Host-microbe interactions in the chemosynthetic *Riftia pachyptila* symbiosis

Tjorven Hinzke[1,2,3]*, Manuel Kleiner[3,4], Corinna Breusing[5], Horst Felbeck[6], Robert Häsler[7], Stefan M. Sievert[8], Rabea Schlüter[9], Philip Rosenstiel[7], Thorsten B. H. Reusch[10], Thomas Schweder[1,2], Stephanie Markert[1,2]*

[1] Institute of Marine Biotechnology e. V., Greifswald, Germany

[2] Institute of Pharmacy, Department of Pharmaceutical Biotechnology, University of Greifswald, Germany

[3] Energy Bioengineering Group, University of Calgary, Canada

[4] Department of Plant & Microbial Biology, North Carolina State University, Raleigh, USA

[5] Monterey Bay Aquarium Research Institute, Moss Landing, USA

[6] Scripps Institution of Oceanography, University of California San Diego, San Diego, USA

[7] Institute of Clinical Molecular Biology (IKMB), Kiel University, Kiel, Germany

[8] Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, USA

[9] Imaging Center of the Department of Biology, University of Greifswald, Germany

[10] Marine Evolutionary Ecology, GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

*email addresses:

tjorven.hinzke@outlook.com

stephanie.markert@uni-greifswald.de

Key words:

Metaproteomics, holobiont, mutualism, mutualistic association, deep sea, hydrothermal vents

Abstract

The deep-sea tubeworm *Riftia pachyptila* lacks a digestive system, but completely relies on bacterial endosymbionts for nutrition. Although the symbiont has been studied in detail on the molecular level, such analyses were unavailable for the animal host, because sequence information was lacking. To identify host-symbiont interaction mechanisms, we therefore sequenced the *Riftia* transcriptome, which enabled comparative metaproteomic analyses of symbiont-containing versus symbiont-free tissues, both under energy-rich and energy-limited conditions. We demonstrate that metabolic interactions include nutrient allocation from symbiont to host by symbiont digestion, and substrate transfer to the symbiont by abundant host proteins. Our analysis further suggests that *Riftia* maintains its symbiont by protecting the bacteria from oxidative damage, while also exerting symbiont population control. Eukaryote-like symbiont proteins might facilitate intracellular symbiont persistence. Energy limitation apparently leads to reduced symbiont biomass and increased symbiont digestion. Our study provides unprecedented insights into host-microbe interactions that shape this highly efficient symbiosis.

Introduction

All animals are associated with microorganisms (Bang et al., 2018; Bosch and McFall-Ngai, 2011), and consequently, animal-microbe interactions shape life on our planet. While research has concentrated for decades on pathogenic associations, beneficial, i.e. mutualistic symbioses are increasingly moving into the center of attention (McFall-Ngai et al., 2013).

Mutualistic relationships are often based on nutritional benefits for both partners: Symbionts supply their host with nutrients otherwise lacking in the hosts' diet, while the host in turn provides the symbionts with metabolites, shelter and optimal growth conditions (Moya et al., 2008). To establish and stably maintain their alliance, the partners have to interact on the molecular level. The hosts' immune system needs to control the symbiont population without erasing it altogether (Feldhaar and Gross, 2009), for example by restricting the symbionts to certain organs and/or by down-regulating its own immune response (reviewed in Nyholm and Graf, 2012). Symbionts, on the other hand, often employ strategies resembling those of pathogens to colonize and persist in their host. For example, similar protein secretion systems are employed by both, symbionts and pathogens, for interactions with the host (Dale and Moran, 2006; Hentschel et al., 2000; McFall-Ngai, 2008; Moya et al., 2008).

In many animals, host-microbe interactions are difficult to assess, due to the high number of potentially involved microbes and the presence of long- and short-term associations, which are hard to distinguish (McFall-Ngai, 2008). Therefore, low-complexity models are important to identify and characterize interaction mechanisms (Webster, 2014). Symbioses of marine invertebrates and their chemoautotrophic symbionts have emerged as suitable study systems. In these symbioses, animal hosts such as gutless annelids and bivalves are often tightly associated with one or few symbiont types, which enable the eukaryotes to prevail in otherwise hostile environments (Dubilier et al., 2008). One of the most conspicuous representatives of these associations, and the first animal in

which chemoautotrophic symbionts were discovered, is the giant tube worm *Riftia pachyptila* (short *Riftia*) that thrives around deep-sea hydrothermal vents of the East Pacific (Cavanaugh et al., 1981; Felbeck, 1981). The host's absolute dependency on its symbiont makes *Riftia* an ideal system to study beneficial host-microbe interactions in a mutualistic symbiosis.

The worm completely lacks a digestive system, but instead receives all necessary nutrients from one phylotype of chemosynthetic endosymbionts (Cavanaugh et al., 1981; Distel et al., 1988; Hand, 1987; Robidart et al., 2008). The host in turn provides the endosymbionts with all necessary inorganic compounds for chemosynthesis (Stewart and Cavanaugh, 2005). This association is remarkably productive: *Riftia* grows extraordinarily fast (> 85 cm increase in tube length per year, Lutz et al., 1994) and reaches body lengths of up to 1.5 m (Jones, 1981).

The uncultured gammaproteobacterial *Riftia* symbiont, tentatively named *Candidatus* Endoriftia persephone (Robidart et al., 2008), densely populates bacteriocytes in the host trophosome, a specialized organ that fills most of the worm's body cavity (Hand, 1987). The bacteria oxidize inorganic reduced compounds such as hydrogen sulfide to generate energy for carbon fixation (Cavanaugh et al., 1981; Fisher et al., 1989; Markert et al., 2011; Petersen et al., 2011; Robidart et al., 2011; Van Dover, 2000). The symbiont can store elemental sulfur, an intermediate of sulfide oxidation, in sulfur globules (Pflugfelder et al., 2005). Trophosome tissue containing high amounts of stored sulfur has a light yellowish color. During sulfide limitation, i.e., when energy availability is restricted, stored sulfur is consumed and the trophosome appears much darker (Pflugfelder et al., 2005; Wilmot Jr. and Vetter, 1990). The energetic status of the symbiosis can thus be directly inferred from the color of the trophosome.

Riftia has been extensively studied, especially with respect to its anatomy, biochemistry, symbiont transmission, and substrate transfer between host, symbionts and the environment (e.g. Drozdov and Galkin, 2012; Liu et al., 2017; Robidart et al., 2011; Sanchez et al., 2007b; Scott et al., 2012; see

Stewart and Cavanaugh, 2005 for a review). The symbiont's metabolism has been studied in detail as well (Stewart and Cavanaugh, 2005), in particular by means of metagenomics and metaproteomics (Gardebrecht et al., 2012; Markert et al., 2007, 2011; Robidart et al., 2008). Yet, little is known about interactions between the two symbiotic partners and, particularly, about the proteins directly involved in these processes.

Our study aimed to illuminate the underlying mechanisms of host-symbiont interactions on the protein level. For this purpose, we employed a state-of-the-art global metaproteomics approach, which required comprehensive sequence data for both partners. While the genome of the *Riftia* symbiont was sequenced previously (Gardebrecht et al., 2012; Robidart et al., 2008), up to now no such information was available for the host. Therefore, we sequenced the transcriptome of the *Riftia* host *de novo*. This enabled us to build a comprehensive protein database, which we used to compare protein abundance patterns in symbiont-containing and symbiont-free *Riftia* tissues. By comparing sulfur-rich and sulfur-depleted specimens, we furthermore examined how host-symbiont interactions vary under high- and low energy conditions. Our analysis sheds light on metabolite exchange processes between both partners, on the host's symbiont maintenance strategies and on the symbiont's molecular mechanisms to persist inside the host.

Results and Discussion

Interaction analysis of a chemosynthetic deep-sea symbiosis

We sequenced the *Riftia* host transcriptome *de novo* and combined it with three existing symbiont genomes to create a comprehensive holobiont database for identification of *Riftia* host and symbiont proteins (see Material and Methods). Our metaproteomic analysis included comparisons between symbiont-containing and symbiont-free tissues of specimens with light and dark trophosomes (hereafter referred to as sulfur-rich, S-rich specimens and tissues, and S-depleted specimens and tissues, respectively). A fully replicated dataset and stringent experimental design enabled us to find statistically significant differences in individual protein abundance between sample types, as well as abundance differences between functional protein groups. For an overview of all identified proteins, see Supplementary Results and Discussion Part 1 (SOM1). We identified symbiosis-specific proteins and molecular interaction processes, including (i) metabolite exchange between host and symbiont, (ii) host strategies of symbiont maintenance, and (iii) symbiont mechanisms to persist inside the host. Furthermore, we found that (iv) S availability affects symbiotic interactions in *Riftia*. For a graphical representation of the main interactions, see Figure 1. Beyond the results presented here, our data sets also provide a valuable resource for future *Riftia* studies and microbe-eukaryote symbiosis research in general.

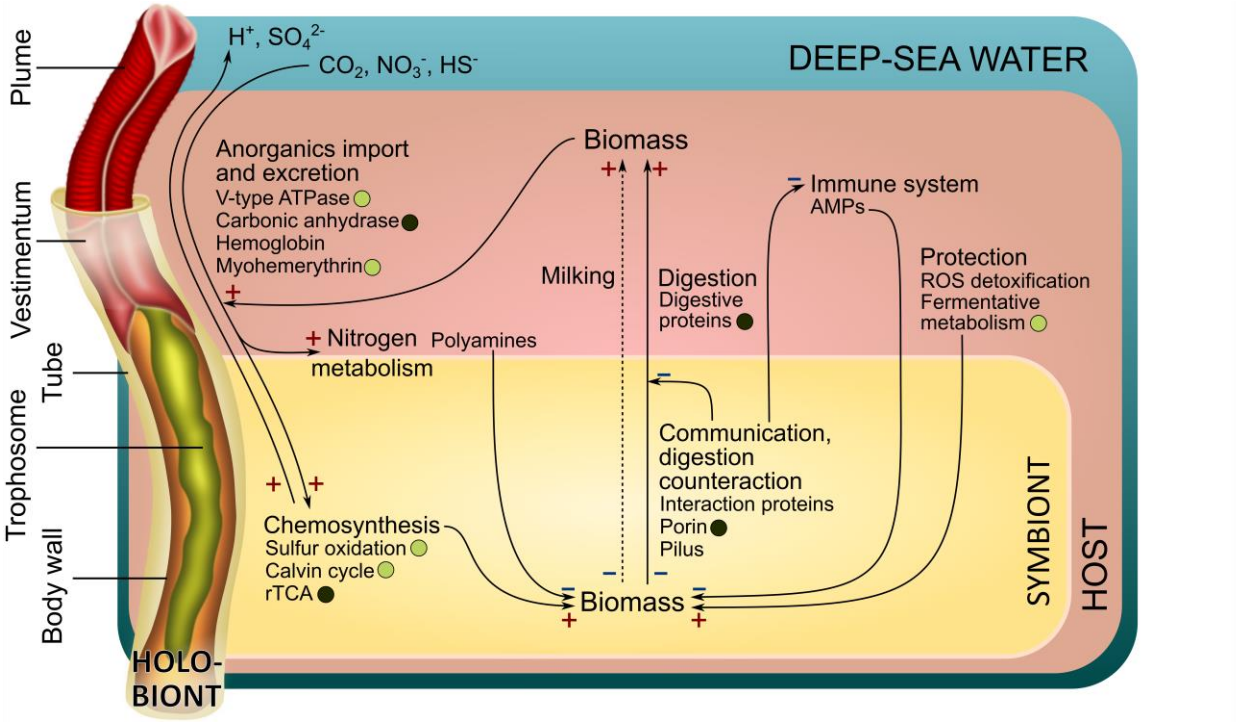


Figure 1: Main interactions in the *Riftia* symbiosis. "+" indicates presumably stimulating interactions, "-" indicates presumably inhibiting interactions. Circles, where present, indicate that the respective proteins are more abundant in S-rich (light circles) or S-depleted (dark circles) specimens, respectively. Milking: Transfer of small organic compounds (see SOM3).

Metabolite exchange between host and symbiont

Riftia digests its symbionts for nutrition

Our results suggest that the main mode of nutrient transfer from symbiont to host is the active digestion of symbiont cells, and that this process might involve endosome-like maturation of symbiont-containing vesicles. We detected a total of 113 host enzymes involved in protein-, amino acid- and glycan degradation as well as in glycolysis and fatty acid beta oxidation. 22 of these proteins were significantly more abundant in trophosome samples than in the other tissues (Table 1). Overall, nearly all of the respective protein groups had higher abundances (i.e. summed-up %orgNSAF) in the symbiont-bearing trophosome than in other tissues, both in S-rich and S-depleted specimens (Figure 2). Many of the protein degradation-related proteins contain signal peptides and are thus likely either

142 contained in lysosomes or secreted into the symbiont-containing vesicles to digest the symbiont cells
143 (Table 1, Supp. Table S1).

