/jbc.M513592200.
- 1044 Dri, A. M., J. Rouviere-Yaniv and P. L. Moreau (1991). Inhibition of cell division in *hupA hupB*
1045 mutant bacteria lacking HU protein. *Journal of Bacteriology* **173**: 2852-2863. DOI:
1046 10.1128/jb.173.9.2852-2863.1991.
- 1047 Duperthuy, M., P. Schmitt, E. Garzón, A. Caro, R. D. Rosa, F. Le Roux, N. Lautrédou-Audouy,
1048 P. Got, B. Romestand, J. De Lorgeril, S. Kieffer-Jaquinod, E. Bachère and D. Destoumieux-
1049 Garzón (2011). Use of OmpU porins for attachment and invasion of *Crassostrea gigas* immune
1050 cells by the oyster pathogen *Vibrio splendidus*. *Proceedings of the National Academy of*
1051 *Sciences of the United States of America* **108**: 2993-2998. DOI: 10.1073/pnas.1015326108.
- 1052 Durand-Heredia, J., E. Rivkin, G. Fan, J. Morales and A. Janakiraman (2012). Identification
1053 of ZapD as a cell division factor that promotes the assembly of FtsZ in *Escherichia coli*. *Journal*
1054 *of Bacteriology* **194**: 3189-3198. DOI: 10.1128/JB.00176-12.
- 1055 Ernst, J. and Z. Bar-Joseph (2006). STEM: a tool for the analysis of short time series gene
1056 expression data. *BMC Bioinformatics* **7**(1): 191. DOI: 10.1186/1471-2105-7-191.
- 1057 Felbeck, H. (1981). Chemoautotrophic Potential of the Hydrothermal Vent Tube Worm, *Riftia*
1058 *pachyptila* Jones (Vestimentifera). *Science* **213**(4505): 336-338. DOI:
1059 10.1126/science.213.4505.336.
- 1060 Felbeck, H. and P. J. Turner (1995). CO₂ transport in catheterized hydrothermal vent
1061 tubeworms, *Riftia pachyptila* (Vestimentifera). *Journal of Experimental Zoology* **272**(2): 95-
1062 102. DOI: 10.1002/jez.1402720203.
- 1063 Fernández, L. and R. E. W. Hancock (2012). Adaptive and mutational resistance: Role of porins
1064 and efflux pumps in drug resistance. *Clinical Microbiology Reviews* **25**: 661-681. DOI:
1065 10.1128/CMR.00043-12.
- 1066 Frees, D., S. N. A. Qazi, P. J. Hill and H. Ingmer (2003). Alternative roles of ClpX and ClpP in
1067 *Staphylococcus aureus* stress tolerance and virulence. *Molecular Microbiology* **48**: 1565-
1068 1578. DOI: 10.1046/j.1365-2958.2003.03524.x.
- 1069 Fujita, M., K. Tanaka, H. Takahashi and A. Amemural (1994). Transcription of the principal
1070 sigma-factor genes, *rpoD* and *rpoS*, in *Pseudomonas aeruginosa* is controlled according to the
1071 growth phase. *Molecular Microbiology* **13**: 1071-1077. DOI: 10.1111/j.1365-
1072 2958.1994.tb00498.x.
- 1073 Gardebrecht, A., S. Markert, S. M. Sievert, H. Felbeck, A. Thürmer, D. Albrecht, A. Wollherr,
1074 J. Kabisch, N. Le Bris, R. Lehmann, R. Daniel, H. Liesegang, M. Hecker and T. Schweder
1075 (2012). Physiological homogeneity among the endosymbionts of *Riftia pachyptila* and *Tevnia*
1076 *jerichonana* revealed by proteogenomics. *The ISME Journal* **6**(4): 766-776. DOI:
1077 10.1038/ismej.2011.137.
- 1078 Graf, D. R. H., C. M. Jones and S. Hallin (2014). Intergenomic Comparisons Highlight
1079 Modularity of the Denitrification Pathway and Underpin the Importance of Community
1080 Structure for N₂O Emissions. *PLOS ONE* **9**(12): e114118. DOI: 10.1371/journal.pone.0114118.
- 1081 Graham, J. M. (2001). Biological centrifugation. *The Basics*, BIOS Scientific Publishers Ltd.,
1082 Oxford, UK.