144 Our findings are in accordance with previous biochemical, autoradiographic and microscopic studies,
145 which suggested symbiont digestion in *Riftia* trophosome (Boetius and Felbeck, 1995; Bright et al.,
146 2000; Hand, 1987; Pflugfelder et al., 2009). Moreover, abundant degradative enzymes and symbiont
147 digestion appear to be common in other mutualistic symbioses as well, including deep-sea mussels
148 (Ponnudurai et al., 2017; Streams et al., 1997, Ponnudurai et al., submitted), shallow-water clams
149 (Caro et al., 2009; König et al., 2015) and the gutless oligochaete *Olavius algarvensis* (Wippler et al.,
150 2016; Woyke et al., 2006).

151 Our metaproteome analysis suggests that symbiont digestion in *Riftia* might involve maturation of
152 symbiont-containing host vesicles in a process resembling the maturation of endosomes. Endosomes
153 form after endocytosis of extracellular compounds and mature from early to late endosomes, which
154 ultimately fuse with lysosomes. The endosome-associated proteins Rab5 and Rab7 showed
155 significantly higher abundances in trophosome samples compared to other host tissues. Rab5 and
156 Rab7 localize to early and late endosomes as well as autophagosomes, respectively, and are markers
157 for these recycling-related organelles (Chavrier et al., 1990; Hyttinen et al., 2013; Vieira et al., 2002).
158 The idea of symbiont degradation via an endosome-like maturation process in *Riftia* is additionally
159 supported by the observation of multilamellar bodies in *Riftia* bacteriocytes in our TEM images
160 (Figure 3). These multilamellar bodies can form in endosomes (Marchetti et al., 2004), but were also
161 suggested to be associated with autophagic cell death in *Riftia* trophosome (Pflugfelder 2009).
162 Although autophagy and apoptosis were suggested to be involved in cell death in *Riftia* trophosome
163 (Pflugfelder et al., 2009), our results contradict this hypothesis. We detected only two autophagy-
164 related proteins (Supp. Table S2) and only 12 of 41 detected apoptosis-related *Riftia* proteins were
165 identified in the trophosome, mostly with similar or significantly lower abundances as compared to
166 other tissues. Caspases, the main apoptotic effectors, were not detected at all on the protein level in

167 trophosome samples (see also SOM2). This is in line with previous microscopic results, which did not
168 indicate apoptosis in the trophosome (Bright and Sorgo, 2003). A non-autophagic, non-apoptotic cell
169 death mechanism was recently described in pea aphid bacteriocytes (Simonet et al., 2018). In the
170 aphids, the proposed mechanism involved hypervacuolation of host bacteriocytes, which was,
171 however, not observed in *Riftia* trophosome. Another caspase-independent cell death mechanism,
172 which involves the protease cathepsin B, has been described in cancer cells (Bröker et al., 2004). As
173 cathepsin B was significantly more abundant in trophosome than in other *Riftia* tissues, we speculate
174 that this protease, amongst other degradative enzymes, may be involved in controlled cell death in
175 *Riftia* trophosome.

176 Besides symbiont digestion, a second mode of nutrient transfer, the release of small organic carbon
177 compounds by intact symbionts, was suggested to be present in *Riftia* (Bright et al., 2000; Felbeck
178 and Jarchow, 1998). Our calculated $\delta^{13}C$ ratios might support this theory (SOM3).

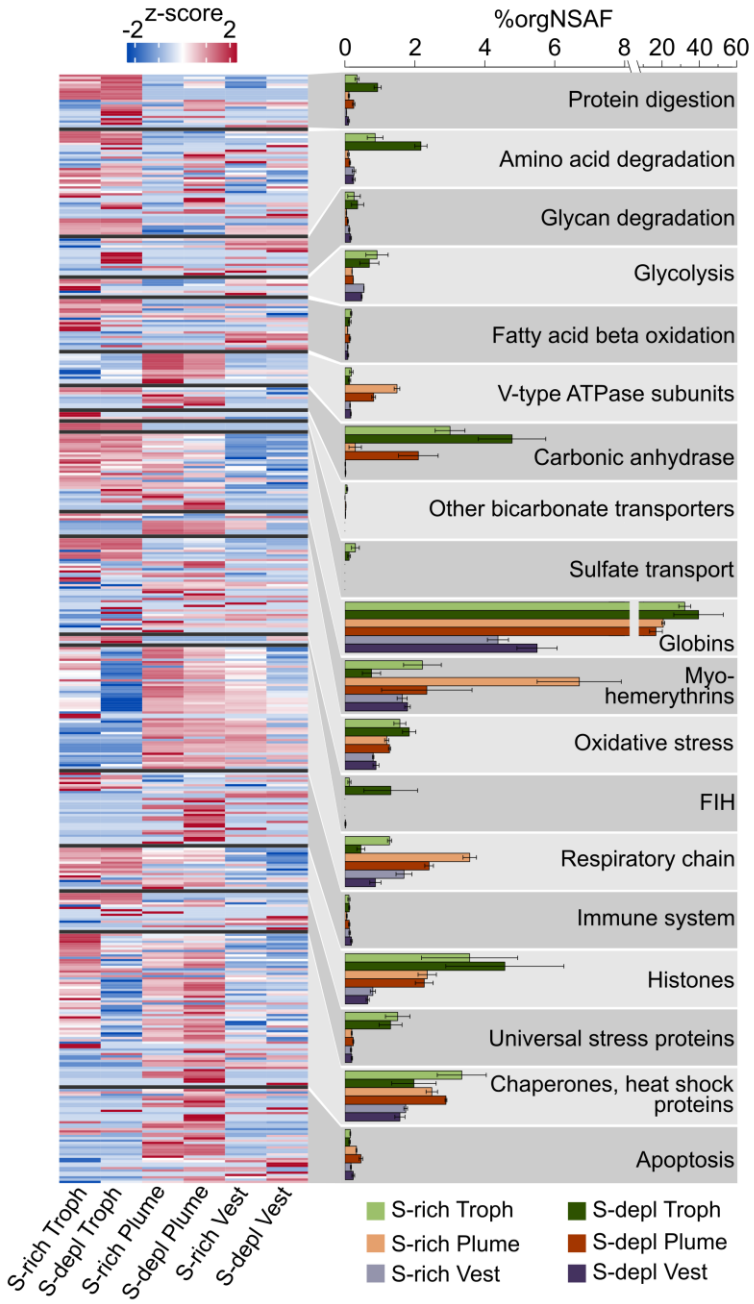


Figure 2: Functional groups of selected *Riftia* host proteins and their relative abundances in tissue samples. The heatmap shows log-normalized, centered and scaled protein abundances. The bar chart shows summed up abundances (%orgNSAF) of the proteins in the respective category. Error bars indicate standard error of the mean. Note the different scaling in the right part of the x-axis. The "Chaperones, heat shock proteins" category also includes chaperonins and Clp proteases. FIH: factor inhibiting hypoxia-inducible factor 1α. S-depl: S-depleted. To view the list of all identified proteins, including their abundances, see Supp. Table S1. (The table can be filtered for the same main or sub categories as presented in this figure; these categories are labelled with "X" in column "Figure 2"). Vest: vestimentum. Troph: trophosome.

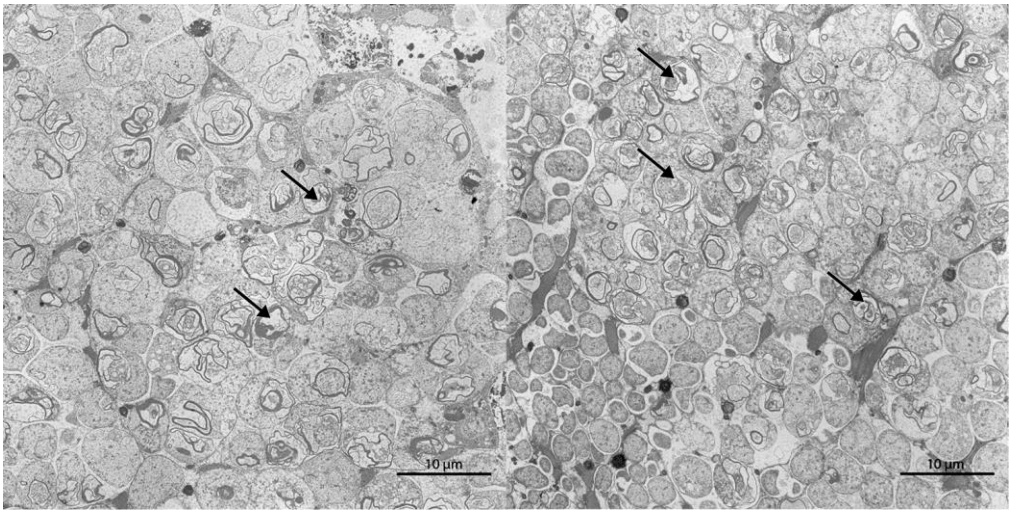


Figure 3: Transmission electron micrographs of *Riftia* trophosome tissue sections. Cell degradation is indicated by the presence of lamellar bodies (black arrows). Brightness and contrast of the micrographs were adjusted for visual clarity. Scale bar: 10 µm

Table 1: Proteins which are putatively involved in symbiont digestion and which had significantly higher abundances in trophosome samples than in other tissues of S-rich and S-depleted specimens.

Accession	Description	Sig in S-rich Troph	Sig in S-depl Troph	Secreted/ membrane?*
Protein digestion				
Host_DN32373_c0_g1_i1::g.193014	Cathepsin Z	x	x	M
Host_DN34261_c0_g1_i1::g.35886	Cathepsin B	x	x	S
Host_DN38047_c1_g1_i1::g.177385	Cathepsin Z	x	x	M
Host_DN41150_c0_g1_i1::g.101468	Cathepsin L1	x	x	S
Host_DN34118_c0_g1_i3::g.155432	Digestive cysteine proteinase 2	x	x	S
Host_DN39514_c3_g1_i1::g.201492	Legumain	x	x	S
Host_DN34848_c0_g1_i1::g.215091	Dipeptidyl peptidase 1	o	x	S
Amino acid degradation				
Host_DN37934_c0_g3_i4::g.212722	4-hydroxyphenylpyruvate dioxygenase	x	x	S
Host_DN35553_c0_g1_i1::g.72896	Maleylacetoacetate isomerase	x	x	-
Host_DN37934_c0_g3_i6::g.212725	4-hydroxyphenylpyruvate dioxygenase	x	x	-
Host_DN40417_c0_g1_i7::g.93374	D-aspartate oxidase	x	x	possibly M
Host_DN41135_c1_g1_i1::g.101501	Homogentisate 1,2-dioxygenase	x	x	-
Host_DN39303_c6_g1_i3::g.66273	Urocanate hydratase OS=Mus musculus GN=Uroc1 PE=1 SV=2	x	x	-

Accession	Description	Sig in S-rich Troph	Sig in S-depl Troph	Secreted/ membrane?*
Host_DN37934_c0_g3_i11::g.212729	4-hydroxyphenylpyruvate dioxygenase	o	x	-
Host_DN39293_c0_g3_i16::g.11113	Histidine ammonia-lyase	o	x	-
Host_DN41135_c1_g1_i2::g.101503	Homogentisate 1,2-dioxygenase	o	x	-
Host_DN40306_c1_g4_i8::g.129962	Aminoacylase-1	o	x	-
Glycan degradation				
Host_DN36692_c1_g2_i4::g.169924	Lysosomal alpha-glucosidase	x	x	M/possibly S
Host_DN36692_c1_g2_i3::g.169923	Glucoamylase 1	o	x	-
Host_DN37016_c0_g1_i1::g.156600	Lysosomal alpha-mannosidase	o	x	S
Fatty acid beta oxidation				
Host_DN34874_c0_g1_i9::g.215370	Propionyl-CoA carboxylase beta chain, mitochondrial	x	o	-
Host_DN41664_c1_g5_i6::g.166806	Peroxisomal bifunctional enzyme	o	x	-

Sig: Significance (x: significant, o: non-significant), Troph: Trophosome, S-depl: S-depleted. *Subcellular localization (M: membrane-associated, S: secreted) was predicted using Phobius, TMHMM, and SignalP. "Possibly": localization prediction based on one tool only.

Riftia dedicates a substantial part of its proteome to provisioning the symbionts with O_2, sulfide and CO_2

We found highly abundant and diverse globins, myohemerythrins, V-type ATPase subunits and carbonic anhydrases in the host proteome (Figure 2), indicating that *Riftia* dedicates a substantial part of its proteome to provisioning the symbiont with all necessary substrates for chemosynthesis.