- 1083 Grimm, F., B. Franz and C. Dahl (2011). Regulation of dissimilatory sulfur oxidation in the
1084 purple sulfur bacterium *Allochromatium vinosum*. *Frontiers in Microbiology* **2**: 51. DOI:
1085 10.3389/fmicb.2011.00051.
- 1086 Gu, Z., R. Eils and M. Schlesner (2016). Complex heatmaps reveal patterns and correlations in
1087 multidimensional genomic data. *Bioinformatics* **32**(18): 2847-2849. DOI:
1088 10.1093/bioinformatics/btw313.
- 1089 Guha, S., S. Udupa, W. Ahmed and V. Nagaraja (2018). Rewired Downregulation of DNA
1090 Gyrase Impacts Cell Division, Expression of Topology Modulators, and Transcription in
1091 *Mycobacterium smegmatis*. *Journal of Molecular Biology* **430**: 4986-5001. DOI:
1092 10.1016/j.jmb.2018.10.001.
- 1093 Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis
1094 program for Windows 95/98/NT. *Nucleic acids symposium series [London]: Information*
1095 *Retrieval Ltd.* **41**(41): c1979-c2000.
- 1096 Hand, S. C. (1987). Trophosome ultrastructure and the characterization of isolated
1097 bacteriocytes from invertebrate-sulfur bacteria symbioses. *Biological Bulletin* **173**: 260-276.
1098 DOI: 10.2307/1541878.
- 1099 Hay, N. A., D. J. Tipper, D. Gygi and C. Hughes (1999). A novel membrane protein influencing
1100 cell shape and multicellular swarming of *Proteus mirabilis*. *Journal of Bacteriology* **181**:
1101 2008-2016.
- 1102 Heath, M. C. (2000). Nonhost resistance and nonspecific plant defenses. *Current Opinion in*
1103 *Plant Biology* **3**: 315-319. DOI: 10.1016/S1369-5266(00)00087-X.
- 1104 Helmann, J. D. and M. J. Chamberlin (1988). Structure and function of bacterial sigma factors.
1105 *Annual Review of Biochemistry* **57**: 839-872. DOI: 10.1146/annurev.biochem.57.1.839.
- 1106 Hengge-Aronis, R. (1993). Survival of hunger and stress: the role of *rpoS* in early stationary
1107 phase gene regulation in *E. coli*. *Cell* **72**(2): 165-168. DOI: 10.1016/0092-8674(93)90655-A.
- 1108 Hensen, D., D. Sperling, H. G. Trüper, D. C. Brune and C. Dahl (2006). Thiosulphate oxidation
1109 in the phototrophic sulphur bacterium *Allochromatium vinosum*. *Molecular Microbiology*
1110 **62**: 794-810. DOI: 10.1111/j.1365-2958.2006.05408.x.
- 1111 Hinzke, T. and S. Markert (2017). Efficient protein extraction for proteomics and
1112 metaproteomics (also suitable for low biomass samples). protocols.io. DOI:
1113 10.17504/protocols.io.kg6ctze
- 1114 Hinzke, T., M. Kleiner and S. Markert (2018). Centrifugation-Based Enrichment of Bacterial
1115 Cell Populations for Metaproteomic Studies on Bacteria–Invertebrate Symbioses. *Microbial*
1116 *Proteomics: Methods and Protocols*. D. Becher. New York, USA, Springer New York: 319-334
1117 DOI: 10.1007/978-1-4939-8695-8_22.
- 1118 Hinzke, T., M. Kleiner, C. Breusing, H. Felbeck, R. Häsler, S. M. Sievert, R. Schlüter, P.
1119 Rosenstiel, T. B. H. Reusch, T. Schweder and S. Markert (2019). Host-Microbe Interactions in
1120 the Chemosynthetic *Riftia pachyptila* Symbiosis. *mBio* **10**(6): e02243-02219. DOI:
1121 10.1128/mBio.02243-19.
- 1122 Ishihama, A. (2000). Functional modulation of *Escherichia coli* RNA polymerase. *Annual*
1123 *Reviews in Microbiology* **54**(1): 499-518. DOI: 10.1146/annurev.micro.54.1.499.