Globins made up about one third of all trophosomal host proteins and one fifth of the total plume proteome (Figure 2), with extracellular hemoglobins being particularly abundant (in sum 32-40%orgNSAF in trophosome and 17-21% in plume samples). *Riftia* has three distinct extracellular hemoglobins composed of globin chains and, in the case of the hexagonal bilayer hemoglobin, globin linker chains (Flores et al., 2005; Zal et al., 1996, 1998). We detected several of these subunits, including isoforms that are (to our knowledge) hitherto undescribed (Supp. Table S1). *Riftia's*

212 extracellular hemoglobins have been shown to bind both O_2 and sulfide (Flores et al., 2005, reviewed
213 in Bailly and Vinogradov, 2005; Hourdez and Weber, 2005). Abundant hemoglobins in the highly
214 vascularized plume therefore ensure efficient uptake of these compounds for transport to the
215 symbionts. The symbionts are microaerophilic (Fisher et al., 1989), and simultaneous reversible O_2-
216 and sulfide-binding to abundant hemoglobins in the trophosome therefore not only provides the
217 bacteria with chemosynthetic substrates and prevents spontaneous sulfide oxidation, but also
218 protects the symbionts from oxygen. (See SOM4 for hemoglobins as a means of protecting the host
219 from sulfide toxicity and for other sulfur metabolic pathways in the host.) In addition to extracellular
220 hemoglobins, we identified four low-abundance (0.002-0.084%orgNSAF) globins that are probably
221 intracellular and might store O_2 (SOM5).

222 Besides hemoglobins, myohemerythrins were detected in all tissues, with particularly high
223 abundances of 6.7%orgNSAF in S-rich plumes. With their comparatively high oxygen-binding
224 capacity (Mangum, 1992), hemerythrins could facilitate oxygen uptake from the environment into
225 the plume, and are possibly also involved in O_2 storage and intracellular transport in *Riftia*. Moreover,
226 the abundance distribution of the nine detected myohemerythrins suggests a tissue-specific function
227 (SOM6).

228 V-type ATPase subunits were found with highest total abundances of up to 1.5%orgNSAF in *Riftia*
229 plumes (Figure 2), and almost all of the detected subunits were significantly more abundant or
230 exclusively detected in the plumes. V-type ATPases have a pivotal function in regulating internal pH
231 and CO_2 uptake (De Cian et al., 2003a) and thus in symbiont provisioning. The high energy demand
232 of V-type ATPase-dependent pH regulation could be met via a relatively higher respiration activity in
233 the plume, as indicated by comparatively high total abundances of respiratory chain proteins (Figure
234 2), ATP synthase and mitochondrial ribosomes in this tissue. Additionally, carbonic anhydrase (CA),
235 another important enzyme for CO_2 uptake, was detected in all tissues. While we observed tissue-
236 specific abundance patterns of individual CAs (Supp. Figure S4, SOM7), overall CA abundance was

highest in the trophosome (Figure 2). CA facilitates CO_2 diffusion into the plume by converting it to HCO_3^- (De Cian et al., 2003a; Goffredi et al., 1999), and likely back-converts the HCO_3^- to CO_2 for fixation by the symbionts in the trophosome. Our analysis suggests that three of the *Riftia* CAs could be membrane-bound (SOM7), which might facilitate CO_2 diffusion into the bacteriocytes by converting HCO_3^- to CO_2 in the direct cell vicinity (De Cian et al., 2003b; Sanchez et al., 2007a). Transport of HCO_3^- to the bacteriocytes could be mediated by bicarbonate exchangers, which we identified in trophosome and plume samples.

While carbon for fixation by the *Riftia* symbiont is likely mainly transported in the form of CO_2/HCO_3^-, the host may additionally pre-fix CO_2 into organic C_4 compounds which are then transported to the symbiont (Felbeck, 1985). We did identify host phosphoenolpyruvate carboxykinase and pyruvate carboxylase, which could be involved in this process (SOM8).

Riftia's nitrogen metabolism depends less on the symbiont than previously assumed

Riftia symbionts supply their host not only with carbon and energy sources, but also with ammonium produced by bacterial nitrate reduction (Figure 4, SOM9). However, with regard to the subsequent metabolization of organic nitrogen, the host might be more self-sufficient than previously thought: Previous biochemical analyses suggested that only the symbiont, but not the host, can *de novo* synthesize pyrimidines (Minic et al., 2001) and produce polyamines (Minic and Hervé, 2003). In contrast to those studies, we found the multifunctional CAD protein (carbamoyl-phosphate synthetase 2, aspartate transcarbamoylase, and dihydroorotase), in the *Riftia* host metatranscriptome, suggesting that the host can catalyze the first steps of pyrimidine synthesis. As we did not detect CAD protein on the protein level, expression levels and associated activities in the host are likely rather low, and most of the pyrimidine demand could be satisfied by digesting symbionts. In addition, we found key genes involved in polyamine synthesis in the hosts' metatranscriptome and partially also detected the respective proteins in the hosts' metaproteome

(Figure 4). Our results suggest that while both *Riftia* symbiosis partners can synthesize spermidine, in fact only the host is able to generate spermine. Host spermidine synthase and spermine synthase were exclusively detected in trophosome samples in our study, suggesting that the polyamines produced by these proteins could have a role in symbiont-host interactions. They could, for example, be involved in restricting the symbiont to its cell compartment, i.e. the bacteriocyte vesicle, as suggested for bacterial pathogens (SOM10). In addition, only the host seems to possess a full urea cycle and might degrade not only its own, but also nitrogen-containing metabolites of the symbiont (SOM9). These results show that the symbiont provides the host with necessary metabolic energy and building blocks for biosynthesis, but that the host has also retained key biosynthetic capacities for N-containing organic compounds.

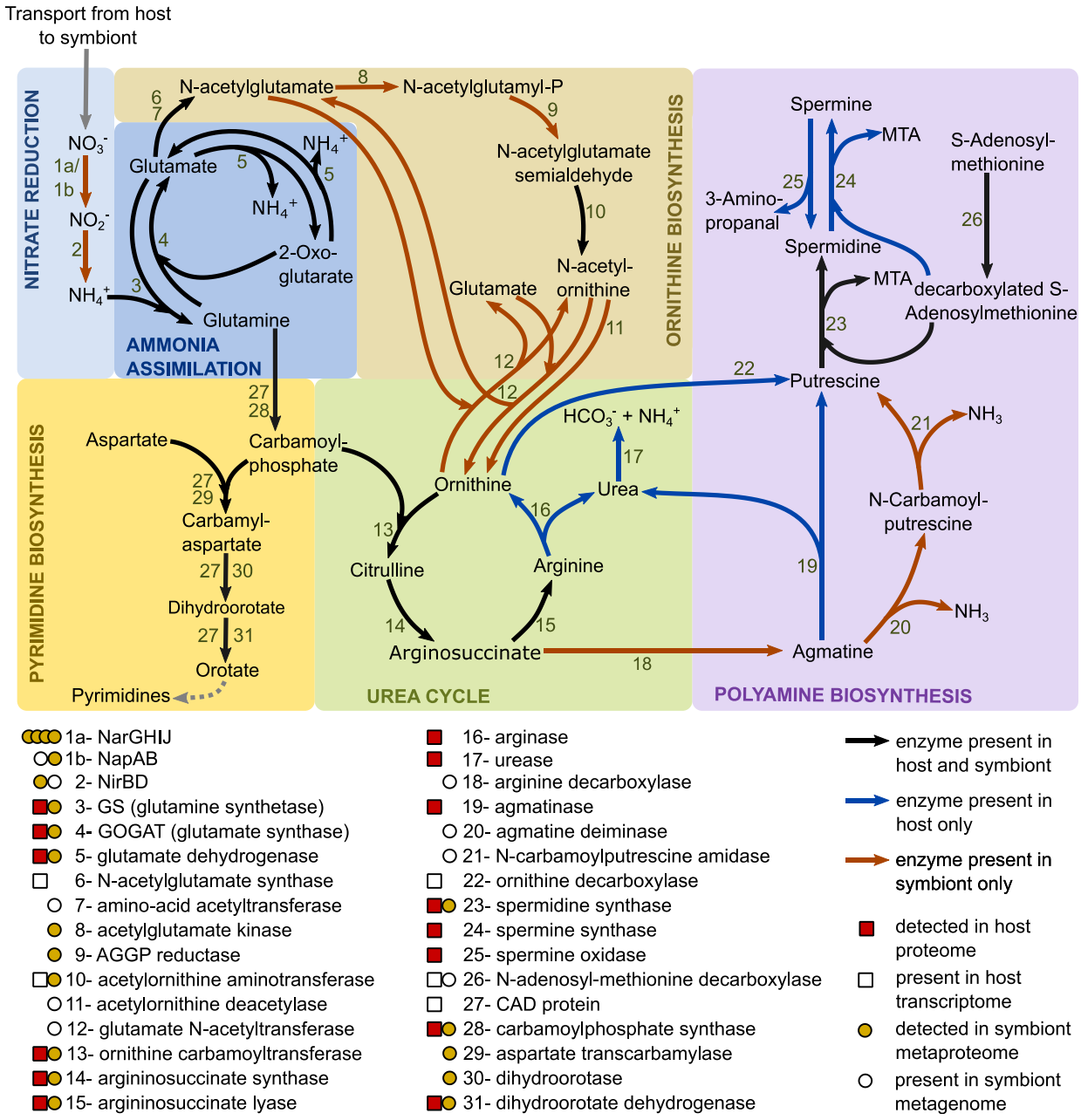


Figure 4: Main nitrogen metabolic pathways in the *Riftia* symbiosis. AGGP reductase: N-acetyl-gamma-glutamyl-phosphate reductase, CAD protein: multifunctional carbamoyl-phosphate synthetase 2, aspartate transcarbamoylase, and dihydroorotase protein, MTA: 5'-methylthioadenosine. Note that the symbiont might also be capable of nitrate respiration (Hentschel and Felbeck, 1993; Markert et al., 2011), which is not depicted here.

Host strategies of symbiont maintenance

Riftia protects its symbiont from oxidative damage and may even generate hypoxic conditions in the trophosome

We found several reactive oxygen species (ROS)-scavenging enzymes (superoxide dismutase, peroxiredoxin, glutathione S-transferase), as well as proteins indicative of anaerobic metabolism and universal stress proteins with significantly higher individual abundance and in higher total amounts (summed %orgNSAF) in the trophosome compared to other tissues (Figure 2, SOM11). *Riftia*'s ROS-detoxifying enzymes probably not only protect the host, but also the microaerophilic symbiont against ROS. Upregulation of host proteins involved in ROS detoxification was previously shown in the *Wolbachia* symbiosis (Brennan et al., 2008; Zug and Hammerstein, 2015). Additionally, malate dehydrogenase was highly abundant in trophosomes. This enzyme is regularly observed in different invertebrates under anaerobic conditions (Hourdez and Lallier, 2007) and is involved in maintaining redox balance during anaerobiosis (Fields and Quinn, 1981). The trophosome might thus rely more on fermentative metabolism than on respiration, as also indicated by the overall lower abundance of host respiratory chain proteins in trophosome compared to other tissues of both, S-rich and S-depleted specimens. We also detected hypoxia-inducible factor 1-alpha inhibitors (factor inhibiting HIF1a; FIH) almost exclusively in trophosome samples, which further supports the idea that free oxygen concentrations in the trophosome are low. This is in line with the high oxygen-binding capacity of *Riftia* hemoglobins (Fisher et al., 1989; Hentschel and Felbeck, 1993), and with the suggestion of fermentative metabolism under hypoxic and even oxic conditions in *Riftia*, based on biochemical results (Arndt et al., 1998). Taken together, lower oxygen concentration in the trophosome, (partial) anaerobic host metabolism, and host ROS-detoxifying enzymes in this tissue would not only protect the symbionts from oxidative damage, but would additionally decrease the competition between the *Riftia* host and its symbionts for oxygen.

The *Riftia* immune system might be involved in symbiont population control

We detected several proteins which are potentially involved in a specific immune reaction of *Riftia* against its symbiont in the trophosome. Two bactericidal permeability-increasing proteins (BPIPs) were detected, one exclusively in the trophosome, the other only in the plume. BPIPs act specifically against Gram-negative bacteria, causing initial growth arrest and subsequent killing due to inner membrane damage (Elsbach and Weiss, 1998). In *Riftia*, BPIPs could be involved in keeping the symbiont population under control, e.g. as part of the digestion process or by preventing the symbionts from leaving their intracellular host vesicles. Likewise, in the *Vibrio*-squid symbiosis, BPIPs have been implied in restricting the symbiont population to the light organ (Chen et al., 2017). In addition to BPIPs, a pathogen-related protein (PRP) was present in all replicates of S-rich trophosome, but absent from all other tissues. In plants, pathogen-related proteins accumulate during defense responses against pathogens (reviewed in Van Loon and Van Strien, 1999). Pathogen-related proteins have also been described in nematodes (Asojo et al., 2005) and humans (Eberle et al., 2002), although their function remains elusive.