- 1124 Janitza, S., E. Celik and A. L. Boulesteix (2018). A computationally fast variable importance
1125 test for random forests for high-dimensional data. *Advances in Data Analysis and*
1126 *Classification* **12**: 885-915. DOI: 10.1007/s11634-016-0276-4.
- 1127 Kang, D., G. Liu, A. Lundström, E. Gelius and H. Steiner (1998). A peptidoglycan recognition
1128 protein in innate immunity conserved. *Proceedings of the National Academy of Sciences of*
1129 *the United States of America* **95**: 10078-10082. DOI: 10.1073/pnas.95.17.10078.
- 1130 King, A. M., G. Pretre, T. Bartpho, R. W. Sermiswan, C. Toma, T. Suzuki, A. Eshghi, M.
1131 Picardeau, B. Adler and G. L. Murray (2014). High-Temperature Protein G Is an Essential
1132 Virulence Factor of *Leptospira interrogans*. *Infection and Immunity* **82**: 1123-1131. DOI:
1133 10.1128/iai.01546-13.
- 1134 Kleiner, M., X. Dong, T. Hinzke, J. Wippler, E. Thorson, B. Mayer and M. Strous (2018).
1135 Metaproteomics method to determine carbon sources and assimilation pathways of species in
1136 microbial communities. *Proceedings of the National Academy of Sciences of the United States*
1137 *of America* **115**: E5576-E5584. DOI: 10.1073/pnas.1722325115.
- 1138 Klint, J., U. Rasmussen and B. Bergman (2007). FtsZ may have dual roles in the filamentous
1139 cyanobacterium *Nostoc/Anabaena* sp. strain PCC 7120. *Journal of Plant Physiology* **164**: 11-
1140 18. DOI: 10.1016/j.jplph.2005.08.021.
- 1141 Klose, J., K. Aistleitner, M. Horn, L. Krenn, V. Dirsch, M. Zehl and M. Bright (2016).
1142 Trophosome of the deep-sea tubeworm *Riftia pachyptila* inhibits bacterial growth. *PLoS ONE*
1143 **11**: e0146446. DOI: 10.1371/journal.pone.0146446.
- 1144 Köhler, S., J. Teyssier, A. Cloeckert, B. Rouot and J. P. Liautard (1996). Participation of the
1145 molecular chaperone DnaK in intracellular growth of *Brucella suis* within U937-derived
1146 phagocytes. *Molecular Microbiology* **20**: 701-712. DOI: 10.1111/j.1365-2958.1996.tb02510.x.
- 1147 Komaki, K. and H. Ishikawa (2000). Genomic copy number of intracellular bacterial symbionts
1148 of aphids varies in response to developmental stage and morph of their host. *Insect*
1149 *Biochemistry and Molecular Biology* **30**: 253-258. DOI: 10.1016/S0965-1748(99)00125-3.
- 1150 Kondorosi, E. and A. Kondorosi (2004). Endoreduplication and activation of the anaphase-
1151 promoting complex during symbiotic cell development. *FEBS Letters* **567**: 152-157. DOI:
1152 10.1016/j.febslet.2004.04.075.
- 1153 Koronakis, V., J. Eswaran and C. Hughes (2004). Structure and Function of TolC: The Bacterial
1154 Exit Duct for Proteins and Drugs. *Annual Review of Biochemistry* **73**: 467-489. DOI:
1155 10.1146/annurev.biochem.73.011303.074104.
- 1156 Kowalczykowski, S. C. (2000). Initiation of genetic recombination and recombination-
1157 dependent replication. *Trends in Biochemical Sciences* **25**: 156-165. DOI: 10.1016/S0968-
1158 0004(00)01569-3.
- 1159 Kruse, T., B. Blagoev, A. Løbner-Olesen, M. Wachi, K. Sasaki, N. Iwai, M. Mann and K. Gerdes
1160 (2006). Actin homolog MreB and RNA polymerase interact and are both required for
1161 chromosome segregation in *Escherichia coli*. *Genes and Development* **20**: 113-124. DOI:
1162 10.1101/gad.366606.
- 1163 Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of
1164 Bacteriophage T4. *Nature* **227**: 680-685. DOI: 10.1038/227680a0.
- 1165 Levine, C., H. Hiasa and K. J. Mariani (1998). DNA gyrase and topoisomerase IV: Biochemical
1166 activities, physiological roles during chromosome replication, and drug sensitivities.

- 1167 *Biochimica et Biophysica Acta - Gene Structure and Expression* **1400**: 29-43. DOI:
1168 10.1016/S0167-4781(98)00126-2.
- 1169 Li, C., J. Wei, Y. Li, X. He, Q. Zhou, J. Yan, J. Zhang, Y. Liu, Y. Liu and H.-b. Shu (2013).
1170 Transmembrane Protein 214 (TMEM214) Mediates Endoplasmic Reticulum Stress-induced
1171 Caspase 4 Enzyme Activation and Apoptosis. *Journal of Biological Chemistry* **288**: 17908-
1172 17917. DOI: 10.1074/jbc.M113.458836.
- 1173 López-Sánchez, M. J., A. Neef, R. Patiño-Navarrete, L. Navarro, R. Jiménez, A. Latorre and A.
1174 Moya (2008). Blattabacteria, the endosymbionts of cockroaches, have small genome sizes and
1175 high genome copy numbers. *Environmental Microbiology* **10**: 3417-3422. DOI:
1176 10.1111/j.1462-2920.2008.01776.x.
- 1177 Lorenzen, D. R., D. Gunther, J. Pandit, T. Rudel, E. Brandt and T. F. Meyer (2000). *Neisseria*
1178 *gonorrhoeae* porin modifies the oxidative burst of human professional phagocytes. *Infection*
1179 *and Immunity* **68**: 6215-6222. DOI: 10.1128/IAI.68.11.6215-6222.2000.
- 1180 Lynch, M. and H. Kuramitsu (2000). Expression and role of superoxide dismutases (SOD) in
1181 pathogenic bacteria. *Microbes and Infection* **2**: 1245-1255. DOI: 10.1016/S1286-
1182 4579(00)01278-8.
- 1183 Lyngstadaas, A., A. Løbner-Olesen and E. Boye (1995). Characterization of three genes in the
1184 *dam*-containing operon of *Escherichia coli*. *MGG Molecular & General Genetics* **247**: 546-
1185 554. DOI: 10.1007/BF00290345.
- 1186 Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert and K. Hornik. (2018). cluster: Cluster
1187 Analysis Basics and Extensions. R package version 2.0.7-1.
- 1188 Markert, S., C. Arndt, H. Felbeck, D. Becher, S. M. Sievert, M. Hugler, D. Albrecht, J. Robidart,
1189 S. Bench, R. A. Feldman, M. Hecker and T. Schweder (2007). Physiological proteomics of the
1190 uncultured endosymbiont of *Riftia pachyptila*. *Science* **315**(5809): 247-250. DOI:
1191 10.1126/science.1132913.
- 1192 Markert, S., A. Gardebrecht, H. Felbeck, S. M. Sievert, J. Klose, D. Becher, D. Albrecht, A.
1193 Thurmer, R. Daniel, M. Kleiner, M. Hecker and T. Schweder (2011). Status quo in physiological
1194 proteomics of the uncultured *Riftia pachyptila* endosymbiont. *Proteomics* **11**(15): 3106-3117.
1195 DOI: 10.1002/pmic.201100059.
- 1196 Mathur, J. and M. K. Waldor (2004). The *Vibrio cholerae* ToxR-regulated porin OmpU confers
1197 resistance to antimicrobial peptides. *Infection and Immunity* **72**: 3577-3583. DOI:
1198 10.1128/IAI.72.6.3577-3583.2004.
- 1199 Mergaert, P., T. Uchiumi, B. Alunni, G. Evanno, A. Cheron, O. Catrice, A. E. Mausset, F. Barloy-
1200 Hubler, F. Galibert, A. Kondorosi and E. Kondorosi (2006). Eukaryotic control on bacterial cell
1201 cycle and differentiation in the *Rhizobium*-legume symbiosis. *Proceedings of the National*
1202 *Academy of Sciences of the United States of America* **103**: 5230-5235. DOI:
1203 10.1073/pnas.0600912103.
- 1204 Mitchell, J. H., J. M. Leonard, J. Delaney, P. R. Girguis and K. M. Scott (2019). Hydrogen Does
1205 Not Appear To Be a Major Electron Donor for Symbiosis with the Deep-Sea Hydrothermal Vent
1206 Tubeworm *Riftia pachyptila*. *Applied and Environmental Microbiology* **86**(1): e01522-01519.
1207 DOI: 10.1128/aem.01522-19.
- 1208 Monds, R. D., T. K. Lee, A. Colavin, T. Ursell, S. Quan, T. F. Cooper and K. C. Huang (2014).
1209 Systematic Perturbation of Cytoskeletal Function Reveals a Linear Scaling Relationship
1210 between Cell Geometry and Fitness. *Cell Reports* **9**: 1528-1537. DOI:
1211 10.1016/j.celrep.2014.10.040.

- 1212 Mosleh, I. M., L. A. Huber, P. Steinlein, C. Pasquali, D. Günther and T. F. Meyer (1998).
1213 *Neisseria gonorrhoeae* porin modulates phagosome maturation. *Journal of Biological*
1214 *Chemistry* **273**: 35332-35338. DOI: 10.1074/jbc.273.52.35332.
- 1215 Mueller, R. S., V. J. Denef, L. H. Kalnejais, K. B. Suttle, B. C. Thomas, P. Wilmes, R. L. Smith,
1216 D. K. Nordstrom, R. B. McCleskey, M. B. Shah, N. C. VerBerkmoes, R. L. Hettich and J. F.
1217 Banfield (2010). Ecological distribution and population physiology defined by proteomics in a
1218 natural microbial community. *Molecular Systems Biology* **6**: 374. DOI: 10.1038/Msb.2010.30.
- 1219 Mukherjee, A., A. O. Sokunbi and A. Grove (2008). DNA protection by histone-like protein HU
1220 from the hyperthermophilic eubacterium *Thermotoga maritima*. *Nucleic Acids Research* **36**:
1221 3956-3968. DOI: 10.1093/nar/gkn348.
- 1222 Murga, M., I. Jaco, Y. Fan, R. Soria, B. Martinez-Pastor, M. Cuadrado, S. M. Yang, M. A. Blasco,
1223 A. I. Skoultchi and O. Fernandez-Capetillo (2007). Global chromatin compaction limits the
1224 strength of the DNA damage response. *Journal of Cell Biology* **178**: 1101-1108. DOI:
1225 10.1083/jcb.200704140.
- 1226 Neubauer, H. and F. Götz (1996). Physiology and interaction of nitrate and nitrite reduction in
1227 *Staphylococcus carnosus*. *Journal of Bacteriology* **178**: 2005-2009. DOI:
1228 10.1128/jb.178.7.2005-2009.1996.
- 1229 Nöllmann, M., N. J. Crisona and P. B. Arimondo (2007). Thirty years of *Escherichia coli* DNA
1230 gyrase: From *in vivo* function to single-molecule mechanism. *Biochimie* **89**: 490-499. DOI:
1231 10.1016/j.biochi.2007.02.012.
- 1232 Nussbaumer, A. D., C. R. Fisher and M. Bright (2006). Horizontal endosymbiont transmission
1233 in hydrothermal vent tubeworms. *Nature* **441**: 345-348. DOI: 10.1038/nature04793.
- 1234 Nyholm, S. V., J. J. Stewart, E. G. Ruby and M. J. McFall-Ngai (2009). Recognition between
1235 symbiotic *Vibrio fischeri* and the haemocytes of *Euprymna scolopes*. *Environmental*
1236 *Microbiology* **11**(2): 483-493. DOI: 10.1111/j.1462-2920.2008.01788.x.
- 1237 Ohniwa, R. L., K. Morikawa, J. Kim, T. Ohta, A. Ishihama, C. Wada and K. Takeyasu (2006).
1238 Dynamic state of DNA topology is essential for genome condensation in bacteria. *EMBO*
1239 *Journal* **25**: 5591-5602. DOI: 10.1038/sj.emboj.7601414.
- 1240 Ozawa, K., K. Kuwabara, M. Tamatani, K. Takatsuji, Y. Tsukamoto, S. Kaneda, H. Yanagi, D.
1241 M. Stern, Y. Eguchi, Y. Tsujimoto, S. Ogawa and M. Tohyama (1999). 150-kDa Oxygen-
1242 regulated Protein (ORP150) Suppresses Hypoxia-induced Apoptotic Cell Death. *Journal of*
1243 *Biological Chemistry* **274**: 6397-6404. DOI: 10.1074/jbc.274.10.6397.
- 1244 Payne, W. (1973). Reduction of nitrogenous oxides by microorganisms. *Bacteriological*
1245 *Reviews* **37**(4): 409.
- 1246 Pende, N., N. Leisch, H. R. Gruber-Vodicka, N. R. Heindl, J. Ott, T. Den Blaauwen and S.
1247 Bulgheresi (2014). Size-independent symmetric division in extraordinarily long cells. *Nature*
1248 *Communications* **5**: Article 4803. DOI: 10.1038/ncomms5803.
- 1249 Petersen, I., R. Schlüter, K. J. Hoff, V. Liebscher, G. Bange, K. Riedel and J. Pané-Farré (2020).
1250 Non-invasive and label-free 3D-visualization shows *in vivo* oligomerization of the
1251 staphylococcal alkaline shock protein 23 (Asp23). *Scientific Reports* **10**(1): 125. DOI:
1252 10.1038/s41598-019-56907-9.
- 1253 Petersen, J. M., F. U. Zielinski, T. Pape, R. Seifert, C. Moraru, R. Amann, S. Hourdez, P. R.
1254 Girguis, S. D. Wankel, V. Barbe, E. Pelletier, D. Fink, C. Borowski, W. Bach and N. Dubilier