We also found that histones had overall higher abundance in *Riftia* trophosome than in other tissues. Four of these histones were significantly more abundant in trophosomes than in other tissues, and three additional histones were exclusively detected in trophosome samples (Supp. Table S1). Besides being crucial for DNA interactions, histones and histone-derived peptides can have antimicrobial effects (Cho et al., 2009; Park et al., 1998; Rose et al., 1998). A blastp search of the detected *Riftia* histones against the Antimicrobial Peptide (AMP) Database APD3 (Wang et al., 2016) gave hits for four of the *Riftia* histones (Supp. Table S3), stimulating the speculation that these histones may have antimicrobial properties. While AMP-like histone-derived peptides in the plume might be involved in defense against environmental microbes, the high abundance of histones in the trophosome could point to a function in host-symbiont interaction. Host-derived AMPs could, for example, be involved in controlling the symbiont's cell cycle. In their life cycle, the symbionts apparently differentiate from

329 actively dividing stem cells into growing, but non-dividing larger cells (Bright and Sorgo, 2003). As
330 various AMPs were shown to inhibit cell division or septum formation and to cause filamentous cell
331 morphologies (reviewed in Brogden, 2005), we speculate that *Riftia* AMPs may inhibit cell division
332 as well, e.g. via interaction with symbiont GroEL. Interaction between a host AMP and a symbiont
333 GroEL has been proposed to lead to cell elongation of bacterial weevil symbionts (Login et al., 2011).
334 A role of histones and histone-derived peptides in immune system responses has been described or
335 suggested in various other organisms, including catfish (Park et al., 1998), Komodo dragons (Bishop
336 et al., 2017), toads (Cho et al., 2009) and humans (Rose et al., 1998).

337 Beyond individual immune system proteins, we did not observe a general immune response of *Riftia*
338 against its symbiont (which is not surprising, as the symbionts are contained inside host vesicles).
339 This indicates that the host immune system does not play a major role in controlling symbiont
340 population size. More likely, symbiont population control might to a large part be a result of digestion
341 of symbionts (a "mowing" process), which effectively prevents the symbionts from escaping their
342 compartments and/or overgrowing the host. Nevertheless, the immune system might be involved in
343 phage protection and symbiont recognition during establishment of the symbiosis (SOM12).

344

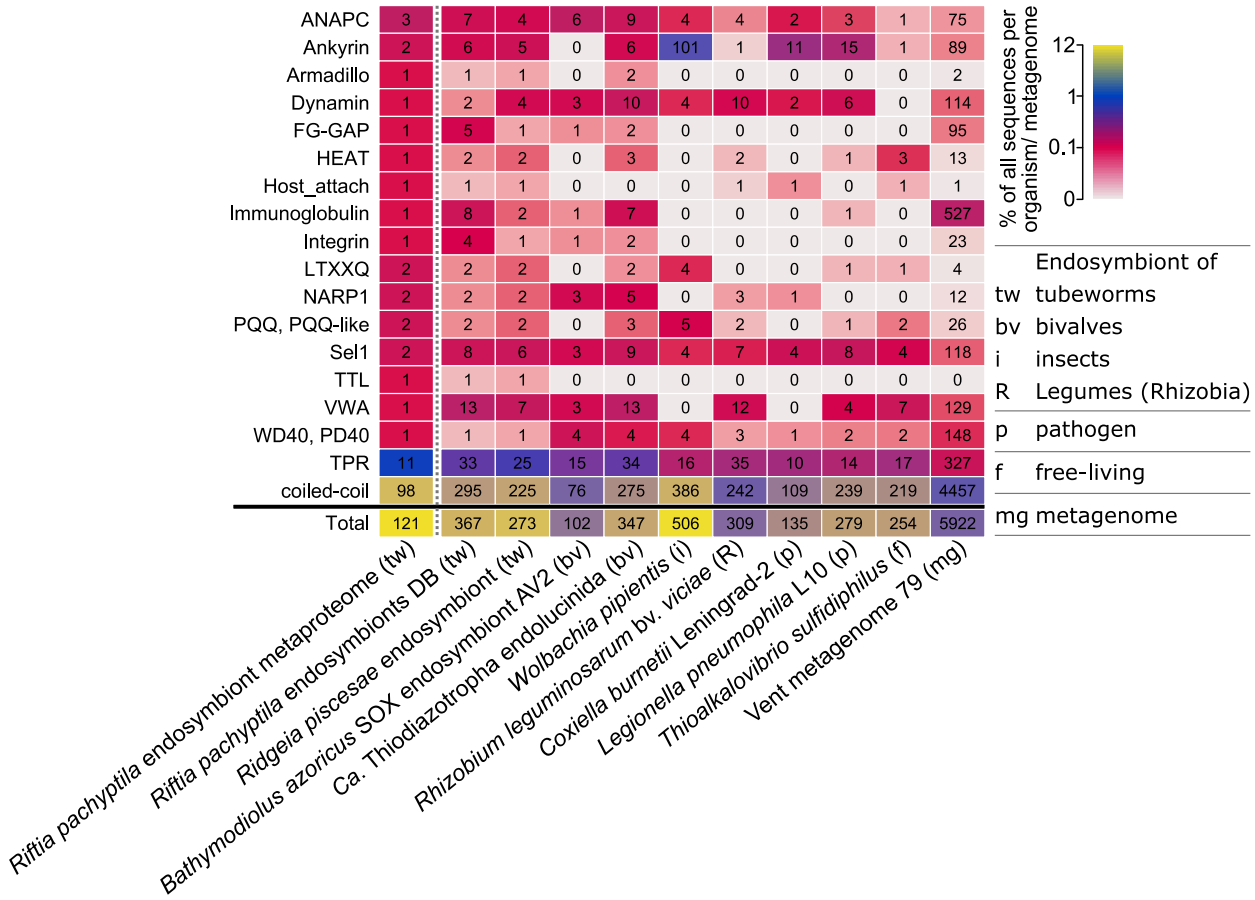
345 Symbiont persistence mechanisms

346 Eukaryote-like protein structures in the symbiont might be involved in host
347 communication

348 The metagenome of the *Riftia* symbiont *Ca.* E. persephone encodes several protein groups with
349 possible roles in symbiont-host interactions, including eukaryote-like protein (ELP) structures, as
350 revealed by our SMART analysis (Supp. Table S4). We detected more than 100 of these symbiont
351 proteins in the trophosome samples (Figure 5), which points to a symbiosis-relevant function.

352 Among the ELPs detected in the symbiont metaproteome were two ankyrin repeat-containing
353 proteins, which contain a signal peptide and are therefore likely secreted (predicted by Phobius,
354 http://phobius.sbc.su.se/). Ankyrin repeats were found to mediate protein-protein interactions (Li
355 et al., 2006). In the sponge *Cymbastela concentrica*, symbiont ankyrins were proposed to interact with
356 the eukaryote's phagocytosis system: The symbiont ankyrins were heterologously expressed in *E.*
357 *coli* and led to inhibition of phagocytosis by amoebae (Nguyen et al., 2014). Likewise, a secreted
358 *Legionella pneumophila* ankyrin protein apparently interferes with host endosome maturation (Pan
359 et al., 2008). The *Ca.* E. persephone ankyrin repeat-containing proteins could therefore directly
360 interact with host proteins as well, e.g. to modulate endosome maturation, and thus to interfere with
361 symbiont digestion by the host. Similarly, proteins with tetratricopeptide repeat (TPR)/Sel1
362 domains, which we also detected in the *Ca.* E. persephone metaproteome, have been shown to impact
363 phagocytosis by amoeba (Reynolds and Thomas, 2016).

364 The *Riftia* symbiont furthermore encodes eukaryote-like proteins of the tubulin-tyrosine ligase
365 family (TTL proteins). These proteins post-translationally modify tubulin and thus interact with the
366 eukaryotic cytoskeleton (Prota et al., 2013). We found one TTL protein in the *Ca*. E. persephone
367 metaproteome. Other protein groups which are involved in protein-protein interactions in
368 eukaryotes, e.g. with cytoskeletal proteins, and which we detected in *Ca.* E. persephone, include
369 armadillo repeat proteins (Coates, 2003) and HEAT repeat-containing proteins (Yoshimura and
370 Hirano, 2016). As several of the protein structures analyzed here are also found in other mutualistic
371 symbionts and pathogens (SOM13, Supp. Table S4), it is conceivable that parallels exist between
372 interaction processes of mutualistic and pathogenic associations, and that the *Riftia* symbiont
373 employs a strategy similar to that of pathogens to communicate with its host on the molecular level.



374

375 Figure 5: Selected domains with eukaryote-like structures and with putative functions in symbiont-
376 host interactions in the *Riftia* symbiont and in selected other organisms and metagenomes. Color
377 scale: percentage of genes/proteins containing the respective domain relative to all gene/protein
378 sequences in this organism or metagenome. Numbers: total number of genes/proteins containing the
379 respective domain. For an overview of all analyzed organisms and domains see Supp. Figure S5. For
380 details on the organisms see Supp. Table S5. For further information about the selected protein
381 groups see Supp. Table S4. '*Riftia pachyptila* endosymbiont metaproteome' refers to the *Riftia*
382 symbiont proteins detected in this study.

383

384 Symbiont membrane proteins may export effector proteins into host cells and lead to

385 strain adaptation

386 We detected various outer membrane-related proteins in the *Ca.* E. persephone proteome, including

387 a porin (Sym_EGV52132.1), which was one of the most abundantly expressed symbiont proteins, and

388 12 type IV pilus (T4P) system proteins (PilQ, PilF, PilC, PilBTU, PilM, PilN, PilP, FimV, PilH, PilY1). Five

389 additional T4P structure proteins were encoded in the metagenome (*pilVWXE*, *pilO*). These proteins

390 are in direct contact with the host cells, and therefore likely involved in interactions between both
391 symbiosis partners, including such processes that facilitate the symbiont's persistence inside the host
392 cells.

393 The abundant symbiont porins could transport effector molecules, e.g. to modulate digestion by the
394 host. A role of porins in effector transport during symbiosis has been hypothesized for the *Vibrio*
395 *fischeri* OmpU, a channel protein that is important for symbiont recognition by the squid host
396 (Nyholm et al., 2009).

397 The T4P system is a complex structure, which, in *Pseudomonas aeruginosa*, comprises more than 40
398 proteins, including structural and regulatory proteins (Leighton et al., 2015). It can have several
399 functions in different species: adhesion, secretion and natural transformation (Davidson et al., 2014;
400 Hager et al., 2006; Leighton et al., 2015; Stone and Kwaik, 1999). As the *Ca.* E. persephone T4P system
401 is likely not involved in adhesion to host cells during symbiosis (although it might be during the initial
402 infection), it could participate in protein secretion and/or natural transformation. The *Riftia*
403 symbiont's T4P system could export putative effector proteins (e.g. ankyrins, SET domain proteins,
404 SOM13, SOM14) for host interactions. Interestingly, in the pathogen *Francisella tularensis* ssp.
405 *novicida*, a T4P structure is involved in secretion of infection-moderating proteins (Hager et al.,
406 2006).

407 Besides their putative function in effector protein export, symbiont membrane proteins may also lead
408 to bacterial strain adaptation. The *Riftia* symbiont population is polyclonal, i.e. consists of several
409 distinct strains (Polzin et al., 2019). T4P system-mediated exchange of genetic material between
410 different symbiont strains would add to this diversity in the symbiosis and might additionally enable
411 exchange of symbiosis-related genes within the free-living *Ca.* E. persephone population. Natural
412 transformation in symbionts has only recently been shown for *V. fischeri* in culture (Pollack-Berti et
413 al., 2010) and the earthworm symbiont *Verminephrobacter eiseniae*, which likely employs a T4P

414 structure for DNA uptake (Davidson et al., 2014). As microbial cell densities are comparatively high
415 in eukaryote-prokaryote mutualisms, natural transformation in these systems might actually be
416 more common than previously recognized. The proposed DNA uptake by the *Riftia* symbiont may not
417 only facilitate exchange between symbiont strains, but may also promote horizontal gene transfer
418 between host and symbiont, e.g. of eukaryote-like proteins. This hypothesis, as well as the
419 speculation that *Ca.* E. persephone might be capable of conjugation (SOM14) certainly warrant
420 further investigations.