- 1255 (2011). Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* **476**: 176-180.
1256 DOI: 10.1038/nature10325.
- 1257 Pflugfelder, B., C. R. Fisher and M. Bright (2005). The color of the trophosome: elemental
1258 sulfur distribution in the endosymbionts of *Riftia pachyptila* (Vestimentifera; Siboglinidae).
1259 *Marine Biology* **146**: 895-901. DOI: 10.1007/s00227-004-1500-x.
- 1260 Polz, M. F., H. Felbeck, R. Novak, M. Nebelsick and J. A. Ott (1992). Chemoautotrophic, Sulfur-
1261 Oxidizing Symbiotic Bacteria on Marine Nematodes: Morphological and Biochemical
1262 Characterization. *Microbial Ecology*: 313-329. DOI: 10.1007/BF00167789.
- 1263 Polzin, J., P. Arevalo, T. Nussbaumer, M. F. Polz and M. Bright (2019). Polyclonal symbiont
1264 populations in hydrothermal vent tubeworms and the environment. *Proceedings of the Royal
1265 Society B: Biological Sciences* **286**: 20181281. DOI: 10.1098/rspb.2018.1281.
- 1266 Ponnudurai, R., M. Kleiner, L. Sayavedra, J. M. Petersen, M. Moche, A. Otto, D. Becher, T.
1267 Takeuchi, N. Satoh and N. Dubilier (2017). Metabolic and physiological interdependencies in
1268 the *Bathymodiolus azoricus* symbiosis. *The ISME journal* **11**(2): 463. DOI:
1269 10.1038/ismej.2016.124.
- 1270 Prinsloo, E., M. M. Setati, V. M. Longshaw and G. L. Blatch (2009). Chaperoning stem cells: A
1271 role for heat shock proteins in the modulation of stem cell self-renewal and differentiation?
1272 *BioEssays* **31**: 370-377. DOI: 10.1002/bies.200800158.
- 1273 R Core Team. (2018). R: A language and environment for statistical computing. R Foundation
1274 for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- 1275 Reimold, C., H. J. Defeu Soufo, F. Dempwolff and P. L. Graumann (2013). Motion of variable-
1276 length MreB filaments at the bacterial cell membrane influences cell morphology. *Molecular
1277 Biology of the Cell* **24**: 2340-2349. DOI: 10.1091/mbc.e12-10-0728.
- 1278 Renoz, F., A. Champagne, H. Degand, A.-M. Faber, P. Morsomme, V. Foray and T. Hance
1279 (2017). Toward a better understanding of the mechanisms of symbiosis: a comprehensive
1280 proteome map of a nascent insect symbiont. *PeerJ* **5**: e3291. DOI: 10.7717/peerj.3291.
- 1281 Robidart, J. C., S. R. Bench, R. A. Feldman, A. Novoradovsky, S. B. Podell, T. Gaasterland, E.
1282 E. Allen and H. Felbeck (2008). Metabolic versatility of the *Riftia pachyptila* endosymbiont
1283 revealed through metagenomics. *Environmental Microbiology* **10**(3): 727-737. DOI:
1284 10.1111/j.1462-2920.2007.01496.x.
- 1285 Robinson, M. D., D. J. McCarthy and G. K. Smyth (2010). edgeR: a Bioconductor package for
1286 differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.
1287 DOI: 10.1093/bioinformatics/btp616.
- 1288 Royet, J., D. Gupta and R. Dziarski (2011). Peptidoglycan recognition proteins : modulators of
1289 the microbiome and inflammation. *Nature Reviews Immunology* **11**: 837-851. DOI:
1290 10.1038/nri3089.
- 1291 Saalfeld, S. (2010). Enhance Local Contrast (CLAHE) - a Fiji plugin; available at:
1292 [https://imagej.net/Enhance_Local_Contrast_\(CLAHE\)](https://imagej.net/Enhance_Local_Contrast_(CLAHE)).
- 1293 Sattler, M. C., C. R. Carvalho and W. R. Clarindo (2016). The polyploidy and its key role in
1294 plant breeding. *Planta* **243**: 281-296. DOI: 10.1007/s00425-015-2450-x.
- 1295 Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch,
1296 C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P.

- 1297 Tomancak and A. Cardona (2012). Fiji: an open-source platform for biological-image analysis.
1298 *Nature Methods* **9**(7): 676-682. DOI: 10.1038/nmeth.2019.
- 1299 Søndergaard, D., C. N. S. Pedersen and C. Greening (2016). HydDB: A web tool for
1300 hydrogenase classification and analysis. *Scientific Reports* **6**: 34212. DOI: 10.1038/srep34212.
- 1301 Sorgo, A., F. Gaill, J.-P. Lechaire, C. Arndt and M. Bright (2002). Glycogen storage in the *Riftia*
1302 *pachyptila* trophosome: contribution of host and symbionts. *Marine Ecology Progress Series*
1303 **231**: 115-120. DOI: 10.3354/meps231115.
- 1304 Soupene, E., M. Foussard, P. Boistard, G. Truchet and J. Batut (1995). Oxygen as a key
1305 developmental regulator of *Rhizobium meliloti* N₂-fixation gene expression within the alfalfa
1306 root nodule. *Proceedings of the National Academy of Sciences of the United States of America*
1307 **92**: 3759-3763. DOI: 10.1073/pnas.92.9.3759.
- 1308 Steck, T. R. and K. Drlica (1984). Bacterial chromosome segregation: Evidence for DNA gyrase
1309 involvement in decatenation. *Cell* **36**: 1081-1087. DOI: 10.1016/0092-8674(84)90058-8.
- 1310 Stockdreher, Y., M. Sturm, M. Josten, H. G. Sahl, N. Dobler, R. Zigann and C. Dahl (2014).
1311 New proteins involved in sulfur trafficking in the cytoplasm of *Allochromatium vinosum*.
1312 *Journal of Biological Chemistry* **289**: 12390-12403. DOI: 10.1074/jbc.M113.536425.
- 1313 Stocks, S. M. (2004). Mechanism and use of the commercially available viability stain,
1314 BacLight. *Cytometry Part A* **61**: 189-195. DOI: 10.1002/cyto.a.20069.
- 1315 Stuger, R., C. L. Woldringh, C. C. V. D. Weijden, N. O. E. Vischer, M. Bakker, R. J. M. V.
1316 Spanning, J. L. Snoep and H. V. Westerhoff (2002). DNA supercoiling by gyrase is linked to
1317 nucleoid compaction. *Molecular Biology Reports* **29**: 79-82. DOI:
1318 10.1023/A:1020318705894.
- 1319 Susin, M. F., R. L. Baldini, F. Gueiros-Filho and S. L. Gomes (2006). GroES/GroEL and
1320 DnaK/DnaJ have distinct roles in stress responses and during cell cycle progression in
1321 *Caulobacter crescentus*. *Journal of Bacteriology* **188**: 8044-8053. DOI: 10.1128/JB.00824-
1322 06.
- 1323 Takata, H., T. Hanafusa, T. Mori, M. Shimura, Y. Iida, K. Ishikawa, K. Yoshikawa, Y. Yoshikawa
1324 and K. Maeshima (2013). Chromatin Compaction Protects Genomic DNA from Radiation
1325 Damage. *PLoS ONE* **8**: e75622. DOI: 10.1371/journal.pone.0075622.
- 1326 Tessmer, I., T. Moore, R. G. Lloyd, A. Wilson, D. A. Erie, S. Allen and S. J. B. Tendler (2005).
1327 AFM studies on the role of the protein RdgC in bacterial DNA recombination. *Journal of*
1328 *Molecular Biology* **350**: 254-262. DOI: 10.1016/j.jmb.2005.04.043.
- 1329 Thanedar, S. and W. Margolin (2004). FtsZ Exhibits Rapid Movement and Oscillation Waves
1330 in Helix-like Patterns in *Escherichia coli*. *Current Biology* **14**: 1167-1173. DOI:
1331 10.1016/j.cub.2004.06.048.
- 1332 The Global Proteome Machine Organization. The Global Proteome Machine: cRAP protein
1333 sequences. Available at: <http://thegpm.org/crap/> [Accessed November 28, 2017].
- 1334 Thévenaz, P., D. Sage and M. Unser (2012). Bi-Exponential Edge-Preserving Smoother. *IEEE*
1335 *Transactions on Image Processing* **21**(9): 3924--3936. DOI: 10.1109/TIP.2012.2200903.
- 1336 Tracy, B. P., S. M. Gaida and E. T. Papoutsakis (2010). Flow cytometry for bacteria: enabling
1337 metabolic engineering, synthetic biology and the elucidation of complex phenotypes. *Current*
1338 *Opinion in Biotechnology* **21**(1): 85-99. DOI: <https://doi.org/10.1016/j.copbio.2010.02.006>.