421 S availability affects symbiotic interactions in *Riftia*

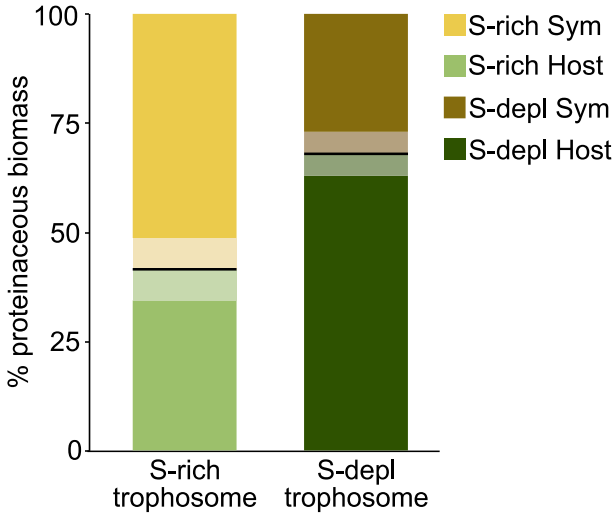
422 S-depleted *Riftia* hosts digest more symbionts than S-rich specimens

423 We compared the metaproteomes of *Riftia* specimens with and without stored sulfur (i.e., energy-
424 rich vs. energy-depleted specimens) to examine how energy availability impacts symbiotic
425 interactions. Metabolite transfer is apparently especially influenced by the energy regime: The host
426 supposedly relies more on symbiont digestion in times of S shortage. Proteinaceous symbiont
427 biomass was notably lower in S-depleted trophosomes (32%) than in S-rich trophosomes (58%;
428 Figure 6). Simultaneously, overall abundances for several groups of host digestive enzymes were
429 higher in S-starved trophosomes (Figure 2), and a number of individual host proteins were
430 significantly more abundant in these S-depleted samples, such as enzymes involved in protein
431 digestion (including cathepsin B), amino acid degradation, the late-endosome protein Rab7 and
432 histones (Supp. Table S1). One reason for this supposed increase in symbiont digestion in S-depleted
433 trophosomes could be a lower nutritional value of the energy-depleted symbionts. S-depleted
434 symbionts have lower abundances of enzymes involved in sulfur oxidation, probably due to lower S
435 availability. Therefore, less energy might be available for biosynthesis under S depletion, rendering
436 the symbiont less “nutritious” for the host. S-depleted hosts may thus have less energy available,
437 despite increased symbiont digestion. This idea is supported by the observation that host proteins

438 involved in the energy-generating glycolysis, TCA cycle, respiratory chain, ATP synthesis and
439 biosynthetic pathways were less abundant in S-depleted trophosomes than in S-rich trophosomes.
440 Potentially, increased symbiont digestion under S-depleted conditions is necessary for the host to
441 satisfy its basal metabolic demand. Concomitant with the postulated lower nutritional value of S-
442 depleted symbionts, the Calvin cycle key enzyme RubisCO had an about 10-fold lower abundance in
443 S-depleted symbionts. Abundance of the rTCA cycle key enzyme ATP citrate lyase (EGV51152.1), on
444 the other hand, was slightly higher in S-depleted symbionts than in S-rich symbionts, albeit only 1.4-
445 fold. Under S-depleted conditions, symbionts apparently rely relatively more on the rTCA cycle,
446 which is more energy-efficient than the Calvin cycle (Markert et al., 2007). The Calvin cycle could be
447 used in addition to the rTCA cycle under favorable conditions to maximize overall carbon fixation.
448 Moreover, symbiont enzymes involved in translation were overall more abundant in S-rich
449 trophosomes than in S-depleted trophosomes. Less protein biosynthesis in S-depleted symbionts
450 would not only impact the nutritional value of these symbionts, but additionally directly decrease the
451 proteinaceous symbiont biomass. The reason for the lower proteinaceous biomass of symbionts in
452 S-depleted trophosomes is therefore probably two-fold: The host digests more symbionts and the
453 symbionts produce less biomass compared to energy-rich trophosomes.

454 These findings are in contrast to previous results (Scott et al., 2012), which showed no significant
455 differences in autotrophic activity and symbiont abundance between *Riftia* specimens from high- vs.
456 low-sulfide habitats. Possibly, increased symbiont digestion is a short-term adaptation to fluctuating
457 environmental conditions, whereas under long-term low-S conditions the symbiosis might adapt e.g.
458 by reduced growth rates. Decrease in symbiont abundance or total protein under energy-limiting
459 conditions has also been noted in *Bathymodiolus* (Stewart et al., 2005) and *C. orbicularis* bivalves
460 (Caro et al., 2009) as well as in *O. algarvensis* oligochaetes (Wippler et al., 2016). Relying on the
461 symbionts as nutrient source also under unfavorable conditions thus appears to be a common

462 symbiosis mechanism, which would ensure survival of the host and a subset of the symbiont
463 population, ultimately prolonging survival of the individual holobiont.



464

465 Figure 6: % proteinaceous biomass contributions of host and symbiont as calculated from our
466 metaproteomics results (Kleiner et al., 2017). Bold lines indicate the mean, semitransparent areas
467 indicate standard error of the mean. Sym: Symbiont, S-depl: S-depleted

468

469 ## S availability influences CO_2 uptake, pH regulation and O_2 regime in the *Riftia* host

470 S-depleted hosts seem to invest relatively more biosynthetic capacities in CO_2 uptake and less in pH
471 regulation, and their trophosomes are supposedly less hypoxic than those of S-rich hosts (SOM11,
472 SOM15). At the same time, S availability appeared to have little influence on non-symbiont-related
473 processes in the host, as only very few (i.e. < 10) individual proteins significantly differed in
474 abundance between S-rich and S-dark plume and vestimentum samples. This indicates that the host's
475 metabolism is very well buffered against changes in environmental conditions.

476 ## Higher digestion pressure might result in symbiont countermeasures

477 In S-depleted *Riftia* specimens, a putative *Ca.* E. persephone dodecin was significantly more abundant
478 than in S-rich specimens. This protein might be involved in protecting the symbiont against oxygen
479 and/or digestion stress (SOM14). A symbiont porin, which was also significantly more abundant in

480 S-depleted specimens, might be involved in counteracting the supposedly higher digestion pressure

481 (see above and SOM14).

Conclusion

To fully understand the biology of organisms, it is crucial to study them together with their symbiotic partners as holobionts (Gilbert et al., 2012). Given its low complexity, high specificity and extreme dependence of the host on the symbiont, the association of *Riftia* and its bacterial partner serves as an excellent system to study mutualistic host-microbe interactions. While *Riftia* lives in a unique and remote environment, many of the interactions we identified, like symbiont digestion by the host, high host investment in substrate transfer to the symbiont, host-directed symbiont population control, and eukaryote-like symbiont proteins that could interact with the hosts' molecular machinery, seem to be critical in other symbiotic associations as well, including insects, mussels and oligochaetes. These interactions might therefore represent common principles among evolutionarily diverse mutualistic animal-microbe associations.

Our study provides access to the *Riftia* host transcriptome and protein sequences and thus paves the way for future research on host-microbe interactions in *Riftia* and other systems. Promising research directions include the elucidation of protein functions, e.g. of *Riftia* immune system proteins and symbiont eukaryote-like proteins by heterologous gene expression and biochemical assays in model systems. Moreover, our work stimulates future in-depth studies of the molecular mechanisms involved in recognition of both partners during the initial infection of *Riftia* larvae by free-living symbionts. Putative differences between *Riftia*'s short- and long-term adaptation strategies in response to changing environmental conditions also warrant further investigation.

Material and Methods

Sampling

Riftia tissue samples were obtained during several research cruises in 2008, 2014 and 2017 with RV Atlantis to the deep-sea hydrothermal vent fields on the East Pacific Rise at 9°50' N, 104°17' W. *Riftia* specimens were collected by the human occupied vehicle Alvin or the remotely operated vehicle Jason in approximately 2,500 m water depth. Sampling dates for all *Riftia* tissue samples for proteomics, transcriptomics and transmission electron microscopy (TEM) are summarized in Supp. Table S6. Different specimens were used for proteomics, transcriptomics and TEM. *Riftia* specimens were dissected onboard and tissue samples stored at -80 °C. The lamellae of the tentacular crown were shaved off to provide "plume" samples, trophosome samples were dissected from whole trophosome, body wall samples were retrieved and washed after removal of the trophosome, and vestimental samples were cut off from the lateral portions of the vestimentum. Specimens were classified into sulfur-rich (S-rich), S-depleted and medium S according to their trophosome color (yellow/light green, dark green/black, or medium green, respectively).

Extraction of whole-tissue RNA

RNA was extracted from a total of 22 tissue samples from 9 specimens (6 x trophosome, 6 x body wall, 5 x plume, 5 x vestimentum, see Figure 1). Tissue samples were homogenized by bead-beating with lysing matrix D (MP Biomedicals) in 1 ml TRIzol® (Thermo Fisher Scientific; 3x 6.5m/s for 30 s, 3 min cooling on ice in between). After 5 min acclimatization to room temperature, samples were applied onto QIAShredder columns (Qiagen) and centrifuged (16,000 x g, 3 min, 4 °C). Afterwards, RNA was isolated from the aqueous flow-through according to the TRIzol extraction protocol, with the modification that samples were centrifuged for 20 min at 12,000 x g and 4 °C for phase separation. Glycogen was added for RNA precipitation. RNA was washed twice with 75% ethanol and purified using the Norgen RNA Clean-Up and Concentration kit according to the manufacturer's Protocol A,

528 including DNA removal with DNase (Qiagen). Quality of extracted RNA was assessed using Nanodrop
529 (Thermo Fisher Scientific) and Bioanalyzer (Agilent) analyses.

530 Transcriptome sequencing and assembly

531 *Transcriptome sequencing*

532 Transcriptome sequencing was performed employing the TruSeq stranded mRNA (poly A-based)
533 library protocol (Illumina) on a HiSeq 4000 (Illumina), according to the manufacturer's guidelines.

534 *Transcriptome assembly*

535 High-throughput paired-end Illumina sequencing resulted in an average of about 26 million reads
536 per end per library (min 16,045,121 reads per end, max 31,318,532 reads per end, 95% CI
537 1,673,590). After de-multiplexing and quality-checking of reads in FastQC v0.11.5 (Andrews, 2010),
538 we trimmed low quality bases and adapters with Trimmomatic v0.32 (Bolger et al., 2014) using the
539 settings ILLUMINACLIP:AllAdapters.fa:2:30:10 SLIDINGWINDOW:4:20, and LEADING:5 TRAILING:5
540 HEADCROP:15 MINLEN:75. Although bacterial mRNA does not possess a polyA tail, previous
541 research has shown that bacterial reads can still be present in polyA-enriched RNA-Seq libraries
542 (Egas et al., 2012). To filter out potential symbiont contaminations from our host transcriptomes, we
543 used the Bowtie 2 v2.2.9 aligner (Langmead and Salzberg, 2012) in very-sensitive mode to map the
544 quality-filtered paired-end reads against the published genomes of the endosymbionts of *Riftia*
545 ("Riftia1", NCBI locus tag prefix RIFP1SYM, and "Riftia2", locus tag prefix RIFP2SYM) and *Tevnia*
546 *jerichonana* (Gardebrecht et al., 2012). Unmapped paired-end reads were subsequently extracted
547 using SAMtools v1.4.1 (Li et al., 2009). Potential environmental sequence contaminations from
548 sample handling were excluded with DeconSeq v0.4.3 (Schmieder and Edwards, 2011) using
549 coverage and identity thresholds of 0.90 and 0.95, respectively. The decontaminated host reads were
550 normalized and assembled with Trinity v2.3.2 (Grabherr et al., 2011). To optimize the transcriptome
551 assembly we performed four different assemblies with different parameters and input files: 1) only

552 paired reads, 2) paired and unpaired reads, 3) only paired reads plus jaccard-clip option (to reduce
553 chimeras), 4) paired and unpaired reads plus jaccard-clip option.

554 To assess the completeness of the different assemblies we compared our transcriptomes against the
555 BUSCO v2.0 eukaryote and metazoan orthologous datasets (Simão et al., 2015). Overall, the best
556 results in terms of transcriptome completeness and quality were obtained by the assembly approach
557 using paired and unpaired reads plus jaccard-clip option (Supp. Table S7). This dataset was used for
558 all further analyses.

559 *Open reading frame (ORF) prediction*

560 TransDecoder v3.0.1 (Haas et al., 2013) was used to identify coding regions in the assembled
561 transcripts. To improve ORF prediction, we examined all candidate ORFs for homology to known
562 proteins by searching the Swiss-Prot (http://www.uniprot.org) and Pfam (Finn et al., 2016)
563 databases (downloaded January 3, 2017) with BLASTP (Altschul et al., 1990, e-value 1e-05) and
564 HMMER3 (Eddy, 2009), respectively. ORFs that were longer than 100 amino acids and/or had a
565 database entry were retained. The FASTA headers of the TransDecoder output files were modified
566 with a custom PERL script to include the BLASTP protein annotations.

567 Database generation

568 A common database for protein identification of *Riftia* host and symbiont was generated. To this end,
569 host protein sequences were clustered at 95% identity with CD-HIT v. 4.6 (Huang et al., 2010). For
570 symbiont sequences, the three proteomes of the *Riftia*1, *Riftia*2 and *Tevnia* symbiont (Gardebrecht et
571 al., 2012) were used. *Riftia*1 was used as basis for clustering the symbiont protein sequences with
572 CD-Hit-2D (Huang et al., 2010). Subsequently, the combined symbiont database was clustered at 95%
573 identity. Identifier prefixes were added to distinguish between host and symbiont sequences for
574 Calis-p (Kleiner et al., 2018, see below). Host and symbiont databases were concatenated and the

575 cRAP database containing common laboratory contaminants (The Global Proteome Machine
576 Organization) was added. The final database contained 71,194 sequences.