- 1339 Troll, J. V., D. M. Adin, A. M. Wier, N. Paquette, N. Silverman, W. E. Goldman, F. J.
1340 Stadermann, E. V. Stabb and M. J. Mcfall-ngai (2009). Peptidoglycan induces loss of a nuclear
1341 peptidoglycan recognition protein during host tissue development in a beneficial animal-
1342 bacterial symbiosis. *Cellular Microbiology* **11**: 1114-1127. DOI: 10.1111/j.1462-
1343 5822.2009.01315.x.
- 1344 Truglio, J. J., D. L. Croteau, B. Van Houten and C. Kisker (2006). Prokaryotic Nucleotide
1345 Excision Repair: The UvrABC System. *Chemical Reviews* **106**: 233-252. DOI:
1346 10.1021/cro40471u.
- 1347 Twine, S. M., N. C. S. Mykytczuk, M. D. Petit, H. Shen, J. W. Conlan and J. F. Kelly (2006). *In*
1348 *vivo* proteomic analysis of the intracellular bacterial pathogen, *Francisella tularensis*, isolated
1349 from mouse spleen. *Biochemical and Biophysical Research Communications* **345**: 1621-1633.
1350 DOI: 10.1016/j.bbrc.2006.05.070.
- 1351 van der Hoeven, R. and S. Forst (2009). OpnS, an outer membrane porin of *Xenorhabdus*
1352 *nematophila*, confers a competitive advantage for growth in the insect host. *Journal of*
1353 *Bacteriology* **191**: 5471-5479. DOI: 10.1128/JB.00148-09.
- 1354 Verbrugghe, E., A. Van Parys, B. Leyman, F. Boyen, F. Haesebrouck and F. Pasmans (2015).
1355 HtpG contributes to *Salmonella typhimurium* intestinal persistence in pigs. *Veterinary*
1356 *Research* **46**: 118. DOI: 10.1186/s13567-015-0261-5.
- 1357 Vizcaíno, J. A., A. Csordas, N. del-Toro, J. A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-
1358 Riverol, F. Reisinger, T. Ternent, Q.-W. Xu, R. Wang and H. Hermjakob (2016). 2016 update
1359 of the PRIDE database and its related tools. *Nucleic Acids Research* **44**(D1): D447-D456. DOI:
1360 10.1093/nar/gkv1145.
- 1361 Wachi, M. and M. Matsushashi (1989). Negative control of cell division by *mreB*, a gene that
1362 functions in determining the rod shape of *Escherichia coli* cells. *Journal of Bacteriology* **171**:
1363 3123-3127. DOI: 10.1128/jb.171.6.3123-3127.1989.
- 1364 Wang, J., Y. Wu, G. Yang and S. Aksoy (2009). Interactions between mutualist *Wigglesworthia*
1365 and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypanosome transmission.
1366 *Proceedings of the National Academy of Sciences of the United States of America* **106**(29):
1367 12133-12138. DOI: 10.1073/pnas.0901226106.
- 1368 Wang, S., H. Arellano-Santoyo, P. A. Combs and J. W. Shaevitz (2010). Actin-like cytoskeleton
1369 filaments contribute to cell mechanics in bacteria. *Proceedings of the National Academy of*
1370 *Sciences of the United States of America* **107**: 9182-9185. DOI: 10.1073/pnas.0911517107.
- 1371 Wang, Y., S. Maharana, M. D. Wang and G. V. Shivashankar (2014). Super-resolution
1372 microscopy reveals decondensed chromatin structure at transcription sites. *Scientific Reports*
1373 **4**(1): 4477. DOI: 10.1038/srep04477.
- 1374 Weiss, D. S. (2004). Bacterial cell division and the septal ring. *Molecular Microbiology* **54**(3):
1375 588-597. DOI: 10.1111/j.1365-2958.2004.04283.x.
- 1376 Weissgerber, T., R. Zigann, D. Bruce, Y.-j. Chang, J. C. Detter, C. Han, L. Hauser, C. D. Jeffries,
1377 M. Land and A. C. Munk (2011). Complete genome sequence of *Allochromatium vinosum* DSM
1378 180 T. *Standards in Genomic Sciences* **5**(3): 311-330. DOI: 10.4056/sigs.2334270.
- 1379 Weissgerber, T., M. Sylvester, L. Kröniger and C. Dahl (2014). A comparative quantitative
1380 proteomic study identifies new proteins relevant for sulfur oxidation in the purple sulfur
1381 bacterium *Allochromatium vinosum*. *Applied and Environmental Microbiology* **80**: 2279-
1382 2292. DOI: 10.1128/aem.04182-13.

- 1383 Welte, C., S. Hafner, C. Krätzer, A. Quentmeier, C. G. Friedrich and C. Dahl (2009). Interaction
1384 between Sox proteins of two physiologically distinct bacteria and a new protein involved in
1385 thiosulfate oxidation. *FEBS Letters* **583**: 1281-1286. DOI: 10.1016/j.febslet.2009.03.020.
- 1386 Wippler, J., M. Kleiner, C. Lott, A. Gruhl, P. E. Abraham, R. J. Giannone, J. C. Young, R. L.
1387 Hettich and N. Dubilier (2016). Transcriptomic and proteomic insights into innate immunity
1388 and adaptations to a symbiotic lifestyle in the gutless marine worm *Olavius algarvensis*. *BMC*
1389 *Genomics* **17**: 942. DOI: 10.1186/s12864-016-3293-y.
- 1390 Woyke, T., D. Tighe, K. Mavromatis, A. Clum, A. Copeland, W. Schackwitz, A. Lapidus, D. Wu,
1391 J. P. Mccutcheon, B. R. Mcdonald, N. A. Moran, J. Bristow and J. F. Cheng (2010). One
1392 bacterial cell, one complete genome. *PLoS ONE* **5**(4): e10314. DOI:
1393 10.1371/journal.pone.0010314.
- 1394 Wright, M. N. and A. Ziegler (2015). ranger: A Fast Implementation of Random Forests for
1395 High Dimensional Data in C++ and R. *Journal of Statistical Software* **77**: 1-17. DOI:
1396 10.18637/jss.v077.i01.
- 1397 Yoshikawa, Y., T. Mori, N. Magome, K. Hibino and K. Yoshikawa (2008). DNA compaction
1398 plays a key role in radioprotection against double-strand breaks as revealed by single-molecule
1399 observation. *Chemical Physics Letters* **456**: 80-83. DOI: 10.1016/j.cplett.2008.03.009.
- 1400 Zambrano, N., P. P. Guichard, Y. Bi, B. Cayrol, S. Marco and V. Arluison (2009). Involvement
1401 of HFq protein in the post-transcriptional regulation of *E. coli* bacterial cytoskeleton and cell
1402 division proteins. *Cell Cycle* **8**: 2470-2472. DOI: 10.4161/cc.8.15.9090.
- 1403 Zerulla, K. and J. Soppa (2014). Polyploidy in haloarchaea: Advantages for growth and
1404 survival. *Frontiers in Microbiology* **5**: 274. DOI: 10.3389/fmicb.2014.00274.
- 1405 Zybilov, B., A. L. Mosley, M. E. Sardu, M. K. Coleman, L. Florens and M. P. Washburn (2006).
1406 Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*.
1407 *Journal of Proteome Research* **5**: 2339-2347. DOI: 10.1021/pro60161n.
- 1408
- 1409

1410 Acknowledgements

1411 We thank captains and crews of R/V Atlantis and DSV Alvin who supported sampling during
1412 cruises AT26-23 and AT37-12. We are grateful to Jana Matulla and Annette Meuche for
1413 excellent technical assistance in sample preparation for proteomics and electron microscopy,
1414 respectively. Thanks to Ruby Ponnudurai and Frank Unfried for help with CARD-FISH, and to
1415 Alexander Graf, Mathis Appelbaum, Judith Zimmermann and Silke Wetzel for advice on
1416 epifluorescence microscopy and staining. We greatly appreciate Elisa Kasbohm's help with
1417 random forest analyses. We are very grateful to Jörg Bernhardt for stitching the transmission
1418 electron micrographs to produce a panorama image with high resolution. This work was
1419 supported by the German Research Foundation DFG (grant MA 6346/2-1 to S.M.), a fellowship
1420 of the Institute of Marine Biotechnology Greifswald (T.H., M.M.), a German Academic
1421 Exchange Service (DAAD) grant (T.H.), the NC State Chancellor's Faculty Excellence Program
1422 Cluster on Microbiomes and Complex Microbial Communities (M.K.), the USDA National
1423 Institute of Food and Agriculture, Hatch project 1014212 (M.K.), the U.S. National Science
1424 Foundation (grants OCE-1131095 and OCE-1559198 to S.M.S), and The WHOI Investment in
1425 Science Fund (to S.M.S).

1426 Author contributions

1427 S.M. conceived the study, S.M., T.H., M.K. designed the experiments, T.H. and H.F. took
1428 samples. T.H. prepared samples for CARD-FISH and metaproteomics analysis, analyzed data,
1429 conducted statistical analyses and prepared figures. T.H. and S.M. wrote the manuscript with
1430 input from all coauthors. T.S. was involved in project coordination, S.M.S. obtained funding
1431 for the research cruises and coordinated sampling as chief scientist. C.H. and F.B. performed
1432 MS measurements of metaproteomics samples, D.B. coordinated MS measurements. M.M.
1433 and J.P.-F. contributed to fluorescence microscopy and R.S. conducted electron microscopy
1434 analyses. P.H. performed flow cytometry analyses, and U.V. coordinated flow cytometry
1435 measurements.

1436 **Conflicts of interest**

1437 **The authors declare no conflicts of interest.**

1438

1439