577 Proteomics sample preparation and analysis

578 For metaproteomics analysis, we used three biological replicates per tissue (trophosome,
579 vestimentum, plume) and condition (specimens with S-rich and S-depleted trophosomes), which
580 resulted in a total of 18 samples. Tissues were disrupted by bead-beating for 45 s at 6.0 m/s with
581 lysing matrix D tubes (MP Biomedicals) in SDT buffer (4% (w/v) sodium dodecyl sulfate (SDS), 100
582 mM Tris-HCl pH 7.6, 0.1 M dithiothreitol (DTT)), followed by heating to 95 °C for 10 min. Tryptic
583 peptides were generated following the FASP protocol of Wiśniewski et al. (2009) with minor
584 modifications as described by Hamann et al. (Hamann et al., 2016). Peptide concentrations were
585 determined with the Pierce Micro BCA assay (Thermo Scientific Pierce) according to the
586 manufacturer's instructions. The tryptic digest was desalted on-line during LC-MS/MS analysis.

587 All samples were analyzed by 1D-LC-MS/MS as in Hinzke et al. (2019), using 4 h gradients. Samples
588 were analyzed in a randomized block design (Oberg and Vitek, 2009) and run in technical triplicates.
589 Two technical replicate runs were acquired with a 50 cm analytical column, one with a 75 cm
590 analytical column. To standardize the stable isotope fingerprinting (SIF) analysis (Kleiner et al.,
591 2018), human hair was measured in technical duplicate alongside the *Riftia* samples in the replicate
592 run using a 75 cm column.

593 Proteomics data evaluation

594 *Protein identification, quantification and statistical analyses*

595 For protein identification, MS/MS spectra of combined technical triplicate runs were searched
596 against the combined host and symbiont database using the Sequest HT node in Proteome Discoverer
597 version 2.0.0.802 (Thermo Fisher Scientific) as in Kleiner et al. (2018). For protein abundance
598 estimates, normalized spectral abundance factors (NSAFs, Zybailov et al., 2006) were calculated per

599 sample and organism (%orgNSAF, Mueller et al., 2010). Statistical evaluation was performed based
600 on spectral counts using the edgeR package (Robinson et al., 2010) in R (R Core Team, 2017). The
601 edgeR package uses an overdispersed Poisson model for analysis of count data. Overdispersion is
602 moderated across proteins using empirical Bayes methods (Robinson et al., 2010). We employed a
603 false-discovery rate (FDR) of 0.05 to assign statistical significance to protein abundance differences.
604 For graphical representation, heatmaps were generated with the R package ComplexHeatmaps (Gu
605 et al., 2016) and intersection plots with the R package UpsetR (Lex et al., 2014). Protein biomasses of
606 host and symbiont were calculated as in Kleiner et al. (2017).

607 $\delta^{13}C$ values of *Riftia* symbiont and host were calculated from mass spectrometry data with Calis-p
608 (Kleiner et al., 2018) using one technical replicate LC-MS/MS run (75 cm analytical column). Human
609 hair was used as reference material.

610 *Protein annotations, functional characterization and categorization*

611 Besides the annotations included in the database, proteins where further characterized using the
612 online tools described in Supp. Table S8. Proteins were manually categorized into functional groups
613 based on their annotations and the information in the Uniprot (The UniProt Consortium, 2017), NCBI
614 (https://www.ncbi.nlm.nih.gov/) and InterPro (Finn et al., 2017) databases. We used the
615 Transporter Automatic Annotation Pipeline (TransAAP)
616 (http://www.membranetransport.org/transportDB2/TransAAP_login.html) of the TransportDB2
617 (Elbourne et al., 2017) and TCDB (Saier Jr et al., 2016) with gblast 2
618 (http://www.tcdb.org/labsoftware.php) to annotate transporters in the *Riftia*1 symbiont
619 metagenome database. To detect possible antimicrobial peptides (AMPs) among the host proteins,
620 we searched the detected host proteins against the antimicrobial peptide database APD3 (Wang et
621 al., 2016) using BLASTP (Altschul et al., 1990) in BLAST+ 2.7.1 (Camacho et al., 2009). Results were
622 filtered for %identity >75% and e-value < 0.005. We screened the *Riftia* proteome for homologs of

known autophagy-related *Drosophila melanogaster* proteins (as listed in Chang and Neufeld, 2010) by Blast-searching (BLASTP, Altschul et al., 1990) in BLAST+ 2.8.1, Camacho et al., 2009) the *Riftia* host proteome against the respective *Drosophila* amino acid sequences (Supp. Table S2).

SMART analysis of eukaryote-like and potential interaction domains

We used the SMART tool (Letunic and Bork, 2018) to screen the *Riftia* symbiont protein database for proteins and domains which could be involved in symbiont-host interactions. Structures which did not meet the threshold required by SMART were excluded, whereas overlapping features were included. We manually filtered the SMART annotations to find putative interaction-relevant structures based on the Pfam and SMART database information. To compare the *Riftia* symbiont with other host-associated (mutualistic or pathogenic) and free-living organisms, we also included domains not present in the *Riftia* annotations, but possibly relevant for host-bacteria interactions in other organisms based on the literature. All annotations we included are given in Supp. Table S4. The organisms and their proteome accession numbers we used for comparison can be found in Supp. Table S5. Proteins with structures that did not pass the threshold criterion in SMART were removed.

Multiple sequence alignments

We used the alignment tool MUSCLE provided by EMBL (https://www.ebi.ac.uk/Tools/msa/muscle/) for multiple sequence alignment of protein sequences. Alignments were verified visually.

Transmission electron microscopy (TEM)

The trophosome sample for TEM was fixed at room temperature for 1 h in fixative containing 4% paraformaldehyde, 1% glutaraldehyde, 10% sucrose in 50 mM HEPES (glutaraldehyde was added directly before use) and stored at 4 °C. The sample was washed three times with washing buffer (100 mM cacodylate buffer [pH 7.0], 1 mM $CaCl_2$, 0.09 M sucrose) for 10 min each step and treated with 1 % osmium tetroxide in washing buffer for 1 h at room temperature. After three additional washing

steps in washing buffer for 10 min each, the sample was dehydrated in a graded series of ethanol (30%, 50%, 70%, 90%, and 100%) on ice for 30 min each step. Afterwards, the material was subjected to stepwise infiltration with the acrylic resin LR White according to Hammerschmidt et al. (2005). Sections were cut with a diamond knife on an ultramicrotome (Reichert Ultracut, Leica UK Ltd), stained with 4 % aqueous uranyl acetate for 5 min and finally examined with a transmission electron microscope LEO 906 (Carl Zeiss Microscopy GmbH) at an acceleration voltage of 80 kV. The micrographs were edited using Adobe Photoshop CS6.

Data availability

The mass spectrometry proteomics data and the database have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al., 2016) partner repository with the dataset identifier PXD012439. Transcriptomics raw data have been deposited to the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) with the BioProject accession number PRJNA534438 (https://www.ncbi.nlm.nih.gov/sra/PRJNA534438). The datasets will be released upon acceptance of the manuscript in a peer-reviewed journal.

Acknowledgements

We thank the captains and crews of RV Atlantis, DSV Alvin, and ROV Jason for their excellent support during the cruises AT15-28, AT26-10, AT26-23, and AT37-12, which were funded through grants of the US National Science Foundation. We are grateful to Ruby Ponnudurai for sampling, to Jana Matulla and Annette Meuche for excellent technical assistance, to Marc Strous for supporting this project by providing access to the proteomics equipment, to Xiaoli Dong for help with database annotations and to Maryam Ataeian, Jackie Zorz and Angela Kouris for help with MS measurements. Sandy Gerschler did preliminary SMART analyses. Målin Tietjen and Lizbeth Sayavedra gave valuable input for RNA sample preparation. This work was supported by the German Research Foundation DFG (grant MA 6346/2-1 to S.M., grant BR 5488/1-1 to C.B.), the German Academic Exchange Service DAAD (T.H.), a fellowship of the Institute of Marine Biotechnology Greifswald (T.H.), the Canada Foundation for Innovation, the Government of Alberta and a Natural Sciences and Engineering Research Council of Canada NSERC through a Banting fellowship (M.K.), the US National Science Foundation (grants OCE-0452333, OCE-1136727, OCE-1131095, and OCE-1559198 to S.M.S) , and *The WHOI Investment in Science Fund* (S.M.S). P.R. is supported by a grant from DFG CCGA Comprehensive Center for Genome Analysis, Kiel and the DFG CRC1182 “Origin and function of metaorganisms”. R. H. and T.B.H.R. were supported by the Deutsche Forschungsgemeinschaft (DFG) CRC1182 ‘Origin and Function of Metaorganisms’, Subprojects B2, Z3 & INF.

Author contributions

T.H., S.M. and M.K. designed experiments, T.H. prepared and analyzed samples for metaproteomics with input from M.K., compiled metaproteomics database with input from M.K. and S.M., performed statistical analyses, prepared samples for RNA sequencing with input from C.B., prepared figures, wrote manuscript. T.S. was involved in project coordination. T.R., R.H. and P.R. coordinated transcriptome sequencing. H.F. helped with sampling. S.M.S. obtained funding for the research

685 cruises and coordinated sampling as chief scientist. C.B. assembled and annotated transcriptomic

686 data. All authors contributed to the final manuscript.

687

688 Competing interests

689 The authors declare no competing financial or non-financial interests.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–10. doi:10.1016/S0022-2836(05)80360-2.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Arndt, C., Schiedek, D., and Felbeck, H. (1998). Metabolic responses of the hydrothermal vent tube worm *Riftia pachyptila* to severe hypoxia. *Mar. Ecol. Prog. Ser.* 174, 151–158. doi:10.3354/meps174151.

Asojo, O. A., Goud, G., Dhar, K., Loukas, A., Zhan, B., Deumic, V., et al. (2005). X-ray structure of *Na*-ASP-2 , a pathogenesis-related-1 protein from the nematode parasite, *Necator americanus*, and a vaccine antigen for human hookworm infection. *J. Mol. Biol.* 346, 801–814. doi:10.1016/j.jmb.2004.12.023.

Bailly, X., and Vinogradov, S. (2005). The sulfide binding function of annelid hemoglobins: relic of an old biosystem? *J. Inorg. Biochem.* 99, 142–150. doi:10.1016/j.jinorgbio.2004.10.012.

Bang, C., Dagan, T., Deines, P., Dubilier, N., Duschl, W. J., Fraune, S., et al. (2018). Metaorganisms in extreme environments: do microbes play a role in organismal adaptation? *Zoology* 127, 1–19. doi:10.1016/j.zool.2018.02.004.

Bishop, B. M., Juba, M. L., Russo, P. S., Devine, M., Barksdale, S. M., Scott, S., et al. (2017). Discovery of novel antimicrobial peptides from *Varanus komodoensis* (Komodo dragon) by large-scale analyses and de-novo-assisted sequencing using electron-transfer dissociation mass spectrometry. *J. Proteome Res.* 1641470–14. doi:10.1021/acs.jproteome.6b00857.

Boetius, A., and Felbeck, H. (1995). Digestive enzymes in marine invertebrates from hydrothermal vents and other reducing environments. *Mar. Biol.* 122, 105–113. doi:10.1007/BF00349283.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170.

Bosch, T. C. G., and McFall-Ngai, M. J. (2011). Metaorganisms as the new frontier. *Zoology* 114, 185–190. doi:10.1016/j.zool.2011.04.001.

Brennan, L. J., Keddie, B. A., Braig, H. R., and Harris, H. L. (2008). The endosymbiont *Wolbachia pipientis* induces the expression of host antioxidant proteins in an *Aedes albopictus* cell line. *PLoS One* 3, e2083. doi:10.1371/journal.pone.0002083.

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

Bright, M., and Sorgo, A. (2003). Ultrastructural reinvestigation of the trophosome in adults of *Riftia pachyptila* (Annelida, Siboglinidae). *Invertebr. Biol.* 122, 345–366. doi:10.1111/j.1744-7410.2003.tb00099.x.

Brogden, K. A. (2005). Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat. Rev. Microbiol.* 3, 238–250. doi:10.1038/nrmicro1098.

Bröker, L. E., Huisman, C., Span, S. W., Rodriguez, A., Kruyt, F. A., and Giaccone, G. (2004). Cathepsin B mediates caspase-independent cell death induced by microtubule stabilizing agents in non-

small cell lung cancer cells. *Cancer Res.* 64, 27–30. doi:10.1158/0008-5472.CAN-03-3060.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421. doi:10.1186/1471-2105-10-421.

Caro, A., Got, P., Bouvy, M., Troussellier, M., and Gros, O. (2009). Effects of long-term starvation on a host bivalve (*Codakia orbicularis*, Lucinidae) and its symbiont population. *Appl. Environ. Microbiol.* 75, 3304–3313. doi:10.1128/AEM.02659-08.

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

Chang, Y.-Y., and Neufeld, T. P. (2010). Autophagy takes flight in *Drosophila*. *FEBS Lett.* 584, 1342–1349. doi:10.1016/j.febslet.2010.01.006.

Chavrier, P., Parton, R., Hauri, H., Simons, K., and Zerial, M. (1990). Localization of low molecular weight GTP binding proteins to exocytic and endocytic compartments. *Cell* 62, 317–329. doi:10.1016/0092-8674(90)90369-P.

Chen, F., Krasity, B. C., Peyer, S. M., Koehler, S., Ruby, E. G., Zhang, X., et al. (2017). Bactericidal permeability-increasing proteins shape host-microbe interactions. *MBio* 8, e00040-17. doi:10.1128/mBio.00040-17.

Cho, J. H., Sung, B. H., and Kim, S. C. (2009). Buforins: histone H2A-derived antimicrobial peptides from toad stomach. *Biochim. Biophys. Acta* 1788, 1564–1569. doi:10.1016/j.bbamem.2008.10.025.

Coates, J. C. (2003). Armadillo repeat proteins: beyond the animal kingdom. *Trends Cell Biol.* 13, 463–471. doi:10.1016/S0962-8924(03)00167-3.

Dale, C., and Moran, N. A. (2006). Molecular interactions between bacterial symbionts and their hosts. *Cell* 126, 453–465. doi:10.1016/j.cell.2006.07.014.

Davidson, S. K., Dulla, G. F., Go, R. A., Stahl, D. A., and Pinel, N. (2014). Earthworm symbiont *Verminephrobacter eiseniae* mediates natural transformation within host egg capsules using type IV pili. *Front. Microbiol.* 5, 546. doi:10.3389/fmicb.2014.00546.

De Cian, M.-C., Andersen, A. C., Bailly, X., and Lallier, F. H. (2003a). Expression and localization of carbonic anhydrase and ATPases in the symbiotic tubeworm *Riftia pachyptila*. *J. Exp. Biol.* 206, 399–409. doi:10.1242/jeb.00074.

De Cian, M.-C., Andersen, A. C., Toullec, J.-Y., Biegala, I., Caprais, J.-C., Shillito, B., et al. (2003b). Isolated bacteriocyte cell suspensions from the hydrothermal-vent tubeworm *Riftia pachyptila*, a potent tool for cellular physiology in a chemoautotrophic symbiosis. *Mar. Biol.* 142, 141–151. doi:10.1007/s00227-002-0931-5.

Distel, D. L., Lane, D. J., Olsen, G. J., Giovannoni, S. J., Pace, B., Pace, N. R., et al. (1988). Sulfur-oxidizing bacterial endosymbionts: analysis of phylogeny and specificity by 16S rRNA sequences. *J. Bacteriol.* 170, 2506–2510. doi:10.1128/jb.170.6.2506-2510.1988.

Drozdov, A. L., and Galkin, S. V. (2012). Morphology of gametes and insemination in the vestimentiferan Riftia pachyptila. *Open J. Mar. Sci.* 2, 96–102. doi:10.4236/ojms.2012.23013.

Dubilier, N., Bergin, C., and Lott, C. (2008). Symbiotic diversity in marine animals: the art of

harnessing chemosynthesis. *Nat. Rev. Microbiol.* 6, 725–740. doi:10.1038/nrmicro1992.

Eberle, H. B., Serrano, R. L., Füllekrug, J., Schlosser, A., Lehmann, W. D., Lottspeich, F., et al. (2002). Identification and characterization of a novel human plant pathogenesis-related protein that localizes to lipid-enriched microdomains in the Golgi complex. *J. Cell Sci.* 115, 827–838.

Eddy, S. R. (2009). A new generation of homology search tools based on probabilistic inference. *Genome Informatics 2009*, 205–211. doi:10.1142/9781848165632_0019.

Egas, C., Pinheiro, M., Gomes, P., Barroso, C., and Bettencourt, R. (2012). The transcriptome of *Bathymodiolus azoricus* gill reveals expression of genes from endosymbionts and free-living deep-sea bacteria. *Mar. Drugs* 10, 1765–1783. doi:10.3390/md10081765.

Elbourne, L. D. H., Tetu, S. G., Hassan, K. A., and Paulsen, I. T. (2017). TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res.* 45, D320–D324. doi:10.1093/nar/gkw1068.

Elsbach, P., and Weiss, J. (1998). Role of the bactericidal/permeability-increasing protein in host defence. *Curr. Opin. Immunol.* 10, 45–49. doi:10.1016/S0952-7915(98)80030-7.

Felbeck, H. (1981). Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila* Jones (Vestimentifera). *Science* 213, 336–338. doi:10.1126/science.213.4505.336.

Felbeck, H. (1985). CO_2 fixation in the hydrothermal vent tube worm *Riftia pachyptila* (Jones). *Physiol. Zool.* 58, 272–281. doi:10.1086/physzool.58.3.30155998.

Felbeck, H., and Jarchow, J. (1998). Carbon release from purified chemoautotrophic bacterial symbionts of the hydrothermal vent tubeworm *Riftia pachyptila*. *Physiol. Zool.* 71, 294–302. doi:10.1086/515931.

Feldhaar, H., and Gross, R. (2009). Insects as hosts for mutualistic bacteria. *InternationalJournalofMedicalMicrobiology* 299, 1–8. doi:10.1016/j.ijmm.2008.05.010.

Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., et al. (2017). InterPro in 2017–beyond protein family and domain annotations. *Nucleic Acids Res.* 45, D190–D199. doi:10.1093/nar/gkw1107.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi:10.1093/nar/gkv1344.

Fisher, C. R., Childress, J. J., and Minnich, E. (1989). Autotrophic carbon fixation by the chemoautotrophic symbionts of *Riftia pachyptila*. *Biol. Bull.* 177, 372–385. doi:10.2307/1541597.

Flores, J. F., Fisher, C. R., Carney, S. L., Green, B. N., Freytag, J. K., Schaeffer, S. W., et al. (2005). Sulfide binding is mediated by zinc ions discovered in the crystal structure of a hydrothermal vent tubeworm hemoglobin. *Proc. Natl. Acad. Sci. U. S. A.* 102, 2713–2718. doi:10.1073/pnas.0407455102.

Gardebrecht, A., Markert, S., Sievert, S. M., Felbeck, H., Thürmer, A., Albrecht, D., et al. (2012). Physiological homogeneity among the endosymbionts of *Riftia pachyptila* and *Tevnia jerichonana* revealed by proteogenomics. *ISME J.* 6, 766–776. doi:10.1038/ismej.2011.137.

Gilbert, S. F., Sapp, J., and Tauber, A. I. (2012). A symbiotic view of life: we have never been

individuals. *Q. Rev. Biol.* 87, 325–341. doi:10.1086/668166.

Goffredi, S. K., Girguis, P. R., Childress, J. J., and Desaulniers, N. T. (1999). Physiological functioning of carbonic anhydrase in the hydrothermal vent tubeworm *Riftia pachyptila*. *Biol. Bull.* 196, 257–264. doi:10.2307/1542950.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi:10.1038/nbt.1883.

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849. doi:10.1093/bioinformatics/btw313.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi:10.1038/nprot.2013.084.

Hager, A. J., Bolton, D. L., Pelletier, M. R., Brittnacher, M. J., Gallagher, L. A., Kaul, R., et al. (2006). Type IV pili-mediated secretion modulates *Francisella virulence*. *Mol. Microbiol.* 62, 227–237. doi:10.1111/j.1365-2958.2006.05365.x.

Hamann, E., Gruber-Vodicka, H., Kleiner, M., Tegetmeyer, H. E., Riedel, D., Littmann, S., et al. (2016). Environmental Breviatea harbour mutualistic *Arcobacter* epibionts. *Nature* 534, 254–258. doi:10.1038/nature18297.

Hammerschmidt, S., Wolff, S., Hocke, A., Rosseau, S., Müller, E., and Rohde, M. (2005). Illustration of pneumococcal polysaccharide capsule during adherence and invasion of epithelial cells. *Infect. Immun.* 73, 4653–4667. doi:10.1128/IAI.73.8.4653.

Hand, S. C. (1987). Trophosome ultrastructure and the characterization of isolated bacteriocytes from invertebrate-sulfur bacteria symbioses. *Biol. Bull.* 173, 260–276. doi:10.2307/1541878.

Hentschel, U., and Felbeck, H. (1993). Nitrate respiration in the hydrothermal vent tubeworm *Riftia pachyptila*. *Nature* 366, 338–340. doi:10.1038/366338a0.

Hentschel, U., Steinert, M., and Hacker, J. (2000). Common molecular mechanisms of symbiosis and pathogenesis. *TRENDS Microbiol.* 8, 226–231. doi:10.1016/S0966-842X(00)01758-3.

Hinzke, T., Kouris, A., Hughes, R.-A., Strous, M., and Kleiner, M. (2019). More is not always better: evaluation of 1D and 2D-LC-MS/MS methods for metaproteomics. *Front. Microbiol.* 10, 238. doi:10.3389/fmicb.2019.00238.

Hourdez, S., and Lallier, F. H. (2007). Adaptations to hypoxia in hydrotehrmal-vent and cold-seep invertebrates. *Rev. Environ. Sci. Bio/Technology* 6, 143–159. doi:10.1007/s11157-006-9110-3.

Hourdez, S., and Weber, R. E. (2005). Molecular and functional adaptations in deep-sea hemoglobins. *J. Inorg. Biochem.* 99, 130–141. doi:10.1016/j.jinorgbio.2004.09.017.

Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26, 680–682. doi:10.1093/bioinformatics/btq003.

Hyttinen, J. M. T., Niittykoski, M., Salminen, A., and Kaarniranta, K. (2013). Maturation of autophagosomes and endosomes: a key role for Rab7. *Biochim. Biophys. Acta* 1833, 503–510.

doi:10.1016/j.bbamcr.2012.11.018.

Jones, M. L. (1981). *Riftia pachyptila* Jones: observations on the vestimentiferan worm from the Galápagos Rift. *Science* 213, 333–336. doi:10.1126/science.213.4505.333.

Kleiner, M., Dong, X., Hinzke, T., Wippler, J., Thorson, E., Mayer, B., et al. (2018). Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* 115, E5576–E5584. doi:10.1073/pnas.1722325115.

Kleiner, M., Thorson, E., Sharp, C. E., Dong, X., Liu, D., Li, C., et al. (2017). Assessing species biomass contributions in microbial communities via metaproteomics. *Nat. Commun.* 8, 1558. doi:10.1038/s41467-017-01544-x.

König, S., Le Guyader, H., and Gros, O. (2015). Thioautotrophic bacterial endosymbionts are degraded by enzymatic digestion during starvation: Case study of two lucinids *Codakia orbicularis* and *C. orbiculata*. *Microsc. Res. Tech.* 78, 173–179. doi:10.1002/jemt.22458.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi:10.1038/nmeth.1923.

Leighton, T. L., Buensuceso, R. N. C., Howell, P. L., and Burrows, L. L. (2015). Biogenesis of *Pseudomonas aeruginosa* type IV pili and regulation of their function. *Environ. Microbiol.* 17, 4148–4163. doi:10.1111/1462-2920.12849.

Letunic, I., and Bork, P. (2018). 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* 46, D493–D496. doi:10.1093/nar/gkx922.

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). UpSet: visualization of intersecting sets. *IEEE Transcations Vis. Comput. Graph.* 20, 1983–1992. doi:10.1109/TVCG.2014.2346248.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352.

Li, J., Mahajan, A., and Tsai, M.-D. (2006). Ankyrin repeat: a unique motif mediating protein-protein interactions. *Biochemistry* 45, 15168–15178. doi:10.1021/bi062188q.

Liu, H., Wang, H., Cai, S., and Zhang, H. (2017). A novel ω3-desaturase in the deep sea giant tubeworm *Riftia pachyptila*. *Mar. Biotechnol.* 19, 345–350. doi:10.1007/s10126-017-9753-9.

Login, F. H., Balmand, S., Vallier, A., Vincent-Monégat, C., Vigneron, A., Weiss-Gayet, M., et al. (2011). Antimicrobial peptides keep insect endosymbionts under control. *Science* 334, 362–365. doi:10.1126/science.1209728.

Lutz, R. A., Shank, T. M., Fornari, D. J., Haymon, R. M., Lilley, M. D., Von Damm, K. L., et al. (1994). Rapid growth at deep-sea vents. *Nature* 371, 663–664. doi:10.1038/371663a0.

Mangum, C. P. (1992). "Physiological function of the hemerythrins," in *Blood and Tissue Oxygen Carriers. Advances in Comparative and Environmental Physiology, vol 13*, ed. M. C.P. (Berlin, Heidelberg: Springer), 173–192. doi:10.1007/978-3-642-76418-9_7.

Marchetti, A., Mercanti, V., Cornillon, S., Alibaud, L., Charette, S. J., and Cosson, P. (2004). Formation of multivesicular endosomes in *Dictyostelium*. *J. Cell Sci.* 117, 6053–6059. doi:10.1242/jcs.01524.

890 Markert, S., Arndt, C., Felbeck, H., Becher, D., Sievert, S. M., Hügler, M., et al. (2007). Physiological
891 proteomics of the uncultured endosymbiont of *Riftia pachyptila*. *Science* 315, 247–250.
892 doi:10.1126/science.1132913.

893 Markert, S., Gardebrecht, A., Felbeck, H., Sievert, S. M., Klose, J., Becher, D., et al. (2011). Status quo in
894 physiological proteomics of the uncultured *Riftia pachyptila* endosymbiont. *Proteomics* 11,
895 3106–3117. doi:10.1002/pmic.201100059.

896 McFall-Ngai, M. (2008). Are biologists in 'future shock'? Symbiosis integrates biology across
897 domains. *Nat. Rev. Microbiol.* 6, 789–792. doi:10.1038/nrmicro1982.

898 McFall-Ngai, M., Hadfield, M. G., Bosch, T. C. G., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., et al.
899 (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad.*
900 *Sci.* 110, 3229–3236. doi:10.1073/pnas.1218525110.

901 Minic, Z., and Hervé, G. (2003). Arginine metabolism in the deep sea tube worm *Riftia pachyptila*
902 and its bacterial endosymbiont. *J. Biol. Chem.* 278, 40527–40533.
903 doi:10.1074/jbc.M307835200.

904 Minic, Z., Simon, V., Penverne, B., Gaill, F., and Hervé, G. (2001). Contribution of the bacterial
905 endosymbiont to the biosynthesis of pyrimidine nucleotides in the deep-sea tube worm *Riftia*
906 *pachyptila*. *J. Biol. Chem.* 276, 23777–23784. doi:10.1074/jbc.M102249200.

907 Moya, A., Peretó, J., Gil, R., and Latorre, A. (2008). Learning how to live together: genomic insights
908 into prokaryote–animal symbioses. *Nat. Rev. Genet.* 9, 218–229. doi:10.1038/nrg2319.

909 Mueller, R. S., Denef, V. J., Kalnejais, L. H., Suttle, K. B., Thomas, B. C., Wilmes, P., et al. (2010).
910 Ecological distribution and population physiology defined by proteomics in a natural
911 microbial community. *Mol. Syst. Biol.* 6, 374. doi:10.1038/msb.2010.30.

912 Nguyen, M. T. H. D., Liu, M., and Thomas, T. (2014). Ankyrin-repeat proteins from sponge symbionts
913 modulate amoebal phagocytosis. *Mol. Ecol.* 23, 1635–1645. doi:10.1111/mec.12384.

914 Nyholm, S. V., Stewart, J. J., Ruby, E. G., and McFall-Ngai, M. J. (2009). Recognition between symbiotic
915 *Vibrio fischeri* and the haemocytes of *Euprymna scolopes*. *Environ. Microbiol.* 11, 483–493.
916 doi:10.1111/j.1462-2920.2008.01788.x.

917 Nyholm, S. V, and Graf, J. (2012). Knowing your friends: invertebrate innate immunity fosters
918 beneficial bacterial symbioses. *Nat. Rev. Microbiol.* 10, 815–827. doi:10.1038/nrmicro2894.

919 Oberg, A. L., and Vitek, O. (2009). Statistical design of quantitative mass spectrometry-based
920 proteomic profiling experiments. *J. Proteome Res.* 8, 2144–2156. doi:10.1021/pr8010099.

921 Pan, X., Lührmann, A., Satoh, A., Laskowski-Arce, M., and R, R. C. (2008). Ankyrin repeat proteins
922 comprise a diverse family of bacterial type IV effectors. *Science* 320, 1651–1654.
923 doi:10.1126/science.1158160.

924 Park, I. Y., Park, C. B., Kim, M. S., and Kim, S. C. (1998). Parasin I, an antimicrobial peptide derived
925 from histone H2A in the cat fish, *Parasilurus asotus*. *FEBS Lett.* 437, 258–262.
926 doi:10.1016/S0014-5793(98)01238-1.

927 Petersen, J. M., Zielinski, F. U., Pape, T., Seifert, R., Moraru, C., Amann, R., et al. (2011). Hydrogen is an
928 energy source for hydrothermal vent symbioses. *Nature* 476, 176–180.
929 doi:10.1038/nature10325.

930 Pflugfelder, B., Cary, S. C., and Bright, M. (2009). Dynamics of cell proliferation and apoptosis reflect
931 different life strategies in hydrothermal vent and cold seep vestimentiferan tubeworms. *Cell*
932 *Tissue Res.* 337, 149–165. doi:10.1007/s00441-009-0811-0.

933 Pflugfelder, B., Fisher, C. R., and Bright, M. (2005). The color of the trophosome: elemental sulfur
934 distribution in the endosymbionts of *Riftia pachyptila* (Vestimentifera; Siboglinidae). *Mar. Biol.*
935 146, 895–901. doi:10.1007/s00227-004-1500-x.

936 Pollack-Berti, A., Wollenberg, M. S., and Ruby, E. G. (2010). Natural transformation of *Vibrio fischeri*
937 requires tfoX and tfoY. *Environ. Microbiol.* 12, 2302–2311. doi:10.1111/j.1462-
938 2920.2010.02250.x.

939 Polzin, J., Arevalo, P., Nussbaumer, T., Polz, M. F., and Bright, M. (2019). Polyclonal symbiont
940 populations in hydrothermal vent tubeworms and the environment. *Proc. R. Soc. B Biol. Sci.*
941 286, 20181281. doi:10.1098/rspb.2018.1281.

942 Ponnudurai, R., Kleiner, M., Sayavedra, L., Petersen, J. M., Moche, M., Otto, A., et al. (2017). Metabolic
943 and physiological interdependencies in the *Bathymodiolus azoricus* symbiosis. *ISME J.* 11, 463–
944 477. doi:10.1038/ismej.2016.124.

945 Prota, A. E., Magiera, M. M., Kuijpers, M., Bargsten, K., Frey, D., Wieser, M., et al. (2013). Structural
946 basis of tubulin tyrosination by tubulin tyrosine ligase. *J. Cell Biol.* 200, 259–270.
947 doi:10.1083/jcb.201211017.

948 R Core Team (2017). R: A language and environment for statistical computing. http://www.R-
949 project.org/. *R Found. Stat. Comput.*

950 Reynolds, D., and Thomas, T. (2016). Evolution and function of eukaryotic-like proteins from
951 sponge symbionts. *Mol. Ecol.* 25, 5242–5253. doi:10.1111/mec.13812.

952 Robidart, J. C., Bench, S. R., Feldman, R. A., Novoradovsky, A., Podell, S. B., Gaasterland, T., et al.
953 (2008). Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed through
954 metagenomics. *Environ. Microbiol.* 10, 727–737. doi:10.1111/j.1462-2920.2007.01496.x.

955 Robidart, J. C., Roque, A., Song, P., and Girguis, P. R. (2011). Linking hydrothermal geochemistry to
956 organismal physiology: physiological versatility in *Riftia pachyptila* from sedimented and
957 basalt-hosted vents. *PLoS One* 6, e21692. doi:10.1371/journal.pone.0021692.

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

961 Rose, F. R. A. J., Bailey, K., Keyte, J. W., Chan, W. C., Greenwood, D., and Mahida, Y. R. (1998). Potential
962 role of epithelial cell-derived histone H1 proteins in innate antimicrobial defense in the human
963 gastrointestinal tract. *Infect. Immun.* 66, 3255–3263.

964 Saier Jr, M. H., Reddy, V. S., Tsu, B. V, Ahmed, M. S., Li, C., and Moreno-Hagelsieb, G. (2016). The
965 transporter classification database (TCDB): recent advances. *Nucleic Acids Res.* 44, D372–
966 D379. doi:10.1093/nar/gkv1103.

967 Sanchez, S., Andersen, A. C., Hourdez, S., and Lallier, F. H. (2007a). Identification, sequencing, and
968 localization of a new carbonic anhydrase transcript from the hydrothermal vent tubeworm
969 *Riftia pachyptila*. *FEBS J.* 274, 5311–5324. doi:10.1111/j.1742-4658.2007.06050.x.

970 Sanchez, S., Hourdez, S., and Lallier, F. H. (2007b). Identification of proteins involved in the

971 functioning of *Riftia pachyptila* symbiosis by Subtractive Suppression Hybridization. *BMC*
972 *Genomics* 8, 337. doi:10.1186/1471-2164-8-337.

973 Schmieder, R., and Edwards, R. (2011). Fast identification and removal of sequence contamination
974 from genomic and metagenomic datasets. *PLoS One* 6, e17288.
975 doi:10.1371/journal.pone.0017288.

976 Scott, K. M., Boller, A. J., Dobrinski, K. P., and Le Bris, N. (2012). Response of hydrothermal vent
977 vestimentiferan *Riftia pachyptila* to differences in habitat chemistry. *Mar. Biol.* 159, 435–442.
978 doi:10.1007/s00227-011-1821-5.

979 Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO:
980 assessing genome assembly and annotation completeness with single-copy orthologs.
981 *Bioinformatics* 31, 3210–3212. doi:10.1093/bioinformatics/btv351.

982 Simonet, P., Gaget, K., Balmand, S., Ribeiro Lopes, M., Parisot, N., Buhler, K., et al. (2018).
983 Bacteriocyte cell death in the pea aphid/*Buchnera* symbiotic system. *Proc. Natl. Acad. Sci. U. S.*
984 *A.* 115, E1819–E1828. doi:10.1073/pnas.1720237115.

985 Stewart, F. J., and Cavanaugh, C. M. (2005). "Symbiosis of thioautotrophic bacteria with *Riftia*
986 *pachyptila*," in *Molecular Basis of Symbiosis. Progress in Molecular and Subcellular Biology, vol*
987 *41*, ed. J. Overmann (Springer, Berlin, Heidelberg), 197–225.

988 Stewart, F. J., Newton, I. L. G., and Cavanaugh, C. M. (2005). Chemosynthetic endosymbioses:
989 adaptations to oxic-anoxic interfaces. *TRENDS Microbiol.* 13, 439–448.
990 doi:10.1016/j.tim.2005.07.007.

991 Stone, B. J., and Kwaik, Y. A. B. U. (1999). Natural competence for DNA transformation by *Legionella*
992 *pneumophila* and its association with expression of type IV pili. *J. Bacteriol.* 181, 1395–1402.

993 Streams, M. E., Fisher, C. R., and Fiala-Médioni, A. (1997). Methanotrophic symbiont location and
994 fate of carbon incorporated from methane in a hydrocarbon seep mussel. *Mar. Biol.* 129, 465–
995 476. doi:10.1007/s002270050187.

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

998 The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*
999 45, D158–D169. doi:10.1093/nar/gkw1099.

1000 Van Dover, C. L. (2000). *The Ecology of Deep-Sea Hydrothermal Vents*. Princeton, New Jersey:
1001 Princeton University Press.

1002 Van Loon, C. L., and Van Strien, E. A. (1999). The families of pathogenesis-related proteins, their
1003 activities, and comparative analysis of PR-1 type proteins. *Physiol. Mol. Plant Pathol.* 55, 85–97.
1004 doi:10.1006/pmpp.1999.0213.

1005 Vieira, O. V, Botelho, R. J., and Grinstein, S. (2002). Phagosome maturation: aging gracefully.
1006 *Biochem. J.* 366, 689–704. doi:10.1042/BJ20020691.

1007 Wang, G., Li, X., and Wang, Z. (2016). APD3: the antimicrobial peptide database as a tool for research
1008 and education. *Nucleic Acids Res.* 44, D1087–D1093. doi:10.1093/nar/gkv1278.

1009 Webster, N. S. (2014). Cooperation, communication, and co-evolution: grand challenges in microbial
1010 symbiosis research. *Front. Microbiol.* 5, 164. doi:10.3389/fmicb.2014.00164.

Wilmot Jr., D. B., and Vetter, R. D. (1990). The bacterial symbiont from the hydrothermal vent tubeworm *Riftia pachyptila* is a sulfide specialist. *Mar. Biol.* 106, 273–283. doi:10.1007/BF01314811.

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

Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362. doi:10.1038/nmeth.1322.

Woyke, T., Teeling, H., Ivanova, N. N., Huntemann, M., Richter, M., Gloeckner, F. O., et al. (2006). Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* 443, 950–955. doi:10.1038/nature05192.

Yoshimura, S. H., and Hirano, T. (2016). HEAT repeats – versatile arrays of amphiphilic helices working in crowded environments? *J. Cell Sci.* 129, 3963–3970. doi:10.1242/jcs.185710.

Zal, F., Lallier, H., Green, B. N., and Vinogradov, S. N. (1996). The multi-hemoglobin system of the hydrothermal vent tube worm *Riftia pachyptila*. II. Complete polypeptide chain composition investigated by maximum entropy analysis of mass spectra. *J. Biol. Chem.* 271, 8875–8881. doi:10.1074/jbc.271.15.8869.

Zal, F., Leize, E., Lallier, F. H., Toulmond, A., Van Dorsselaer, A., and Childress, J. (1998). S-sulfohemoglobin and disulfide exchange: The mechanisms of sulfide binding by *Riftia pachyptila* hemoglobins. *Proc. Natl. Acad. Sci. USA* 95, 8997–9002. doi:10.1073/pnas.95.15.8997.

Zug, R., and Hammerstein, P. (2015). *Wolbachia* and the insect immune system: what reactive oxygen species can tell us about the mechanisms of *Wolbachia*-host interactions. *Front. Microbiol.* 6, 1201. doi:10.3389/fmicb.2015.01201.

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